

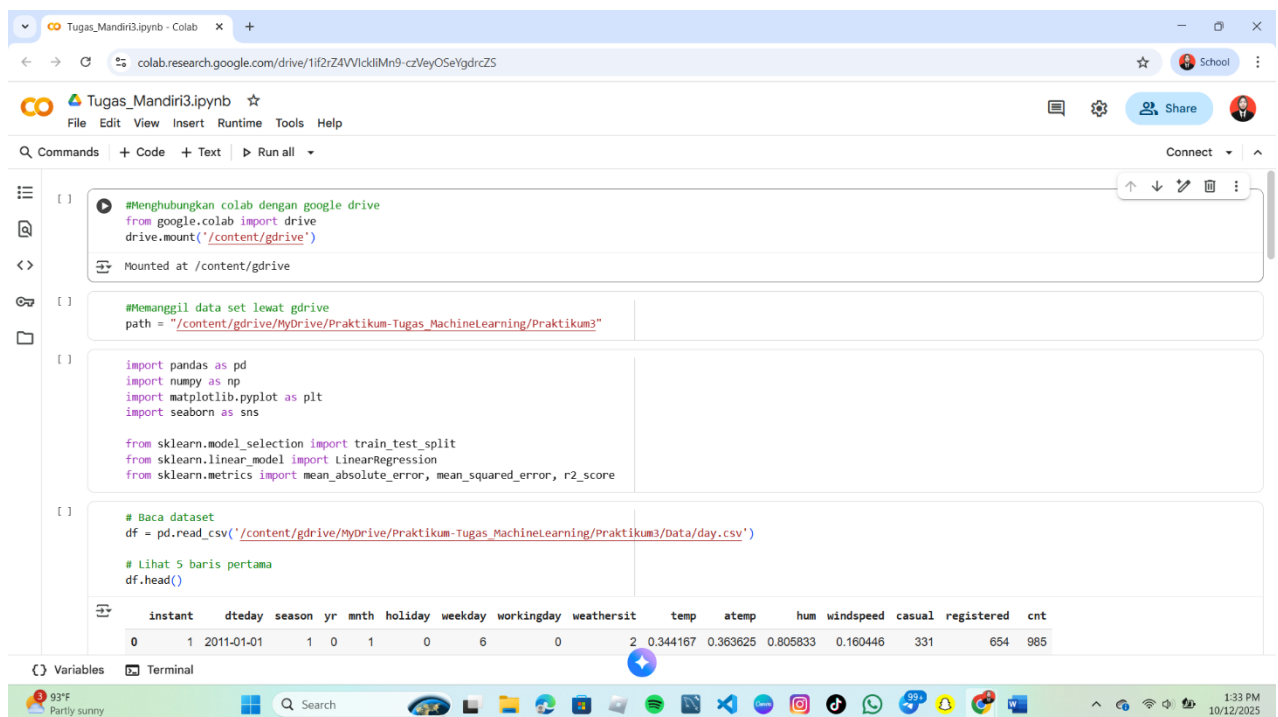
# Tugas 3: Tugas Praktikum Mandiri 3 – Machine Learning

Yurida Yahsyia 1 - 0110224100 <sup>1</sup>\*

<sup>1</sup> Teknik Informatika, STT Terpadu Nurul Fikri, Depok

\*E-mail: [0110224100@student.nurulfikri.ac.id](mailto:0110224100@student.nurulfikri.ac.id) – email mahasiswa 1

## 1. Latihan Mandiri 3



```
#Menghubungkan colab dengan google drive
from google.colab import drive
drive.mount('/content/gdrive')

Mounted at /content/gdrive

#Memanggil data set lewat gdrive
path = "/content/gdrive/MyDrive/Praktikum-Tugas_MachineLearning/Praktikum3"

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

# Baca dataset
df = pd.read_csv('/content/gdrive/MyDrive/Praktikum-Tugas_MachineLearning/Praktikum3/Data/day.csv')

# Lihat 5 baris pertama
df.head()
```

| instant | dteday | season     | yr | mnth | holiday | weekday | workingday | weathersit | temp | atemp    | hum      | windspeed | casual   | registered | cnt |     |
|---------|--------|------------|----|------|---------|---------|------------|------------|------|----------|----------|-----------|----------|------------|-----|-----|
| 0       | 1      | 2011-01-01 | 1  | 0    | 1       | 0       | 6          | 0          | 2    | 0.344167 | 0.363625 | 0.805833  | 0.160446 | 331        | 654 | 985 |

### 1. - from google.colab import drive

Memanggil modul bawaan Colab.

