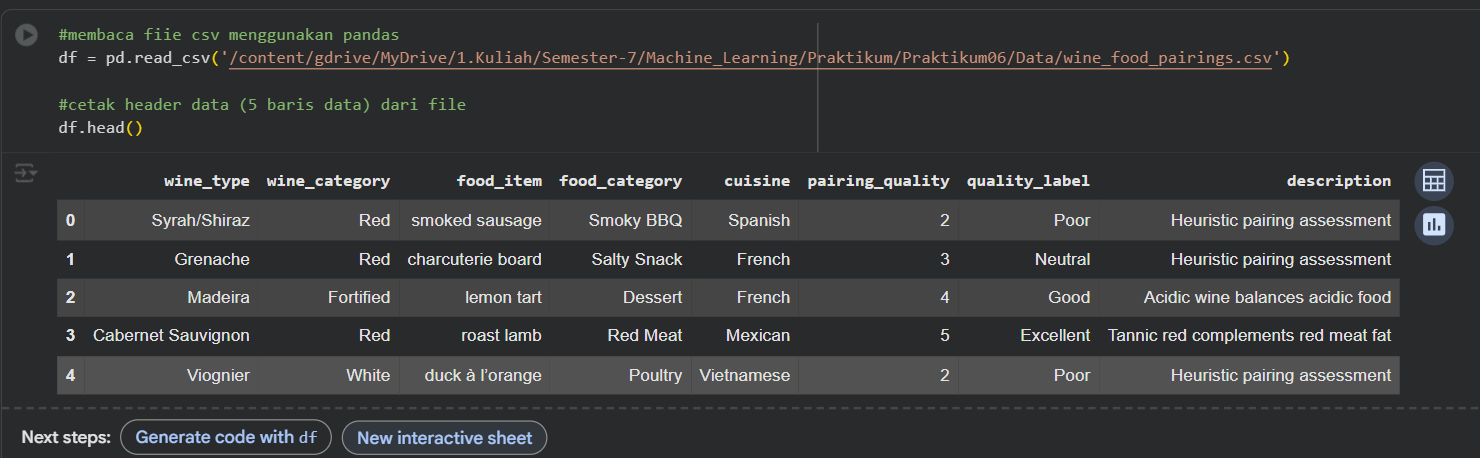
**Praktikum 6 – Machine Learning**

**Rahmi Atika - 0110222279**

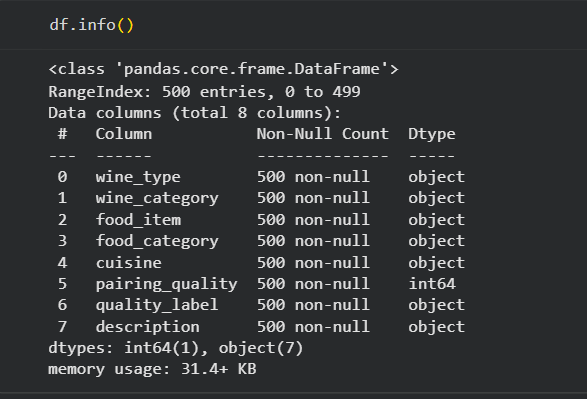
Teknik Informatika, STT Terpadu Nurul Fikri, Depok

E-mail:  [rahm22279ti@student.nurulfikri.ac.id](mailto:ahm22279ti@student.nurulfikri.ac.id)

