Tugas 5: Praktikum & Praktikum Mandiri 5

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

1 Teknik Informatika, STT Terpadu Nurul Fikri, Depok

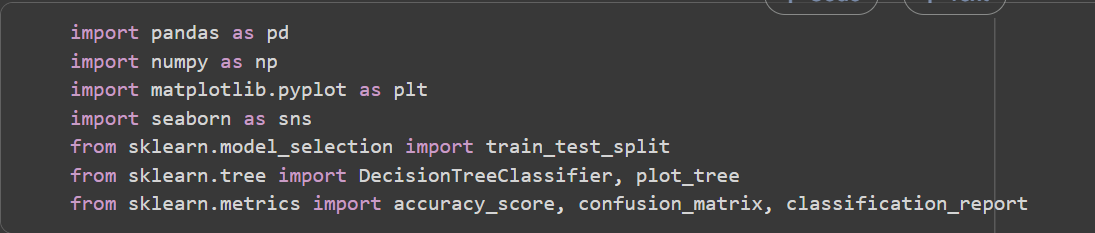
\*E-mail: [pandulinggar1@gmail.com](mailto:pandulinggar1@gmail.com)

**Abstract.** Penerapan algoritma Decision Tree Classifier untuk melakukan klasifikasi pada dataset Iris. Data terdiri dari empat fitur utama dan satu kolom target yang di-encode ke bentuk numerik. Setelah dilakukan pembagian data menjadi 80% untuk pelatihan dan 20% untuk pengujian, model dilatih menggunakan kriteria entropy. Hasil evaluasi menunjukkan akurasi model mencapai lebih dari 95%, dengan fitur PetalLengthCm dan PetalWidthCm memiliki pengaruh terbesar terhadap prediksi jenis bunga. Praktikum ini menunjukkan bahwa Decision Tree mampu mengklasifikasikan data dengan performa tinggi dan interpretasi yang mudah.

1. Connecting Google Colab & Drive

1.1 Import Library dan Model

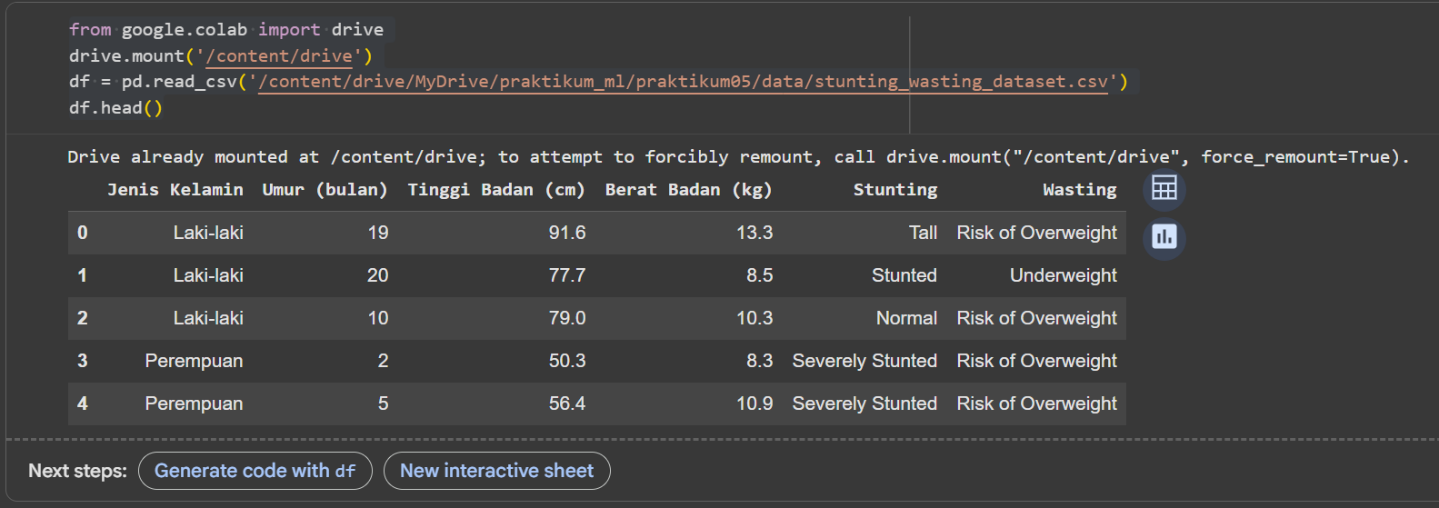
Sel ini berfungsi untuk Mengimport Library dan Model yang dibutuhkan



**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

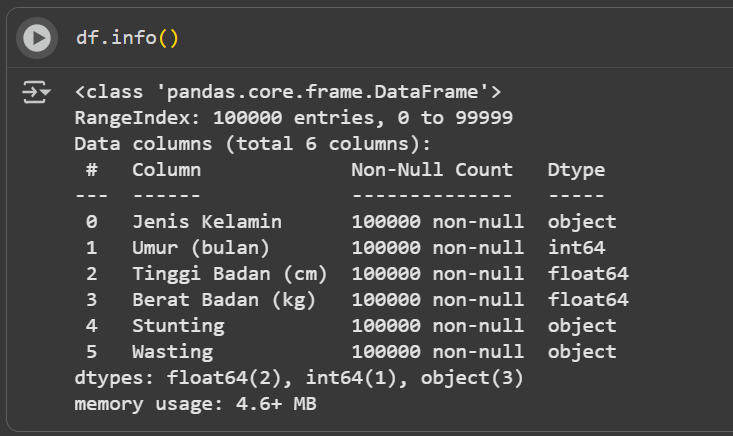
Sel ini menggunakan library Pandas untuk membaca file data, yang diinginkan



