	Nama	Mujahid Ansori Majid
	NIM	1197050093
	Matakuliah	Pembelajaran Mesin – D
	Tugas	Quiz 9 dan 10
	Dosen Pengampu	Wildan Budiawan Zulfikar, S.T, M.Kom

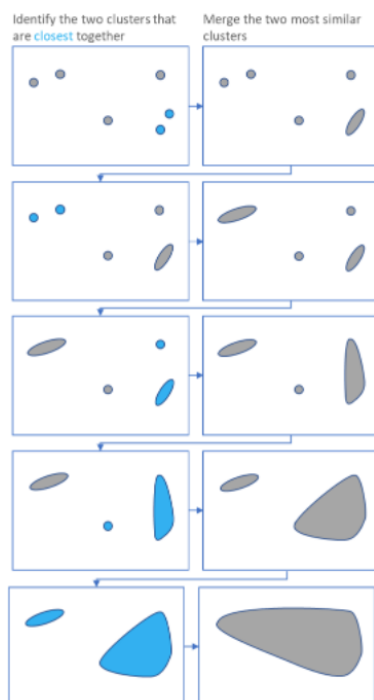
### Soal 1

Buatlah resume mengenai Hierarchical Clustering (Agglomerative dan Divisive) dan DBCSAN menggunakan kata-kata anda sendiri!

### Jawab

Hierarchical Clustering adalah algoritma yang mengelompokkan objek serupa ke dalam kelompok yang disebut cluster. Titik akhir adalah satu set cluster, di mana setiap cluster berbeda satu sama lain, dan objek dalam setiap cluster secara luas mirip satu sama lain. Hierarchical Clustering dapat dilakukan dengan matriks jarak atau data mentah. Ketika data mentah disediakan, perangkat lunak akan secara otomatis menghitung matriks jarak di latar belakang.

Hierarchical Clustering dimulai dengan memperlakukan setiap pengamatan sebagai kluster yang terpisah. Kemudian, berulang kali menjalankan dua langkah berikut: (1) mengidentifikasi dua cluster yang paling dekat bersama, dan (2) menggabungkan dua cluster yang paling mirip. Proses iteratif ini terus berlanjut sampai semua cluster digabungkan menjadi satu. Hal ini diilustrasikan dalam diagram di bawah ini.



## Agglomerative dan divisive algoritma

Hierarchical Clustering biasanya bekerja dengan menggabungkan cluster serupa secara berurutan, seperti yang ditunjukkan di atas. Ini dikenal sebagai pengelompokan hierarki aglomerasi. Secara teori dapat juga dilakukan dengan mengelompokkan semua observasi ke dalam satu cluster terlebih dahulu, kemudian secara berturut-turut memisahkan cluster-cluster tersebut. Ini dikenal sebagai pengelompokan hierarkis yang memecah belah. Pengelompokan divisif jarang dilakukan dalam praktik.

## DBSCAN

Density-based clustering mengacu pada metode pembelajaran tanpa pengawasan yang mengidentifikasi kelompok/cluster yang berbeda dalam data, berdasarkan gagasan bahwa sebuah cluster dalam ruang data adalah wilayah yang berdekatan dengan kepadatan titik tinggi, dipisahkan dari cluster lain seperti itu oleh wilayah yang berdekatan dengan kepadatan titik rendah.

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) adalah algoritma dasar untuk pengelompokan berbasis kepadatan. Ia dapat menemukan cluster dengan berbagai bentuk dan ukuran dari sejumlah besar data, yang mengandung noise dan outlier.

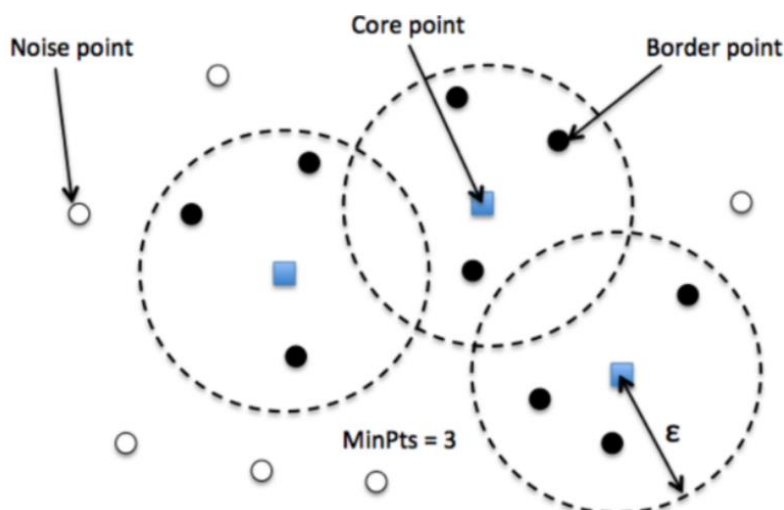
Algoritma DBSCAN menggunakan dua buah parameter

