

**ANALISIS HUBUNGAN UKURAN SEPATU DENGAN TINGGI
BADAN MENGGUNAKAN METODE REGRESI LINEAR**



HAIKAL AHMAD ROBBANI 0110120208

SISTEM INFORMASI

STT NURUL FIKRI

2023

A. Latar Belakang

Salah satu metode statistika yang sering digunakan adalah metode regresi linear. Regresi linear adalah metode yang digunakan untuk memprediksi suatu hubungan linear antara satu atau beberapa variabel. Terdapat dua jenis variabel dalam regresi linear, yaitu variabel bebas (independent) dan variabel tak bebas (dependen). Tujuan metode regresi linear adalah mendapatkan gambaran dari hubungan antara variabel bebas dan tidak bebas.

Sepatu merupakan perlengkapan yang digunakan oleh banyak orang dalam kehidupan. Tak jarang bagi kebanyakan orang saat ini sepatu merupakan salah satu kebutuhan utama dalam menunjang kegiatan aktivitas sehari-hari mereka. Oleh karena itu diperlukan ukuran sepatu yang sesuai agar dapat memberikan kenyamanan kepada penggunanya. Seringkali kita banyak melihat beberapa orang memiliki ukuran sepatu yang sama akan tetapi memiliki tinggi badan yang berbeda. Oleh karena itu muncul suatu hipotesa adanya suatu hubungan ukuran sepatu seseorang dengan tinggi badan mereka.

Melalui penggunaan metode regresi linear, dapat dilakukan analisis untuk mengetahui apakah terdapat hubungan atau korelasi antara ukuran sepatu dan tinggi badan, serta seberapa kuat hubungan tersebut. Metode regresi linear memungkinkan kita untuk menentukan persamaan matematis yang menggambarkan hubungan antara dua variabel. Dalam kasus ini, variabel bebas atau independen adalah ukuran sepatu, sedangkan variabel tak bebas atau dependen adalah tinggi badan.

Untuk mendukung analisis ini, maka diperlukan data-data mengenai ukuran sepatu dan tinggi badan. Oleh karena itu, penulis melakukan pengumpulan data menggunakan google form yang disebarakan kepada masyarakat. Informasi mengenai hubungan ukuran sepatu dengan tinggi badan dapat digunakan di berbagai bidang. Salah satunya, produsen merk sepatu dapat memanfaatkan informasi ini untuk membantu menghasilkan produk sepatu yang sesuai dengan tinggi badan konsumen. Oleh karena itu, analisis mengenai hubungan ukuran sepatu dengan tinggi badan menggunakan metode regresi linear merupakan pendekatan yang penting untuk dilakukan. Dalam analisis tersebut, pengukuran kinerja menggunakan parameter Mean Absolute Error, Mean Squared Error, dan Root mean squared error juga dapat membantu dalam mengevaluasi seberapa akurat persamaan regresi linear yang telah dibentuk. Analisis ini dipraktekkan dengan membuat program untuk menghitung garis regresi linear, memvisualisasikan data dan garis regresi yang di bentuk. Program ini dibuat dengan menggunakan bahasa pemrograman python

B. Rumusan Masalah

1. Apa hubungan ukuran sepatu dengan tinggi badan seseorang?
2. Bagaimana cara pengumpulan data ?
3. Bagaimana bentuk garis regresi linear yang dibentuk?
4. Bagaimana visualisasi data dan garis regresi linear yang di bentuk?
5. Berapa nilai hasil uji pengukuran kinerja model?

C. Tujuan

1. Mengetahui hubungan ukuran sepatu dengan tinggi badan seseorang
2. Mengetahui cara mengumpulkan data
3. Mengetahui bentuk garis regresi linear yang dibentuk
4. Mengetahui visualisasi data yang dibentuk
5. Mengetahui nilai hasil uji pengukuran kinerja model

BAB II

Teori

- **Regresi linear**

Regresi linear merupakan salah cara dari model statistic yang dilakukan untuk memprediksi antara 2 variabel dengan cara supervised learning. Model ini merupakan yang paling umum dan sederhana dalam penggunaan dari machine learning. Regresi Linier hanya bisa digunakan untuk data yang bersifat interval dan rasio yang biasanya bersifat diskrit dan kontinu, dan merupakan analisis bivariate dan multivariate. Regresi Linier melibatkan 2 variabel dimana salah satunya adalah variabel independen (x) dan satu lagi adalah variabel dependen (y). Independen berarti variabel ini sebagai variabel utama yang mungkin akan mempengaruhi nilai variabel kedua (dependen). Dependen berarti nilai variabel ini akan tergantung dari nilai variabel independennya, jika korelasi tinggi maka dependensi juga tinggi.