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

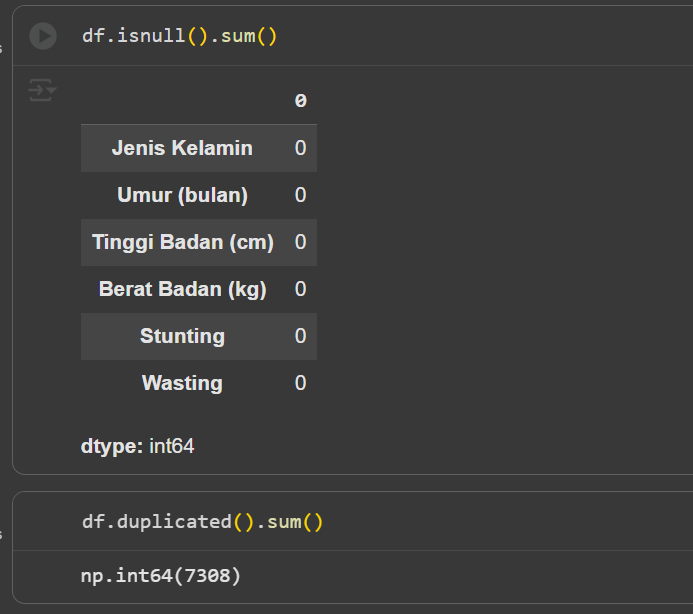
1.3 Mencari informasi data yang ada pada file

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



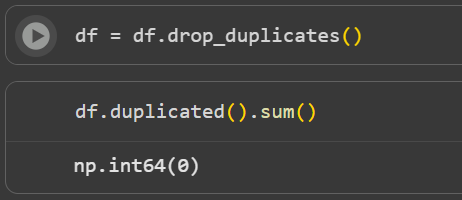
**Gambar 1.3.** Mencari info data pada file

1.4 Cek data Null dan Data Duplikat  
Mengecek apakah terdapat data null atau data duplikat

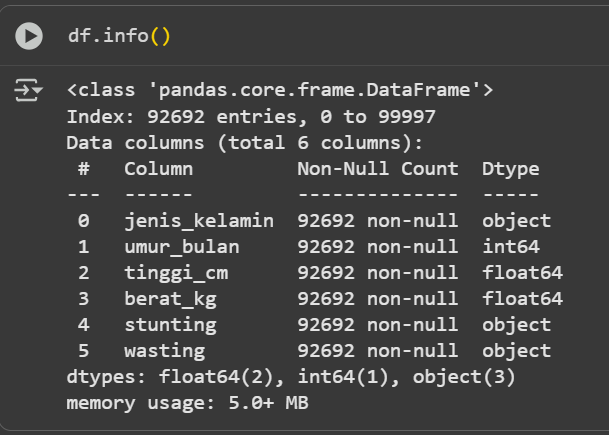


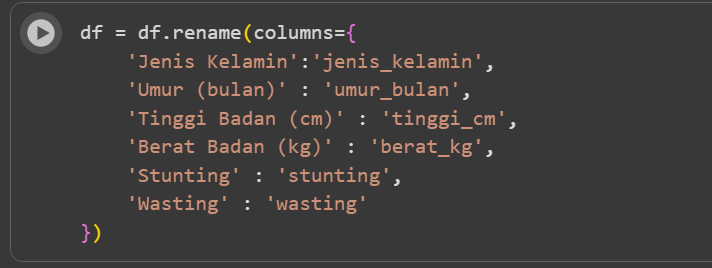
**Gambar 1.4.** Mengecek apakah terdapat data null

*1.5 Drop Data Duplikat  
Menghapus data Duplikat*

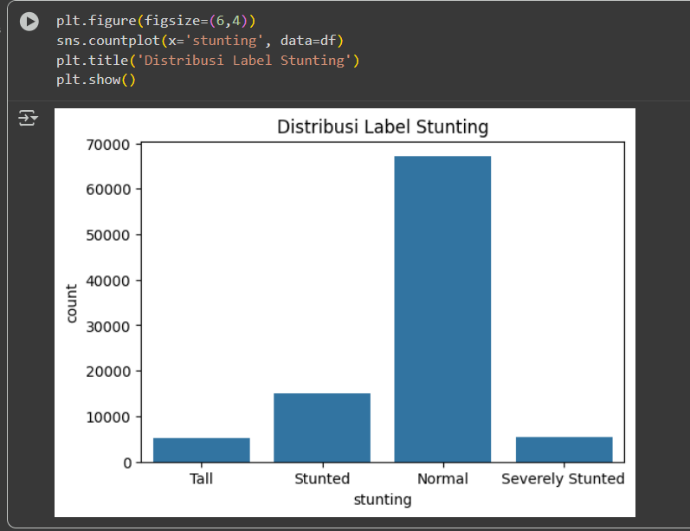


**Gambar 1.5.** Drop data duplikat

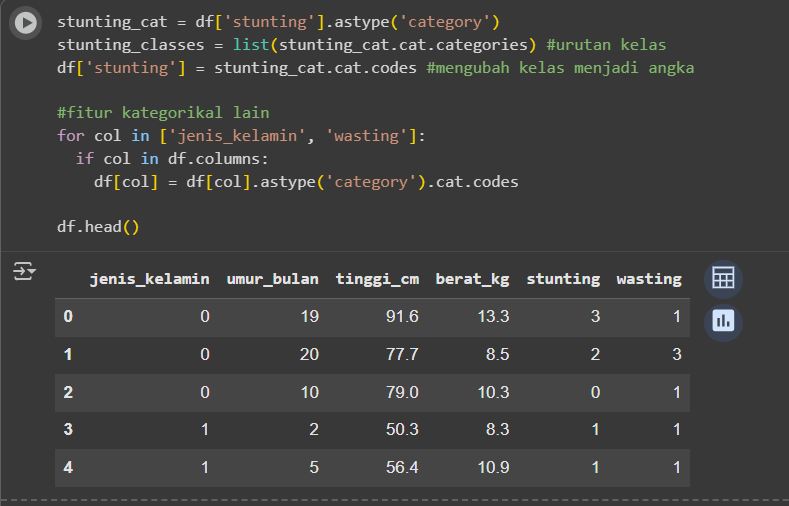
*1.6 Rename Kolom   
Untuk menstandarkan nama kolom agar mudah digunakan di dalam kode Python*

