Kode program yang akan diajarkan di sini bisa diunduh di tautan [berikut ini](https://colab.research.google.com/drive/1qZcUnq94vln6I5tHV-qouhUpUEiRx8Gm?usp=sharing). Untuk membukanya upload berkasnya ke Google Colab.

**Tujuan**

Pada codelab kali ini kita akan menggunakan *cross\_validation\_score* pada *classifier decision\_tree*. Dataset yang digunakan adalah dataset iris.

**Tahapan Latihan**

Tahapan yang dilakukan pada codelab ini sebagai berikut:

1. Impor library yang dibutuhkan.
2. Pisahkan antara atribut dan label pada dataset.
3. Buat model *decision tree*.
4. Hitung hasil*cross validation* dari model dengan fungsi *cross\_val\_score()*.

**Codelab**

Para pengembang ML biasanya meng-*import  s*emua *library* yang dibutuhkan di *cell* pertama. Dalam tahap latihan ini, Anda akan melakukan *import library*pada *cell*yang berkaitan saja. Tujuannya adalah agar Anda memahami fungsi setiap *library* yang digunakan dalam model ML yang Anda buat.

Dataset yang akan kita gunakan adalah dataset iris yang dipakai pada submodul sebelumnya.

1. import sklearn
2. from sklearn import datasets
4. # Load iris dataset
5. iris = datasets.load\_iris()

Kemudian kita bagi antara atribut dan label pada dataset.

1. # mendefinisikan atribut dan label pada dataset
2. x=iris.data
3. y=iris.target

Kita akan membuat model machine learning pertama kita yaitu decision tree, menggunakan *library scikit learn*. Model machine learning juga sering disebut sebagai classifier. Lebih lanjut, variabel clf adalah singkatan dari *classifier*.

1. from sklearn import tree
3. # membuat model dengan decision tree classifier
4. clf = tree.DecisionTreeClassifier()

Setelah dataset dan model siap, kita bisa menggunakan cross validation untuk mengevaluasi performa dari model machine learning. Fungsi *cross\_val\_score()* seperti di bawah menerima 4 parameter yaitu, ‘clf’ yang merupakan model machine learning, ‘X’ yang merupakan atribut dari dataset, ‘y’ yang merupakan label dari dataset, dan ‘cv’ yang merupakan jumlah *fold*yang akan dipakai pada *cross validation.*

1. from sklearn.model\_selection import cross\_val\_score
3. # mengevaluasi performa model dengan cross\_val\_score
4. scores = cross\_val\_score(clf, x, y, cv=5)

*Cross\_val\_score* mengembalikan nilai berupa larik atau array yang terdiri dari akurasi pengujian setiap *fold*dari dataset. Untuk mencetak dan mengetahui hasilnya, tambahkan kode scores di bawah kode sebelumnya. Tampilannya seperti gambar di bawah ini.

Elemen pertama dari larik menunjukkan nilai 0.96666 yang berarti ketika *fold* pertama dijadikan *validation set* dan fold lainnya dijadikan train set, hasil dari pengujian tersebut adalah akurasi sebesar 0.96666.

Melihat akurasi dari seluruh pengujian fold yang memiliki nilai tinggi dan konsisten pada tiap fold, kita mendapatkan gambaran bahwa model kita memiliki performa yang sangat baik.

Secara umum jika hasil dari pengujian tiap fold pada *cross validatio*n memiliki nilai yang bervariasi dari 0.85 sampai 0.99, maka model tersebut dapat dikatakan baik.