

K-means

Cara Kerja:

1. Tentukan jumlah kluster k
2. Jika menggunakan metode inisialisasi secara *random*, pilih k *centroid* secara acak. Namun, jika menggunakan metode inisialisasi secara *k-means++*, pilih k *centroid* yang memiliki jarak terjauh satu sama lain.
3. Untuk setiap data, hitung jarak ke setiap *centroid* dan tetapkan data tersebut ke *centroid* terdekat.
4. Perbarui *centroid* untuk setiap *cluster* dengan menghitung rata-rata dari semua titik data yang ada pada *cluster* tersebut.
5. Ulangi langkah 3 dan 4 hingga *centroid* tidak berubah secara signifikan atau jumlah iterasi terpenuhi.

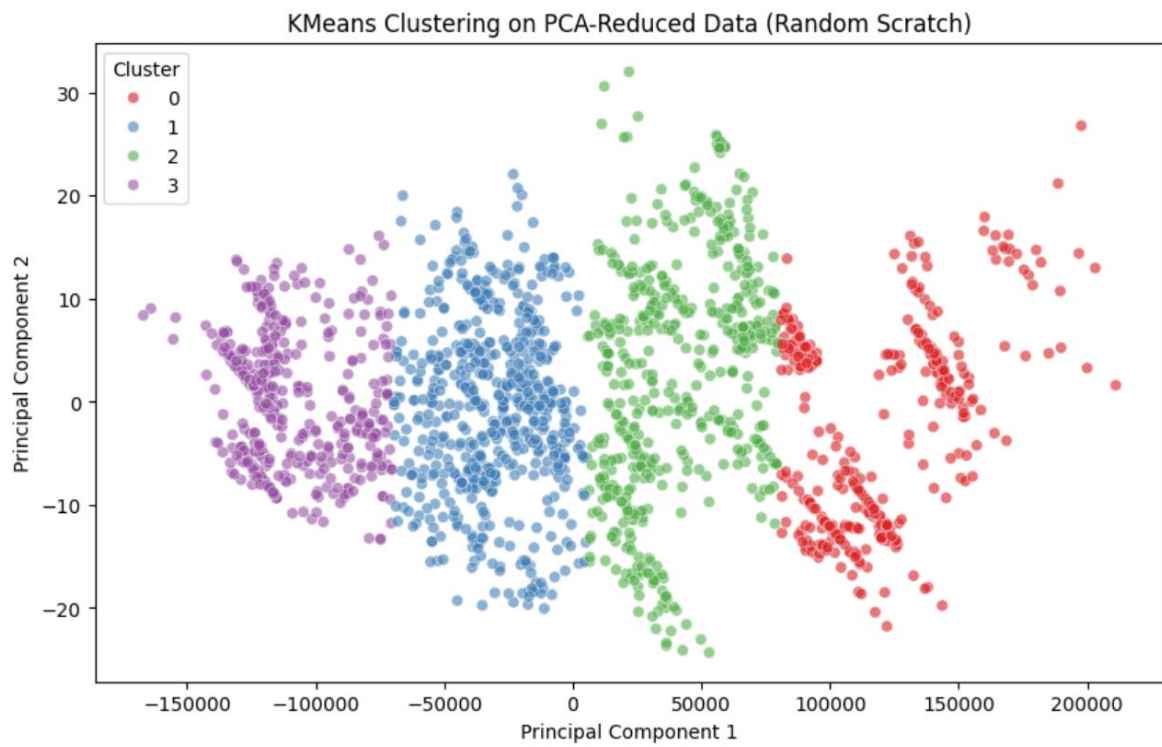
Perbandingan Hasil:



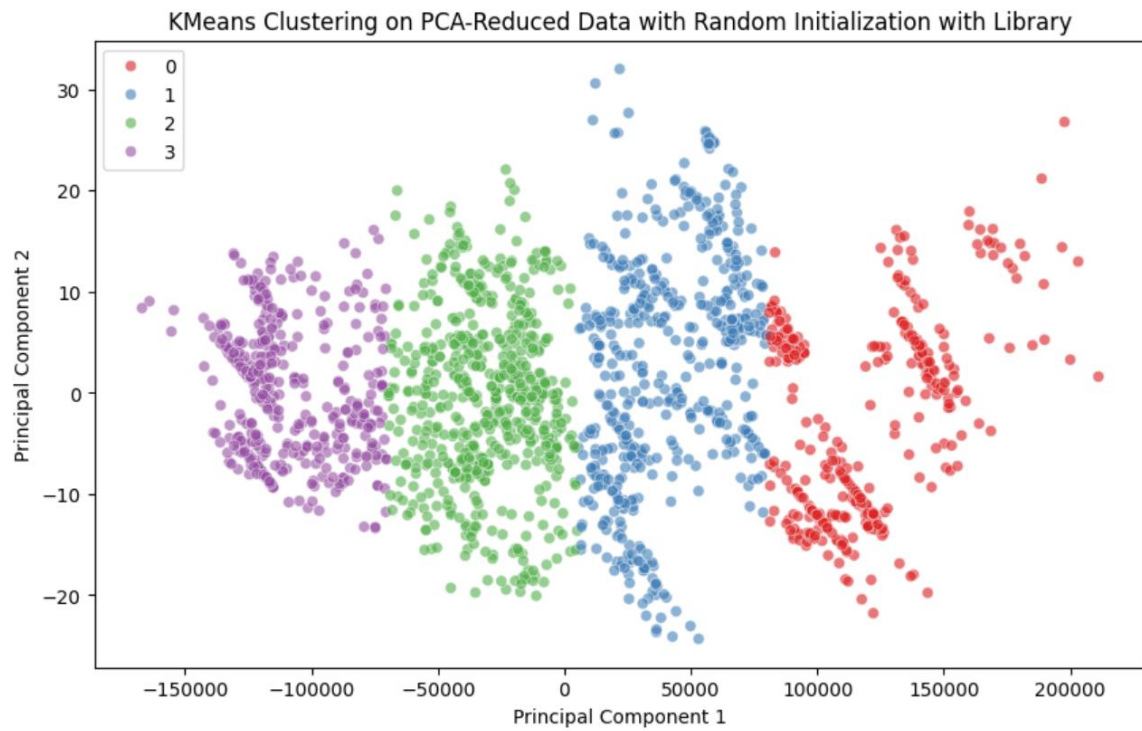
Gambar 1.1



Gambar 1.2



Gambar 1.3



Gambar 1.4

Gambar 1.1 dan gambar 1.2 adalah *k-means clustering* menggunakan inisialisasi *k-means++*. Gambar 1.3 dan gambar 1.4 adalah *k-means clustering* menggunakan inisialisasi *random*. Pada kedua hal ini, tidak terdapat perbedaan pada hasil *clustering* yang ada.