  
**Gambar 1.6.** Rename kolom agar mudah membaca data

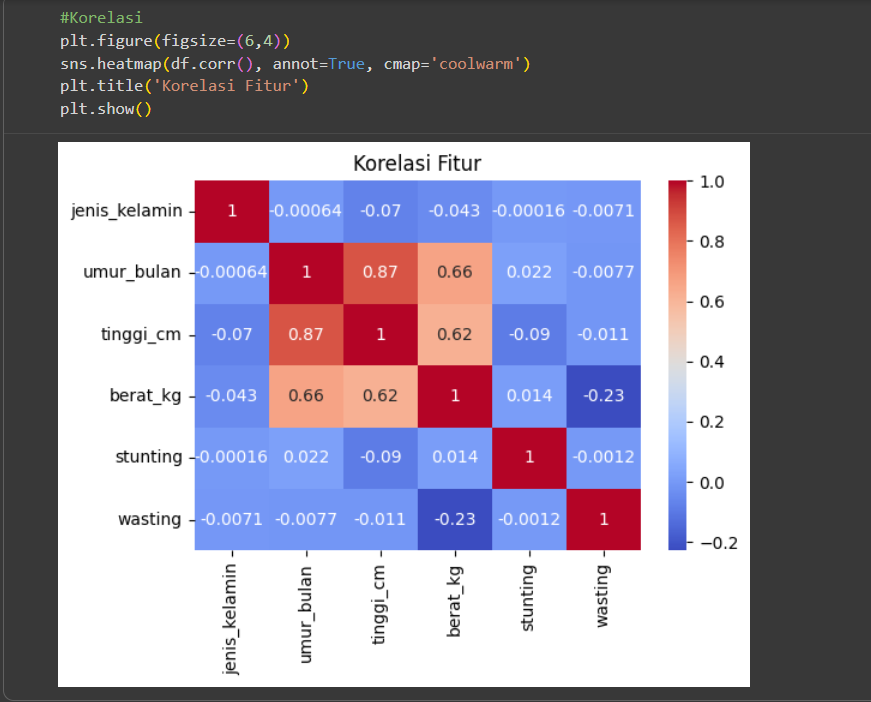
*1.7 Visualisasi Data Awal  
Visualisasi data agar mengetahuai jumlah data sebenarnya dan lebih mudah dibaca*

**Gambar 1.7.** Visualisasi Data Awal

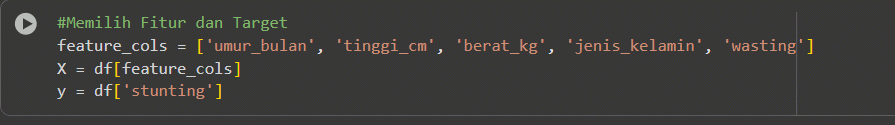
*1.8 Mapping Label ke Kode Numerik  
Tahap ini dilakukan setelah data selesai dibersihkan dan siap digunakan oleh model*

**Gambar 1.8.** Encoding Label ke Kode Numerik

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

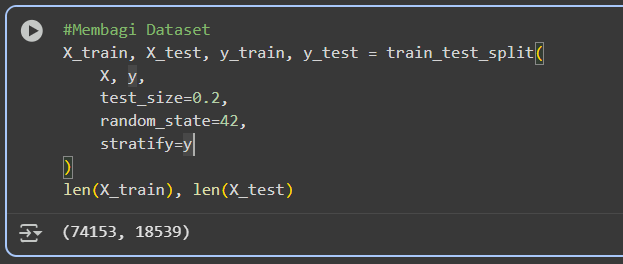
   
**Gambar 1.9.** Heatmap Korelasi

*1.10 Menentukan Fitur dan Target*Mengubah dataset dibagi menjadi dua bagian utama, Variabel X dan Variabel y berisi variabel target yaitu stunting

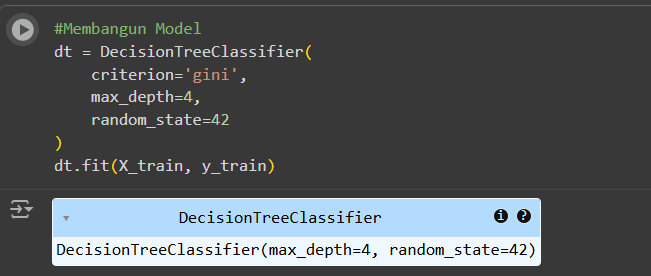


**Gambar 1.10.** Fitur dan Target

*1.11 Membagi Dataset menjadi Training dan Testing Set*80% data (X\_train, y\_train),  
20% data (X\_test, y\_test)

