



## **Tutorial 2: Scikit-learn (Sklearn)**

### **Ketentuan**

1. Dokumen *template* tutorial dengan format .ipynb dan dataset (jika dibutuhkan) telah disediakan di SCeLe.
2. Tutorial ini dirancang agar mahasiswa dapat menggunakan *library* yang akan digunakan dalam pengerjaan lab dan tugas yang akan datang.
3. **Tidak ada submisi** dalam pelaksanaan tutorial. Silakan berdiskusi jika mengalami kesulitan. Meski demikian, diharapkan setelah pengerjaan tutorial ini mahasiswa dapat melakukan *task* yang diberikan di dalam tutorial **secara mandiri**.

### **A. Scikit-learn Secara Umum**

**Scikit-learn** sudah disinggung di lab sebelumnya, namun belum ada penjelasan tentang apa itu Scikit-learn. Scikit-learn (Sklearn) merupakan sebuah *library* yang sangat bermanfaat dalam pembelajaran mesin untuk bahasa pemrograman Python. Sklearn sendiri dibangun menggunakan beberapa *library* yang kita juga sudah kenal sebelumnya, yaitu Numpy, Scipy, dan Matplotlib.

Sklearn dibuat terkhusus untuk memodelkan data, yang meliputi **membangun** (build), **melatih** (train), dan **mengevaluasi** (evaluate) model dari sebuah kumpulan data. Untuk itu, tutorial kali akan berfokus bagaiman melakukan ketiga hal tersebut yang akan bermanfaat dalam pengerjaan lab dan tugas yang akan datang.

Model yang dapat dibangun menggunakan Sklearn seperti Decision Tree, Linear/Logistic Regression, Support Vector Machine (SVM), dan masih banyak model yang lain.

### **B. Penggunaan Scikit-learn**

Silakan buka berkas Tutorial2.ipynb menggunakan Jupyter Notebook, Google Colab, atau aplikasi lain. Terdapat instruksi yang dapat dijalankan untuk mempelajari bagaimana cara membangun, melatih, dan mengevaluasi sebuah model dari data yang disediakan.

### C. Deskripsi Dataset *wine*:

[NOTE] Deskripsi Dataset ini dilansir dari tautan berikut:

<https://archive.ics.uci.edu/ml/datasets/wine>

Dataset berisi hasil analisis kimiawi dari anggur yang tumbuh dalam daerah yang sama di Italia namun diturunkan dari 3 (tiga) kultivar yang berbeda. Analisis tersebut menghasilkan sebanyak 13 (tiga belas) atribut yang ditemukan pada tiga tipe anggur. Berikut merupakan deskripsi dari setiap atribut pada dataset:

- `alcohol_type` (target): kultivar alkohol dari tiga kultivar yang tersedia
- `alcohol_percentage`: total kandungan alkohol (dalam persentase) dari kultivar tersebut
- `malic_acid`: jumlah komponen organik yang menimbulkan rasa asam pada anggur
- `ash`: total zat yang tersisa setelah proses pengabuan sisa penguapan anggur (terdiri dari alkali, alkali tanah, dan fosfor non logam)
- `alcanity_of_ash`: total kandungan kation (selain ammonia) dan asam organik dalam alkohol dari total zat sisa pengabuan
- `magnesium`: total magnesium dalam alkohol
- `total_phenols`: total komponen fenol (flavonoid dan proantosianidin) dalam kultivar tersebut
- `flavonoids`: jenis fenol yang memberikan komponen warna kuning pada alkohol
- `nonflavanoid_phenols`: jenis fenol memberikan komponen warna lain pada alkohol
- `proanthocyanidins`: jenis fenol yang memberikan komponen warna merah pada alkohol
- `color_intensity`: properti warna yang menandakan intensitas/kecerahan dari alkohol
- `hue`: properti warna berupa panjang gelombang warna yang dominan untuk menentukan seberapa merah atau biru sebuah warna
- `OD280/OD315_of_diluted_wines`: tingkat konsentrasi protein pada anggur
- `proline`: jenis asam amino dalam anggur, menandakan total kandungan zat prolin dalam kultivar tersebut