Tugas 3: Praktikum & Praktikum Mandiri 3

**Pandu Linggar Kumara - 0110221277,   
Link GitHub - https://github.com/PanduLgg/M\_Learning.git**

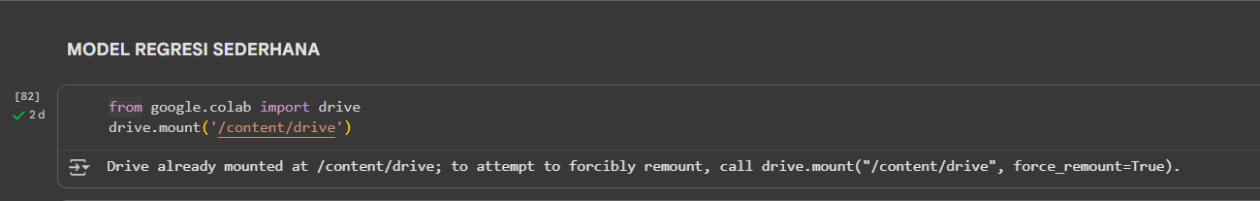
1 Teknik Informatika, STT Terpadu Nurul Fikri, Depok

\*E-mail: [pandulinggar1@gmail.com](mailto:pandulinggar1@gmail.com)

**Abstract.** Eksperimen dimulai dari studi kasus sederhana, yaitu prediksi berat badan berdasarkan tinggi dan umur, kemudian dilanjutkan dengan penerapan pada dataset nyata yaitu Bike Sharing Dataset dari Kaggle. Analisis dilakukan menggunakan metode Ordinary Least Squares (OLS) dengan bantuan pustaka statsmodels pada Python untuk memodelkan hubungan antara variabel dependen (cnt, total penyewaan sepeda) dan beberapa variabel independen seperti suhu, kelembapan, musim, serta kecepatan angin. Proses pemodelan mencakup tahap persiapan data, pemilihan fitur, pembuatan model, serta evaluasi menggunakan metrik statistik seperti koefisien determinasi (R²), koefisien regresi, dan nilai signifikansi (p-value). Hasil analisis menunjukkan bahwa regresi linear berganda mampu menjelaskan variasi jumlah penyewaan sepeda berdasarkan faktor lingkungan dan musiman secara efektif.

1. Connecting Google Colab & Drive

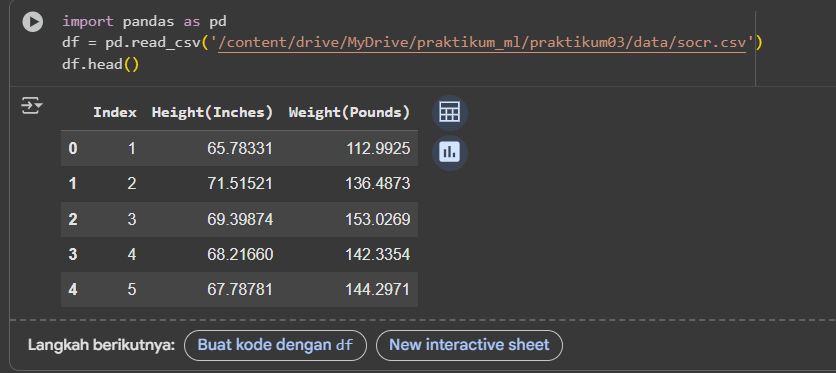
1.1 Menghubungkan lingkungan Google Colab dengan akun Google Drive

Sel ini berfungsi untuk menghubungkan lingkungan Google Colab dengan akun Google Drive