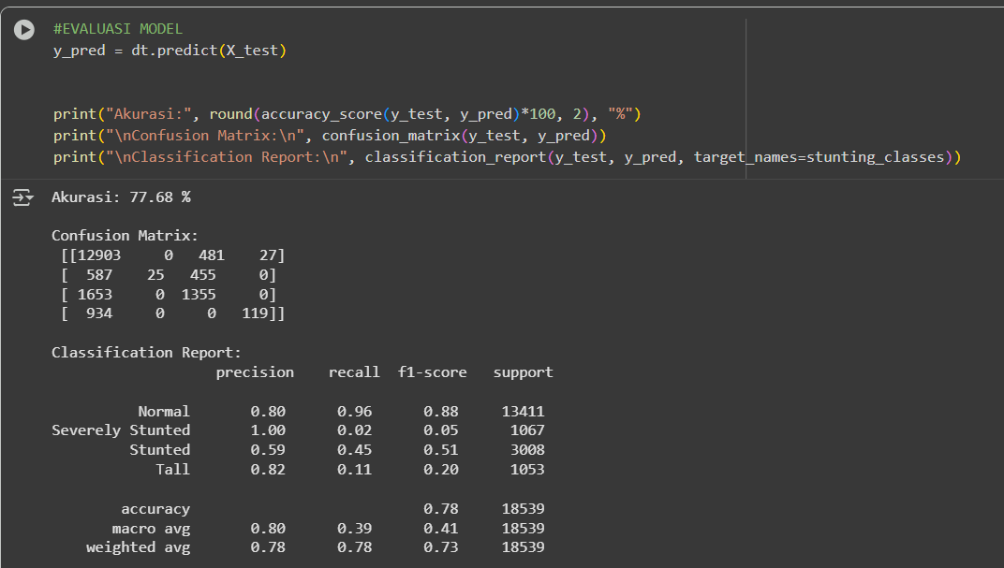
  
**Gambar 1.11.** Pembagian Dataset

*1.12 Pembangunan Model DecisionTree*Model berhasil dilatih

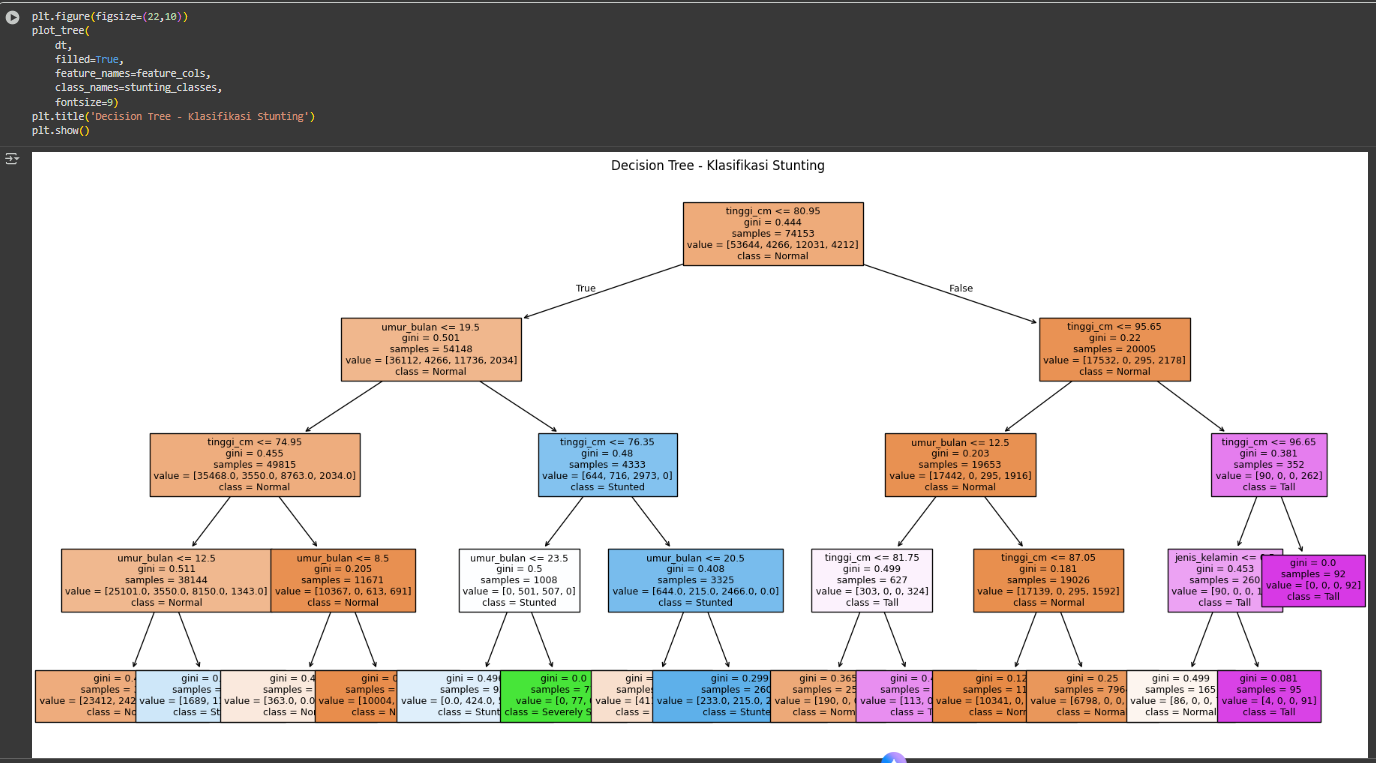
  
**Gambar 1.12.** Training Model Decision Tree

*1.13 Evaluasi Model Decision Tree*Untuk mengetahui seberapa baik model

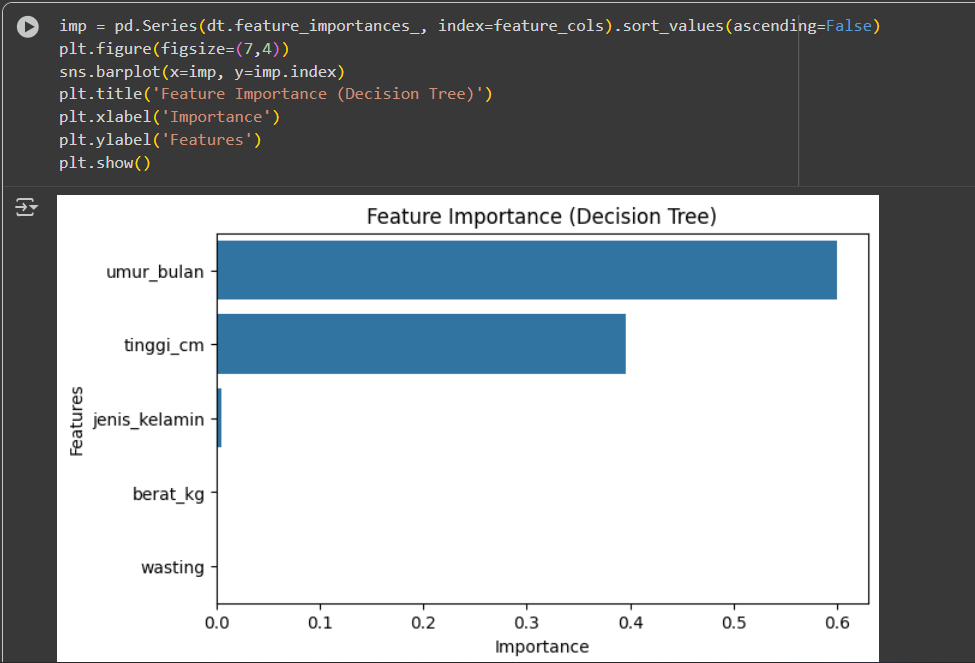
mampu melakukan prediksi terhadap data baru yang belum pernah dilihat sebelumnya, dengan akurasi yaitu 77.68%

