Tugas 3: Membuat Model Prediksi Jumlah Penyewaan Sepeda (Bike Sharing Dataset)

**Ahmad Azka Ridha - 0110222062 1**

1 Teknik Informatika, STT Terpadu Nurul Fikri, Depok

\*E-mail: [ahma22062ti@student.nurulfikri.ac.id](mailto:ahma22062ti@student.nurulfikri.ac.id)

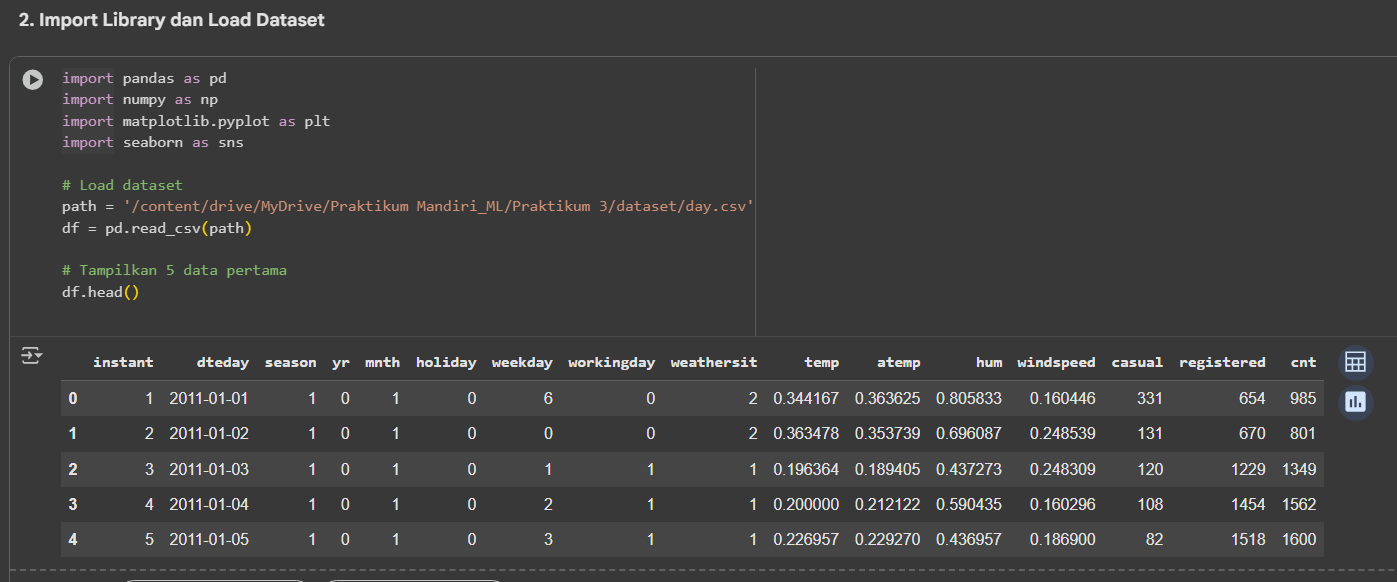
1. Praktikum Mandiri

1.1 menghubungkan dengan Google Drive



Langkah ini digunakan agar Colab dapat mengakses file yang tersimpan di Google Drive.  
Dataset yang telah diunduh dari Kaggle harus disimpan di dalam Drive.