### - drive.mount('/content/gdrive')

Membuat jembatan supaya Colab bisa membaca file yang ada di Drive, dan semua file bisa diakses lewat folder /content/gdrive/MyDrive/.

- **Output:** Menunjukkan bahwa Drive telah berhasil terhubung dan dapat diakses melalui jalur (path) /content/gdrive.

### 2. path = "/content/gdrive/MyDrive/Praktikum/Praktikum3"

path = alamat menuju folder Praktikum2 di dalam Google Drive.

### 3. import pandas as pd

import numpy as np

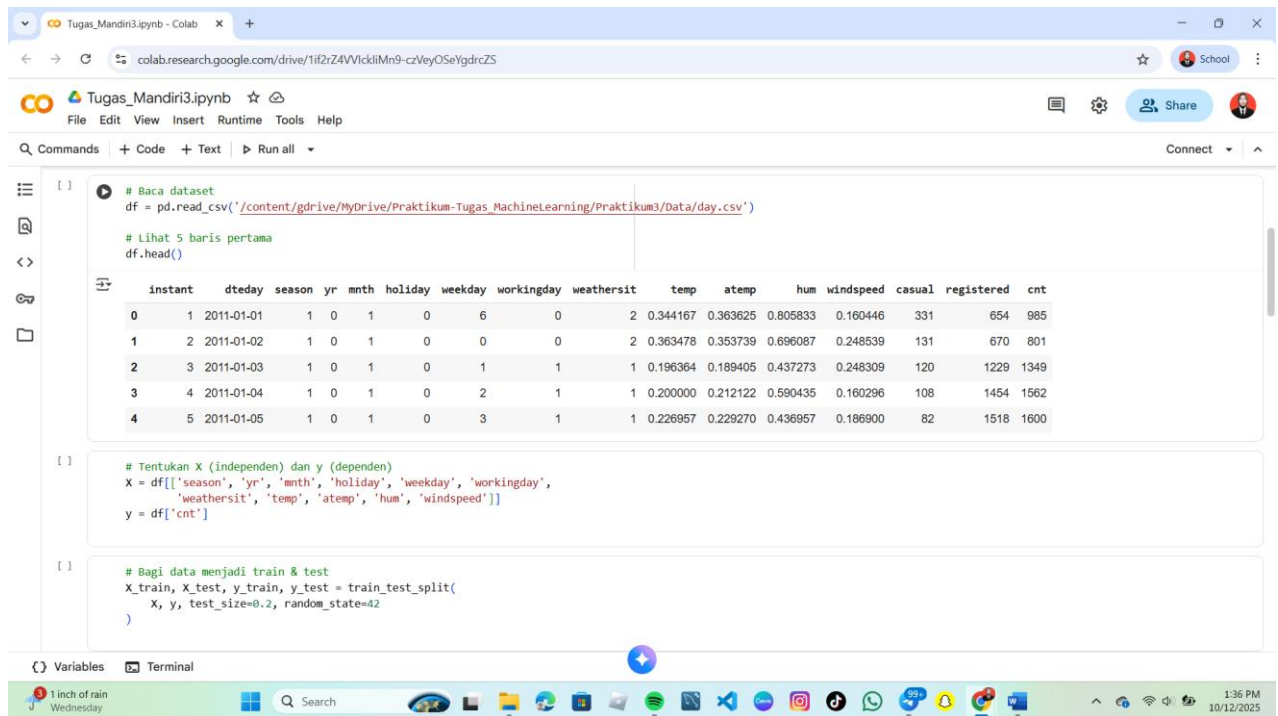
import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

**from sklearn.metrics import mean\_absolute\_error, mean\_squared\_error, r2\_score**  
Mengimpor semua library Python yang diperlukan untuk analisis data, visualisasi, dan pembangunan model Machine Learning.



