

**Nama : Wahyu Hidayat**

**NIM : A11.2021.13336**

**Kelompok : A11.4404**

### **Tugas Akhir Data Mining**

**Dengan judul : Klastering Dengan Hasil Pertandingan Premier League 2022-2023**

**Deskripsi Eksperimen** Klasterisasi Tim Premier League menggunakan Algoritma K-Means

#### **Tujuan:**

Eksperimen ini bertujuan untuk melakukan klasterisasi (pengelompokan) tim-tim sepak bola Premier League berdasarkan hasil pertandingan mereka dengan menggunakan algoritma k-means. Klasterisasi dilakukan untuk mengidentifikasi kelompok tim yang memiliki pola permainan dan performa serupa dalam kompetisi.

#### **Dataset:**

Dataset yang digunakan dalam eksperimen ini berisi hasil pertandingan Premier League pada musim tertentu. Setiap baris dalam dataset mencakup informasi seperti tim-tim yang bertanding, skor pertandingan, posisi klasemen, jumlah gol yang dicetak, jumlah gol yang diterima, dan atribut lainnya yang relevan.

#### **Langkah-langkah Eksperimen:**

- 1. Memuat Data:** Pertama, data dari file '2023\_matchday\_results.csv' dimuat menggunakan library Pandas dan disimpan dalam variabel `dataset`.
- 2. Preprocessing Data:** Data yang relevan untuk klasterisasi dipilih dari dataset menggunakan slicing pada DataFrame Pandas. Pada contoh ini, kolom ke-4 dan ke-7 (indeks 3 dan 6 karena indeks dimulai dari 0).
- 3. Elbow Method:** Eksperimen dimulai dengan menggunakan metode "Elbow" untuk menentukan jumlah cluster yang optimal. Looping dilakukan dari 1 hingga 10 untuk mencoba berbagai jumlah klaster.
- 4. Visualisasi Elbow Method:** Setelah selesai melakukan loop, hasil inertia untuk setiap jumlah klaster ditampilkan dalam bentuk grafik menggunakan library Matplotlib. Tujuan dari visualisasi ini adalah untuk mencari "siku" dalam grafik.
- 5. Pemilihan Jumlah Klaster:** Berdasarkan visualisasi Elbow Method, kita dapat memilih jumlah klaster yang optimal. Pada contoh ini, jumlah klaster yang diambil adalah 5 (ditentukan secara manual).

**6. Klasterisasi:** Setelah jumlah klaster yang optimal ditentukan, model KMeans dilatih ulang menggunakan jumlah klaster yang dipilih (5 pada contoh ini) dengan inisialisasi 'k-means++'.

**7. Visualisasi Klaster:** Data hasil klasterisasi ditampilkan dalam bentuk scatter plot menggunakan Matplotlib.

**8. Evaluasi dengan Davies-Bouldin Index:** Mengukur kualitas klasterisasi, kita menggunakan Davies-Bouldin Index dengan bantuan library scikit-learn. Nilai indeks ini dihitung dan ditampilkan.

### **Kesimpulan:**

Eksperimen ini akan memberikan wawasan tentang bagaimana algoritma k-means dapat digunakan untuk mengelompokkan tim-tim sepak bola Premier League berdasarkan hasil pertandingan mereka. Dengan menggunakan metode klasterisasi ini, kita dapat memahami lebih baik pola dan karakteristik tim-tim dalam kompetisi, serta mendapatkan wawasan yang berguna bagi manajer tim, pelatih, atau analis sepak bola.

**Link Github :** [https://github.com/WahyuHidayat17/TA\\_DataMining](https://github.com/WahyuHidayat17/TA_DataMining)

**Link Dokumentasi :**

[https://drive.google.com/drive/folders/1ZNM9lCbUb4S\\_BZqEUh00vX-D1OP6a0b4](https://drive.google.com/drive/folders/1ZNM9lCbUb4S_BZqEUh00vX-D1OP6a0b4)