Tugas 4: Praktikum & Praktikum Mandiri 4

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

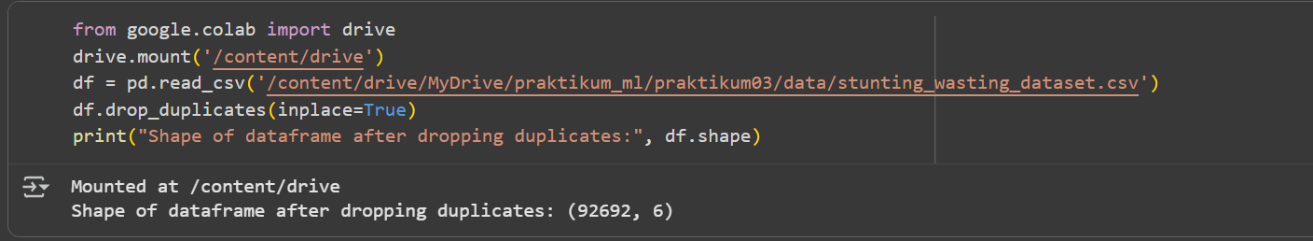
1 Teknik Informatika, STT Terpadu Nurul Fikri, Depok

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

**Abstract.** Algoritma Logistic Regression untuk memprediksi status stunting pada balita dengan menggunakan variabel umur, tinggi badan, berat badan, dan jenis kelamin. Proses diawali dengan tahap pra-pemrosesan data seperti pengubahan data kategorik menjadi numerik, standarisasi data, serta pembagian dataset menjadi data latih dan data uji. Model kemudian dilatih menggunakan scikit-learn dan dievaluasi menggunakan metrik akurasi, presisi, recall, F1-score, dan ROC-AUC. Berdasarkan hasil pengujian, model mencapai akurasi sekitar 90% dengan nilai ROC-AUC sebesar 0,97 yang menunjukkan performa yang sangat baik. Dari hasil analisis, tinggi badan berpengaruh paling besar dalam menurunkan kemungkinan stunting, sedangkan faktor umur dan jenis kelamin berpengaruh positif terhadap risiko stunting. Praktikum ini membantu mahasiswa memahami bagaimana Logistic Regression dapat digunakan untuk memecahkan masalah nyata di bidang kesehatan masyarakat.

1. Connecting Google Colab & Drive

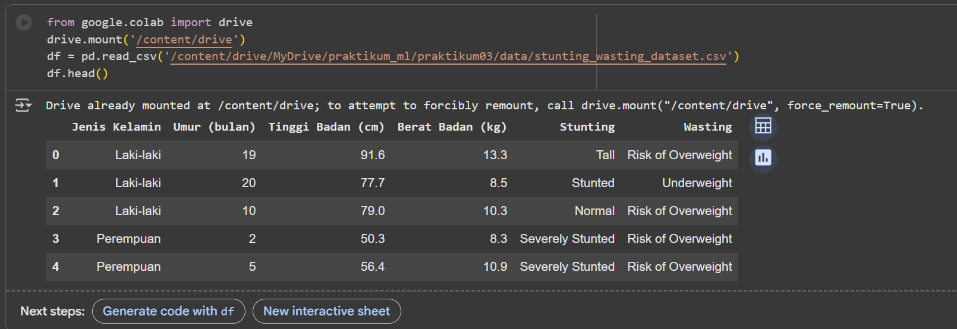
1.1 Menghubungkan lingkungan Google Colab dengan akun Google Drive

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

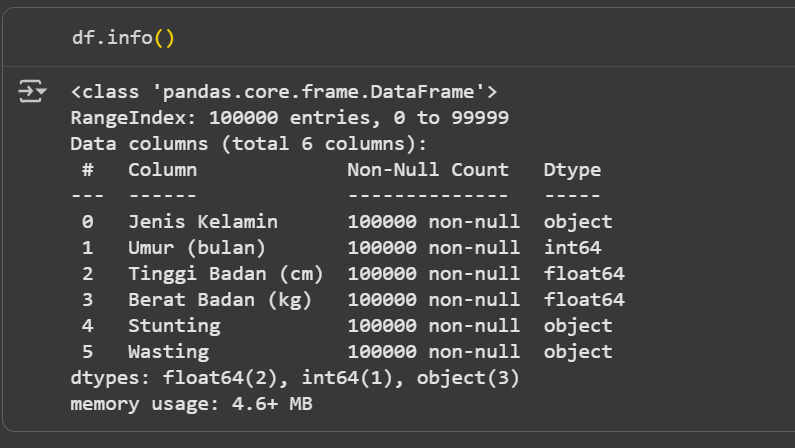
Sel ini menggunakan library Pandas untuk membaca file data, yang diinginkan