Model Persamaan Regresi Linear Sederhana adalah seperti berikut ini :

$$Y = a + bX$$

Dimana :

Y = Variabel Response atau Variabel Akibat (Dependent)

X = Variabel Predictor atau Variabel Faktor Penyebab (Independent)

a = konstanta

b = koefisien regresi (kemiringan); besaran Response yang ditimbulkan oleh Predictor.

Nilai-nilai a dan b dapat dihitung dengan menggunakan Rumus dibawah ini :

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

selain itu, model regresi linear juga dapat dilakukan di pemrograman python dengan cara menggunakan library yang sudah disediakan oleh scikit.learn

```
from sklearn.linear_model import LinearRegression
```

- **Pengukuran kinerja menggunakan parameter Mean Absolute Error, Mean Squared Error dan Root mean squared error.**

1. MEA

Mean Absolute Error atau MEA, adalah rata-rata selisih mutlak nilai sebenarnya (aktual) dengan nilai prediksi (peramalan). MAE digunakan untuk mengukur keakuratan suatu model statistik dalam melakukan prediksi atau peramalan. rumus menghitung MAE adalah :

$$MAE = \frac{1}{n} \sum_{i=1}^n |A_i - F_i|$$

dimana:

n adalah ukuran sampel

A_i adalah nilai data aktual ke- i

F_i adalah nilai data peramalan ke- i

Dalam pemrograman python, mae dapat dihitung dengan menggunakan kode berikut :

```
Mae = mean_absolute_error(nilai sebenarnya, nilai prediksi)
```

mean_absolute_error adalah fungsi menghitung MAE yang disediakan dari library scikit-learn

2. MSE

MSE, atau mean squared error digunakan untuk mengecek seberapa dekat perkiraan atau forecast dengan nilai sebenarnya. Metode Mean Squared Error secara umum digunakan untuk mengecek estimasi berapa nilai kesalahan pada prediksi. Nilai Mean Squared Error yang rendah atau nilai mean squared error mendekati nol menunjukkan bahwa hasil peramalan sesuai dengan data aktual dan bisa dijadikan untuk perhitungan peramalan di periode mendatang.

$$MSE = \frac{\sum_{t=1}^n (A_t - F_t)^2}{n}$$

A_t = Nilai Aktual permintaan

F_t = Nilai hasil peramalan

n = banyaknya data

dalam pemrograman python, cara menghitung MSE adalah dengan kode berikut

```
mse = mean_squared_error(nilai_aktual , nilai prediksi)
```

dimana :

mse = variable yang berikan nilai mse

mean squared error = fungsi untuk menghitung mse yang diambil dari library scikit-learn

3. RMSE

RMSE atau Root Mean Square Error (RMSE) adalah metode pengukuran model dari regresi linear dengan mengukur perbedaan nilai dari prediksi sebuah model sebagai estimasi atas nilai yang diobservasi. RMSE berguna untuk mengukur tingkat keakuratan model dalam memprediksi nilai.

Root Mean Square Error adalah hasil dari akar kuadrat Mean Square Error
cara menghitung :

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (A_t - F_t)^2}{n}}$$

Dimana :

A_t = Nilai data Aktual

F_t = Nilai hasil peramalan

N = banyaknya data

Σ = Summation (Jumlahkan keseluruhan nilai)

Cara melakukan perhitungan rmse di python dengan kode berikut :

`Rmse = np.sqrt(mse)`

Rmse = variable yang bernama rmse

Np. = NumPy, library python yang berguna untuk melakukan operasi matematika pada suatu array