1.2 Import Library dan Load Dataset

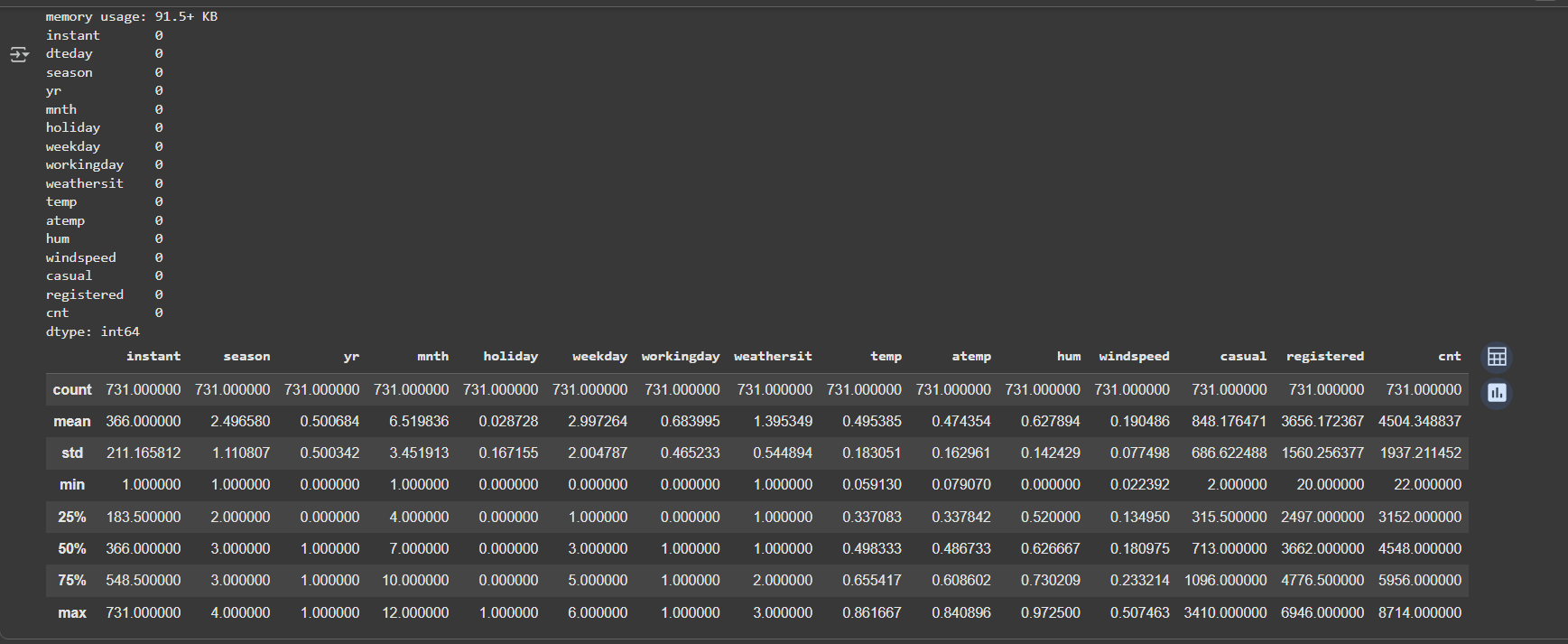
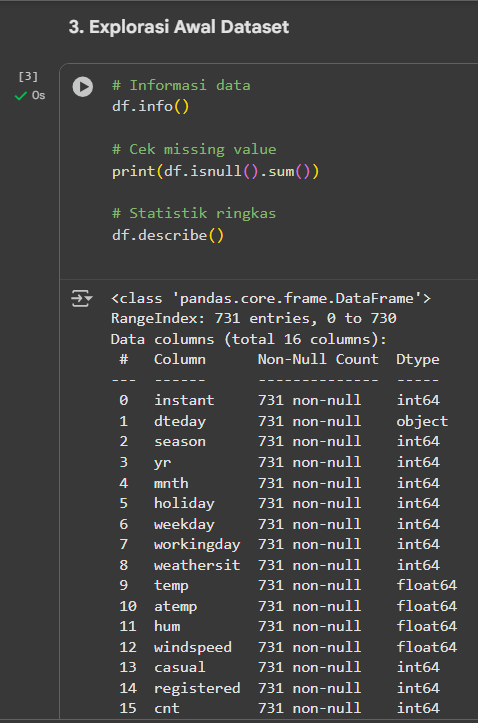


 Library seperti **pandas** dan **numpy** digunakan untuk membaca serta memproses data.

 **matplotlib** dan **seaborn** digunakan untuk visualisasi grafik.

 Dataset “day.csv” dibaca dan ditampilkan 5 baris pertama untuk melihat struktur datanya.

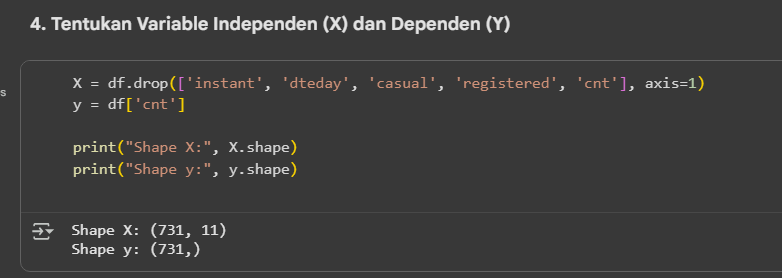
1.3 Explorasi Awal Dataset



* df.info menampilkan tipe data tiap kolom (numerik, kategorikal, dsb).
* df.isnull.sum memeriksa apakah terdapat nilai kosong (missing value).
* df.describe menampilkan statistik deskriptif seperti rata-rata, nilai minimum, dan maksimum.

Hasilnya menunjukkan bahwa dataset bersih dan tidak memiliki nilai kosong.

