

Seleksi Asisten Lab AI '23

Unsupervised Learning: K-Means

1. Cara Kerja Algoritma K-Means.

Konsep inti: K-Means adalah algoritma yang bertujuan untuk mempartisi pengamatan ke dalam k cluster di mana setiap pengamatan termasuk dalam cluster dengan mean (centroid) terdekat.

Proses kerjanya iteratif, dimana awalnya secara acak akan dipilih k titik data dari dataset untuk dijadikan sebagai centroid awal. Untuk setiap titik data, dihitung jaraknya ke setiap k centroid. Kemudian, titik data tersebut akan ke cluster dari centroid terdekat. Setelah itu, akan dihitung ulang posisi centroid untuk setiap cluster. Centroid baru adalah titik rata-rata (mean) dari semua titik data yang termasuk dalam cluster tersebut.

2. Analisis Perbandingan Eksperimen.

Pada visual scatter plot, kedua implementasi berhasil mengidentifikasi tiga segmen pelanggan yang sama persis. Bentuk, ukuran, dan anggota dari setiap cluster sama antara model from scratch dan model scikit-learn.

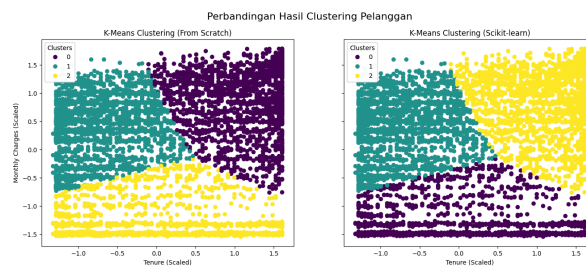


Figure 1: Perbandingan Hasil Clustering

Namun, terdapat perbedaan dalam pemberian label numerik (0, 1, 2) pada cluster yang sama di kedua model. Hal ini adalah perilaku dari algoritma K-Means yang mana pemberian label awal bergantung pada posisi centroid yang diinisialisasi secara acak.