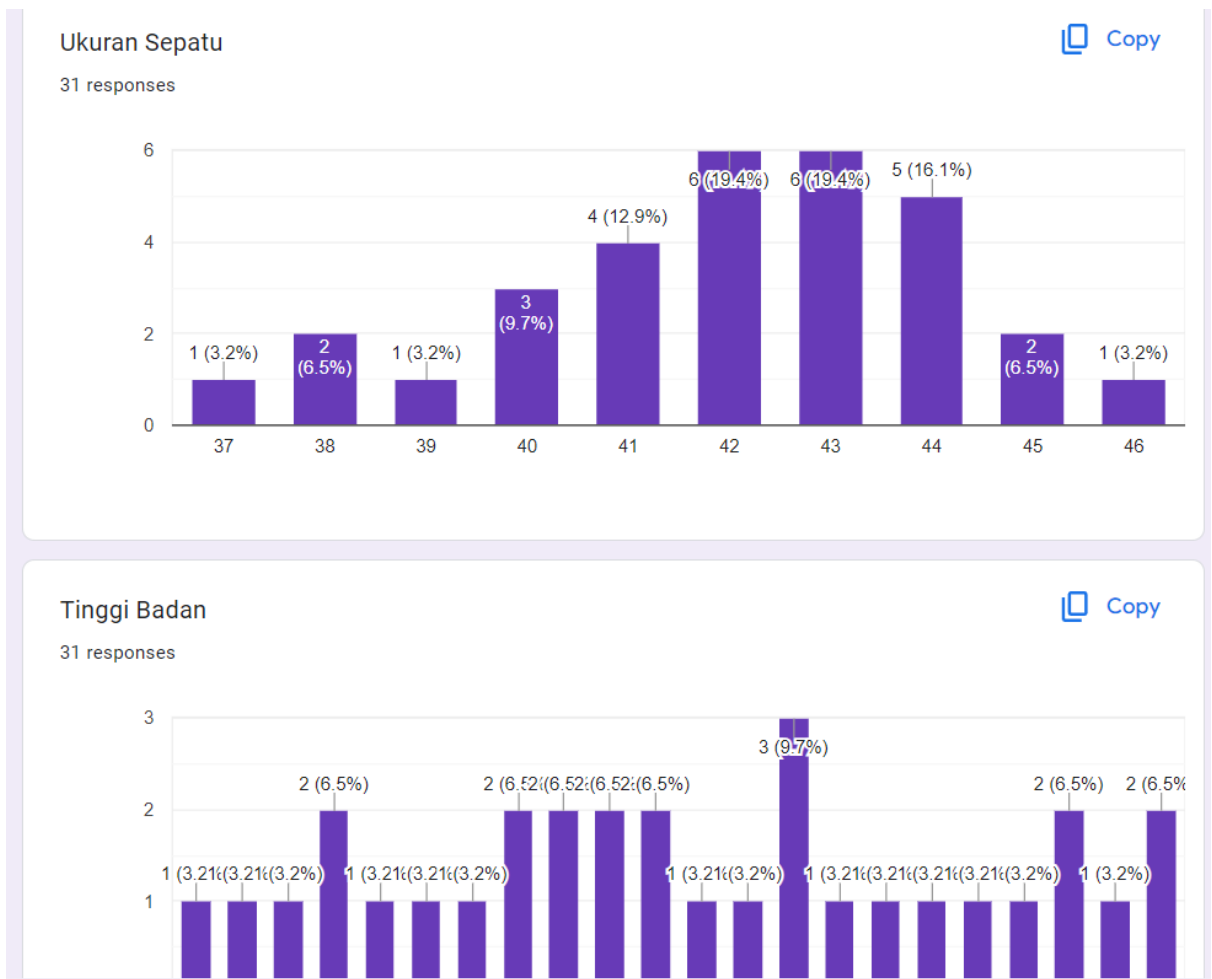
Sqrt = **sqrt** adalah fungsi matematika bawaan dari Python yang digunakan untuk menghitung akar kuadrat dari suatu bilangan.

BAB III

Perancangan

1. Mengumpulkan data

Pada tahapan ini, dilakukan pengumpulan data-data yang nantinya akan menjadi variabel bebas(ukuran sepatu) dan variabel tidak bebas(tinggi badan). Penulis mengumpulkan data dengan metode kuesioner. Sample pengumpulan data ini berdasarkan masyarakat indonesia. Tidak terbatas pada umur dan jenis kelamin.



Hasil dari kuesioner didapatkan sebanyak 30 data koresponden. Selanjutnya data tersebut dimasukkan ke dalam file csv agar mempermudah program melakukan analisis.

