

Nama Dosen : Teguh Iman Hermanto, M.Kom
 Mata Kuliah : Machine Learning 1
 Pembahasan : Simple Linear Regresion
 Pokok Pemb : - Mengenal Regresi Linier
 - Membangun Model Regresi Linier
 - Simulasi Algoritma Regresi Linier
 - Evaluasi Algoritma Regresi Linier

1. Mengenal Regresi Linier

a. Konsep Regresi Linier

Regresi linier berusaha menemukan garis lurus yang terbaik (disebut line of best fit) untuk mendekati data dan memprediksi nilai target. Garis ini dinyatakan dengan persamaan:

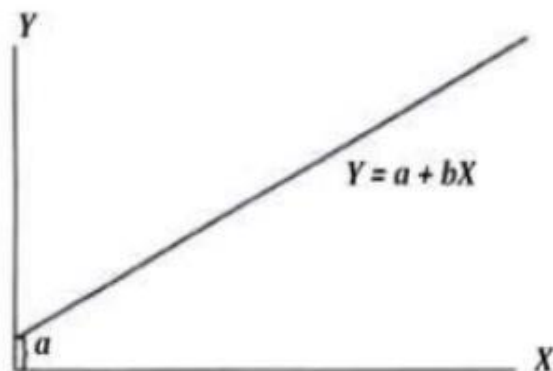
$$Y = a + bX$$

Y = Variabel Response atau Variabel Akibat (Dependent)

X = Variabel Predictor atau Variabel Faktor Penyebab (Independent)

a = Konstanta / intercept

b = Koefisien regresi (kemiringan/slope); besaran response yang ditimbulkan oleh predictor



b. Langkah-langkah regresi linier

- Membuat dataset

Nama	Kalori/hari(X)	Berat Badan (Y)
Adi	530	89
Rudi	300	48
Didi	358	56
Budi	510	72
Intan	302	54
Putu	300	42
Parta	387	60
Esti	527	85
Wike	415	63
Elis	512	74

- Menghitung nilai variabel

Nama	Kalori/hari(X)	X ²	Berat Badan (Y)	Y ²	XY
Adi	530	280900	89	7921	47170
Rudi	300	90000	48	2304	14400
Didi	358	128164	56	3136	20048
Budi	510	260100	72	5184	36720
Intan	302	91204	54	2916	16308
Putu	300	90000	42	1764	12600
Parta	387	149769	60	3600	23220
Esti	527	277729	85	7225	44795
Wike	415	172225	63	3969	26145
Elis	512	262144	74	5476	37888
Σ	4141	1802235	643	43495	279294

- Perhitungan koefisien a dan b

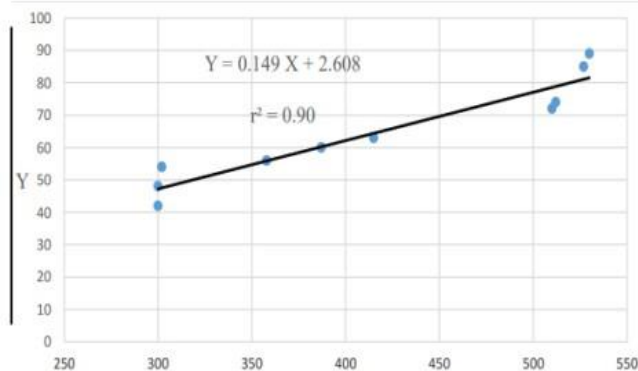
$$a = \frac{(\sum Y_i)(\sum X_i^2) - (\sum X_i)(\sum X_i Y_i)}{n \sum X_i^2 - (\sum X_i)^2}$$

$$= \frac{(643)(1802235) - (4141)(279294)}{10(1802235) - (4141)^2} = \frac{2280651}{874469} \cong 2,608$$

$$b = \frac{n (\sum X_i Y_i) - (\sum X_i)(\sum Y_i)}{n \sum X_i^2 - (\sum X_i)^2}$$

$$= \frac{10(279294) - (4141)(643)}{10(1802235) - (4141)^2} = \frac{130227}{874469} \cong 0,14892 \approx 0,149$$

