



KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS ANDALAS
FAKULTAS TEKNOLOGI INFORMASI
DEPARTEMEN SISTEM INFORMASI
KELOMPOK KEILMUAN REKAYASA DATA DAN BUSINESS INTELLIGENCE
Kampus Universitas Andalas, Limau Manis, Padang - 25163

INSTRUKSI MODUL 7:

K-MEANS

Sebuah lembaga internasional yang fokus pada pembangunan berkelanjutan dan kesejahteraan anak-anak, membutuhkan analisis mengenai hubungan antara pendapatan per kapita dan kesehatan anak sehingga dapat diketahui kondisi ekonomi dari negara tersebut. Hasil analisis tersebut akan digunakan untuk mengelompokkan negara sesuai dengan kondisi ekonomi mereka sehingga dapat ditentukan kebijakan/solusi yang tepat untuk masing-masing negara. Lembaga ini memiliki data [Country-data.csv](#) dengan atribut:

1. country (nama negara)
2. child_mort (angka kematian anak di bawah 5 tahun per 1000 kelahiran)
3. exports (nilai ekspor barang dan jasa per kapita yang dihitung sebagai persentase GDP per kapita suatu negara)
4. health (total pengeluaran untuk kesehatan per kapita yang dihitung sebagai persentase GDP per kapita)
5. imports (nilai impor barang dan jasa per kapita yang dihitung sebagai persentase GDP per kapita suatu negara)
6. income (penghasilan bersih per orang)
7. inflation (pengukuran tingkat pertumbuhan tahunan dari Total PDB)
8. life_expec (angka harapan hidup)
9. total_fer (jumlah rata-rata anak yang akan dilahirkan oleh wanita bila tingkat kesuburan berdasarkan kelompok usia tetap sama seperti sekarang)
10. gdpp (GDP per kapita)

Instruksi:

1. Import library dan dataset yang akan digunakan
2. Identifikasi dan pilih atribut yang relevan. Pemilihan atribut dapat melalui indexing atau drop atribut yang tidak diperlukan.
3. Lakukan pengecekan missing value dan duplikat. Bila terdapat missing value dan duplikat, hapus data tersebut untuk menghindari bias dan overfitting.



KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS ANDALAS
FAKULTAS TEKNOLOGI INFORMASI
DEPARTEMEN SISTEM INFORMASI
KELOMPOK KEILMUAN REKAYASA DATA DAN BUSINESS INTELLIGENCE
Kampus Universitas Andalas, Limau Manis, Padang - 25163

4. Ubah dataframe ke dalam bentuk array
5. Lakukan penskalaan untuk menyetarakan skala setiap atribut
6. Berikutnya hitung jumlah kluster optimal menggunakan metode elbow
7. Lakukan inisialisasi dan training model K-Means sesuai dengan jumlah kluster optimal yang berhasil diidentifikasi
8. Tambahkan atribut kluster ke dalam dataframe untuk memudahkan proses visualisasi
9. Lakukan visualisasi hasil klustering menggunakan scatter plot. Berikut merupakan beberapa ketentuan scatter plot yang perlu diikuti:
 - a. Warna centroid untuk setiap kluster adalah magenta
 - b. Centroid menggunakan marker d
 - c. Ukuran centroid 150
10. Lakukan visualisasi menggunakan boxplot untuk melihat sebaran pendapatan per kapita untuk masing masing kluster
11. Berdasarkan hasil visualisasi tersebut buatlah kesimpulan terkait jumlah kluster dan karakteristik tiap-tiap kluster agar dapat membantu lembaga internasional mengambil kebijakan yang tepat untuk setiap negara.

Catatan:

1. Lampirkan tangkapan **layar penuh (full screen)** untuk setiap tahapan instruksi
1. Bila ketahuan melakukan plagiarisme maka dianggap **tidak** mengumpulkan instruksi