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

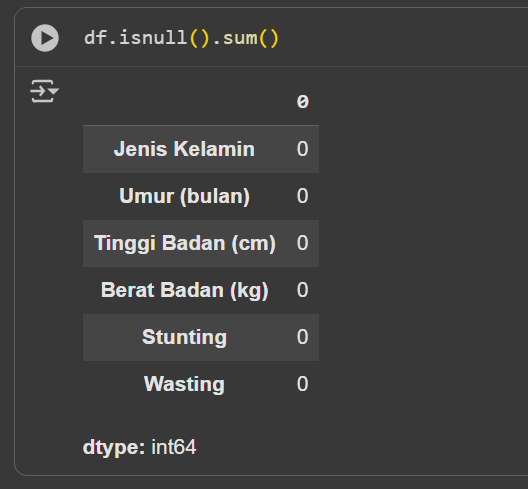
1.3 Mencari informasi data yang ada pada file

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



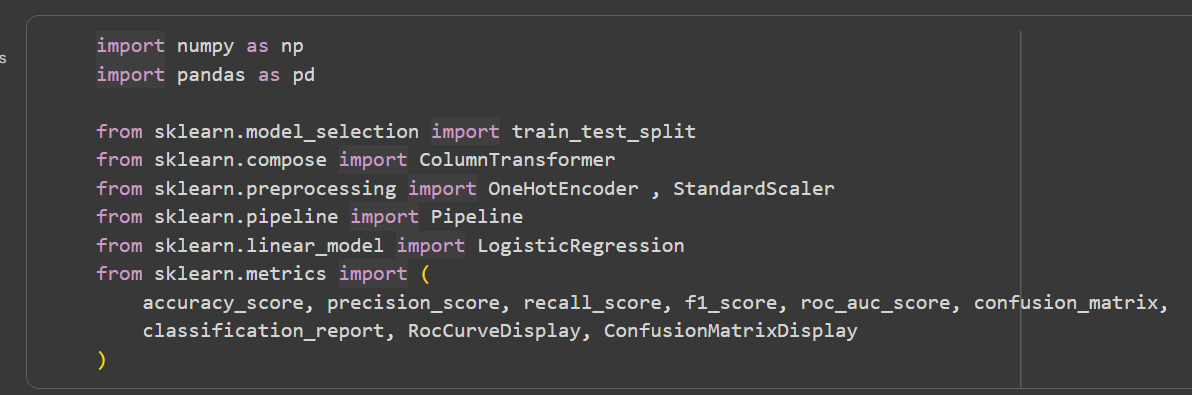
**Gambar 1.3.** Mencari info data pada file

1.4 Cek data Null  
Mengecek apakah terdapat data null



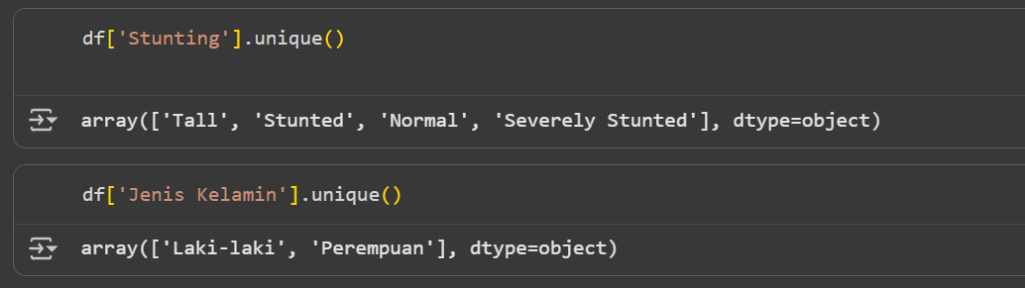
**Gambar 1.4.** Mengecek apakah terdapat data null

*1.5 Import Library dan Model  
Mengimport beberapa library dan Model yang dibutuhkan*