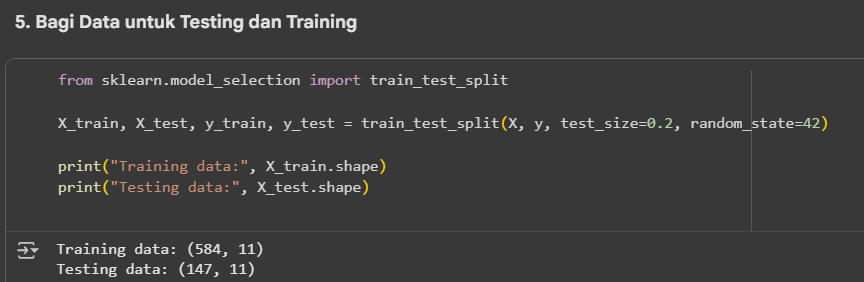
1.4 Menentukan Variabel Independen (X) dan Dependen (Y)



* **Variabel dependen (Y)** adalah cnt, yaitu jumlah total penyewaan sepeda.
* **Variabel independen (X)** adalah kolom-kolom lain yang memengaruhi cnt, **kecuali:**
  + instant → hanya nomor urut
  + dteday → tanggal
  + casual dan registered → sudah merupakan komponen dari cnt

Jadivariabel X yang digunakan adalah:  
season, yr, mnth, holiday, weekday, workingday, weathersit, temp, atemp, hum, windspeed

*1.5 Membagi Data Menjadi Training dan Testing*

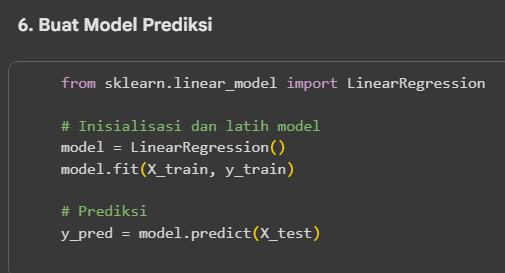
**

 Dataset dibagi menjadi dua bagian:

* **Data Training (80%)** → untuk melatih model.
* **Data Testing (20%)** → untuk menguji model.

 Parameter random\_state=42 memastikan pembagian data selalu sama setiap kali dijalankan (reproducible).

*1.6 Membuat Model Prediksi (Linear Regression)*

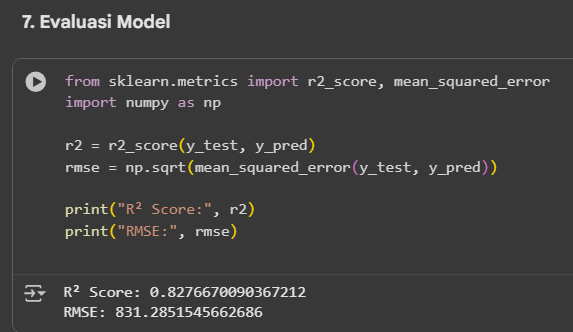
**

 Digunakan algoritma **Linear Regression**, yaitu model sederhana yang memprediksi hubungan linier antara variabel input dan output.

 model.fit() → melatih model dengan data training.

 model.predict() → menghasilkan prediksi pada data testing.

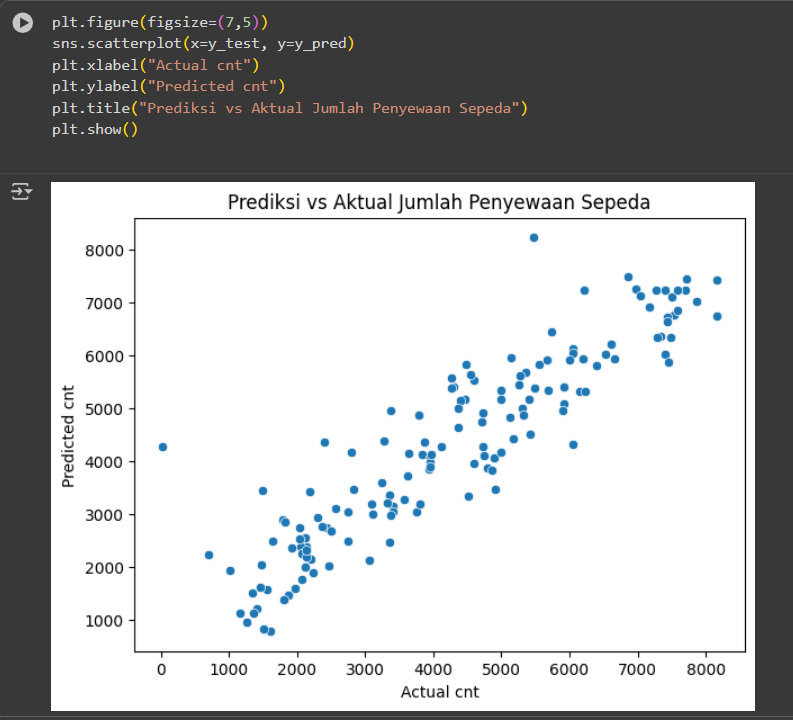
*1.7 Evaluasi Model*

**

 **R² Score** menunjukkan seberapa besar variasi data target dapat dijelaskan oleh model. Nilai mendekati **1** berarti model bagus.

 **RMSE (Root Mean Square Error)** menunjukkan rata-rata error prediksi. Semakin kecil nilai RMSE, semakin akurat model.

*1.8 Visualisasi Hasil*



Grafik ini memperlihatkan perbandingan antara nilai aktual (y\_test) dan nilai prediksi (y\_pred).  
Semakin dekat titik-titik pada garis diagonal, semakin akurat model dalam memprediksi data.