Nama : Putri Amelia Suhendi

Nim : 1103213195

Tugas ml week 6

**1. Elbow Method (WCSS - Within Cluster Sum of Squares)**

**Tujuan:**

Menentukan jumlah klaster yang optimal, dengan melihat **total variansi dalam klaster**.

**Rumus:**

**A black background with white letters and symbols

AI-generated content may be incorrect.**

* xj​: titik data ke-jjj dalam klaster ke-i
* μi: centroid dari klaster i
* Ci​: kumpulan titik dalam klaster i

**Intuisi:**

* Semakin banyak klaster → WCSS menurun (karena tiap klaster lebih kecil).
* **Elbow Point**: Titik di mana penurunan WCSS melambat drastis → indikasi jumlah klaster optimal.

**Contoh:**

Misal:

* Dengan 2 klaster: WCSS = 15000
* Dengan 3 klaster: WCSS = 8000
* Dengan 4 klaster: WCSS = 6000 (melambat) Maka 3 adalah jumlah klaster optimal.

**2. Silhouette Coefficient**

**Tujuan:**

Mengukur **kualitas** penempatan titik dalam klaster → apakah cocok atau salah tempat.

**Rumus per titik:**

A black background with white text

AI-generated content may be incorrect.

* a(i): rata-rata jarak antara titik iii dan semua titik dalam **klasternya sendiri**
* b(i): jarak rata-rata titik iii ke titik di **klaster terdekat** yang bukan miliknya

**Interpretasi:**

* s(i)≈1: sangat cocok dalam klaster-nya
* s(i)≈0: di antara dua klaster
* s(i)<0: kemungkinan salah klaster

**Contoh:**

* Titik A dalam klaster 1, jarak rata-rata ke sesama = 2, jarak ke klaster tetangga = 5:

A black background with white text

AI-generated content may be incorrect.

**3. Calinski-Harabasz Index (CH Index)**

**Tujuan:**

Mengukur **seberapa terpisah dan kompak klaster**, berdasarkan variansi.

**Rumus:**

**A black background with white text

AI-generated content may be incorrect.**

* Numerator = variansi antar centroid
* Denominator = variansi dalam klaster
* N: jumlah total titik, kkk: jumlah klaster

**Interpretasi:**

* Semakin besar = klaster lebih baik terpisah & kompak
* Cocok untuk data **spherical** (misal hasil normalisasi/penggunaan KMeans)

**Contoh:**

* CH = 250 ➡️ cluster terpisah baik
* CH = 30 ➡️ cluster tumpang tindih

**4. Davies-Bouldin Index (DBI)**

**Tujuan:**

Mengukur **rata-rata kemiripan antar klaster** — seberapa **"bertabrakan"** klaster satu dengan yang lain.

**Rumus:**

**A black background with white text

AI-generated content may be incorrect.**

* si​: rata-rata jarak tiap titik ke centroid-nya (klaster i)
* dij​: jarak antar centroid klaster i dan j

**Interpretasi:**

* Semakin rendah = cluster lebih **terpisah dan tajam**
* Semakin tinggi = cluster tumpang tindih

**Contoh:**

* DB = 0.5 → bagus
* DB = 1.5 → kurang terpisah

**5. Dunn Index**

**Tujuan:**

Menilai **rasio pemisahan antar klaster** terhadap **diameter klaster terbesar**

**Rumus:**

**A black background with white text

AI-generated content may be incorrect.**

* d(Ci,Cj): jarak minimum antar klaster i dan j
* δ(Ck): diameter (jarak maksimum) dalam klaster k

**Interpretasi:**

* Nilai besar → klaster jauh antar satu sama lain dan kompak
* Cocok untuk data dengan bentuk **apapun** (tidak harus bundar)

**Contoh:**

* Dunn = 1.2 → bagus (jarak antar klaster > diameter dalam klaster)
* Dunn = 0.3 → buruk (klaster tumpang tindih)

**Kesimpulan Clustering**

1. **Elbow Method (WCSS)**  
    Titik "tekukan" menunjukkan jumlah klaster optimal berada di sekitar **3–4 klaster**. Artinya, setelah titik ini, penambahan klaster tidak banyak meningkatkan pemisahan data.
2. **Silhouette Coefficient**  
    Skor tertinggi muncul pada jumlah klaster yang sama (3 atau 4). Ini berarti mayoritas data cocok ditempatkan di klasternya dan antar klaster cukup terpisah.
3. **Calinski-Harabasz Index**  
    Nilai CH tinggi menunjukkan **variansi antar klaster lebih besar** daripada variansi dalam klaster. Ini menandakan struktur klaster yang **jelas dan efektif**.
4. **Davies-Bouldin Index**  
   Nilai rendah mengindikasikan **klaster yang tidak saling tumpang tindih**, artinya model berhasil memisahkan kelompok pelanggan dengan baik.
5. **Dunn Index**  
   Nilai cukup tinggi menandakan **pemisahan antar klaster bagus**, dan tiap klaster bersifat kompak. Ini penting untuk segmentasi yang tajam (misal: pelanggan loyal vs baru).