Kode program yang akan diajarkan di sini bisa diunduh di tautan [berikut ini](https://colab.research.google.com/drive/1tPLDXgz6DORmUKgdUtm030cVoAFy3wLh?usp=sharing). Untuk membukanya, upload berkasnya ke Google Colab.

**Tujuan**

Pada latihan ini, kita akan melakukan klasifikasi data yang kita miliki dengan teknik Decision Tree menggunakan **dataset iris**, salah satu dataset paling populer yang sering digunakan untuk belajar machine learning.

**Tahapan Latihan**

Dataset iris terdiri dari 4 atribut yaitu panjang sepal, lebar sepal, panjang petal, dan lebar petal. Terdapat 3 kelas target pada dataset ini. Data ini digunakan untuk masalah klasifikasi, di mana kita memprediksi jenis spesies sebuah bunga berdasarkan atribut-atribut yang diberikan.

Tahapan yang ada pada latihan ini antara lain:

1. Ubah dataset ke dalam dataframe.
2. Hapus kolom 'Id' pada dataframe serta pisahkan antara atribut dan label.
3. Bagi dataset menjadi data latih dan data uji.
4. Buat dan latih model Decision Tree.
5. Lakukan pengujian model dengan menggunakan data uji.
6. Lakukan prediksi dengan model yang telah dilatih.
7. Visualisasi model Decision Tree yang telah dilatih.

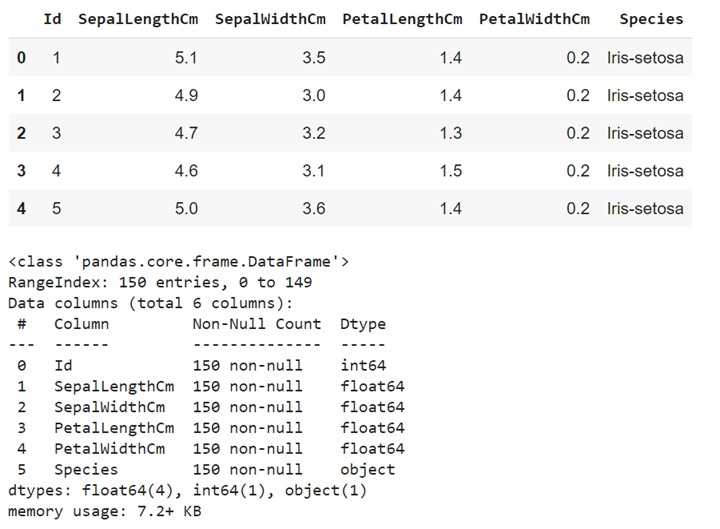
**Codelab**

Pertama kita akan mengimpor library yang dibutuhkan dan mempersiapkan dataset. Dataset dapat anda unduh di [tautan](https://www.kaggle.com/uciml/iris) berikut. Setelah data diunduh, masukkan berkas **Iris.csv** ke dalam Google Colab. Lalu jangan lupa konversi dataset menjadi Pandas dataframe.

1. import pandas as pd
3. # Membaca file iris.csv
4. iris = pd.read\_csv('Iris.csv')

Untuk melihat informasi mengenai data, gunakan fungsi info(). Selain itu, Anda juga bisa melihat lima data teratas pada dataset menggunakan fungsi head().

1. # Melihat informasi dataset
2. iris.info()
4. # melihat informasi dataset pada 5 baris pertama
5. iris.head()

[](https://www.dicoding.com/academies/184/tutorials/8397?from=8387)

Dari output di atas, kita dapat mengidentifikasi kolom yang tidak penting pada dataset yaitu kolom "Id". Untuk menghilangkan kolom tersebut, gunakan fungsi drop().

1. # menghilangkan kolom yang tidak penting
2. iris.drop('Id',axis=1,inplace=True)

Sebelum melatih model kita perlu memisahkan atribut dengan label. Selain itu, kita juga perlu membagi dataset menjadi data latih dan data uji. Jalankan kode berikut untuk menerapkan tahapan di atas.

1. # memisahkan atribut dan label
2. X = iris[['SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm' ]]
3. y = iris['Species']
5. # Membagi dataset menjadi data latih & data uji
6. from sklearn.model\_selection import train\_test\_split
7. X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.1, random\_state=123)

Selanjutnya, definisikan model decision tree yang akan kita gunakan. Kemudian,  latih model menggunakan data latih menggunakan fungsi fit().

