

| | |
|-------|---------------------|
| Nama | : Adhimas Aryo Bimo |
| NIM | : 13523052 |
| Model | : KMeans |

1. Algoritma K-Means Clustering bekerja dengan membagi data ke dalam k kelompok (cluster) berdasarkan jarak antar titik dan pusat cluster (centroid). Prosesnya iteratif sampai pusat cluster stabil (tidak banyak berubah). Berikut langkah-langkahnya:
 - a. Inisialisasi centroid
 - Pilih k titik awal sebagai centroid.
 - Bisa dengan metode random (acak) atau kmeans++ (lebih terarah agar lebih stabil).
 - b. Assign step (pemberian label)
 - Hitung jarak tiap titik ke semua centroid.
 - Setiap titik diberi label sesuai centroid terdekatnya.
 - c. Update step (update centroid)
 - Hitung rata-rata posisi dari semua titik di tiap cluster.
 - Jadikan rata-rata itu sebagai centroid baru.
 - c. Kriteria berhenti
 - Jika centroid baru sama (atau sangat mirip) dengan centroid lama, maka iterasi berhenti.
 - Atau, jika jumlah iterasi sudah mencapai max_iter.
 - d. Output
 - Centroid akhir
 - Label cluster untuk tiap data
 - Nilai loss bisa diukur dengan inertia (jumlah jarak kuadrat antar titik ke centroid).
2. Untuk melihat perbedaanya, saya menggunakan Silhouette Score. Sederhananya untuk mengukur kualitas clustering (seberapa baik cluster yang terbentuk) dengan nilai antara -1 hingga + 1. Semakin mendekati nilai 1 maka semakin baik cluster yang terbentuk. Pada hal ini score tersebut memiliki nilai berikut:


```
Custom KMeans - Silhouette Score: 0.0350
Sklearn KMeans - Silhouette Score: 0.0329
```

Hal ini menandakan Kmeans custom sedikit lebih baik dalam melakukan clustering. Hal yang membedakan dari kedua ini kemungkinan antara random rng nya yang berbeda atau default init_methodnya. Scikit-learn menggunakan init method kmeans++ sedangkan kode saya menggunakan random.