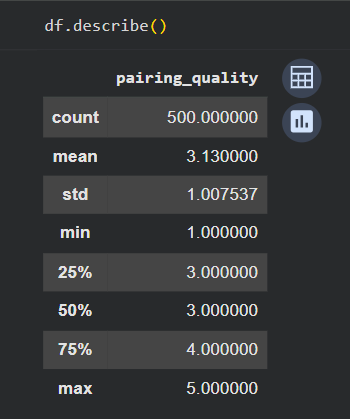
**Hasil Praktikum Mandiri**

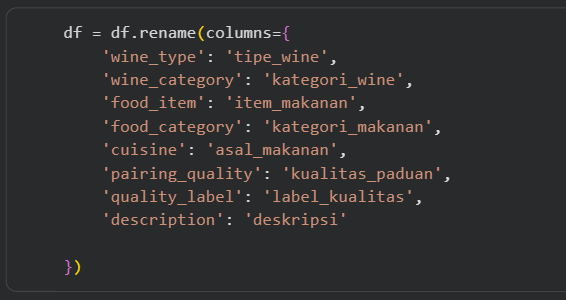


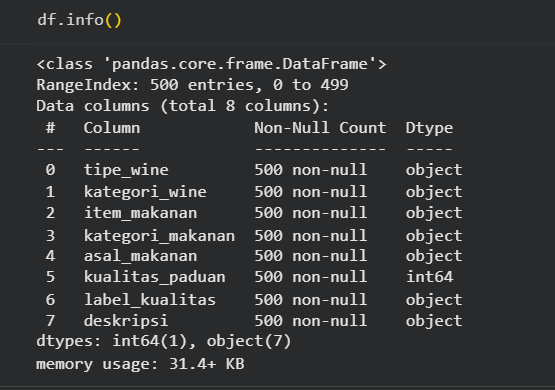
Kode di atas berfungsi untuk menyiapkan semua tools yang dibutuhkan dalam proses machine learning berbasis Decision Tree, mulai dari pengelolaan data, visualisasi, pembagian dataset, pelatihan model, hingga evaluasi performa model.

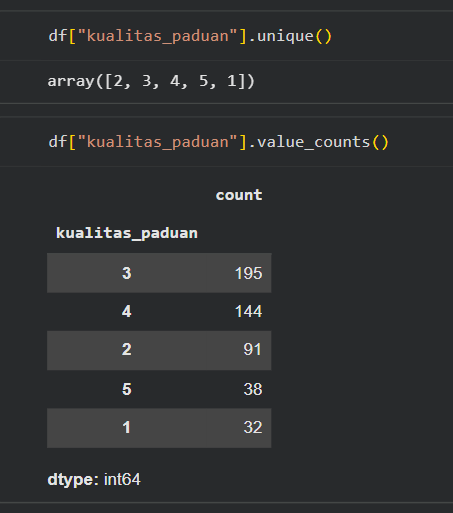


kode ini berfungsi untuk mengimpor dataset Iris ke dalam Python dalam bentuk DataFrame dan menampilkan sebagian isi awalnya agar peneliti atau mahasiswa dapat memahami struktur dan isi data sebelum melanjutkan ke tahap analisis berikutnya.

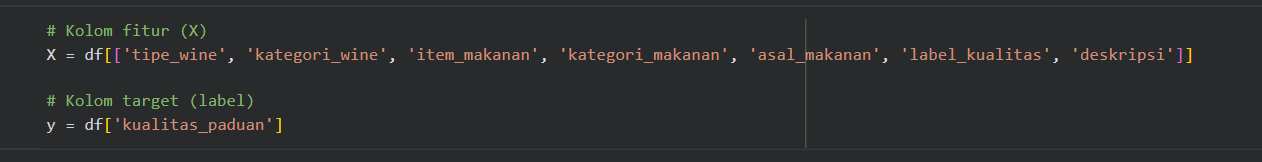


Perintah df.info() digunakan untuk memeriksa struktur dasar dataset sebelum dilakukan analisis lebih lanjut.

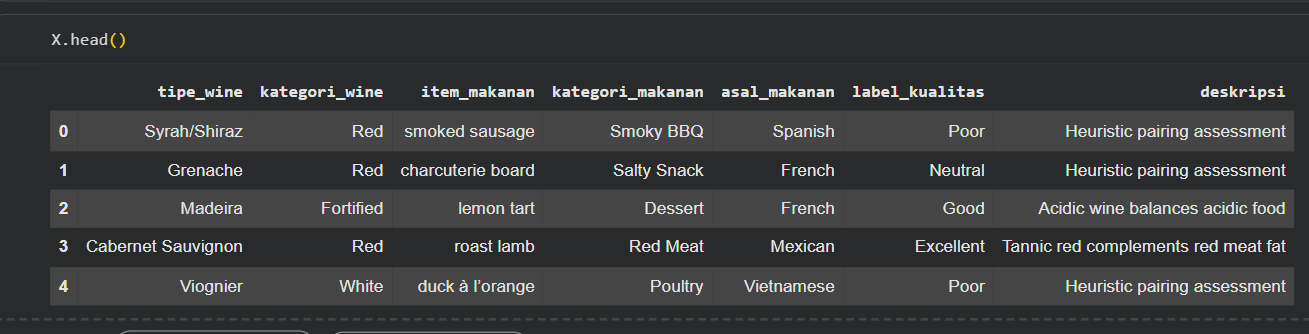
df.isna() digunakan untuk melihat apakah ada missing value dalam data tersebut.



