


<p>Nama :</p> <p>NIM :</p>	 <p>IF5171 Pembelajaran Mesin</p> <p>DSAI</p>	<p>Dosen Pengampu:</p> <p>Dr. Masayu Leylia Khodra, ST, MT</p>
<p>Hari, Tanggal:</p> <p>Kamis, 11 November 2022</p>		<p>Asisten Praktikum:</p> <p>1. Aaz M. Hafidz Azis 2. Dimmas Mulya 3. Mishka Alditya Priatna</p>

Petunjuk Praktikum 4 – Clustering

I. Persiapan

- a. Pada praktikum ini akan dilakukan dengan mengunduh dataset yang akan digunakan selama proses praktikum. Berikut adalah dataset yang akan digunakan yaitu *Mall Customer*.

Keterangan variabel:

Variabel	Keterangan
<i>CustomerId</i>	: ID customer dari responden/ <i>instance</i>
<i>Gender</i>	: Jenis kelamin dari responden/ <i>instance</i>
<i>Age</i>	: Umur dari responden/ <i>instance</i>
<i>Annual Income</i>	: Jumlah pendapatan tahunan
<i>Spending Score</i>	: skor pembelian

Link download:

https://drive.google.com/file/d/1ykJLVy0ohjadhP8m9SMOwjXsHSZ_mkoG/view?usp=share_link

Jika sudah berhasil mengunduh semua dataset yang telah dijabarkan, maka dapat dilakukan ke tahap selanjutnya.

- b. Peserta dibebaskan menggunakan *library* yang mendukung *task*.

II. Pengerjaan

- a. *Exploratory Data Analysis* (memberikan keterangan/penjelasan secara singkat pada gambar/grafik yang ditampilkan pada EDA. Grafik/plot yang digunakan diharapkan memberikan grafik yang berhubungan atau mendukung dengan *task* klasterisasi)
- b. *Feature Engineering* (tahapan ini boleh dilakukan apabila diperlukan untuk mendukung *task*)
- c. *Modeling*
 - i. *K-Means*
 1. Tampilkan visualisasi klaster yang terbentuk dengan melakukan 3 variasi klaster berdasarkan nilai **n_clusters** (Jumlah variasi klaster dibebaskan)
 2. Dari visualisasi yang telah dilakukan, Lakukan analisis apakah terjadi *overlap* atau *outlier*.
 - ii. *Hierarchical Clustering*
 1. Tampilkan visualisasi klaster yang terbentuk dengan melakukan 3 variasi klaster berdasarkan nilai **n_clusters** (Jumlah variasi klaster dibebaskan)
 2. Tampilkan visualisasi *dendogram* satu dari tiga model yang terbentuk, kemudian berikan penjelasan singkat mengenai dendogram yang terbentuk
 - iii. *DBScan*

Tampilkan visualisasi klaster yang terbentuk dengan melakukan 3 variasi klaster berdasarkan nilai **eps** (Jumlah variasi klaster atau parameter dibebaskan)
 - iv. *Meanshift*

Tampilkan visualisasi klaster yang terbentuk dengan melakukan 3 variasi klaster berdasarkan nilai **bandwidth** (perhitungan dapat digunakan dengan bantuan fungsi `estimate_bandwidth` dari library `sklearn.cluster`).
- d. Analisis algoritma
 - i. Melakukan visualisasi perbandingan hasil klaster dari keempat algoritma yang digunakan dengan hasil terbaik menurut kalian. Berikan penjelasan perbandingan dari keempat algoritma yang digunakan.
 - ii. Pilih salah satu model yang telah Anda buat, kemudian buat skema *bussniness approach* yang dapat dipecahkan menggunakan model yang anda buat.

