

Nama/NIM : Rafli Limandijaya/1103210243

Penjelasan Matematika

1. Inertia(untuk Elbow Method)

$$\text{Inertia} = \sum_{i=1}^n \min_{\mu_j \in C} \|x_i - \mu_j\|^2$$

Inertia adalah jumlah kuadrat jarak dari setiap titik data ke pusat klusternya. Digunakan dalam Elbow Method untuk menentukan jumlah kluster optimal. Semakin kecil inertia, semakin baik klusternya — tapi jika terlalu kecil, bisa jadi overfitting.

2. Silhouette Coefficient

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

-a(i): Rata-rata jarak dari titik iii ke titik-titik lain dalam kluster yang sama.

-b(i): Rata-rata jarak dari titik iii ke titik-titik dalam kluster terdekat yang berbeda.

Penjelasan:

Nilai Silhouette berkisar dari -1 hingga 1.

- Nilai mendekati 1 → titik terkluster dengan baik.
- Nilai mendekati 0 → titik berada di batas antara dua kluster.
- Nilai mendekati -1 → titik salah kluster.

3. Calinski-Harabasz Index

$$CH = \frac{\text{Tr}(B_k)}{k - 1} \div \frac{\text{Tr}(W_k)}{n - k}$$

Tr(B_k): *Between-cluster dispersion matrix* (variansi antar kluster).

Tr(W_k): *Within-cluster dispersion matrix* (variansi dalam kluster).

k : jumlah kluter

n : jumlah data

4. Davies-Bouldin Index

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right)$$

σ_i : rata-rata jarak dari titik-titik dalam kluster ke pusat klusternya (intra-kluster).
 $d(c_i, c_j)$: jarak antara pusat kluster i dan j (inter-kluster).

5. Dunn Index

$$D = \frac{\min_{i \neq j} d(C_i, C_j)}{\max_k \delta(C_k)}$$

$D(c_i, c_j)$: jarak minimum antar dua kluster berbeda (inter-kluster).

$\delta(C_k)$: diameter dari kluster k (jarak maksimum antara dua titik dalam kluster itu).

Dunn Index mengukur rasio antara seberapa jauh antar kluster dan seberapa besar masing-masing kluster.

Semakin besar nilai Dunn, maka kluster semakin baik (terpisah dan padat).