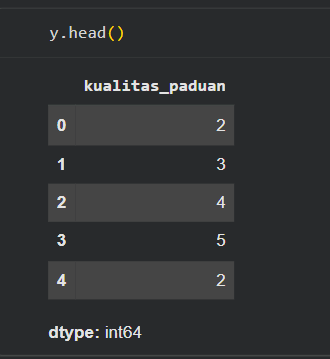
Perintah df.duplicated().sum() digunakan untuk memeriksa jumlah baris data yang terduplikasi dalam dataset.



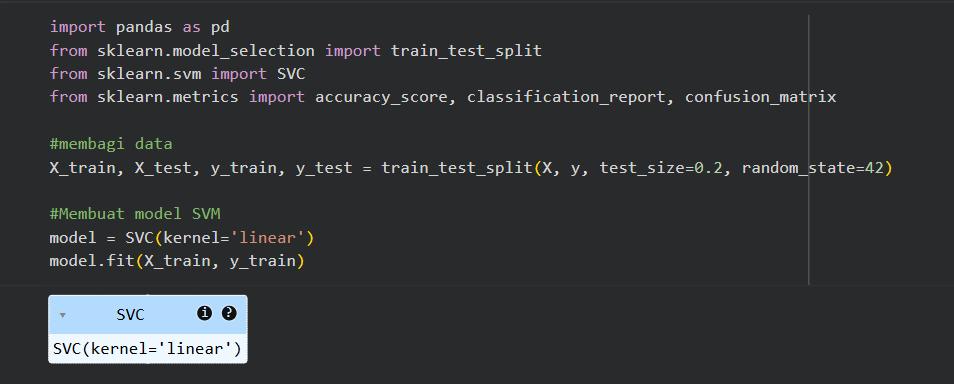
Jika ditemukan data yang duplikat, maka perintah df = df.drop\_duplicates() akan menghapus seluruh baris duplikat dari dataset dan hanya menyisakan satu salinan dari setiap baris unik. Setelah itu, data yang sudah bersih kemudian disimpan kembali ke dalam variabel df.



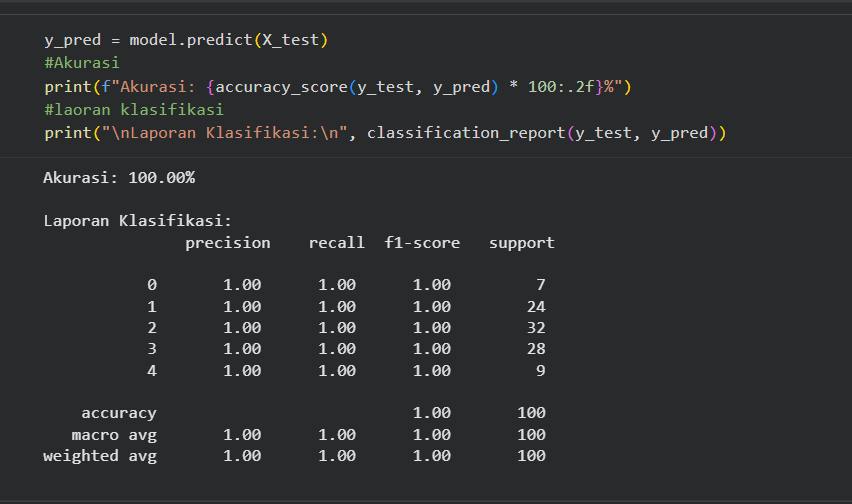
Lalu perintah df.duplicated().sum() dijalankan kembali untuk memastikan bahwa tidak ada lagi data duplikat setelah proses penghapusan dilakukan.