**Gambar 1.1.** Proses ini hanya perlu dilakukan satu kali per sesi.

1.2 Memanggil Data set dari Gdrive dan Membaca file .CSV menggunakan Pandas

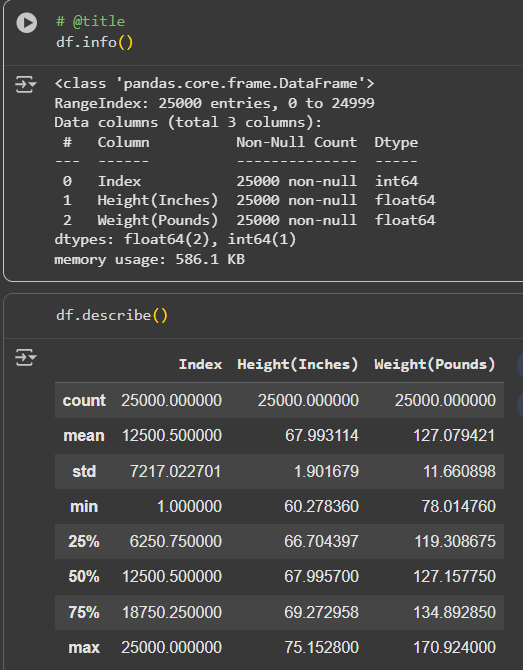
Sel ini menggunakan library Pandas untuk membaca file data, yang diinginkan



**Gambar 1.2.** Proses ini hanya perlu dilakukan satu kali per sesi.

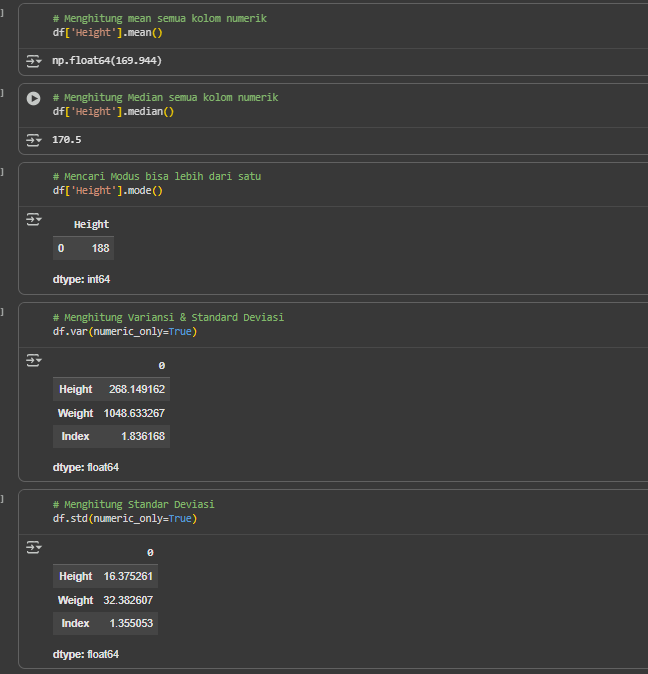
1.3 Mencari informasi data yang ada pada file

Sel ini menampilkan informasi yang ada di dalam file dari mulai tipe data nama kolom, dsb.



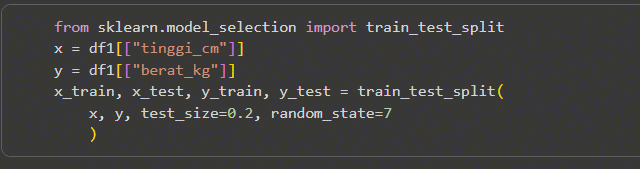
**Gambar 1.3.** Mencari info data pada file

1.4 Data Pre-processing  
kita membuat DataFrame baru bernama df1 yang berisi data tinggi dan berat badan yang telah dikonversi dari satuan inci dan pon menjadi sentimeter dan kilogram



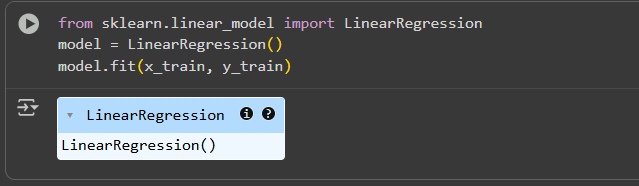
**Gambar 1.4.** Tabel tinggi dan berat dalam satuan metrik yang lebih mudah dibaca