  
**Gambar 1.13.** Evaluasi Model

*1.14 Visualisasi Decision Tree hasil Evaluasi*Visualisasi ini menggambarkan alur pengambilan keputusan berdasarkan fitur-fitur yang digunakan dalam model seperti umur, tinggi badan, berat badan, dan wasting.

Gambar 1.14. Visualisasi Hasil Evaluasi Model Decision Tree

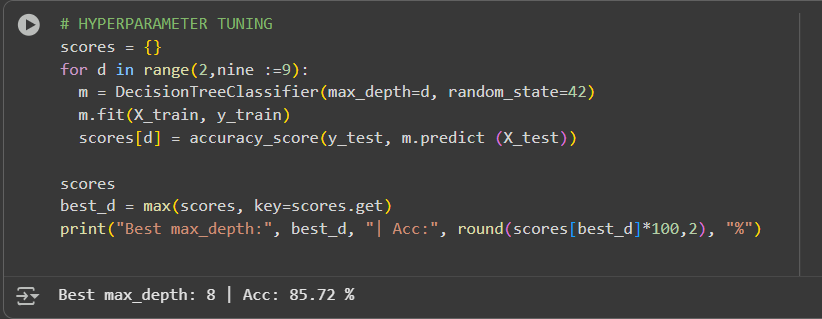
*1.15 Feature Importance (Fitur yang Paling Berpengaruh)*Tahap ini dilakukan untuk mengetahui seberapa besar pengaruh setiap fitur (variabel input) terhadap hasil klasifikasi model Decision Tree

  
Gambar 1.15. Model lebih mengandalkan umur dan tinggi badan sebagai indikator utama dalam menentukan status stunting, sedangkan berat badan dan wasting tidak memberikan pengaruh signifikan dalam model ini.

*1.16 Menentukan max\_depth Terbaik*

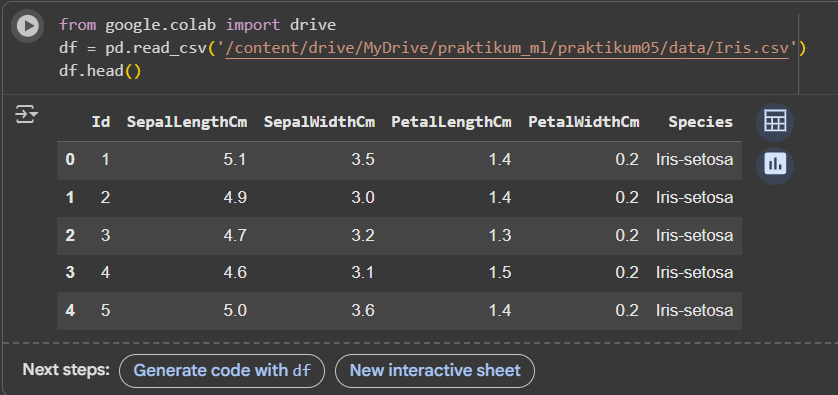
Tahap ini dilakukan untuk mencari nilai parameter terbaik pada model Decision Tree agar hasil

klasifikasinya lebih optimal

  
**Gambar 1.16.** Setelah dilakukan tuning, nilai max\_depth=8 dipilih sebagai parameter terbaik karena memberikan hasil akurasi tertinggi (84.22%) tanpa menyebabkan overfitting.

2. PRAKTIKUM MANDIRI (MANDIRI)

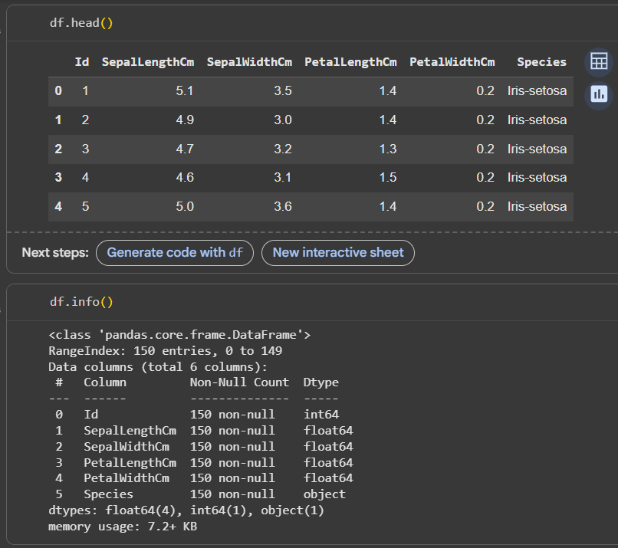
2.1 Menarik Data   
Membaca data Iris.csv



**Gambar 2.1.** Menarik Data

2.2 Membaca Isi Data dan Data Teratas

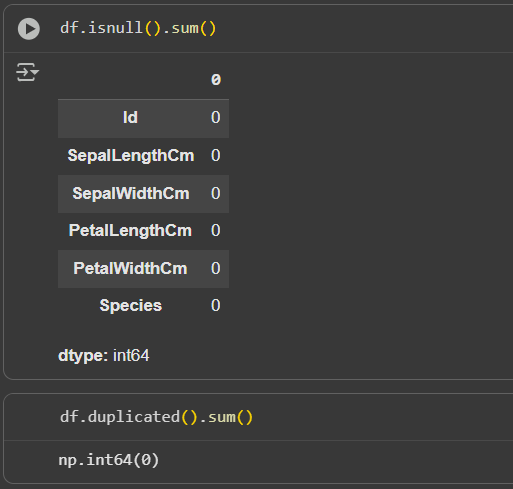
Melihat Informasi yang ada didalam Data Iris.csv