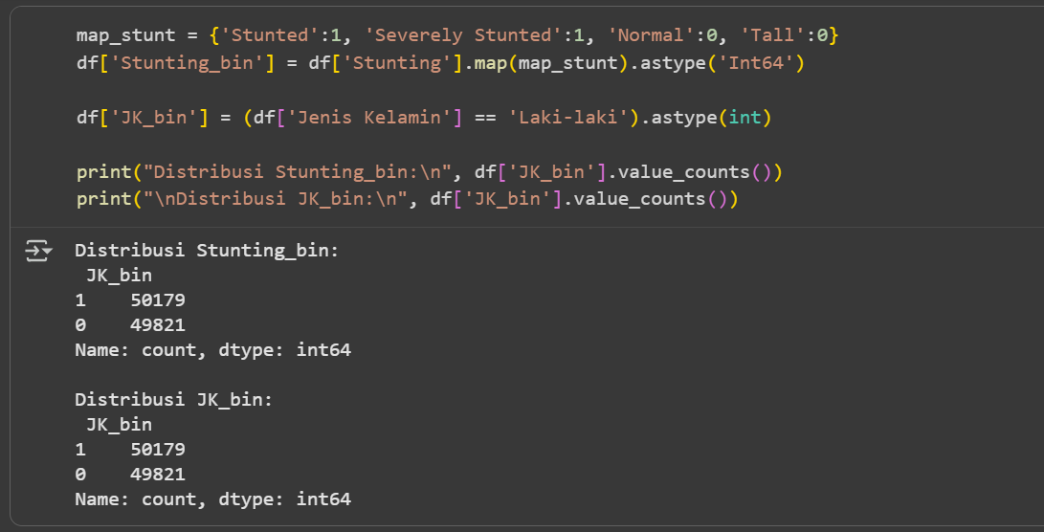


**Gambar 1.5.** Mengimport Library

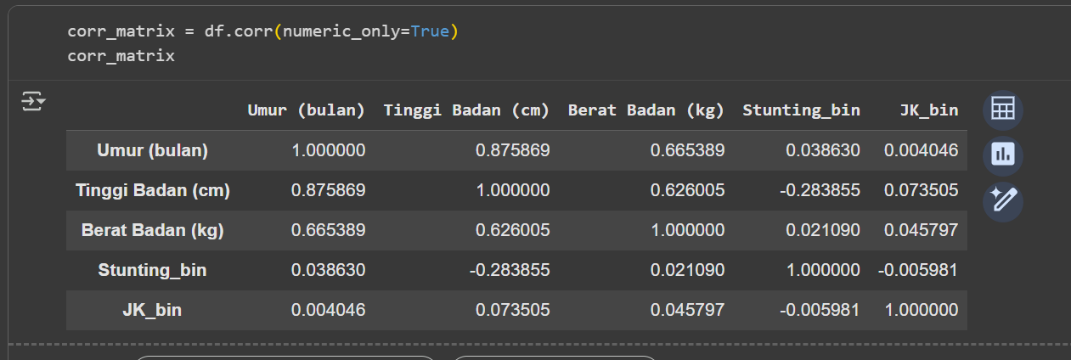
*1.6 Cek Nilai Unik   
Mengecek Data Unit dari mulai kolom Stunting dan Jenis Kelamin*

  
**Gambar 1.6.** Cek Nilai Unik

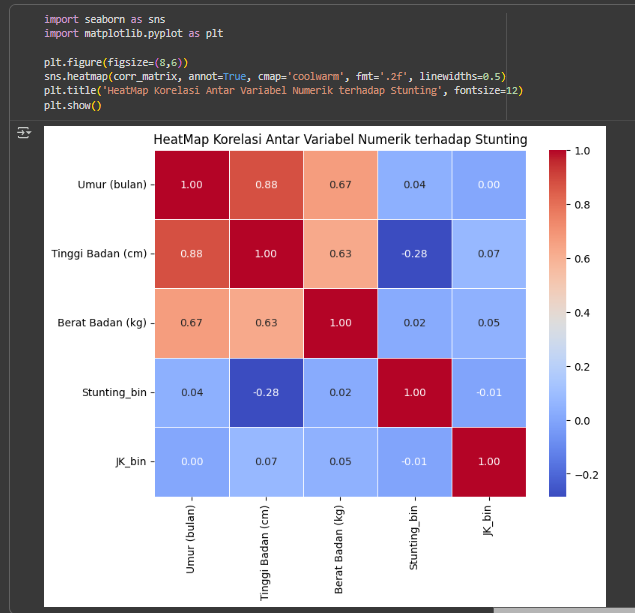
*1.7 Mapping Kolom Kategorik ke Bentuk Numerik  
Mengubah nilai teks menjadi bentuk numerik agar dapat diproses model*

**Gambar 1.7.** Mapping Kolom Kategorik ke Bentuk Numerik

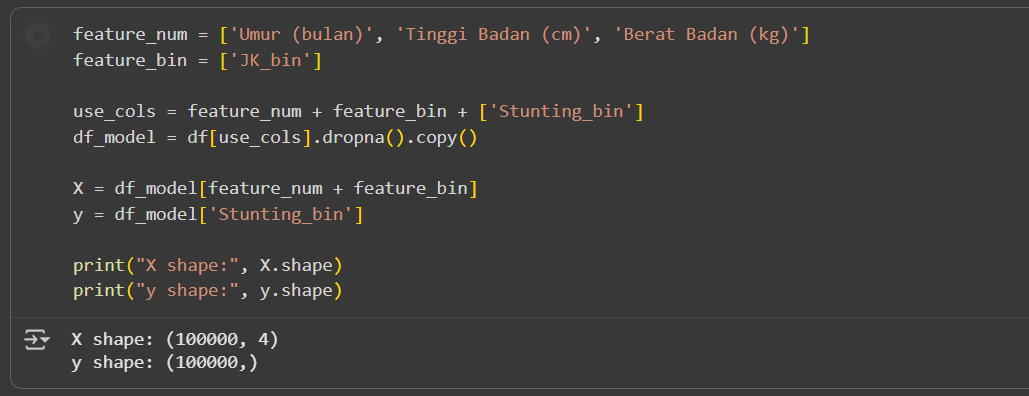
*1.8 Korelasi Antar Variabel Numerik  
menghitung korelasi Pearson antara setiap variabel numerik terhadap target Stunting\_bin*

**Gambar 1.8.** Korelasi Antar Variabel Numerik

*1.9 Visualisasi Heatmap Korelasi  
untuk melihat hubungan antar variabel numerik secara menyeluruh dalam dataset*

