

Nama : Destina Bkti Setyaningsih

Kelas : SD 04-01

NIM : 2311110018

Tugas 4 IPSD

Kasih komen penjelasan di code pertemuan 6, interpretasikan hasil R2 Score dan MSE nya

- a.

```
# Load Library
from sklearn.datasets import load_diabetes
from sklearn.linear_model import LinearRegression, Lasso, Ridge
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
```

Interpretasi : sekumpulan fungsi yang dirancang untuk mempermudah dan mempercepat proses pengembangan tertentu.

- b.

```
“load_diabetes()['target']”
```

Interpretasi : untuk menampilkan dataset diabetes dengan kolom target. Maksud target tersebut buat variable kolom target, tetapi kolom yang berisi nilai yang ingin di prediksi, seperti penyakit diabetes dalam 3 bulan terakhir, dan sebagainya.

- c.

```
“x, y = load_diabetes(return_X_y=True)”
```

Interpretasi : untuk menentukan fitur dan target dalam sebuah dataset.

- d.

```
“lr = LinearRegression()”
```

Interpretasi : untuk membuat model regresi linear dan menyimpan di dalam variable “lr”.

- e.

```
“len(load_diabetes().feature_names)”
```

Interpretasi : untuk menghitung jumlah elemen dari dataset diabetes, dan mengembalikan fitur kolom yang ada didalam dataset tersebut.

- f.

```
“lr.fit(x, y)
y_pred = lr.predict(x)”
```

interpretasi : fungsi (fit) yaitu untuk melatih model pada fitur x dan target y, sedangkan (pred) yaitu untuk membuat prediksi nilai target dengan fitur yang ada, lalu hasil nya akan disimpan dalam variable “y_pred”.

- g.

```
“print(r2_score(y, y_pred))”
```

Interpretasi : untuk mengprint atau mencetak hasil perhitungan dari nilai R^2 dan nilai dari y, dan akan disimpan dalam variable “y_pred”.

h. `print(mean_squared_error(y, y_pred))`

Interpretasi : untuk mencetak hasil perhitungan nilai Mean Squared Error (MSE).

Note:

- Mse -> adalah metrik yang digunakan untuk mengevaluasi model regresi dengan menggunakan nilai rata-rata.
- Mse -> ukuran yang menunjukkan rata-rata dari kuadrat selisih antara nilai yang diprediksi oleh model dan nilai yang sebenarnya.
- R^2 -> ukuran statistik yang digunakan untuk menentukan seberapa baik model regresi dengan cara menjelaskan variabilitas yang ada di dalam data tersebut.

Kodingan 1:

```
load_diabetes()['target']  
x, y = load_diabetes(return_X_y=True)  
lr = LinearRegression()  
len(load_diabetes().feature_names)  
lr.fit(x, y)  
y_pred = lr.predict(x)  
print(r2_score(y, y_pred))  
print(mean_squared_error(y, y_pred))
```

Output :

```
R2_score = 0.5177484222203499  
Mean_squared_error = 2859.69634758675
```

Interpretasi :

- Nilai dari R^2 score sebesar 51%, hal ini menunjukkan bahwa variable target bisa dijelaskan dengan variable fitur, yang artinya model regresi tersebut menjelaskan sedikit > setengah variasi dari data tersebut.
- Nilai MSE sebesar 28%, hal ini yang membuat model tersebut memiliki kesalahan prediksi rata-rata yang cukup besar. Jika nilai MSE nya tinggi maka bisa untuk memperbaiki model, seperti menambah fitur, membuat model yang berbeda, dsb..

Kodingan 2:

```
# Hold Out

# train -> validation -> test

from sklearn.model_selection import train_test_split

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)

lr = LinearRegression()

lr.fit(x_train, y_train)

y_pred = lr.predict(x_test)

print(r2_score(y_test, y_pred))

print(mean_squared_error(y_test, y_pred))
```

Output :

R2_score = 0.4526027629719195

Mean_squared_error = 2900.193628493482

Interpretasi :

- Nilai dari R2_score sebesar 45%, hal ini menunjukkan bahwa variable target bisa dijelaskan dengan variable fitur, yang artinya model regresi tersebut menjelaskan sedikit < setengah variasi dari data tersebut.
- Nilai MSE sebesar 29%, hal ini yang membuat model tersebut memiliki kesalahan prediksi rata-rata yang cukup besar dan jika ingin membandingkan model bisa menggunakan model lainnya. Jika nilai MSE nya tinggi maka bisa untuk memperbaiki model, seperti menambah fitur, membuat model yang berbeda, dsb..