*1.5 Membagi dataset untuk Training dan Test  
Sel ini membagi variabel dependen dan independen dibagi menjadi 2 yaitu train dan test*

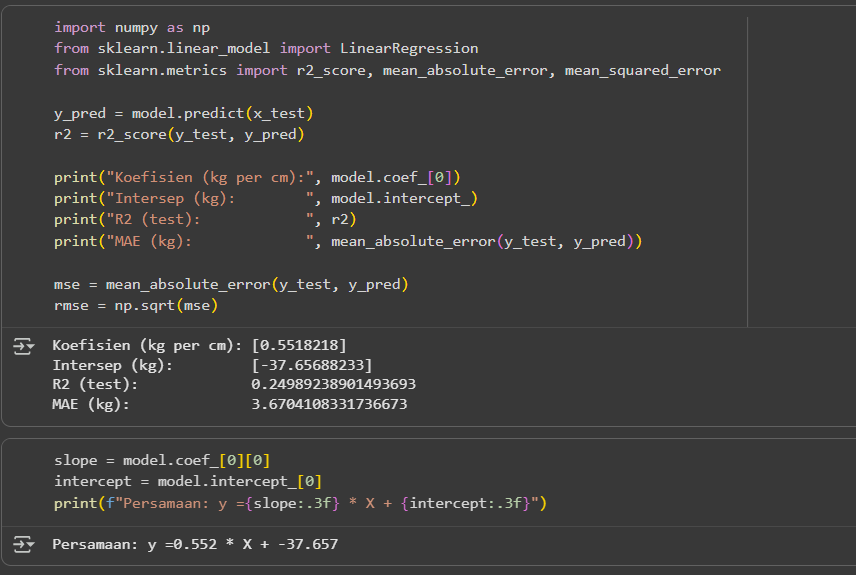


**Gambar 1.5.** Variabel dependen dan independen dibagi menjadi 2

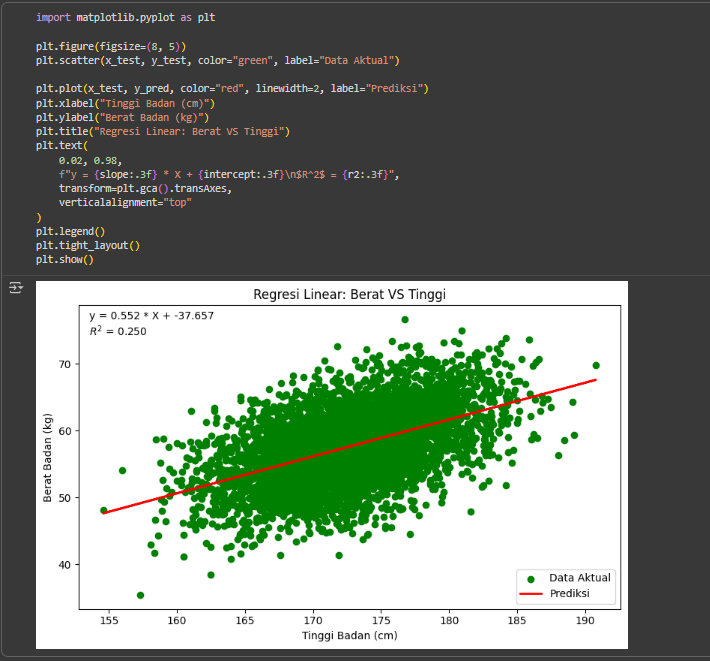
*1.6 Melatih model dengan menggunakan regresi linear   
Untuk menerapkan regresi linear kita perlu menggunakan LinearRegression milik scikit-learn*