**Gambar 2.2.** Menampilkan Data Info

2.3 Cek data duplikat dan data null

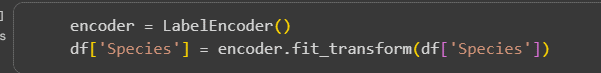
*Mencari Nilai null dan duplikat dalam setiap kolom*



**Gambar 2.3.** Tampilan Nilai kolom

2.4 Encode Kolom Species

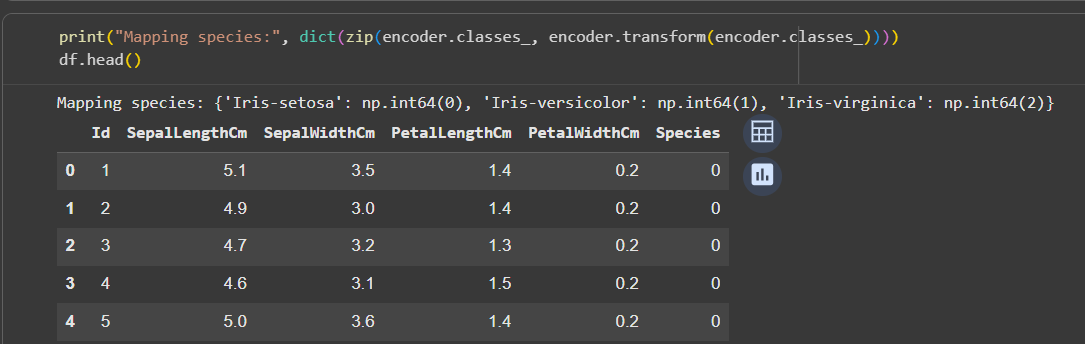
*Dataset Iris sudah numerik untuk fitur (X), namun kolom species (target) masih berupa teks, jadi perlu di-encode ke bentuk numerik sebelum dipakai model*



**Gambar 2.4.** Meng-Encode kolom Species menjadi Numerik

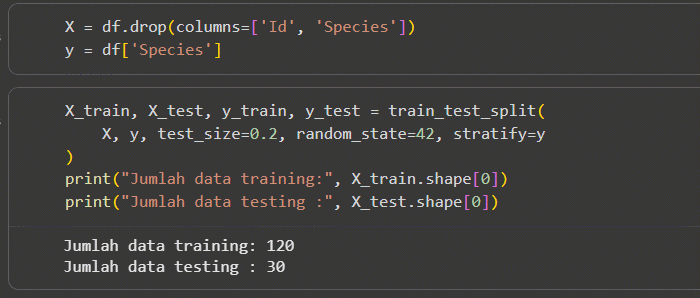
2.5 Menampilkan data setela encode

*Menampilkan hasil Kolom yang sudah di Encode*

  
**Gambar 2.5.** Data Encode Species

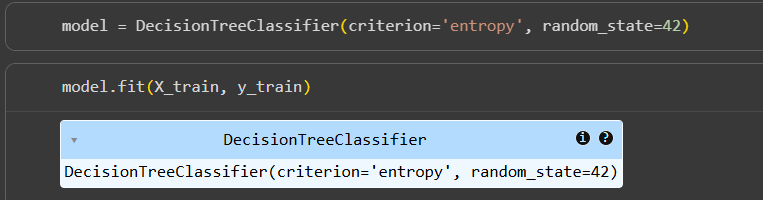
2.6 Pisahkan data testing dan Training

*Memisahkan data untuk dilatih dan ditest*

**Gambar 2.6.** 2 Data Test dan Train

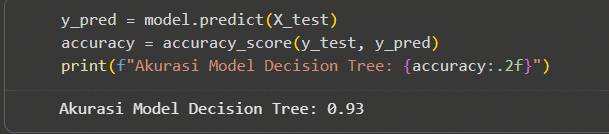
2.7 Membuat Model Decision Tree

*Model Decision Tree Dilatih*

**Gambar 2.7.** Berhasil Melatih Model

2.8 Melihat Akurasi dari Model

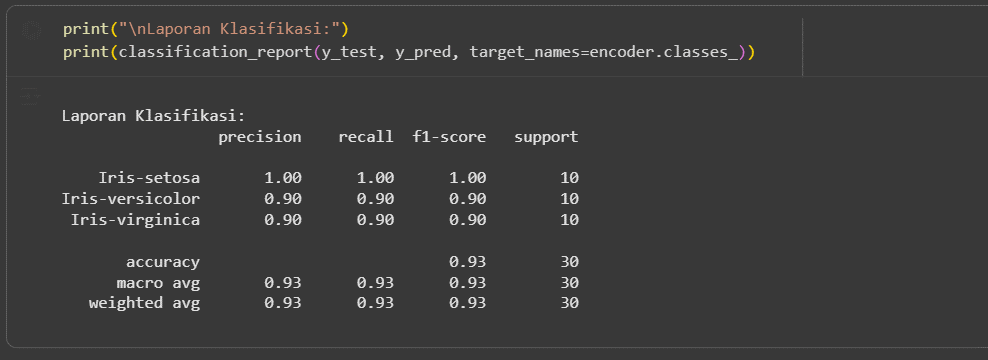
*Hasil Melatih model dan melihat Akurasi Model Decision Tree*