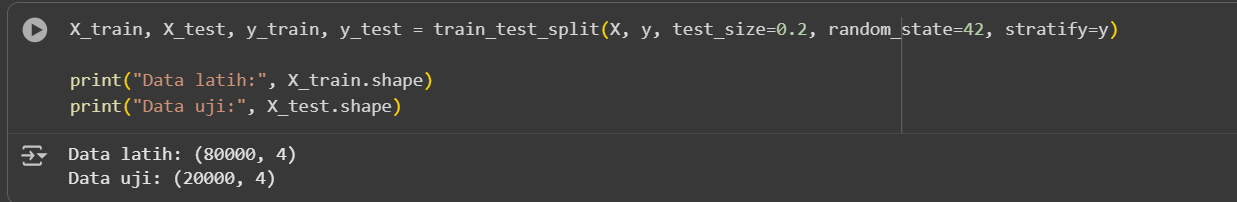
   
**Gambar 1.9.** Heatmap Korelasi

*1.10 Menentukan Fitur dan Target*Mengubah dataset dibagi menjadi dua bagian utama, Variabel X berisi fitur atau variabel independen (Umur, Tinggi Badan, Berat Badan, dan Jenis Kelamin). Variabel y berisi variabel target yaitu Stunting\_bin.

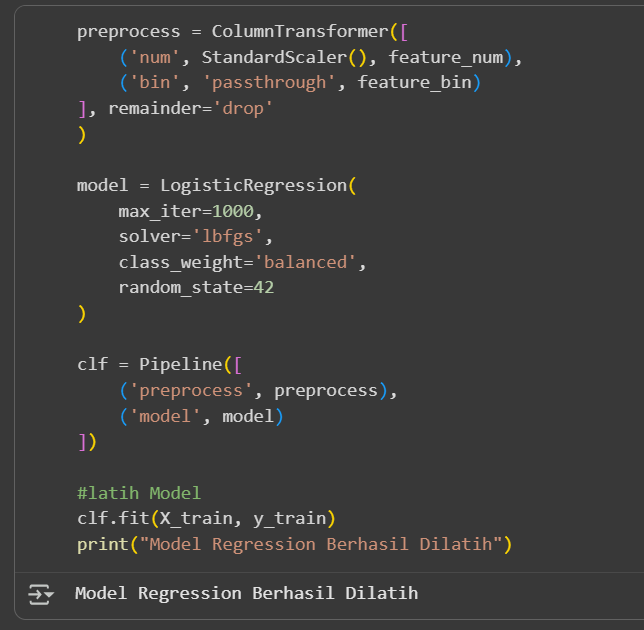


**Gambar 1.10.** Fitur dan Target

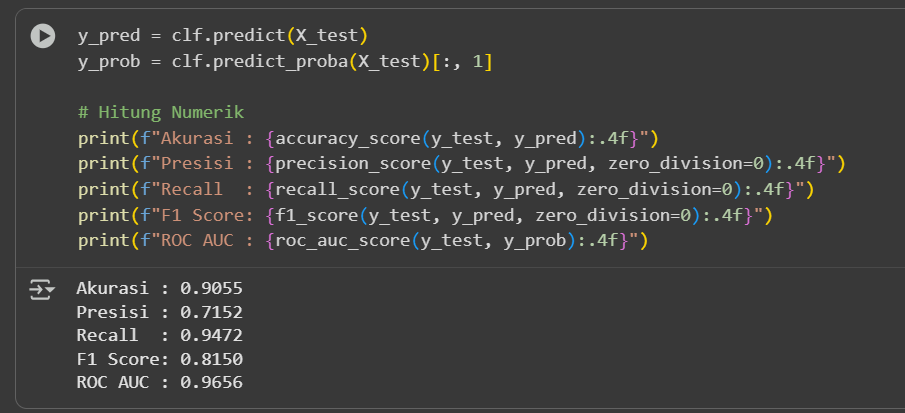
*1.11 Membagi Dataset menjadi Training dan Testing Set*80% data (80.000 baris) digunakan untuk melatih model (X\_train, y\_train),  
20% data (20.000 baris) digunakan untuk menguji performa model (X\_test, y\_test)

