

K-Means

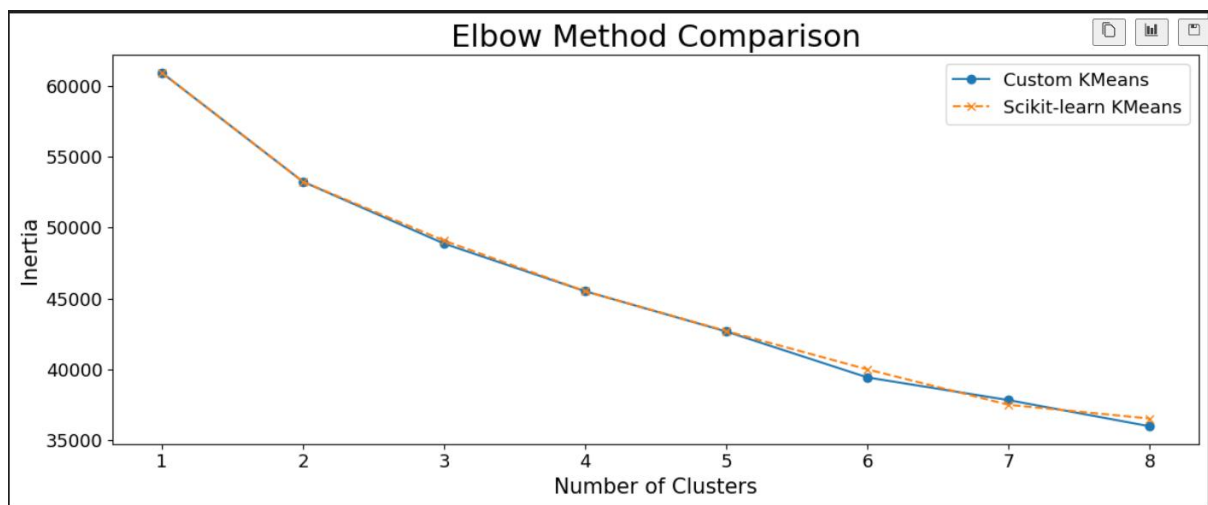
1. Cara kerja Algoritma

Secara garis besar, algoritma K-Means terdiri dari tiga tahap: *centroid initialization*, *cluster assigning*, dan *update centroid*. Pada tahap pertama, akan dipilih centroid/titik pusat berbeda sebanyak k (sebuah hyperparameter) secara acak. Titik pusat adalah titik tengah yang mewakili suatu cluster.

Pada tahap kedua, setiap titik data pada dataset akan dimasukkan ke kelompok milik titik pusat terdekatnya. Penentuan nilai jarak ada beberapa cara: euclidan, Manhattan, atau minkowski. Pada tahap ketiga, titik pusat setiap cluster akan dihitung ulang menjadi senilai dengan rata-rata nilai titik anggota cluster tersebut. Tahap dua dan tiga akan diulang-ulang sampai titik pusat tidak berubah pada tahap tiga.

Untuk menghasilkan cluster yang lebih baik secara lebih konsisten, dibuat teknik inisialisasi yang berbeda, yang dinamakan K-Means++. Pada inisialisasi versi ini, ditetapkan satu titik pusat dari titik data yang ada secara acak terlebih dahulu. Lalu, dipilih titik pusat selanjutnya yang merupakan titik data terjauh dari titik pusat yang ada. Langkah kedua ini diulang terus sebanyak k .

2. Perbandingan Hasil Evaluasi Model



Hasil inertia yang dihasilkan hampir sama. Perbedaan kecil ini terjadi karena perbedaan titik awal yang dipilih