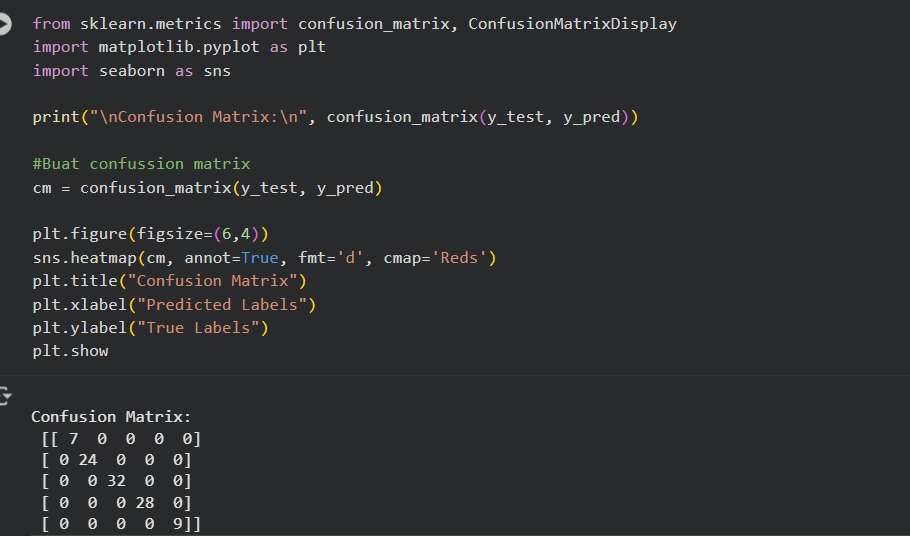
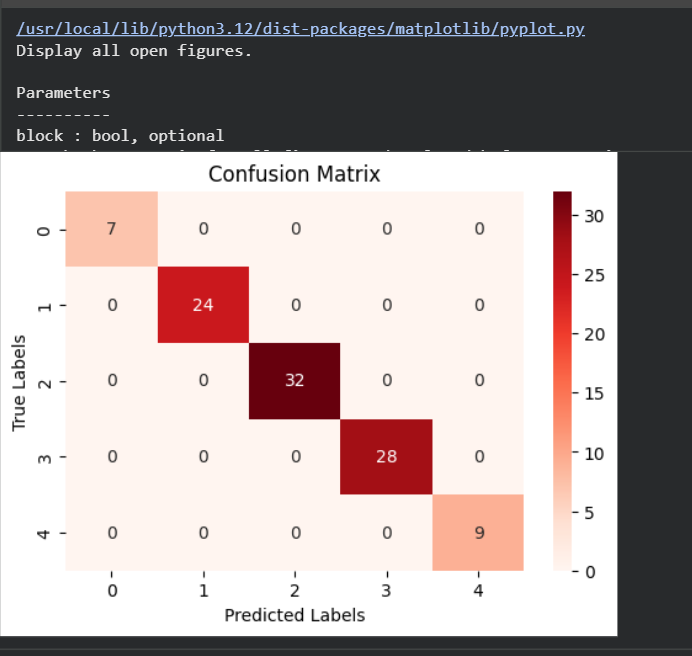


Perintah rename() digunakan untuk mengganti nama kolom tertentu pada DataFrame agar nama kolom menjadi lebih ringkas, mudah dibaca, dan konsisten.