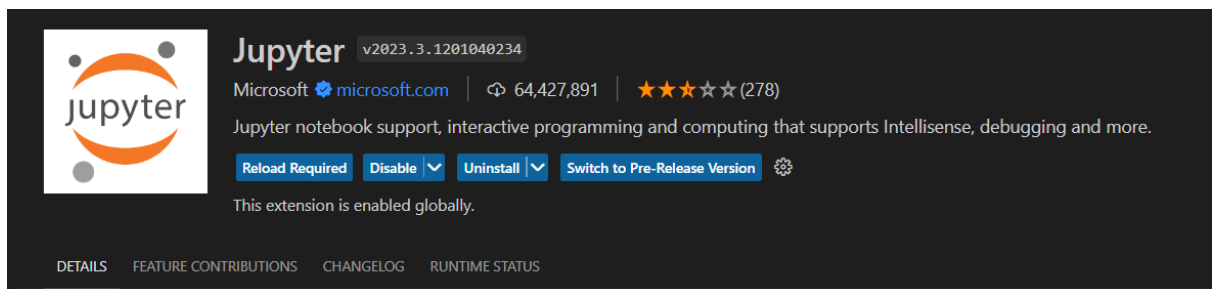
	A	B	C	D	E	F	G
1	Ukuran Sepatu,Tinggi Badan						
2	41,164						
3	43,166						
4	38,161						
5	44,173						
6	45,169						
7	43,182						
8	40,157						
9	38,155						
10	42,172						
11	43,165						
12	42,163						
13	42,164						
14	41,16						
15	46,176						
16	37,154						
17	44,17						
18	45,186						
19	40,165						
20	41,167						
21	44,178						
22	40,16						
23	42,168						
24	44,178						
25	43,186						
26	39,162						
27	43,174						
28	42,17						
29	41,166						
30	42,167						
31	43,17						
32	44,177						
33							

Tampilan dari isi file regresi.csv

2. Menginstal Jupiter notebook di text editor

Dalam menunjang analisis hubungan ukuran sepatu dengan tinggi, maka diperlukan program yang dapat membantu analisis tersebut. penulis membuat program dengan menggunakan bahasa pemrograman python. Untuk itu diperlukan tools tambahan, yaitu jupyter notebook.

Jupyter Notebook merupakan singkatan dari tiga bahasa pemrograman, yakni Julia (Ju), Python (Py), dan R. Jupyter adalah sebuah tools gratis yang paling banyak dipakai oleh data scientist. Aplikasi ini dipakai untuk membuat dan membagikan dokumen yang memiliki kode, hasil hitungan, visualisasi, dan teks. Untuk itu, penulis menggunakan tools ini di teks editor vscode



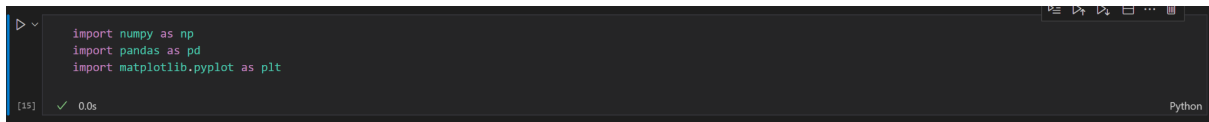
BAB IV

Implementasi dan pengujian

- **Regresi linear**

Implementasi dengan menggunakan python dengan bantuan jupyter notebook di vscode

1. Import library-library yang dibutuhkan

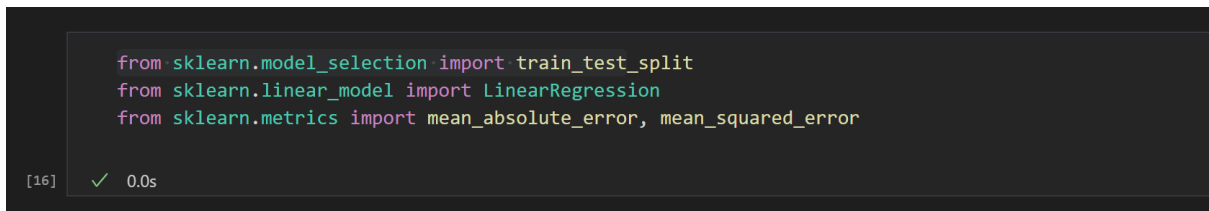


```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

[15] ✓ 0.0s Python

- a. numpy : library yang menyediakan fungsi untuk melakukan komputasi numerik di python
- b. pandas : library yang menyediakan fungsi untuk melakukan manipulasi dan analisis data
- c. matplotlib : library yang berguna untuk membuat visualisasi data

2. mengimport library dari scikit-learn



