

Nama : Andrey Satrio Pramono
Kelas : 6-Sisfo 4
No. : 2
NPM : 4131230218

RESUME & STUDI KASUS MINI CLUSTERING

Bagian 1: Matriks Pemahaman

Aspek	K-Means	Hierarchical (AGNES)	DBSCAN
Ide Dasar	Membagi data ke dalam k kelompok dengan menghitung jarak tiap titik ke rata-rata (centroid).	Menggabungkan objek satu per satu dari bawah ke atas hingga membentuk pohon (dendrogram).	Mengelompokkan titik berdasarkan kepadatan; area padat jadi klaster, area jarang dianggap noise.
Parameter Utama	Jumlah klaster (k) ditentukan di awal.	Metode linkage (single, complete, centroid) dan titik potong dendrogram.	Eps (radius tetangga) dan MinPts (jumlah minimal titik).
Kelebihan Utama	Cepat, efisien, cocok untuk data besar.	Tidak perlu tentukan jumlah klaster di awal; dendrogram jelas menunjukkan hubungan antar objek.	Tahan terhadap noise/outlier, bisa temukan klaster dengan bentuk bebas.
Kelemahan / Kapan Jangan Digunakan	Sensitif terhadap outlier; sulit tentukan k; tidak cocok untuk klaster non-bulat.	Lambat untuk data besar; keputusan awal tidak bisa diubah.	Sulit jika kepadatan antar klaster berbeda; butuh tuning Eps & MinPts.
Bentuk Cluster Ideal	Bulat, ukuran seragam.	Cocok untuk data dengan struktur bertingkat.	Bentuk bebas, termasuk memanjang atau melengkung.

Bagian 2: Penerapan Studi Kasus

Kasus A (E-commerce, jutaan data transaksi, 5 kelompok, butuh komputasi cepat)

Algoritma Pilihan: K-Means.

Alasan: K-Means efisien karena hanya menghitung jarak tiap titik ke centroid, sehingga bisa menangani jutaan data dengan cepat. Jumlah klaster sudah jelas (5), jadi parameter k langsung bisa ditentukan sesuai kebutuhan marketing.

Kasus B (Persebaran titik pusat gempa, bentuk tidak beraturan, banyak noise)

Algoritma Pilihan: DBSCAN.

Alasan: DBSCAN membentuk kluster berdasarkan kepadatan, sehingga bisa mengikuti pola patahan yang berkelok. Titik dengan kepadatan rendah otomatis dianggap noise, jadi gempa kecil tidak mengganggu kluster utama.

Kasus C (Hubungan kekerabatan evolusi 50 spesies burung berdasarkan DNA)

Algoritma Pilihan: Hierarchical (AGNES).

Alasan: AGNES menggabungkan objek mirip secara bertahap, sehingga dendrogram bisa menunjukkan urutan percabangan evolusi dari nenek moyang hingga spesies baru. Metode ini cocok untuk data kekerabatan yang bertingkat.

Bagian 3: Pertanyaan Kritis

1. Perbedaan Supervised vs Unsupervised Learning (clustering)

Supervised learning adalah proses belajar dari data yang sudah diberi label. Misalnya, kita sudah paham ciri mangga matang: kulit kekuningan, agak lembek, dan baunya harum. Ciri-ciri ini menjadi label "mangga matang". Jadi, ketika kita mendapat mangga yang masih hijau dan keras, kita bisa menebak bahwa buah itu belum matang, lalu menunggu sekitar 3 hari agar siap dimakan. Jadi, ada kunci jawaban yang jelas untuk melatih model.

Sebaliknya, Unsupervised learning bekerja tanpa label. Misalnya, Toko online memiliki data perilaku pelanggan (frekuensi belanja, nominal belanja, kategori barang yang dilihat). Tanpa memberi label "pelanggan kaya" atau "pelanggan hemat", algoritma clustering akan secara otomatis mengelompokkan pelanggan ke dalam beberapa segmen (misal: "pemburu diskon", "pembelanja barang mewah", "pembeli rutin kebutuhan pokok") berdasarkan kemiripan perilaku mereka.

2. Dua Langkah Perbaikan Jika Evaluasi K-Means Buruk

- Evaluasi ulang nilai k: hasil buruk bisa muncul karena jumlah kluster yang dipilih tidak sesuai. Untuk memperbaikinya, gunakan metode seperti Elbow Method atau Silhouette Score agar jumlah kluster lebih optimal. Dengan cara ini, kita bisa tahu apakah k yang dipilih sudah tepat atau perlu diubah.
- Tangani outlier dan lakukan normalisasi data: outlier bisa menarik posisi centroid sehingga hasil kluster jadi tidak akurat. Selain itu, jika skala fitur berbeda-beda, hasil perhitungan jarak bisa bias. Karena itu, data perlu dibersihkan dari outlier dan fitur harus dinormalisasi/standardisasi agar semua variabel punya pengaruh yang seimbang. Setelah itu, jalankan ulang K-Means untuk mendapatkan hasil yang lebih stabil.