The screenshot shows a Google Colab notebook titled 'Tugas\_Mandiri3.ipynb'. The code in the first cell reads a CSV file from Google Drive and displays the first five rows of the dataset. The second cell defines independent variables (X) and the dependent variable (y). The third cell splits the data into training and testing sets.

```
[1] # Baca dataset
df = pd.read_csv('/content/gdrive/MyDrive/Praktikum-Tugas_MachineLearning/Praktikum3/Data/day.csv')

# Lihat 5 baris pertama
df.head()

[2] # Tentukan X (independen) dan y (dependen)
X = df[['season', 'yr', 'mnth', 'holiday', 'weekday', 'workingday', 'weathersit', 'temp', 'atemp', 'hum', 'windspeed']]
y = df['cnt']

[3] # Bagi data menjadi train & test
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)
```

|   | instant | dteday     | season | yr | mnth | holiday | weekday | workingday | weathersit | temp     | atemp    | hum      | windspeed | casual | registered | cnt  |
|---|---------|------------|--------|----|------|---------|---------|------------|------------|----------|----------|----------|-----------|--------|------------|------|
| 0 | 1       | 2011-01-01 | 1      | 0  | 1    | 0       | 6       | 0          | 2          | 0.344167 | 0.363625 | 0.805833 | 0.160446  | 331    | 654        | 985  |
| 1 | 2       | 2011-01-02 | 1      | 0  | 1    | 0       | 0       | 0          | 2          | 0.363478 | 0.353739 | 0.696087 | 0.248539  | 131    | 670        | 801  |
| 2 | 3       | 2011-01-03 | 1      | 0  | 1    | 0       | 1       | 1          | 1          | 0.196364 | 0.189405 | 0.437273 | 0.248309  | 120    | 1229       | 1349 |
| 3 | 4       | 2011-01-04 | 1      | 0  | 1    | 0       | 2       | 1          | 1          | 0.200000 | 0.212122 | 0.590435 | 0.160296  | 108    | 1454       | 1562 |
| 4 | 5       | 2011-01-05 | 1      | 0  | 1    | 0       | 3       | 1          | 1          | 0.226957 | 0.229270 | 0.436957 | 0.186900  | 82     | 1518       | 1600 |

1. **df = pd.read\_csv('/content/gdrive/MyDrive/Praktikum-Tugas Machinelearning/Praktikum3/Data/day.csv')**  
**df.head()**

Memuat file dataset day.csv ke dalam DataFrame Pandas bernama df dan menampilkan lima baris teratas untuk memverifikasi data. Jalur (path): Menunjukkan lokasi file di Google Drive.

- **Output:** Tabel ini menampilkan 5 baris pertama data. Setiap kolom adalah variabel yang akan digunakan untuk pemodelan

2. **X = df[['season', 'yr', 'mnth', 'holiday', 'weekday', 'workingday', 'weathersit', 'temp', 'atemp', 'hum', 'windspeed']]**

**y = df['cnt']**

Memisahkan data menjadi variabel independen (fitur) dan variabel dependen (target).

3. **X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)**  
**)**

Membagi dataset yang sudah dipisahkan (X dan y) menjadi dua bagian utama:

1. Data Training (X\_train, y\_train): Digunakan untuk melatih model.
2. Data Testing (X\_test, y\_test): Digunakan untuk menguji seberapa baik kinerja model pada data yang belum pernah dilihat.

```
[ ]  
# Bagi data menjadi train & test  
X_train, X_test, y_train, y_test = train_test_split(  
    X, y, test_size=0.2, random_state=42  
)  
  
[ ]  
# Buat dan latih model Linear Regression  
model = LinearRegression()  
model.fit(X_train, y_train)  
  
LinearRegression()  
  
[ ]  
# Prediksi  
y_pred = model.predict(X_test)  
  
# Hitung metrik evaluasi  
mae = mean_absolute_error(y_test, y_pred)  
mse = mean_squared_error(y_test, y_pred)  
rmse = np.sqrt(mse)  
r2 = r2_score(y_test, y_pred)  
  
print("MAE :", mae)  
print("MSE :", mse)  
print("RMSE:", rmse)  
print("R² Score:", r2)
```

**1. model = LinearRegression()  
model.fit(X\_train, y\_train)**

**output:** Membuat dan melatih model Multiple Linear Regression.

```
[ ]  
# Prediksi  
y_pred = model.predict(X_test)  
  
# Hitung metrik evaluasi  
mae = mean_absolute_error(y_test, y_pred)  
mse = mean_squared_error(y_test, y_pred)  
rmse = np.sqrt(mse)  
r2 = r2_score(y_test, y_pred)  
  
print("MAE :", mae)  
print("MSE :", mse)  
print("RMSE:", rmse)  
print("R² Score:", r2)  
  
MAE : 617.3930656443376  
MSE : 691035.0082022651  
RMSE : 831.2851545662686  
R² Score: 0.8276670090367212  
  
[ ]  
# Lihat koefisien tiap fitur  
coef_df = pd.DataFrame({  
    "Feature": X.columns,  
    "Coefficient": model.coef_  
})  
print(coef_df)  
  
Feature Coefficient  
season 524.722536
```