```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error
```

[16] ✓ 0.0s

- a. train_test_split: library ini menyediakan fungsi untuk membagi data menjadi set pelatihan dan pengujian.
- b. LinearRegression: library ini menyediakan model regresi linier untuk melakukan prediksi.
- c. metrics: library ini menyediakan fungsi untuk mengukur kinerja model.

3. Membaca data dari file csv

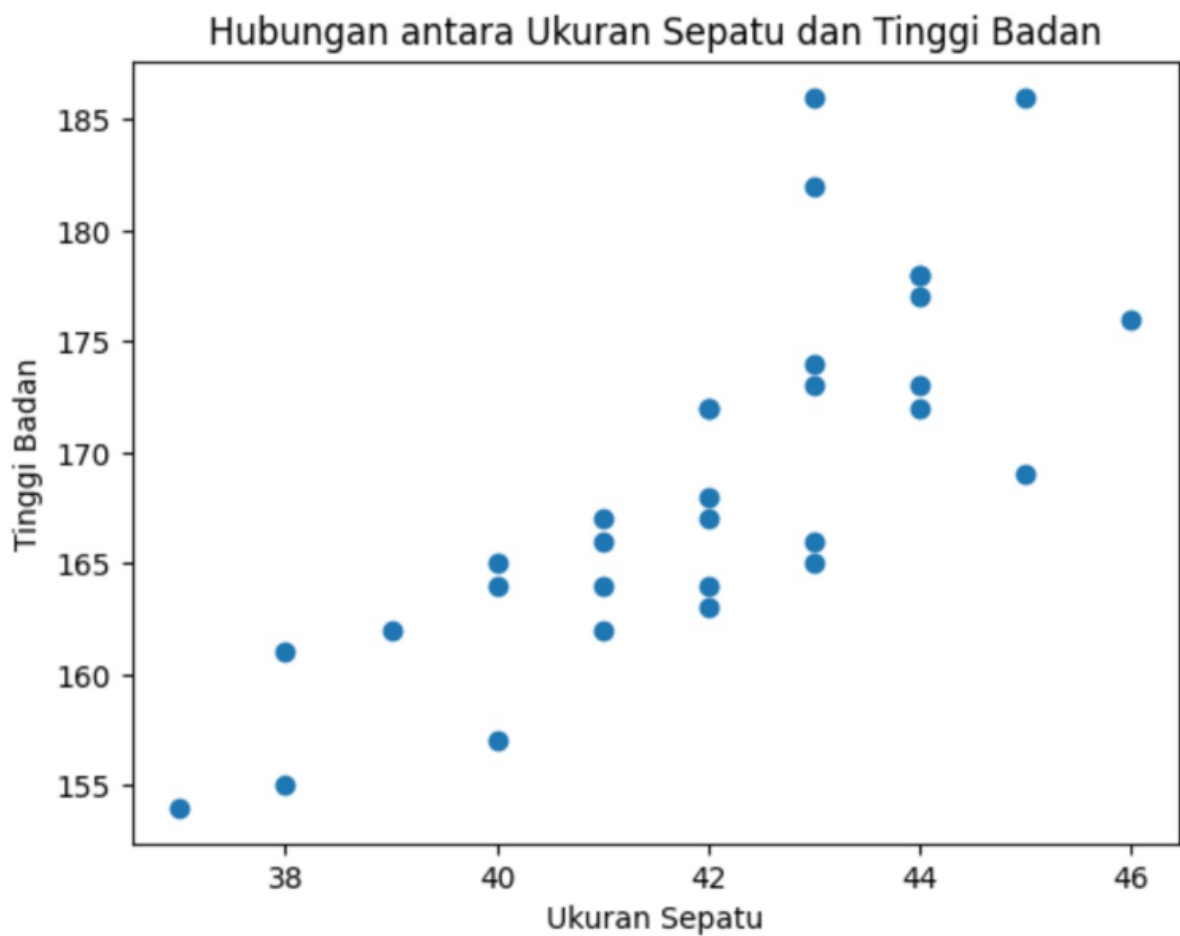
```
# mengambil data dari file csv
data = pd.read_csv('regresi.csv')
shoe_size = data['ukuran_sepatu']
height = data['tinggi_badan']
```

✓ 0.0s

4. Membuat Scatter plot