  
**Gambar 1.6.** LinearRegression

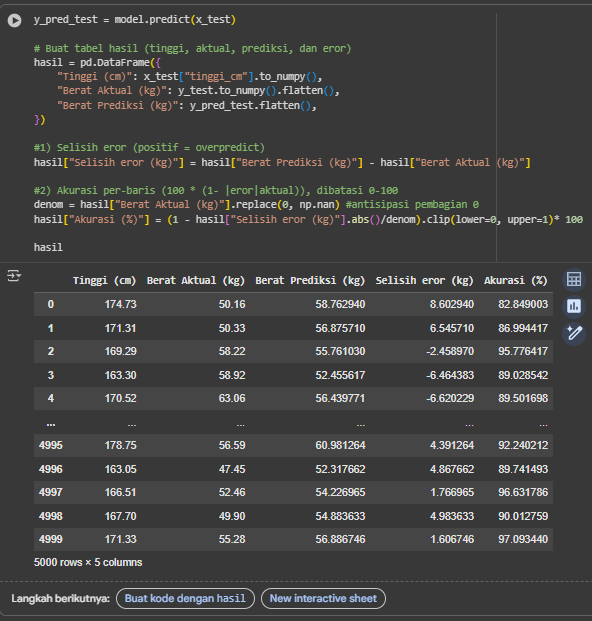
*1.7 Evaluasi Model dan Menghubungkan 2 Variabel  
Menilai seberapa baik model Multiple Linear Regression bekerja dan membentuk persamaan regresi*

**Gambar 1.7.** Evaluasi Model menunjukkan Koef, Intercept, R2, MAE

*1.8 Visualisasi Regresi   
Menghubungkan Berat dan TInggi badan*

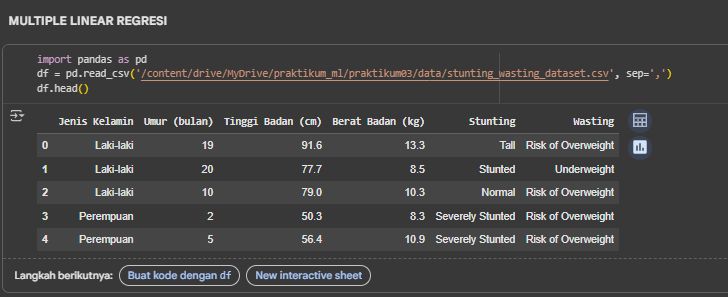
**Gambar 1.8.** grafik menunjukkan bahwa sekitar 25% variasi berat badan dapat dijelaskan oleh tinggi badan

*1.9 Model Prediksi   
Menghitung prediksi dan akurasi data hasil prediksi per-baris*

   
**Gambar 1.8.** Tabel hasil menampilkan 4 kolom utama

2. MULTIPLE LINEAR REGRESI

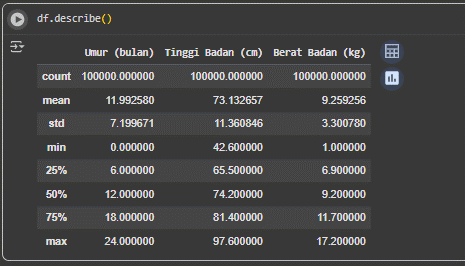
2.1 Membaca Data   
Membaca data stunting\_wasting\_dataset.csv menggunakan library pandas dan menggunakan sep untuk memisahkan koma



**Gambar 2.1.** Menampilkan Kolom data

2.2 Menghitung statistik deskriptif pada kolom numeric dengan describe

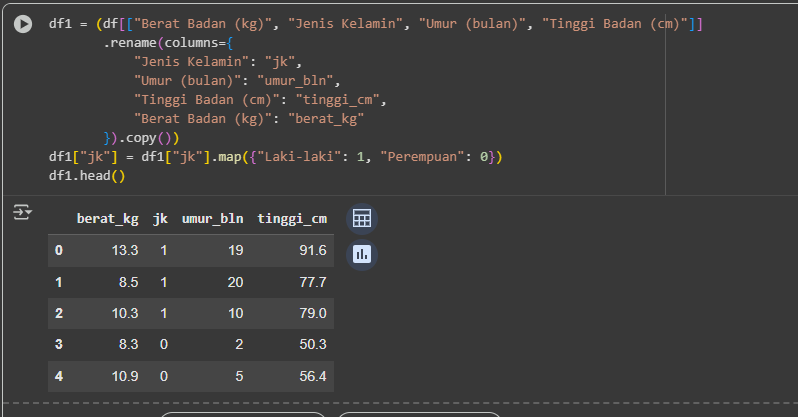
*Metode .describe() secara otomatis menghitung statistik deskriptif dasar untuk semua kolom numerik.Ini memberikan gambaran cepat tentang distribusi data*



**Gambar 2.2.** Tabel yang berisi metrik-metrik

2.3 Data Pre-processing

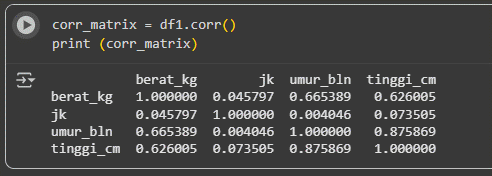
*mengcopy variabel df dan hanya memakai pada variabel/kolom Berat Badan (kg), Jenis Kelamin, Umur (bulan), Tinggi Badan (cm) dengan menamai dengan df1. Namun dilanjutkan dengan mengubah nama kolom dengan function rename.*



**Gambar 2.3.** Kolom hasil filter data yang diinginkan

2.4 Analisis Korelasi

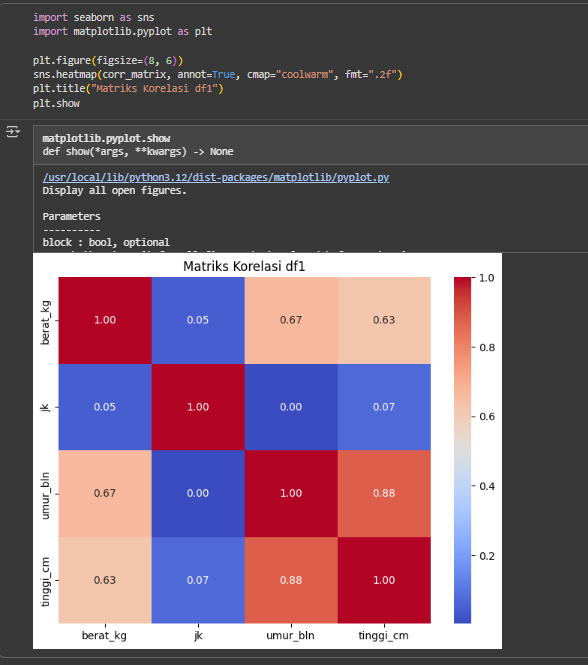
*Tahapan ini melakukan evaluasi terhadap beberapa variabel indepen (x) untuk memprediksi variable dependen (y).*



**Gambar 2.4.** Evaluasi variable

2.5 Heatmap

*Memilih variabel yang dominan untuk dimasukan dalam model prediksi dengan menggunakan fungsi koefisien korelasi .corr(), untuk selanjutnya divisulisasikan dalam bentuk grafik Heatmap.*

  
**Gambar 2.5.** Heatmap Variabel yang paling berperngaruh

Hasil analisis koefisien korelasi menunjukkan bahwa variabel yang paling berpengaruh dalam prediksi

berat badan balita adalah sebagai berikut:

1. Umur: 0.67 → berpengaruh dominan
2. Tinggi: 0.63 → berpengaruh dominan
3. Jenis Kelamin: 0.05 → tidak berpengaruh signifikan

Berdasarkan hasil korelasi tersebut, variabel yang digunakan untuk membangun model regresi adalah:

Variabel independen (X):

o X₁ = Umur

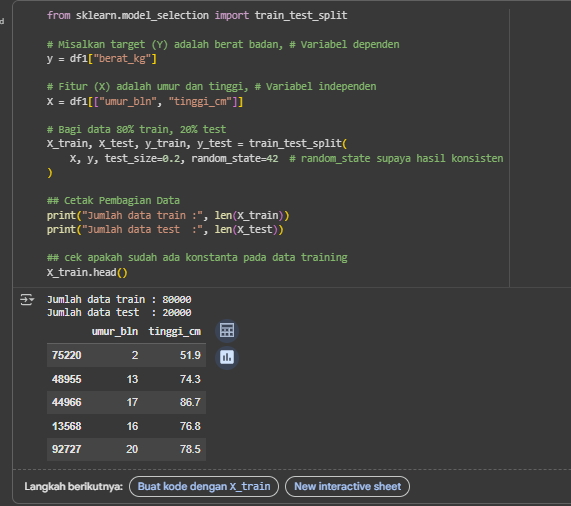
o X₂ = Tinggi

Variabel dependen (Y):

o Y = Berat

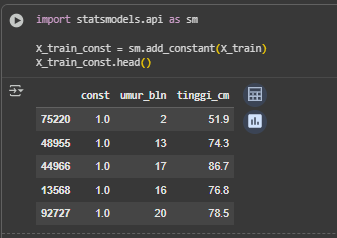
2.6 Membagi dataset untuk Training dan Test

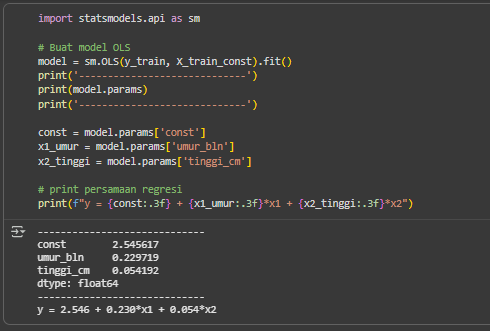
*Pada tahapan ini membagi dataset menjadi 80% data training dan 20% data testing*

  
**Gambar 2.6.** Membagi dataset menjadi 80% data training dan 20% data testing

2.7 Pemodelan

*Melakukan pemodelan dengan Pustaka program OLS, dan jalankan training data, kemudian cetak parameter constan, x1 dan x2 dan tampilkan persamaan regresi nya. Cek apakah data training telah memiliki nilai konstan, jika belum ada tambahkan variabel konstan bernilai 1.0*

  
**Gambar 2.7.1** Menambahkan nilai konstan

   
**Gambar 2.7.2** Persamaan Regrsi X1 dan X2

3. Praktikum Mandiri  
Buat program untuk membagi dataset day.csv menjadi tiga bagian, yaitu:

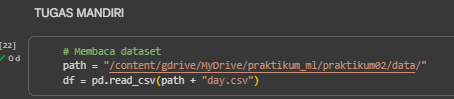
(a) Data Training: 80% dari total dataset

(b) Data Validation: 10% dari data training

(c) Data Testing: 20% dari total dataset

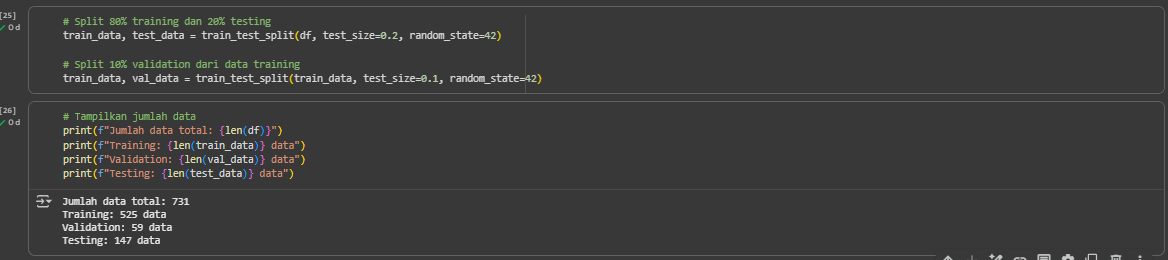
3.1 Membaca Dataset File

*Sel ini membaca dataset yang diinginkan yaitu day.csv*

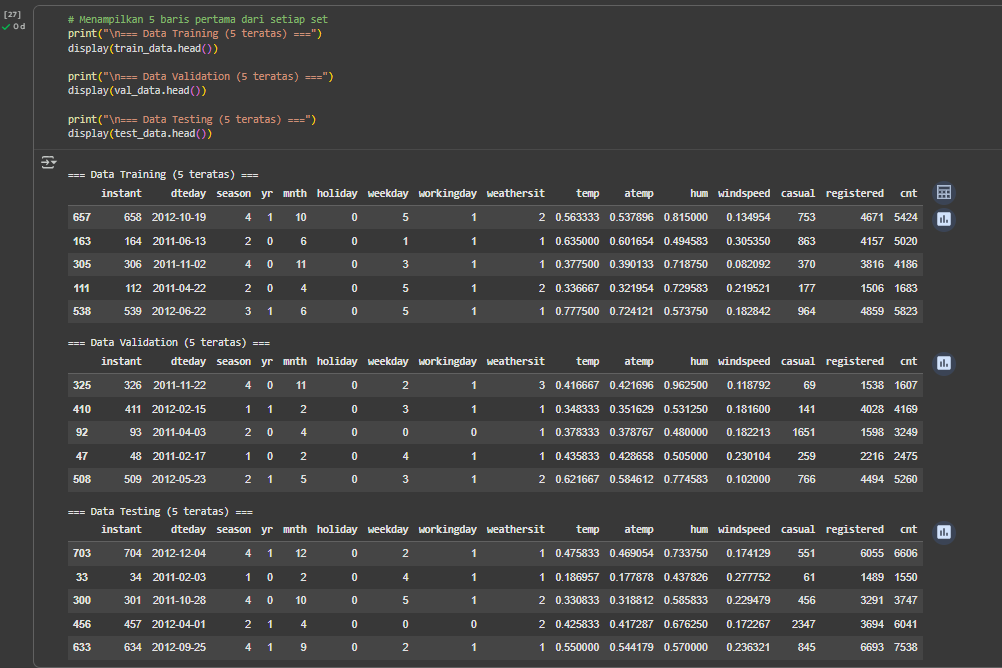


**Gambar 3.1.** Proses ini hanya perlu dilakukan satu kali per sesi.

3.2 Membaca Dataset File   
Sel ini membagi data Training 80%, Data Testing 20%, dan Data validation 10%, lalu menampilkan jumlah datanya

  
**Gambar 3.2.** Menampilkan Jumlah Data hasil pembagian persentase

3.3 Menampilkan 5 baris perta pada setiap Set   
Sel ini menampilkan hasil dari 5 data teratas dari Data Training, Data Validation, dan Data Testing



**Gambar 3.3.** Menampilkan 5 Baris Data Teratas setiap Set

Referensi:

Munir, S., Seminar, K. B., Sudradjat, Sukoco, H., & Buono, A. (2022). The Use of Random Forest Regression for Estimating Leaf Nitrogen Content of Oil Palm Based on Sentinel 1-A Imagery. *Information*, *14*(1), 10. https://doi.org/10.3390/info14010010

Seminar, K. B., Imantho, H., Sudradjat, Yahya, S., Munir, S., Kaliana, I., Mei Haryadi, F., Noor Baroroh, A., Supriyanto, Handoyo, G. C., Kurnia Wijayanto, A., Ijang Wahyudin, C., Liyantono, Budiman, R., Bakir Pasaman, A., Rusiawan, D., & Sulastri. (2024). PreciPalm: An Intelligent System for Calculating Macronutrient Status and Fertilizer Recommendations for Oil Palm on Mineral Soils Based on a Precision Agriculture Approach. *Scientific World Journal*, *2024*(1). https://doi.org/10.1155/2024/1788726