**1. y\_pred = model.predict(X\_test)  
mae = mean\_absolute\_error(y\_test, y\_pred)  
mse = mean\_squared\_error(y\_test, y\_pred)  
rmse = np.sqrt(mse)**

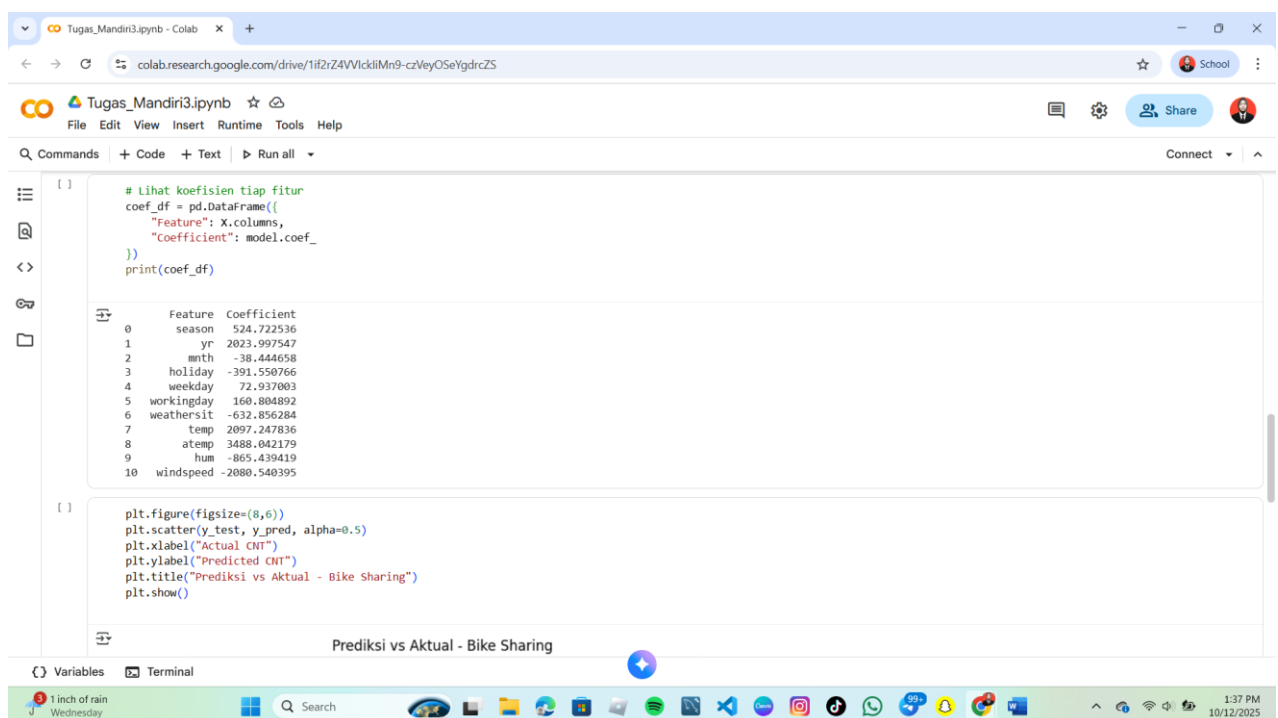
```
r2 = r2_score(y_test, y_pred)
```

```
print("MAE:", mae)
print("MSE:", mse)
print("RMSE:", rmse)
print("R2 Score:", r2)
```

Menggunakan model yang sudah dilatih untuk membuat prediksi pada data uji ( $X_{\text{test}}$ ) dan kemudian menghitung metrik untuk menilai seberapa baik kinerja model.