  
**Gambar 1.11.** Pembagian Dataset

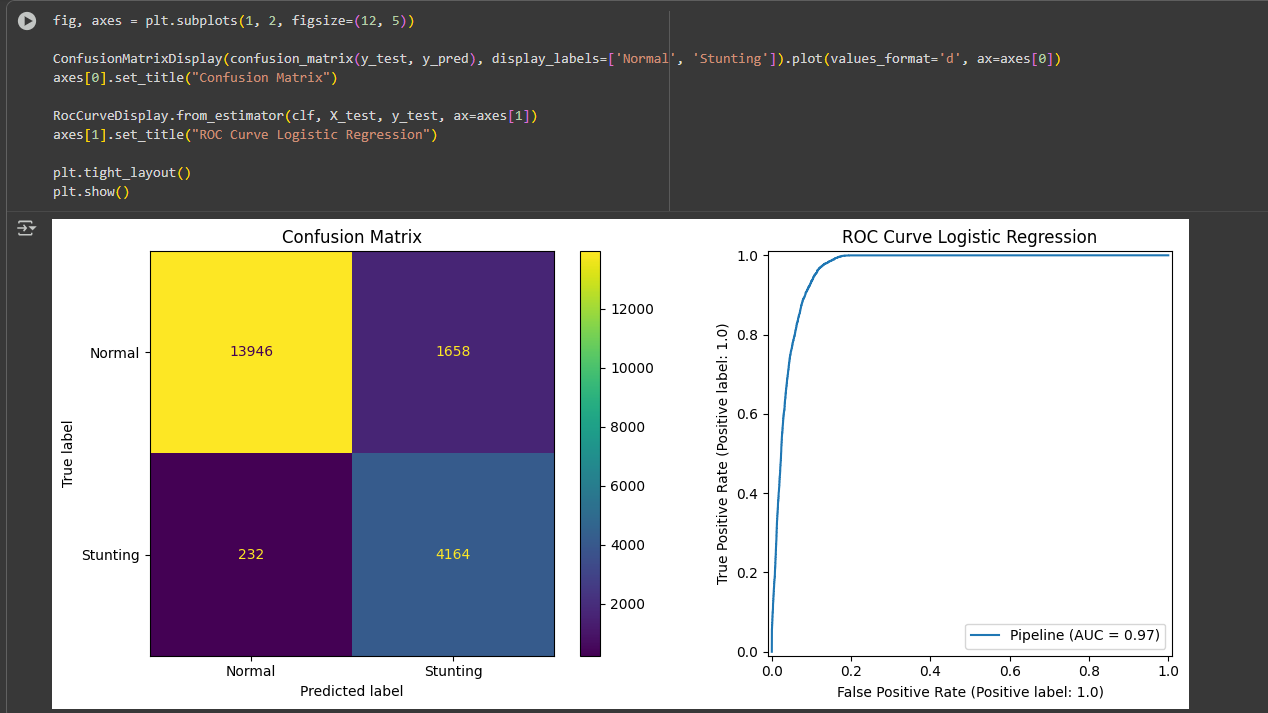
*1.12 Pembangunan Model Logistic Regression*Model berhasil dilatih menggunakan data training (80.000 baris)

  
**Gambar 1.12.** Training Model Logistic Regression

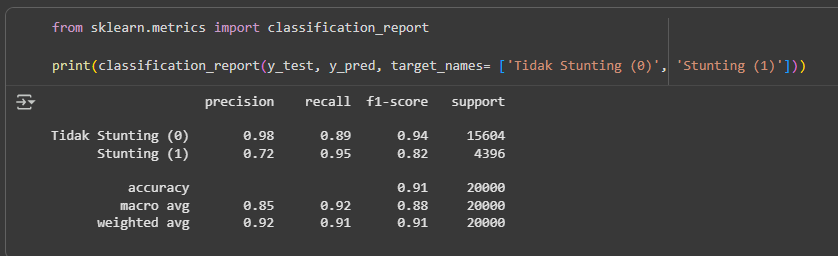
*1.13 Prediksi Model dan Evaluasi Model*menghasilkan label klasifikasi 0 (tidak stunting) atau 1 (stunting) dan memberikan probabilitas peluang stunting.

  
**Gambar 1.13.** Prediksi Model dan Evaluasi Model

*1.14 Visualisasi Hasil Evaluasi*Setelah model Logistic Regression dilatih dan diuji menggunakan data uji sebanyak 20.000 baris, dilakukan evaluasi performa model melalui dua visual utama, yaitu ROC Curve dan Confusion Matrix

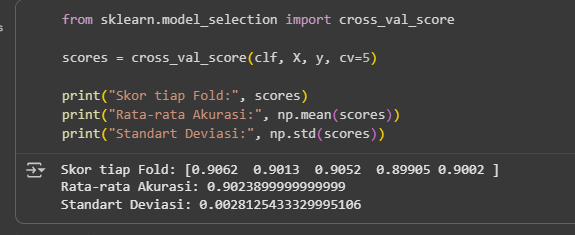
Gambar 1.14. Visualisasi Hasil Evaluasi Model

*1.15 Classification Report*mengevaluasi performa model klasifikasi (seperti Logistic Regression)

  
Gambar 1.15. Gambaran keseimbangan performa model antara mendeteksi data positif (Stunting) dan negatif (Tidak Stunting)

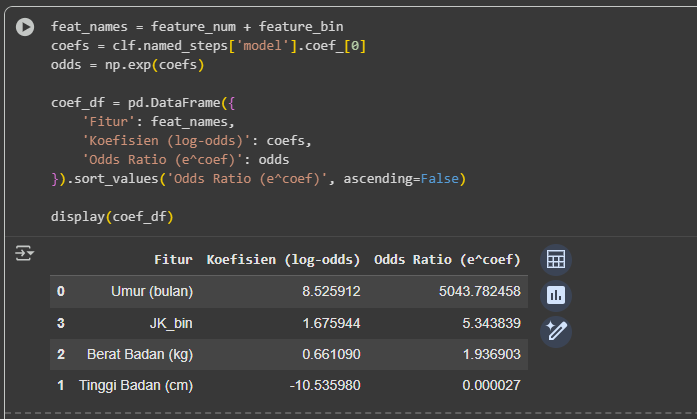
*1.16 Hasil Validasi*

Langkah ini berfungsi untuk memastikan hasil model benar-benar konsisten dan tidak kebetulan bagus hanya di satu pembagian data.