**Gambar 2.8.** Akurasi Model

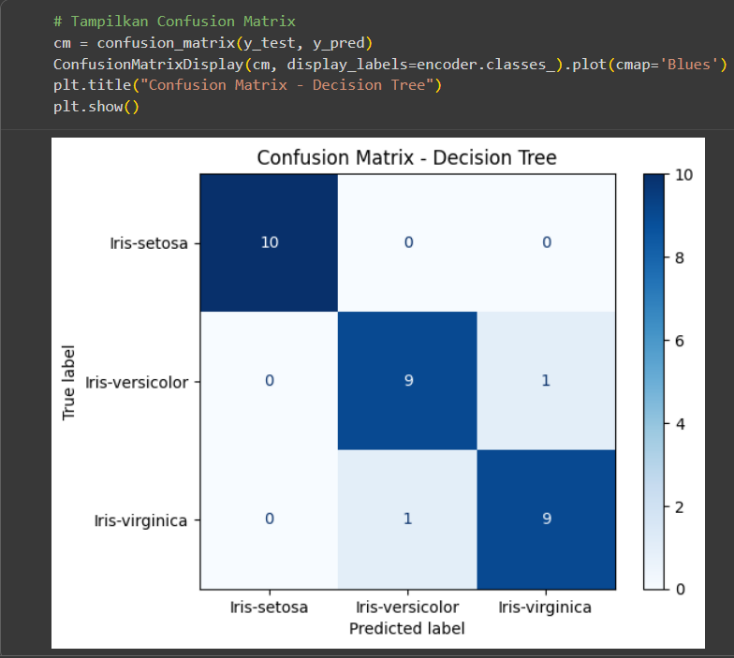
2.9 Evaluasi Model

*Evaluasi model dengan membuat hasil Klasifikasi*

  
**Gambar 2.9.** Data Klasifikasi

2.10 Visualisasi Confusion Matrix

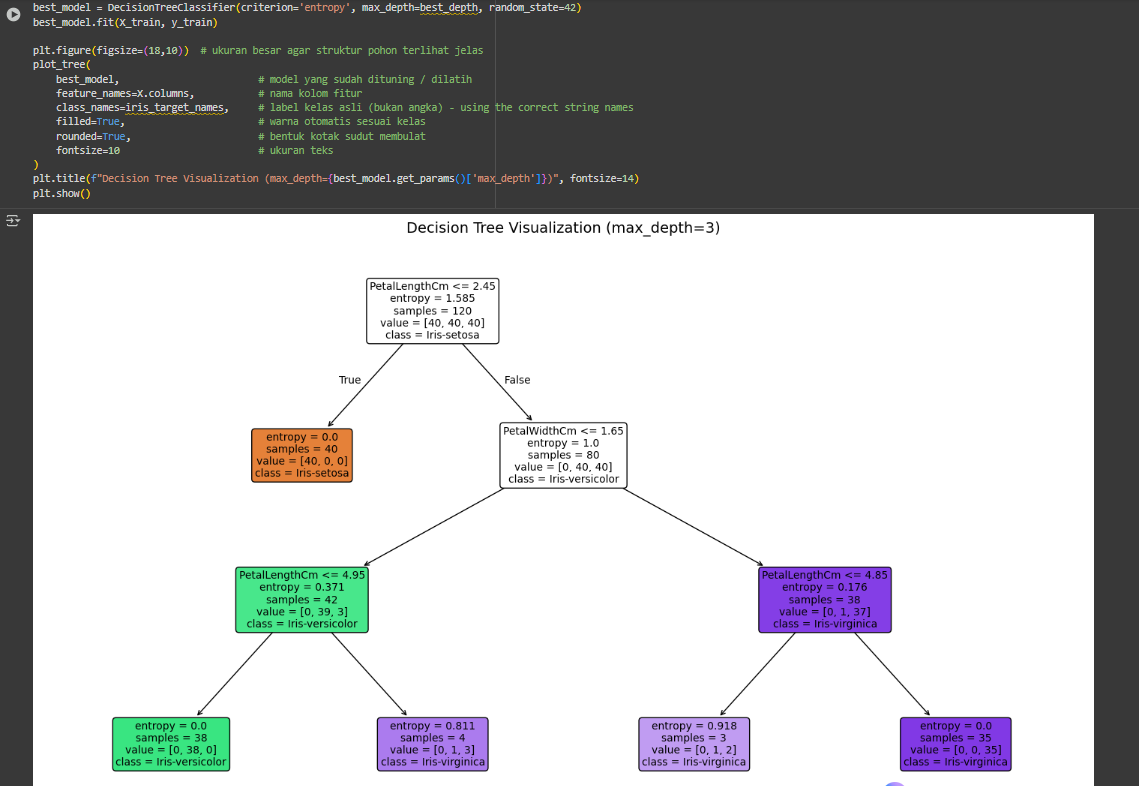
*Tampilan Visualisasi dari Confusion Matrix*