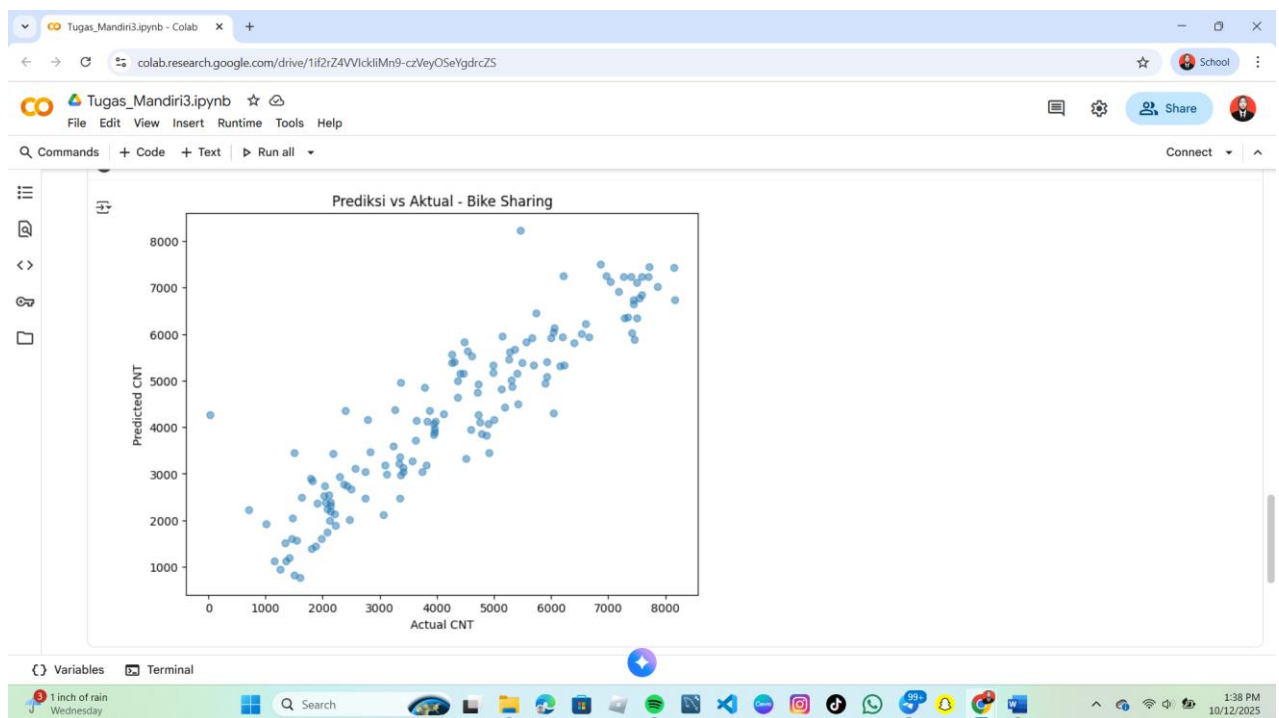
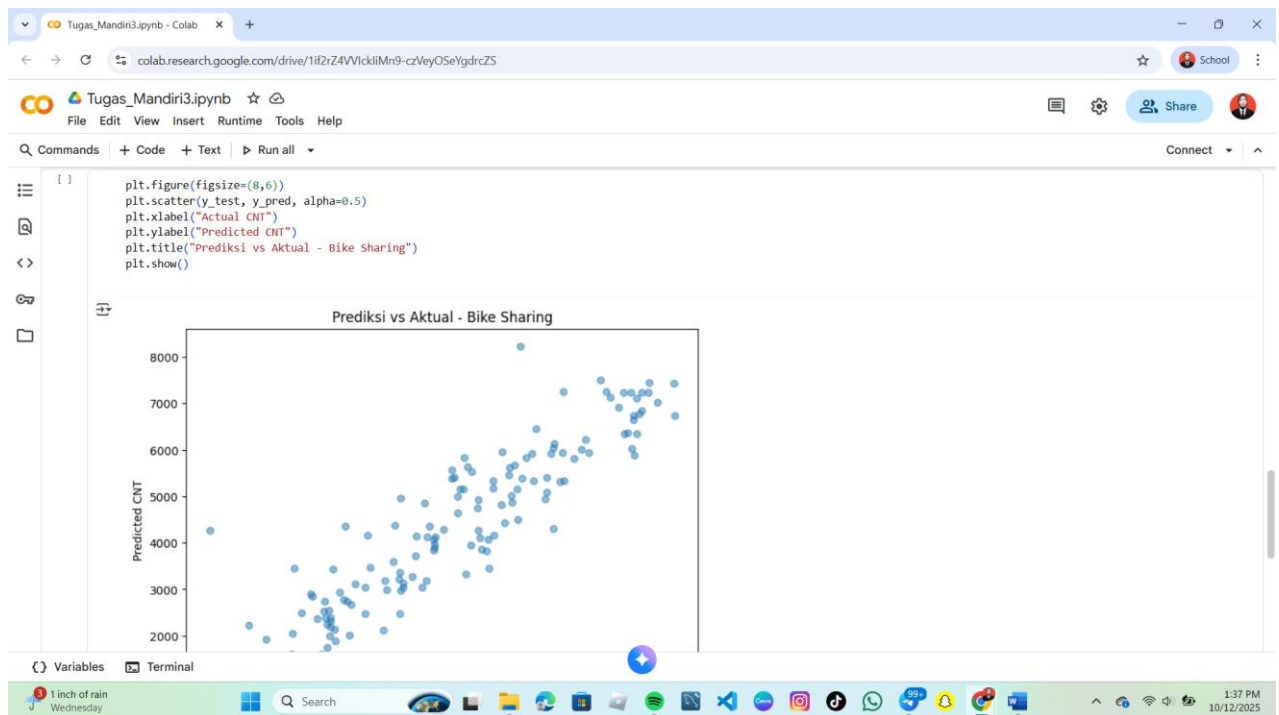
- **Output:**