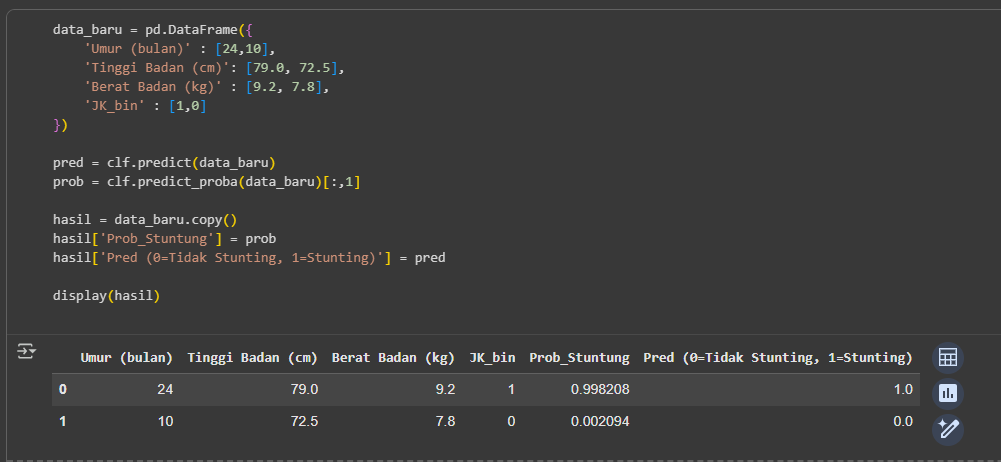
  
**Gambar 1.16.** Model Logistic Regression menunjukkan performa yang stabil dan dapat diandalkan di seluruh data

*1.17 Interpretasi Model Logistic Regression*

Langkah ini berfungsi untuk memastikan hasil model benar-benar konsisten dan tidak kebetulan bagus hanya di satu pembagian data.

  
**Gambar 1.17.** Mengetahui seberapa besar pengaruh setiap fitur (variabel input) terhadap peluang terjadinya stunting

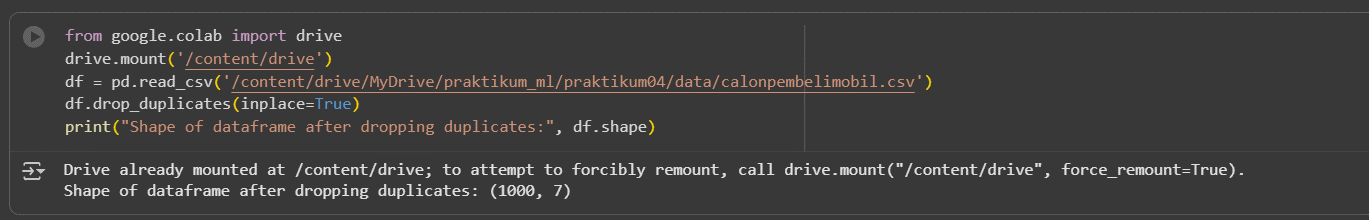
*1.18 (Contoh Kasus)*Untuk mengetahui apakah model dapat melakukan prediksi dengan benar dan logis berdasarkan karakteristik anak.



**Gambar 1.17.** Mengetahui seberapa besar pengaruh setiap fitur (variabel input) terhadap peluang terjadinya stunting

2. LOGISTIC REGRESSION (MANDIRI)

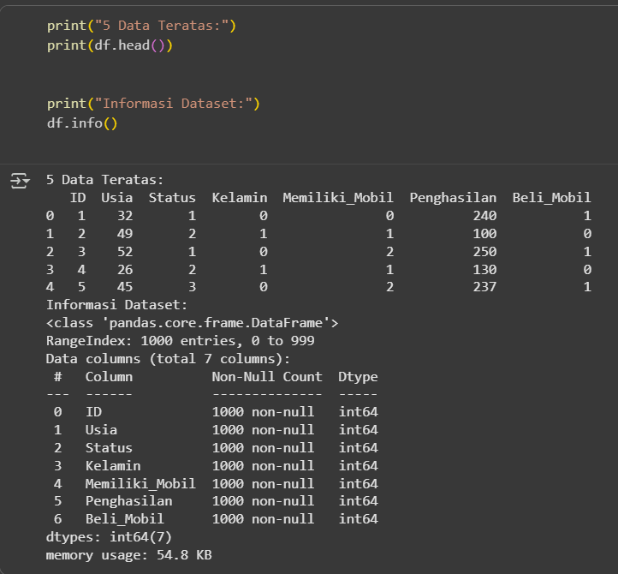
2.1 Menarik Data   
Membaca data calonpembelimobil.csv