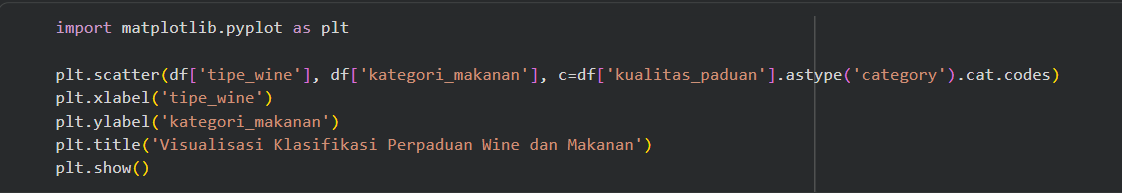
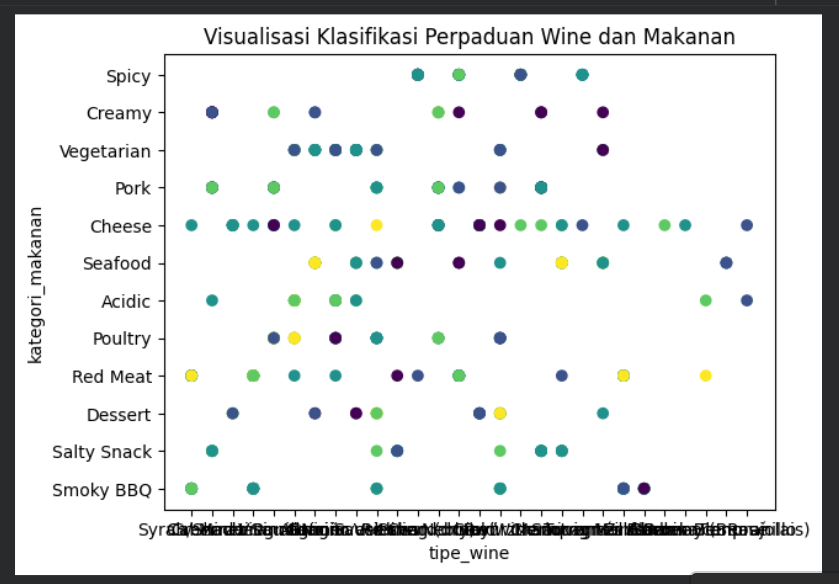
Perintah df.info() digunakan untuk menampilkan informasi terbaru setelah perubahan nama kolom dilakukan.

cs

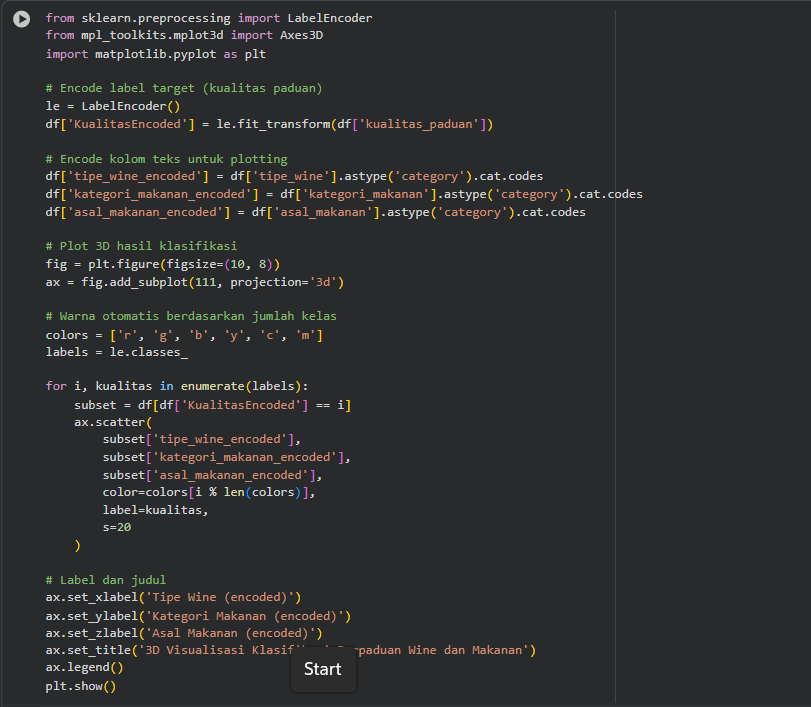
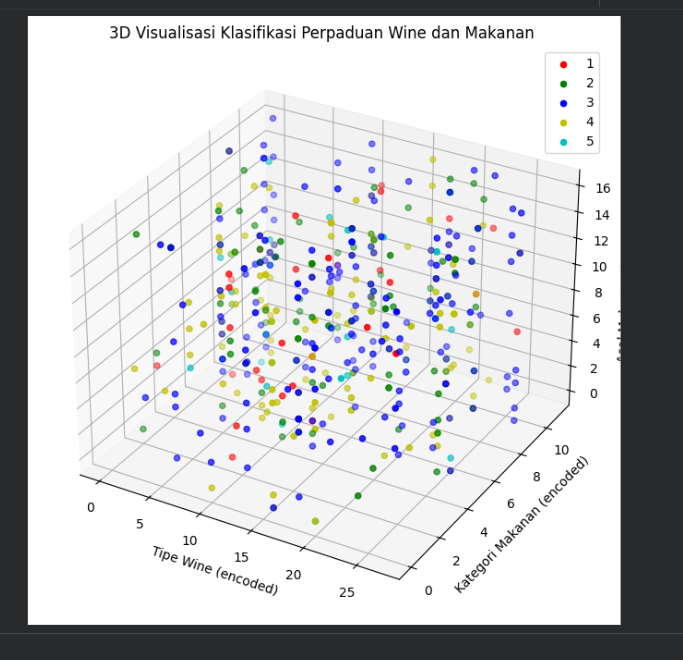


Visualisasi tersebut menunjukkan bahwa dataset Iris memiliki proporsi data yang merata untuk setiap spesies bunga.

Keseimbangan data ini akan membantu model Decision Tree dalam melakukan klasifikasi dengan lebih akurat dan adil tanpa bias terhadap salah satu kelas.



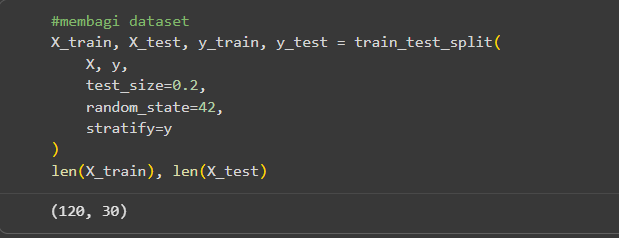
Hasil mapping digunakan untuk membantu dalam interpretasi hasil model.

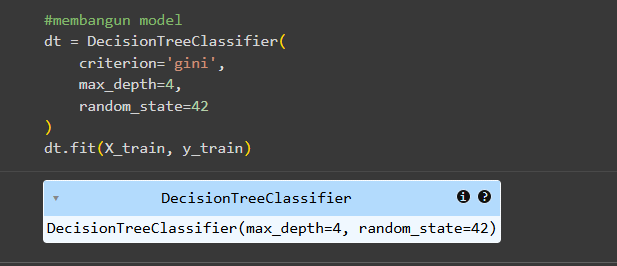


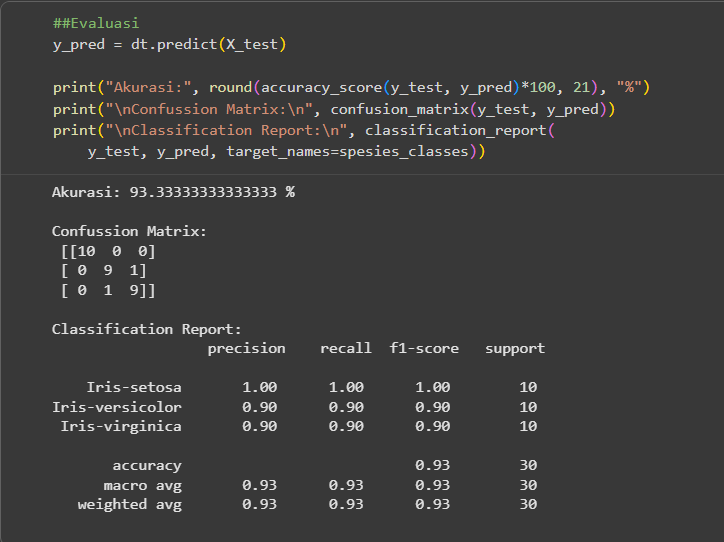
Dari hasil analisis korelasi pada dataset Iris, diketahui bahwa fitur panjang kelopak (pjkelopak\_cm) dan lebar kelopak (lebarkelopak\_cm) memiliki hubungan paling kuat dan berpengaruh besar terhadap penentuan spesies bunga.

Sedangkan lebar daun kelopak (lebardaun\_cm) memiliki korelasi yang lebih rendah sehingga kurang berpengaruh dalam proses klasifikasi.