- **MAE** (Mean Absolute Error): Ini adalah rata-rata kesalahan prediksi model.
- **MSE** (Mean Squared Error): Rata-rata kuadrat dari kesalahan. Nilai yang besar karena satuan dikuadratkan, membuatnya sensitif terhadap kesalahan besar (*outliers*).
- **RMSE** (Root Mean Squared Error): Akar kuadrat dari MSE. Nilainya memiliki satuan yang sama dengan target (cnt).
- **R2 Score**: Nilai yang mendekati 1 menunjukkan model memiliki daya prediksi yang cukup baik.



1. **coef\_df = pd.DataFrame:** Kode ini membuat DataFrame baru (`coef_df`) untuk menyajikan koefisien model secara terstruktur dan mudah dibaca.
2. **"Feature": X.columns:** Mengambil nama-nama fitur (variabel independen) yang digunakan dalam pelatihan model.
3. **"Coefficient": model.coef\_:** Mengambil nilai koefisien (bobot) yang dipelajari oleh model LinearRegression untuk setiap fitur. Koefisien ini menunjukkan hubungan dan bobot setiap fitur terhadap variabel target (cnt).
4. **print(coef\_df):** Mencetak DataFrame tersebut.

**Output:** Tabel menunjukkan dampak setiap fitur terhadap jumlah penyewaan sepeda (cnt).



1. **plt.figure(figsize=(8,6))**: Menentukan ukuran grafik yang akan dibuat, yaitu 8×6 inci.
2. **plt.scatter(y\_test, y\_pred, alpha=0.5)**: Membuat plot sebar (scatter plot). Sumbu X: Menggunakan nilai aktual dari data uji (y\_test). Sumbu Y: Menggunakan nilai prediksi dari model (y\_pred). alpha=0.5 membuat titik sedikit transparan agar tumpukan titik terlihat.
3. **plt.xlabel("Actual CNT")**: Memberi label pada sumbu X.
4. **plt.ylabel("Predicted CNT")**: Memberi label pada sumbu Y.
5. **plt.title**: Memberi judul pada grafik.
6. **plt.show()**: Menampilkan grafik.

**Output:** Pada grafik, titik-titik sangat rapat dan membentuk pola linear diagonal yang jelas. Ini menunjukkan bahwa:

- Ada korelasi positif yang kuat antara prediksi model dan nilai sebenarnya.
  - Model Regresi Linear telah berhasil menangkap hubungan antara fitur (X) dan target (y) dengan baik.
- **Kesimpulan:** Visualisasi ini memvalidasi skor  $R^2$  tinggi yang diperoleh sebelumnya ( $\approx 0.83$ ), menegaskan bahwa model efektif dalam memprediksi jumlah penyewaan sepeda.

**Kesimpulan hasil implementasi algoritma:** Secara keseluruhan, model Multiple Linear Regression berhasil diimplementasikan dan menunjukkan kinerja yang baik ( $R^2 > 0.8$ ). Model ini tidak hanya dapat memprediksi permintaan sepeda harian dengan tingkat akurasi yang dapat diandalkan, tetapi juga memberikan wawasan yang jelas mengenai faktor-faktor utama (terutama suhu dan tren tahunan) yang memengaruhi perilaku penyewaan sepeda.

**LINK GITHUB UPLOAD TUGAS :** <https://github.com/Yurida26/Machine-Learning/tree/ca132bbf3b1d5d7f9002445317330f3dd63c4d40/Praktikum3>