1. from sklearn.tree import DecisionTreeClassifier
3. # membuat model Decision Tree
4. tree\_model = DecisionTreeClassifier()
6. # Melatih model dengan menggunakan data latih
7. tree\_model = tree\_model.fit(X\_train, y\_train)

Setelah model dilatih, uji model menggunakan data uji untuk melihat seberapa baik model yang telah kita buat. Pengujian model ini bisa dilakukan dengan menggunakan fungsi predict().

Berikutnya, gunakan metrik akurasi untuk melihat seberapa baik model yang telah kita latih. Penjelasan terkait metrik akurasi ini akan dibahas pada modul selanjutnya.

1. # Evaluasi Model
2. from sklearn.metrics import accuracy\_score
4. y\_pred = tree\_model.predict(X\_test)
6. acc\_secore = round(accuracy\_score(y\_pred, y\_test), 3)
8. print('Accuracy: ', acc\_secore)

[dos:3b619564fdc0353bdcbd7b0f95dc889f20211126141845.jpeg](https://www.dicoding.com/academies/184/tutorials/8397?from=8387)

Nah, kita bisa mencoba model yang telah kita buat untuk memprediksi spesies dari sebuah bunga Iris. Masih ingat bukan, atribut yang menjadi masukan dari model adalah panjang sepal, lebar sepal, panjang petal, dan lebar petal? Kita akan memasukkan nilai yang sesuai dengan format tersebut secara berurutan dalam satuan centimeter. Pada contoh berikut, kita ingin memprediksi spesies dari sebuah bunga iris  yang memiliki panjang sepal 6,2 centimeter, lebar sepal 3,4 centimeter, panjang petal 5,4 centimeter, dan lebar petal 2,3 centimeter.

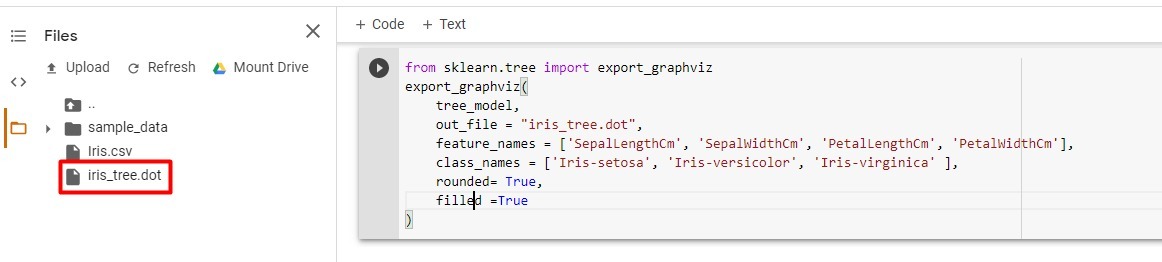
1. # prediksi model dengan tree\_model.predict([[SepalLength, SepalWidth, PetalLength, PetalWidth]])
2. print(tree\_model.predict([[6.2, 3.4, 5.4, 2.3]])[0])

[dos:28d001bfed0394b10d341d2368b4515c20211126142114.jpeg](https://www.dicoding.com/academies/184/tutorials/8397?from=8387)

Selain melakukan prediksi, kita juga bisa melihat visualisasi dari decision tree yang kita buat terhadap data menggunakan **library Graphviz**. Hasil dari graphviz adalah **dot file** yang akan muncul pada folder file di panel sebelah kiri Google Colab (jika Anda menggunakan Google Colab).

1. from sklearn.tree import export\_graphviz
2. export\_graphviz(
3. tree\_model,
4. out\_file = "iris\_tree.dot",
5. feature\_names = ['SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm'],
6. class\_names = ['Iris-setosa', 'Iris-versicolor', 'Iris-virginica' ],
7. rounded= True,
8. filled =True)

Setelah kode di atas berhasil dijalankan, Anda akan mendapatkan output berupa berkas**iris\_tree.dot**, sebagai berikut:

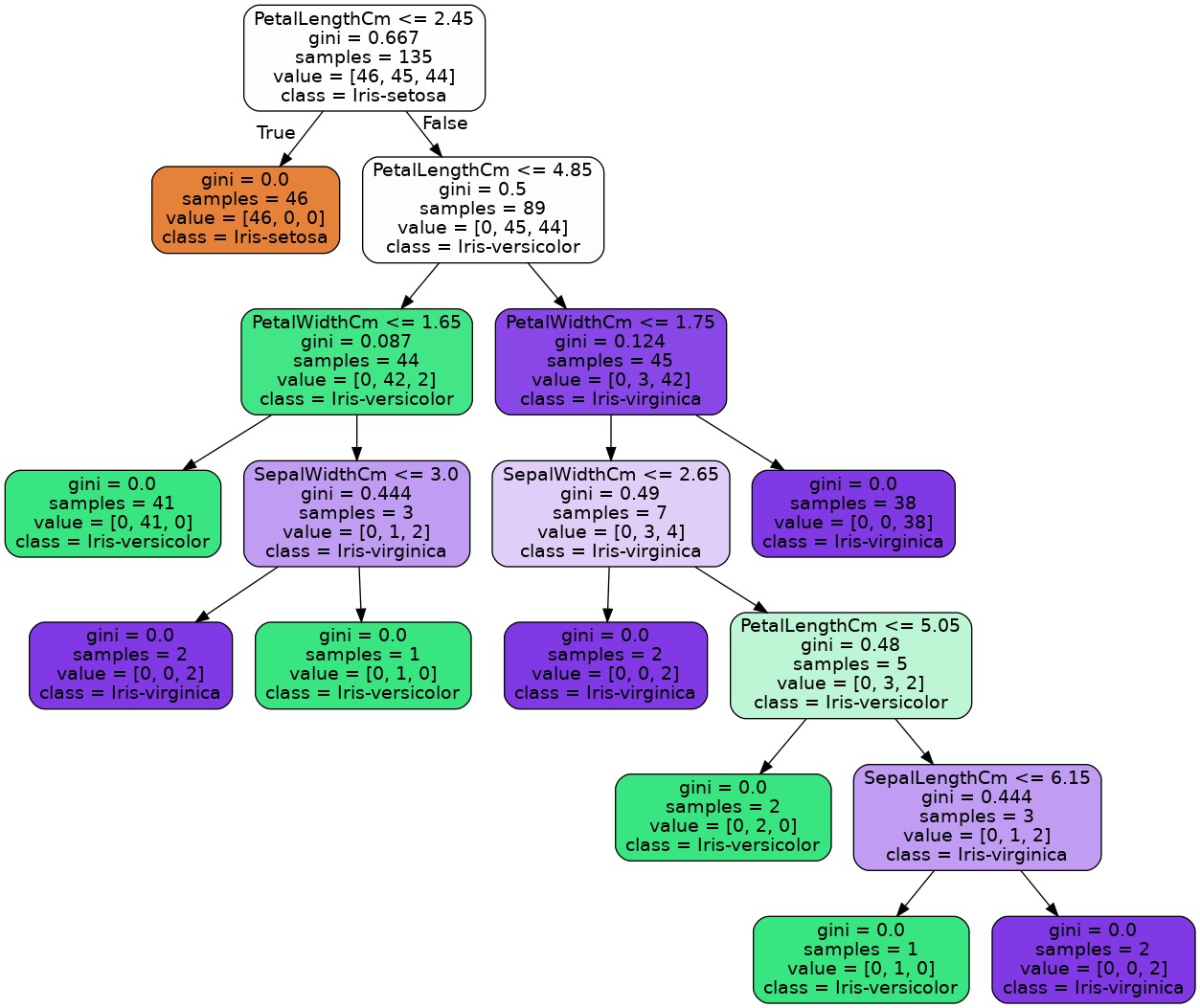
[](https://www.dicoding.com/academies/184/tutorials/8397?from=8387)

Untuk mengunduh berkas **iris\_tree.dot** pada gambar di atas, kita dapat melakukan klik kanan pada berkas tersebut kemudian mengunduhnya.

Jika kita ingin melihat visualisasi decision tree, lakukan konversi dot file ke dalam file **png** menggunakan situs konversi berkas berikut ini : <https://onlineconvertfree.com/converter/images/>.

Catatan : Jangan lupa ganti opsi ke images sebelum menekan tombol convert

Berikut merupakan hasil visualisasi dari model decision tree yang telah kita gunakan:

[](https://www.dicoding.com/academies/184/tutorials/8397?from=8387)

Selamat! Anda telah berhasil membuat sebuah model decision tree untuk klasifikasi spesies bunga Iris. Anda juga telah berhasil menguji model anda untuk memprediksi spesies dari sebuah bunga iris. Untuk belajar lebih mendalam tentang decision tree, kunjungi [tautan](https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052) berikut yah.