```
# Membuat gambaran dari data
plt.scatter(shoe_size, height)
plt.xlabel("Ukuran Sepatu")
plt.ylabel("Tinggi Badan")
plt.title("Hubungan antara Ukuran Sepatu dan Tinggi Badan")
plt.show()
```

[18] ✓ 0.1s



6. membuat data uji dan data latih

```
# membagi data, menjadi data uji dan data latih
X_train, X_test, y_train, y_test = train_test_split(shoe_size, height, test_size=0.2, random_state=0)
```

✓ 0.0s

7. Membuat data latih menjadi bentuk array

```
# membuat array yang berisikan data uji dan data latih
X_train = np.array(X_train).reshape(-1, 1)
y_train = np.array(y_train).reshape(-1, 1)
```

0] ✓ 0.0s

8. Membuat model regresi

```
# membuat model regresi
regressor = LinearRegression()
regressor.fit(X_train, y_train)
```

21] ✓ 0.0s

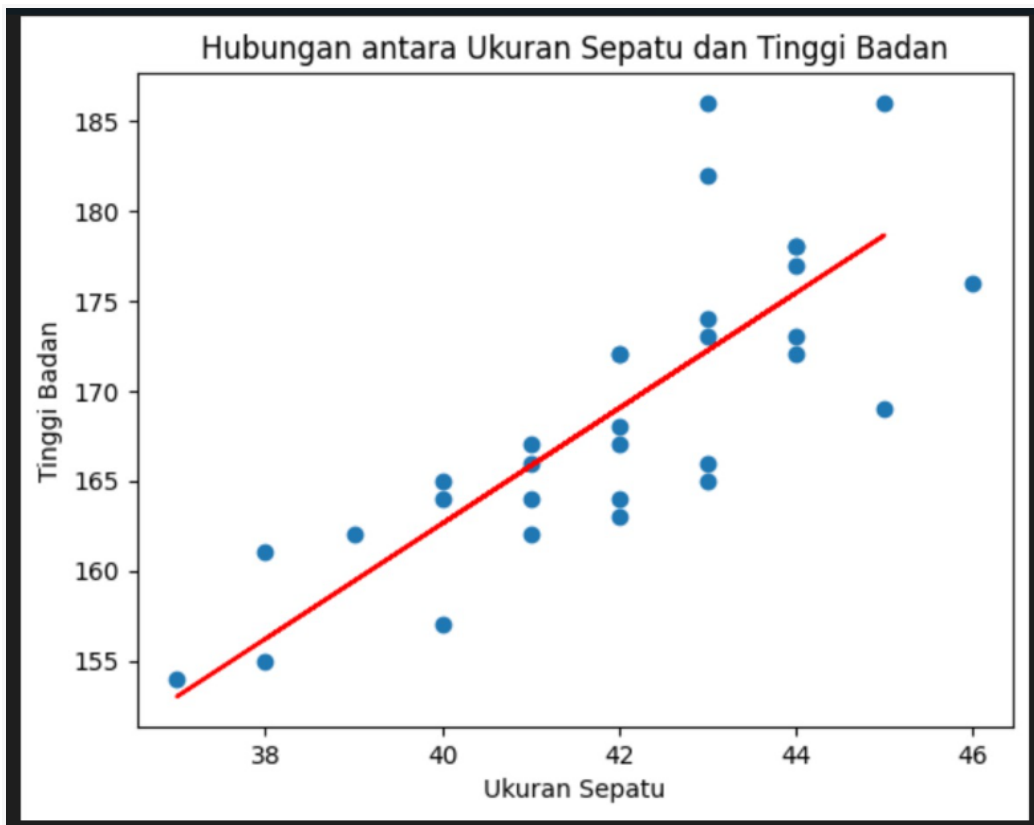
..

- ▼ LinearRegression
- LinearRegression()

9. Membuat garis regresi linear

```
# membuat visualisasi untuk garis regresi linear
plt.scatter(shoe_size, height)
plt.plot(X_train, regressor.predict(X_train), color='red')
plt.xlabel("Ukuran Sepatu")
plt.ylabel("Tinggi Badan")
plt.title("Hubungan antara Ukuran Sepatu dan Tinggi Badan")
plt.show()
```

2] ✓ 0.2s



- Mengukur kinerja model

1. Menjadi variable data test(data uji) menjadi array dan membuat variabel yang menyimpan nilai regresi prediksi

```
# mengukur kinerja model
```

```
X_test = np.array(X_test).reshape(-1, 1)
```

```
y_test = np.array(y_test).reshape(-1, 1)
```

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

3] ✓ 0.0s

2. Melakukan perhitungan kinerja dengan MAE,MSE dan RMSE.

