

Clustering

Lecture Note Pengantar Data
Mining 2006

Sumber :

<http://www.bandmservices.com/>

Definisi Clustering

- Clustering adalah suatu alat untuk analisa data, yang memecahkan permasalahan penggolongan.
- Obyek nya adalah untuk kasus pendistribusian (orang-orang, objek, peristiwa dll.) ke dalam kelompok, sedemikian sehingga derajat tingkat keterhubungan antar anggota cluster yang sama adalah kuat dan lemah antar anggota dari cluster yang berbeda.
- Dengan Cara ini masing-masing cluster menguraikan, dalam kaitan dengan kumpulan/koleksi data, class dimana milik anggota-anggotanya.
- Cluster : Data item dikelompokkan menurut pilihan konsumen atau hubungan logis. Sebagai contoh, data dapat dimaknakan untuk mengidentifikasi segmen pasar atau ketertarikan

Definisi Clustering (Lanjt)

- Cluster adalah alat penemuan. Ia mungkin mengungkapkan hubungan dan struktur di dalam data, yang sebelumnya tidak jelas, meskipun demikian adalah bermanfaat dan masuk akal sekali ketika ditemukan.
- Hasil dari analisis cluster mungkin berperan untuk definisi dari suatu rencana penggolongan yang formal, seperti suatu taksonomi untuk binatang yang terkait, serangga atau tumbuhan; atau menyarankan model statistik yang menguraikan populasi; atau menandai aturan untuk menugaskan kasus yang baru ke class untuk identifikasi dan tujuan yang diagnostik; atau menyediakan ukuran dari definisi, ukuran dan perubahan dalam konsep sebelumnya yang tidak hanya luas;
- Bisnis apapun yang sedang anda lakukan, cepat atau lambat anda akan berhadapan dengan suatu masalah penggolongan. Analisa Cluster mungkin menyediakan metodologi untuk

Definisi Clustering (Lanjt)

- Singkatnya: Algoritma Clustering berusaha untuk menemukan kelompok komponen secara natural, berdasarkan pada beberapa kesamaan.
- Contoh di bawah ini memperagakan **clustering** dari gembok yang jenisnya sama. Ada total 10 gembok yang memiliki tiga warna berbeda. Kita tertarik untuk mengelompokkan gembok yang memiliki tiga warna dan bentuk yang berbeda dalam tiga kelompok yang berbeda.

Definisi Clustering (Lanjt)



- Gembok yang jenisnya sama dikelompokkan ke dalam suatu kelompok seperti terlihat di bawah ini:



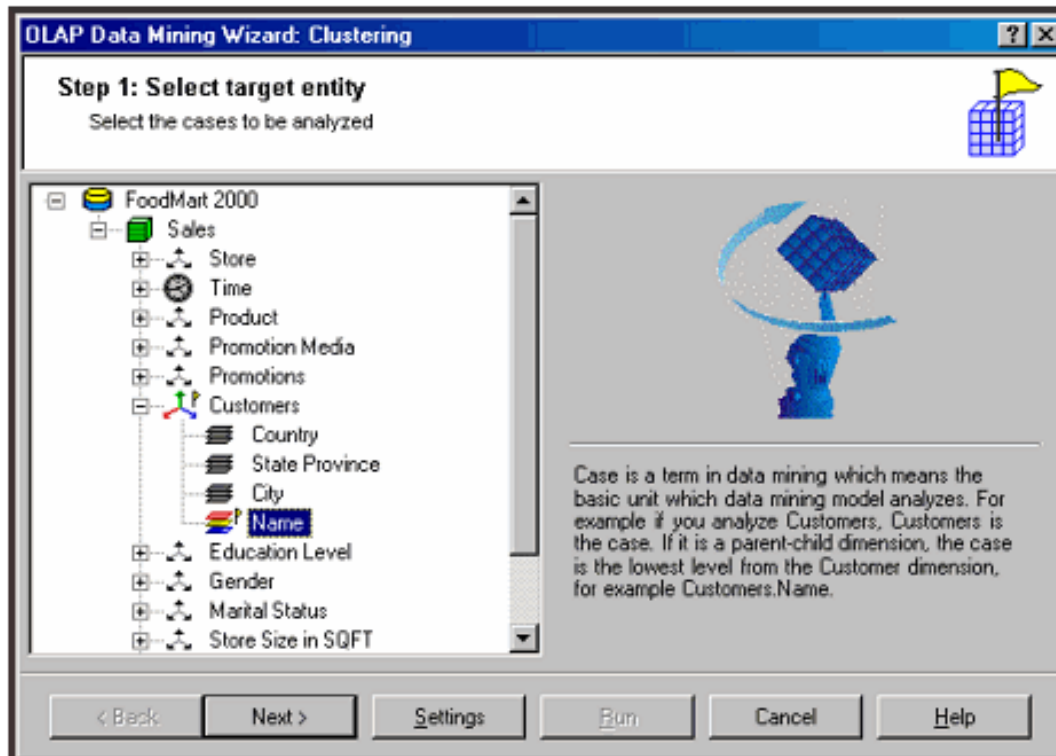
- Sehingga, kita lihat bahwa clustering bermakna menggolongkan data atau membagi satuan data yang besar ke dalam satuan data yang lebih kecil yang memiliki kesamaan. Algoritma clustering tercakup dalam aplikasi BI2M. Lihat contoh dari penggunaan clustering dalam BI2M.