**Gambar 2.1.** Menarik Data

2.2 Membaca Isi Data dan Data Teratas

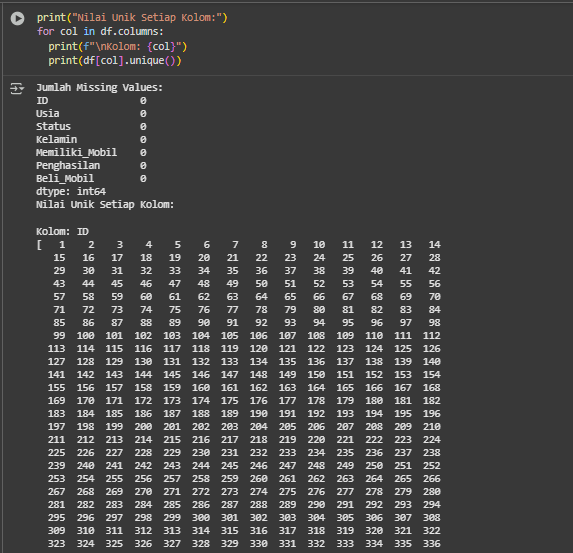
Melihat Informasi yang ada didalam Data calonpembelimobil.csv



**Gambar 2.2.** Menampilkan Data Info

2.3 Mencari Nilai Unik

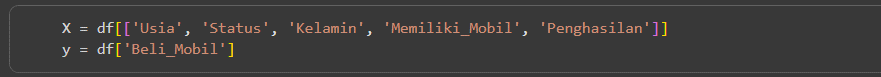
*Mencari Nilai Unik dalam setiap kolom*



**Gambar 2.3.** Tampilan Nilai kolom

2.4 Memisahkan Fitur dan Target

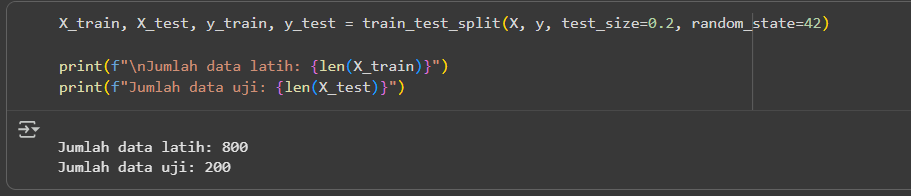
*Tahapan ini memisahkan Nilai Fitur dan Target*



**Gambar 2.4.** Memisahkan FItur dan Target

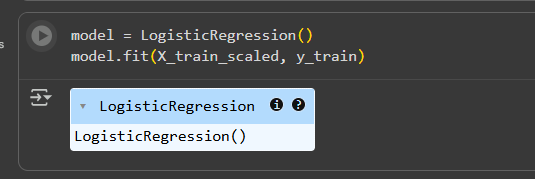
2.5 Membagi Data Latih dan Uji

*Melatih dan menguji 2 data*

  
**Gambar 2.5.** Data Latih dan Uji

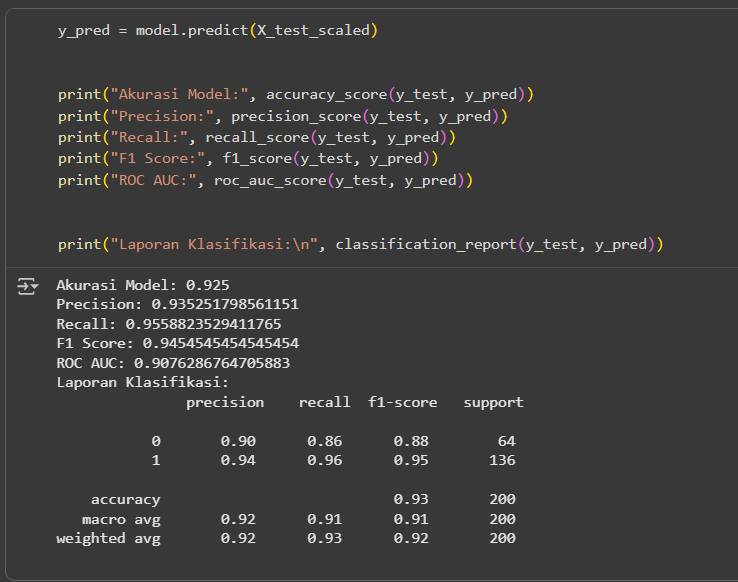
2.6 Buat dan latih model Logistic Regression