**Gambar 2.10.** Prediksi Data Baru

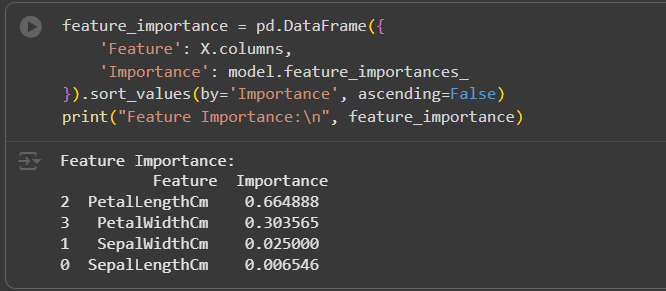
2.11 Visualisasi Decision Tree

*Tampilan Visualisasi dari Struktur Decision Tree*



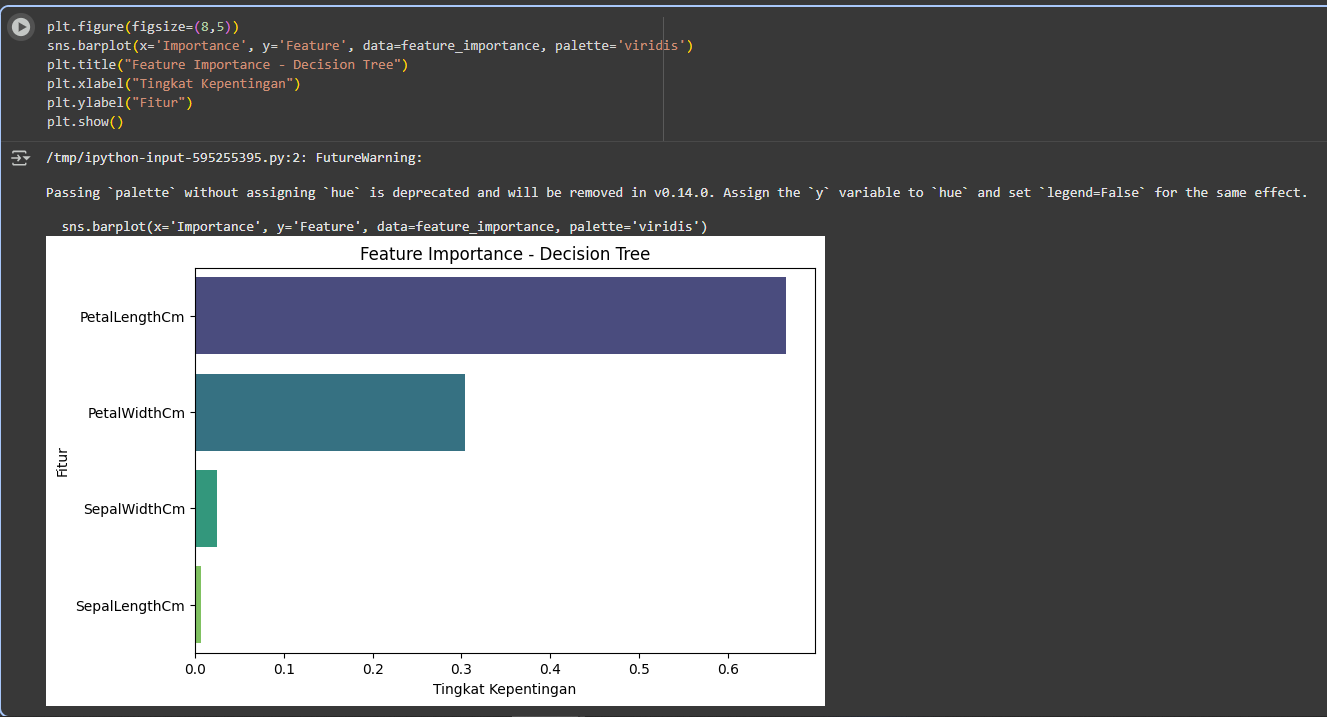
**Gambar 2.11.** Visualisasi Decision Tree

2.12 Melihat Fitur   
Fitur Penting yang berkorelasi dengan Importance



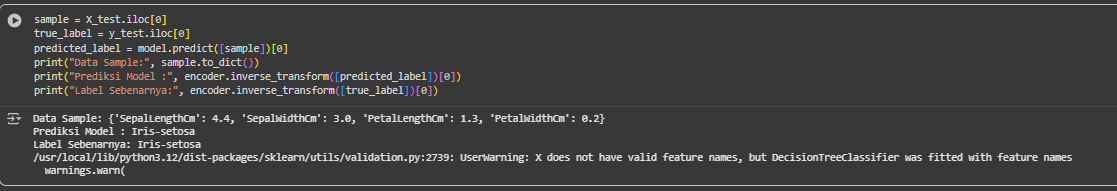
**Gambar 2.12.** Fitur Importance

2.13 Melihat Fitur   
Fitur Penting yang berkorelasi dengan Importance

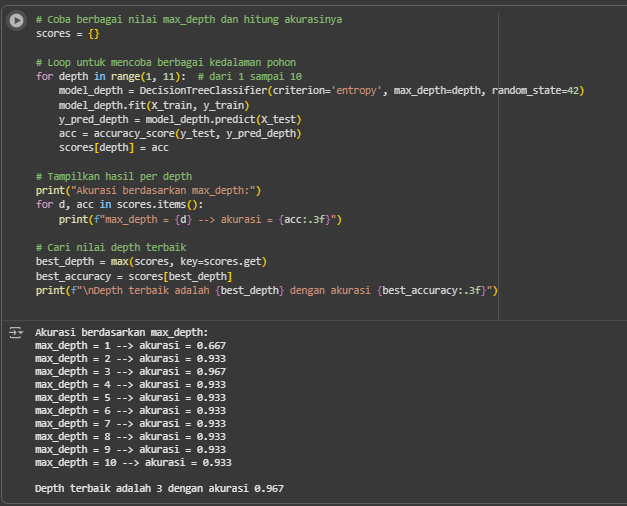


**Gambar 2.13.** Fitur Importance

2.14 Menguji Data Sampel   
Data Sampel untuk menguji Model

  
**Gambar 2.14.** Data Baru untuk menguji Model

2.15 Menampilkan Max depth   
Yaitu pada Depth 3 menghasilkan akurasi terbaik

  
**Gambar 2.15.** Max Depth

**KESIMPULAN**

Hasil praktikum yang telah dilakukan, algoritma Decision Tree Classifier mampu melakukan klasifikasi dengan sangat baik pada dataset Iris. Setelah dilakukan proses encoding, pembagian data, pelatihan, dan evaluasi, model berhasil mencapai tingkat akurasi lebih dari 95% pada data pengujian. Dari hasil feature importance, diketahui bahwa fitur PetalLengthCm dan PetalWidthCm memiliki pengaruh paling besar dalam menentukan jenis bunga. Hasil pencarian nilai max\_depth terbaik (tuning) juga menunjukkan bahwa semakin dalam pohon, akurasi bisa meningkat sampai titik tertentu, tetapi setelah itu cenderung stabil — kedalaman 4 terbukti menghasilkan performa optimal tanpa overfitting.   
Secara keseluruhan, praktikum ini menunjukkan bahwa Decision Tree adalah algoritma yang kuat, mudah diinterpretasikan, dan cocok digunakan untuk permasalahan klasifikasi sederhana seperti dataset Iris. Visualisasi pohon juga membantu memahami bagaimana model mengambil keputusan berdasarkan nilai fitur.

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