```
# melakukan perhitungan mae,mse dan rmse
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)

# menampilkan hasilnya

print("Mean Absolute Error:", mae)
print("Mean Squared Error:", mse)
print("Root Mean Square Error:", rmse)
```

✓ 0.0s

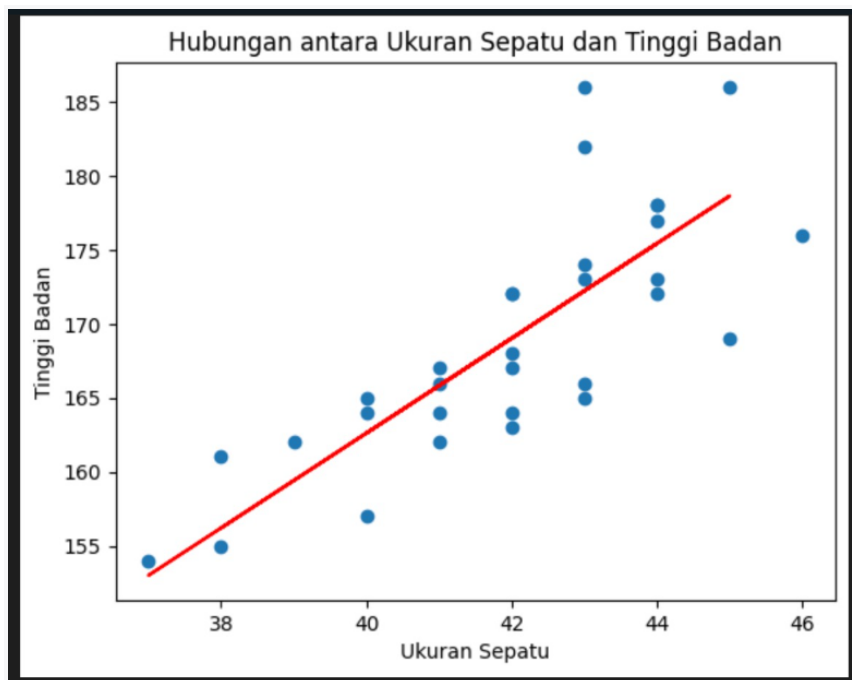
3. Hasilnya

```
Mean Absolute Error: 3.135200974421435
Mean Squared Error: 14.824302972667825
Root Mean Square Error: 3.8502341451745274
```

BAB V

Kesimpulan

Setelah dilakukan implementasi terhadap dua variabel bebas dan tidak bebas, maka didapatkan diagram tebar atau scatter plotnya. dari scatter plot tersebut dapat kita simpulkan bahwa terdapat positive correlation, dimana Y (tinggi badan) cenderung meningkat mengikut penambahan nilai X(ukuran sepatu).



Sementara dari hasil pengujian, di dapatkan MSE,MSA dan RMSE sebagai berikut :

MAE : 3,13

MSE : 14,8

RMSE : 3,85

Berdasarkan hasil pengujian data di atas, dapat disimpulkan bahwa model regresi linear yang dibuat untuk memprediksi tinggi badan berdasarkan ukuran sepatu memiliki performa yang cukup baik. Hal ini dapat dilihat dari nilai MAE yang hanya sebesar 3,13 dan RMSE yang sebesar 3,85, yang menunjukkan bahwa rata-rata kesalahan prediksi tinggi badan hanya sekitar 3,13 cm, dan kesalahan prediksi secara umum hanya berkisar sekitar 3,85 cm. Meskipun demikian, nilai MSE yang sebesar 14,8 menunjukkan bahwa model masih memiliki variansi yang cukup tinggi. Namun secara keseluruhan, model regresi linear dapat digunakan dengan baik untuk memprediksi tinggi badan berdasarkan ukuran sepatu.