*Melatih Model Regression*

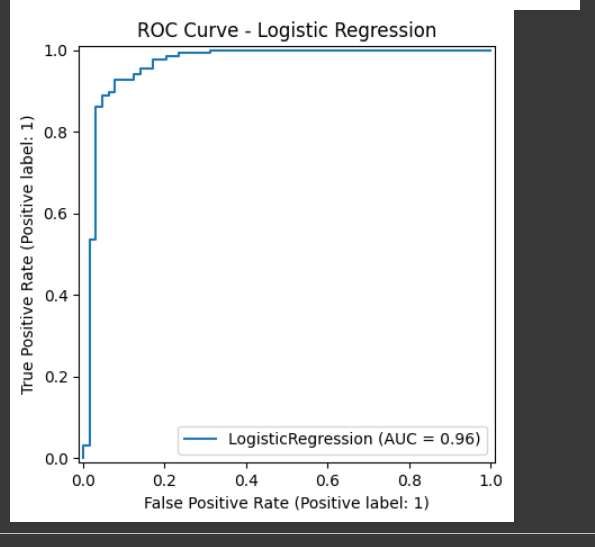
**Gambar 2.6.** Latih Model

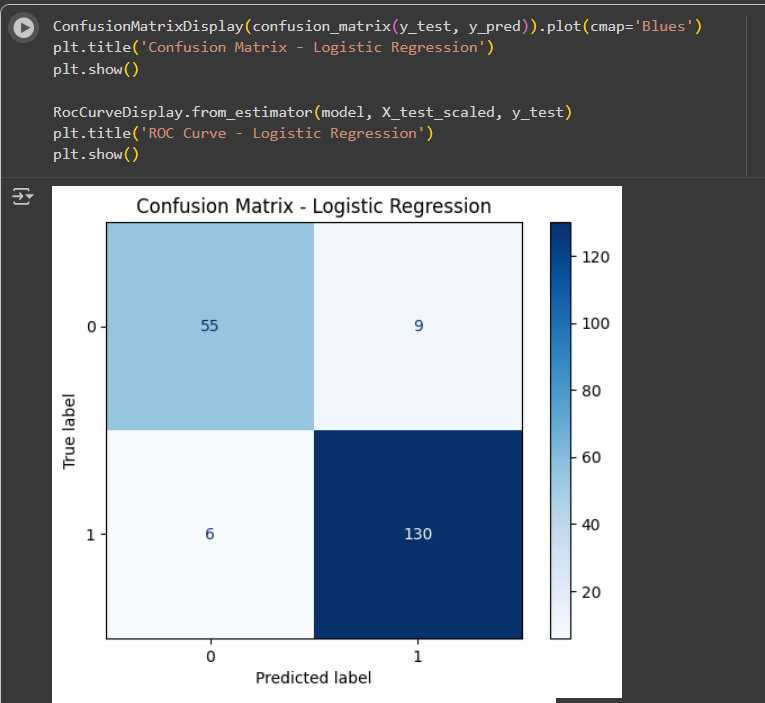
2.7 Evaluasi Model

*Evaluasi Model Regression*

**Gambar 2.7.** Evaluasi Model

2.8 Visualisasi Data

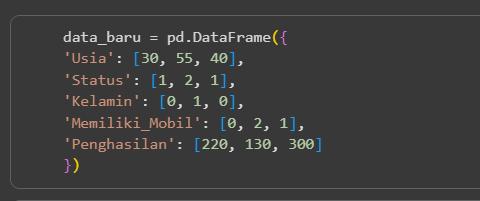
*Visualisasi Data hasil*



**Gambar 2.8.** Visualsisasi Model

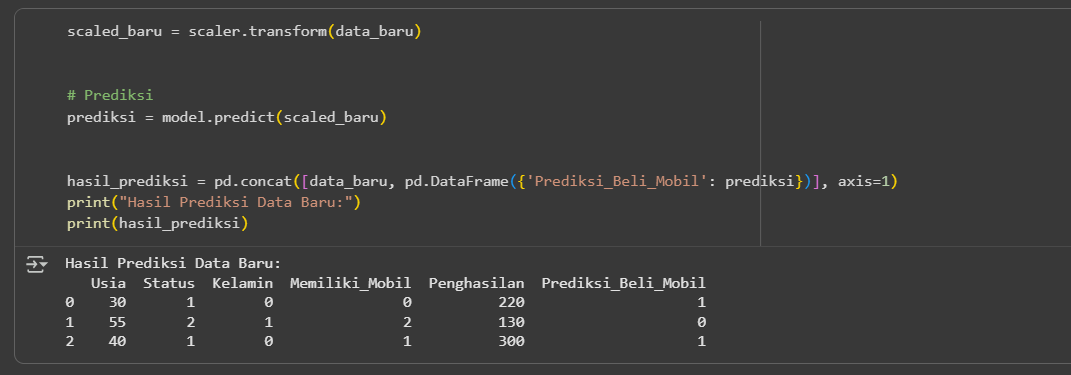
2.9 Buat Data Baru

*Data Baru untuk menguji apakah model dapat memprediksi sesuai dengan data serupa*

  
**Gambar 2.9.** Data Baru

2.10 Buat Data Baru

*Prediksi Data Baru*



**Gambar 2.10.** Prediksi Data Baru

**KESIMPULAN**

Berdasarkan hasil percobaan yang telah dilakukan menggunakan metode **Logistic Regression**, model mampu mengklasifikasikan calon pembeli mobil dengan cukup baik. Nilai akurasi yang diperoleh menunjukkan bahwa model memiliki performa yang baik dalam memprediksi keputusan pembelian mobil berdasarkan variabel seperti umur, pendapatan, status, dan pekerjaan.

Dengan demikian, Logistic Regression dapat digunakan sebagai model prediksi awal untuk menentukan kemungkinan seseorang membeli mobil.

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