- Persamaan garis regresi



Setelah didapat koefisien a dan b, maka persamaan garisnya adalah $Y = 2,608 + 0.149X$

- Hitung estimasi
Melakukan Estimasi terhadap variable predictor atau response

Estimasi berat badan mahasiswa jika asupan kalori adalah 600 kalori/hari :

$$Y = 2,608 + 0,149 X$$

$$Y = 2,608 + (0,149 * 600)$$

$$\text{Estimasi } Y = 92 \text{ Kilogram}$$

Estimasi asupan kalori mahasiswa, jika berat badan mahasiwa adalah 40 kilogram :

$$40 = 2,608 + 0,149 X$$

$$40 - 2,608 = 0,149 X$$

$$37,392 = 0,149 X$$

$$\text{Estimasi } X = 250.59 \text{ kalori/hari}$$

2. Membangun model regresi linier

- Import Library yang dibutuhkan

```
1 import pandas as pd
2 from sklearn.model_selection import train_test_split
3 from sklearn.linear_model import LinearRegression
4 from sklearn.metrics import mean_squared_error, r2_score
5 import matplotlib.pyplot as plt
6 import seaborn as sns
```

- Deskripsi Dataset

```
1 df = pd.read_csv('konsumsi-bbm.csv')
```

```
1 df.head()
```

```
1 df.info()
```

```
1 df.describe()
```

- Exploratory Data Analysis

```
1 plt.figure(figsize=(8, 6))
2 sns.scatterplot(x='Jarak_Tempuh', y='Konsumsi_BBM', data=df)
3 plt.title('Hubungan Jarak Tempuh dan Konsumsi BBM')
4 plt.xlabel('Jarak Tempuh')
5 plt.ylabel('Konsumsi BBM')
6 plt.show()
```

```
1 plt.figure(figsize=(8, 6))
2 sns.boxplot(data=df[['Jarak_Tempuh', 'Konsumsi_BBM']])
3 plt.title('Box Plot Jarak Tempuh dan Konsumsi BBM')
4 plt.show()
```

```
1 plt.figure(figsize=(8, 6))
2 sns.histplot(df['Jarak_Tempuh'], kde=True)
3 plt.title('Distribusi jarak tempuh')
4 plt.xlabel('jarak tempuh')
5 plt.ylabel('Frekuensi')
6 plt.show()
```

- Membangun Model Regresi

```
1 X = df[['Jarak_Tempuh']]
2 y = df['Konsumsi_BBM']
```



```
1 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```



```
1 model = LinearRegression()
```



```
1 model.fit(X_train, y_train)
```



```
1 y_pred = model.predict(X_test)
```

3. Simulasi algoritma regresi linier



```
1 jarak_baru = [60, 70, 80] # Contoh jarak tempuh baru
2 konsumsi_prediksi = model.predict([[jarak] for jarak in jarak_baru])
```



```
1 print('\nSimulasi Konsumsi BBM:')
2 for i in range(len(jarak_baru)):
3     print(f'Jarak Tempuh: {jarak_baru[i]}, Konsumsi BBM Prediksi: {konsumsi_prediksi[i]}')
```



```
1 plt.scatter(X_train, y_train, color='blue', label='Data Pelatihan')
2 plt.plot(X_train, model.predict(X_train), color='red', label='Garis Regresi')
3 plt.scatter(jarak_baru, konsumsi_prediksi, color='green', label='Data Baru (Prediksi)')
4
5 plt.xlabel('jarak')
6 plt.ylabel('konsumsi bbm')
7 plt.title('Grafik Regresi Linier')
8 plt.legend()
9 plt.show()
```

4. Evaluasi algoritma regresi linier



```
1 # Menghitung MSE dan R-squared
2 mse = mean_squared_error(y_test, y_pred)
3 r2 = r2_score(y_test, y_pred)
```



```
1 print(f'Mean Squared Error (MSE): {mse}')
2 print(f'R-squared (R2): {r2}')
```