

Nama/NIM: Rafli Limandijaya/1103210243

Data Clustering Online Retail

Percobaan dilakukan dengan menggunakan 1000 baris dari dataset yang diberikan

Hasil percobaan

	Model	Silhouette	Calinski-Harabasz	Davies-Bouldin
0	KMeans	0.554964	205.679208	0.666719
1	DBSCAN	NaN	215.323417	1.296107

Model terbaik: KMeans

Karena

-Silhouette valid dan tinggi

-Davies-Bouldin lebih kecil

-DBSCAN menghasilkan NaN di Silhouette, ini indikasi keras bahwa clustering-nya tidak optimal (mungkin terlalu banyak noise / satu cluster saja)

1. Silhouette Score:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Dimana

-a(i) = rata-rata jarak ke kluster sendiri

-b(i) = rata-rata jarak ke kluster tetangga terdekat

-Nilai mendekati 1 → cluster bagus.

Dalam percobaan, didapatkan nilai 0.554964 untuk Kmeans dan NaN untuk DBSCAN.

2. Calinski-Harabasz Index:

$$CH = \frac{\text{between-cluster dispersion}}{\text{within-cluster dispersion}} \times \frac{n - k}{k - 1}$$

Dimana semakin tinggi nilainya maka clustering lebih baik.

3. Davies-Bouldin Index:

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{s_i + s_j}{d_{ij}} \right)$$

Dimana

s_i = rata-rata jarak dalam kluster

d_{ij} = jarak antar pusat kluster

Nilai semakin rendah semakin baik