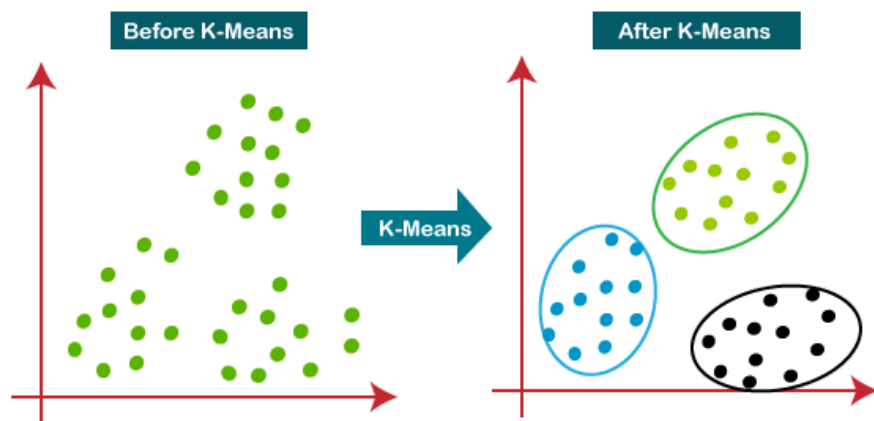


K-Means

Salman Hanif - 13523056

Cara Kerja Algoritma

Algoritma K-Means memiliki konsep pengelompokan dan perhitungan jarak yang sama dengan KNN, tetapi pada kasus K-Means adalah untuk objektif clustering. K-Means adalah unsupervised learning yang bekerja dengan mengelompokkan data berdasarkan jarak fiturnya, yang dekat dengan centroid dari setiap cluster akan menjadi bagiannya.



Pada Algoritma class KNN yang dibuat, fungsi tunggal untuk clustering langsung di fungsi predict, tanpa fit yang biasanya umum di ML. Langkah-langkah cara kerjanya:

1. Secara acak, k titik data dipilih dari dataset untuk dijadikan sebagai centroid awal. Centroid ini akan menjadi titik referensi untuk pembentukan kluster.
2. Kemudian untuk setiap titik data dihitung jaraknya ke setiap centroid. Jarak terdekat ke satu centroid, maka dianggap golongannya.
3. Setelah semua data di-plot klasternya, centroid akan diperbarui menjadi suatu titik imajiner yang lokasinya merupakan rata-rata dari semua titik kluster.
4. langkah 2 dan 3 diulangi hingga posisi centroid terus bergeser dan data terbagi ke kluster yang tepat

Hasil Evaluasi model dari hasil scratch dan dari library. Jelaskan perbedaan

Dalam evaluasi model clustering yang sudah dibuat, Kmeans yang saya buat memang menghasilkan hasil klaster yang sedikit berbeda dengan KMeans dari library sklearn.

Saya merasa KMeans saya masih memiliki beberapa kekurangan dibanding sklearn, contohnya inisialisasi centroid yang saya buat masih memilih data random, bukan metode k-means++. Hal ini akan membuat data tidak konsisten karena pemilihan yang acak.

Improvement yang bisa dilakukan

Improvement yang bisa saya lakukan adalah mengembangkan algoritma k-means++ agar inisialisasi centroid lebih tepat dan stabil. Kemudian pengembangan dengan menentukan jumlah klaster dinamis yang mencari jumlah k optimal untuk setiap dataset.