Contoh Clustering

- Terdapat Database Foodmart2000 dengan OLAP Cube Sales. Kita tertarik untuk menemukan 3 segmen pelanggan dari toko Foodmart dalam rangka menciptakan suatu program untuk menawarkan manfaat yang berbeda untuk pelanggan yang tergantung pada karakteristik pribadi mereka. Tujuannya adalah untuk meningkatkan kesetiaan mereka terhadap toko tsb. Kita akan menggunakan algoritma clustering di database Foodmart2000 dimana segmen pelanggan ada di OLAP cube sales dalam tiga kategori yang berdasarkan pada informasi berikut: Jenis kelamin, status perkawinan, Pendapatan tahunan, Pendidikan, Kartu Anggota, dan Penjual toko.

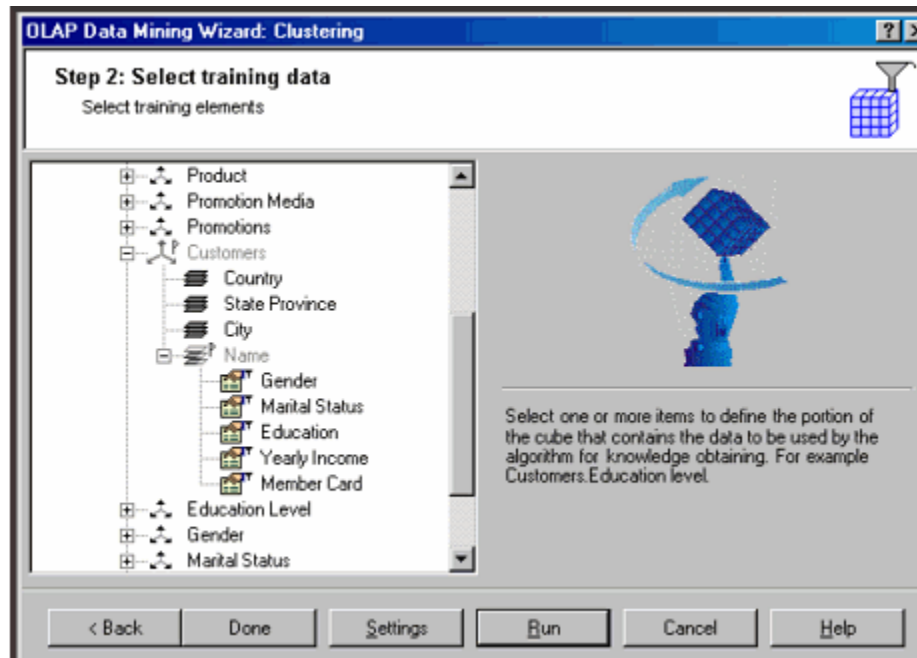
Contoh Clustering (Lanjt)

- **Langkah 1**– Ketika kita akan menggolongkan pelanggan, kita harus memilih Customer sebagai kasus pada halaman yang pertama dari OLAP Data



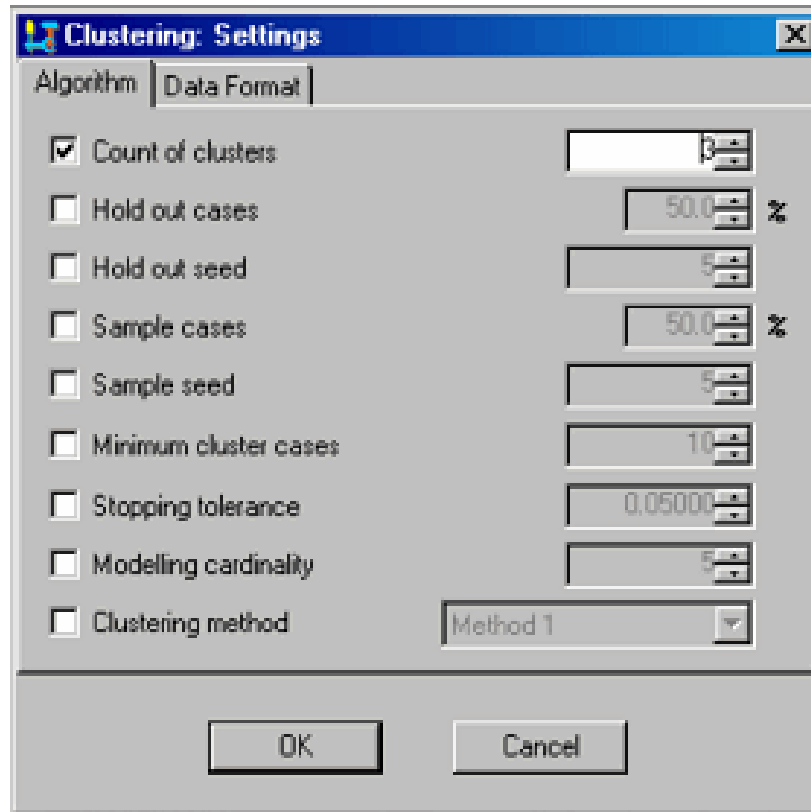
Contoh Clustering (Lanjt)

- Langkah 2– pada langkah ini kita memilih karakteristik yang akan diproses oleh algoritma tsb. Pada basis cluster mereka yang akan dibuat. Pada tugas saat ini kita tertarik akan Jenis kelamin pelanggan/Customers' Gender, Status Perkawinan/marital status, pendidikan/education, Kartu anggota/Member Card, Pendapatan tahunan/yearly income dan penjual toko, itu sebabnya mengapa kita memilihnya.



Contoh Clustering (Lanjt)

- Tentukan banyaknya cluster = 3



Contoh Clustering (Lanjt)

- Hasil dari data mining adalah:

