

K-Means

1. Jelaskan cara kerja dari algoritma tersebut! (boleh dalam bentuk *pseudocode* ataupun narasi)
 ⇒ Pertama, model menerima data masukan dan kemudian menentukan *centroid* secara acak sebanyak k . Titik *centroid* diambil secara acak dari data masukan. Setelah itu dilakukan pembentukan *cluster* dengan mengiterasi seluruh data masukan dan menghitung jaraknya terhadap setiap *centroid* yang ada dan pilih *centroid* yang memiliki jarak paling dekat untuk “masuk” ke *cluster* dari *centroid* tersebut. Setelah seluruh data “masuk” ke suatu *cluster*, selanjutnya melakukan penyesuaian *centroid* dengan menghitung nilai rata-rata dari seluruh data yang terdapat pada *cluster* tersebut. Ulangi hal tersebut sebanyak *maximum iteration* ataupun sampai *converged*, yaitu ketika *centroid* yang lama dan yang baru sudah tidak berubah.
2. Bandingkan hasil evaluasi model from scratch dan *library*, bagaimana hasil perbandingannya? Jika ada perbedaan, jelaskan alasannya!

Waktu Eksekusi	
<i>Scratch</i>	<i>Library</i>
0.405 detik	0.165 detik
Silhouette Score	
<i>Scratch</i>	<i>Library</i>
0.356	0.358

Berdasarkan data hasil percobaan, didapatkan bahwa nilai *silhouette score* tidak berbeda terlalu jauh, menandakan ke-”bagus”an pengelompokan data yang dilakukan kedua model sama bagusnya (atau jeleknya). Namun, dapat terlihat bahwa waktu eksekusi model yang dibuat sendiri lebih lama dibandingkan dengan model *library*, hal tersebut dapat dikarenakan model yang dibuat sendiri tidak mengimplementasikan untuk optimasi kecepatan pemrosesan data.