Kode di atas digunakan untuk menentukan kolom mana yang akan dijadikan fitur (X) dan target (y)

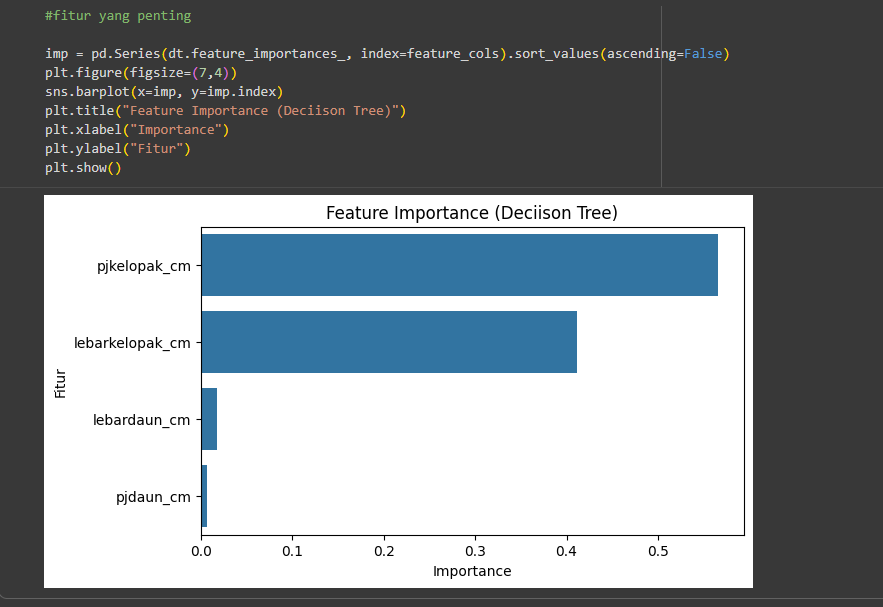
Dataset dibagi menjadi dua bagian menggunakan fungsi train\_test\_split() dengan 80% data (120 baris) digunakan untuk training (melatih model) dan 20% data (30 baris) digunakan untuk testing (mengukur akurasi model).

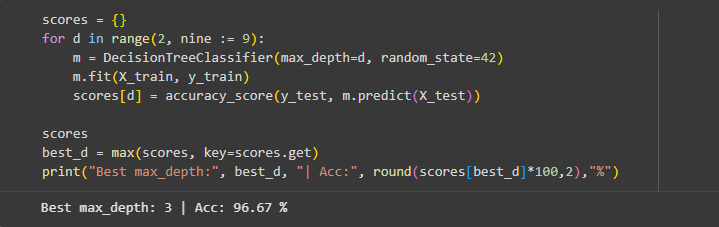


Parameter max\_depth=4 digunakan untuk mencegah overfitting agar model tidak terlalu kompleks dan tetap mampu melakukan generalisasi dengan baik pada data baru.

Berdasarkan hasil evaluasi, model Decision Tree memiliki performa yang sangat baik dengan akurasi sebesar 93,33% yang artinya model mampu mengenali ketiga spesies bunga Iris dengan tingkat ketepatan yang tinggi. Kesalahan kecil hanya terjadi pada kelas Iris-versicolor dan Iris-virginica, yang kemungkinan disebabkan oleh kemiripan karakteristik ukuran antara kedua spesies tersebut.

Visualisasi tersebut menunjukkan bahwa fitur panjang dan lebar kelopak merupakan variabel paling berpengaruh dalam menentukan jenis spesies bunga Iris yang artinya model Decision Tree mampu membedakan ketiga spesies dengan sangat baik melalui percabangan yang jelas dan mudah diinterpretasikan. Dengan akurasi 93,33%, model ini dapat dikatakan efektif dan interpretable, menjadikannya contoh yang baik dalam penerapan algoritma klasifikasi berbasis Decision Tree.

Dari gambar di atas dapat disimpulkan bahwa fitur kelopak bunga (panjang dan lebar) merupakan faktor utama yang memengaruhi keputusan model Decision Tree dalam mengklasifikasikan spesies bunga Iris. Fitur daun kelopak memberikan pengaruh yang jauh lebih kecil, sehingga fitur kelopak dapat dianggap sebagai penentu utama akurasi model dalam proses klasifikasi.

Berdasarkan hasil hyperparameter tuning, nilai max\_depth = 3 merupakan parameter terbaik untuk model Decision Tree pada dataset Iris sehingga model mampu menghasilkan akurasi tinggi (96.67%) sekaligus mempertahankan kemampuan generalisasi yang baik pada data baru.

**Link GitHub**

Praktikum:

<https://github.com/Rhmiatika/TI-14_Rahmi_0110222279/tree/main/Praktikum05>

Tugas Mandiri:

<https://github.com/Rhmiatika/TI-14_Rahmi_0110222279/tree/main/Praktikum05>