- **minPts**: Jumlah minimum titik (ambang batas) yang dikelompokkan bersama agar suatu wilayah dianggap padat.
- **eps ( $\epsilon$ )**: Ukuran jarak yang akan digunakan untuk menemukan titik-titik di sekitar titik mana pun.

Parameter ini dapat dipahami jika kita mengeksplorasi dua konsep yang disebut Density Reachability dan Density Connectivity. Reachability dalam hal kepadatan menetapkan suatu titik yang dapat dijangkau dari yang lain jika terletak dalam jarak tertentu (eps) darinya.

Konektivitas, di sisi lain, melibatkan pendekatan rantai berbasis transitivitas untuk menentukan apakah titik terletak di cluster tertentu. Misalnya, titik p dan q dapat dihubungkan jika  $p \rightarrow r \rightarrow s \rightarrow t \rightarrow q$ , di mana  $a \rightarrow b$  berarti b berada di sekitar a.

Ada tiga jenis titik setelah pengelompokan DBSCAN selesai:



- **Core** – Ini adalah titik yang memiliki setidaknya  $m$  titik dalam jarak  $n$  dari titik itu sendiri.
- **Border** – Ini adalah titik yang memiliki setidaknya satu titik inti pada jarak  $n$ .
- **Noise** - This is a point that is neither a Core nor a Border. And it has less than  $m$  points within distance  $n$  from itself.

#### Langkah-langkah algoritma untuk pengelompokan DBSCAN

- Algoritme dilanjutkan dengan mengambil titik dalam kumpulan data secara sewenang-wenang (sampai semua titik telah dikunjungi).
- Jika setidaknya ada titik 'minPoint' dalam radius " $r$ " ke titik tersebut, maka kami menganggap semua titik ini sebagai bagian dari cluster yang sama.
- Cluster kemudian diperluas dengan mengulangi perhitungan lingkungan secara rekursif untuk setiap titik tetangga

#### Soal 2

##### Market Basket Analysis

Market Basket Analysis adalah teknik penambangan data yang digunakan oleh pengecer untuk meningkatkan penjualan dengan lebih memahami pola pembelian pelanggan. Ini melibatkan analisis kumpulan data besar, seperti riwayat pembelian, untuk mengungkapkan pengelompokan produk, serta produk yang kemungkinan akan dibeli bersama.

Penerapan analisis keranjang pasar dibantu oleh munculnya sistem point-of-sale (POS) elektronik. Dibandingkan dengan catatan tulisan tangan yang disimpan oleh pemilik toko, catatan digital yang dihasilkan oleh sistem POS memudahkan aplikasi untuk memproses dan menganalisis data pembelian dalam jumlah besar.

Implementasi analisis keranjang pasar memerlukan latar belakang statistik dan ilmu data, serta beberapa keterampilan pemrograman komputer algoritmik. Bagi mereka yang tidak memiliki keterampilan teknis yang dibutuhkan, ada alat komersial yang tersedia.

Salah satu contohnya adalah alat Analisis Keranjang Belanja di Microsoft Excel, yang menganalisis data transaksi yang terdapat dalam spreadsheet dan melakukan analisis keranjang pasar. Item yang akan dianalisis harus terkait dengan ID transaksi. Alat Analisis Keranjang Belanja kemudian membuat dua lembar kerja: lembar kerja Kelompok Barang Keranjang Belanja, yang mencantumkan item yang sering dibeli bersamaan, dan lembar kerja Aturan Keranjang Belanja, yang menunjukkan bagaimana item terkait (Misalnya, pembeli Produk A cenderung membeli Produk B).