

Nama : Sulis Afryanti Lubis

NPM : 4522210024

Link Google Collab :

<https://colab.research.google.com/drive/16DeGLMCvZEaErPxAT7jj5DeK7HFDvtLf?usp=sharing>

Langkah 1

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

```
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

```
import pandas as pd

# Path ke file dataset di Google Drive
file_path = '/content/drive/MyDrive/foodstruct_nutritional_facts.csv'

# Baca dataset
df = pd.read_csv(file_path)
```

kode diatas berisikan data set nutritional yang dimasukan kedalam google drive, dan dapat membaca data set yang ada sesuai perintah.

```
df.head()
```

Kode diatas digunakan untuk menampilkan data yang ada

```
# Lihat data duplikat
duplicates = df.duplicated()
print(f'Jumlah data duplikat: {duplicates.sum()}')
```

```
➞ Jumlah data duplikat: 0
```

Kode diatas digunakan untuk melihat data yang duplikat

```
# Lihat jumlah missing values per kolom
print(df.isnull().sum())
# Menghapus baris yang memiliki missing values
# Jika Anda ingin menghapus semua baris dengan missing values
df = df.dropna()
```

Kode diatas digunakan untuk melihat missing value, pada data diatas missing value nya 0

```
# Baca file CSV dari lokasi yang ditentukan
df = pd.read_csv('/content/drive/MyDrive/foodstruct_nutritional_facts.csv')

# Tampilkan hanya data yang dipilih
print(df[['Category Name', 'Carbs']])
```

Kode diatas digunakan untuk memilih kolom yang akan di proses, pada data diatas memilih data Calories dan Carbs

```
[20] df.shape #mengetahui jumlah kolom dan baris
```

```
➞ (1174, 59)
```

Kode diatas digunakan untuk mengetahui jumlah kolom dan baris yang ada pada data

```
plt.scatter(df.Calories, df.Carbs, color='blue')
plt.xlabel("Calories")
plt.ylabel("Carbs")
plt.show()
```

Kode diatas digunakan untuk menampilkan digunakan untuk membuat grafik scatter plot, yaitu grafik yang menampilkan titik-titik pada bidang kartesius.

```

# Asumsikan Calories sebagai data dependent dan Carbs sebagai data Independent
X = df[['Carbs']] # Independent variable (Carbs)
y = df['Calories'] # Dependent variable (Calories)

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Create and train the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions on the test set
y_pred = model.predict(X_test)

# Print the regression coefficients
print("Intercept:", model.intercept_)
print("Coefficient:", model.coef_[0]) # Assuming only one independent variable

# Visualize the regression line
plt.scatter(X_test, y_test, color='blue', label='Actual Data') # Scatter plot of actual data
plt.plot(X_test, y_pred, color='red', linewidth=2, label='Regression Line') # Regression line
plt.title('Linear Regression: Calories vs. Carbs')
plt.xlabel('Carbs')
plt.ylabel('Calories')
plt.legend()
plt.show()

```

Kode diatas digunakan untuk memproses regresi linear.