

ABSTRAK

Clustering adalah suatu alat untuk analisa data, yang memecahkan permasalahan penggolongan. Obyeknya 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. 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.

Skripsi ini berfokus pada identifikasi perbedaan dan menganalisis dampak terhadap penerapan teknik *clustering* ke *database*. Skripsi ini juga mengenalkan metodologi, teknik-teknik serta alat-alat yang akan digunakan dalam mengimplementasikan distribusi basis data (*clustering*).

Kata kunci: *Clustering*, Analisa Data, Analisis *Cluster*

ABSTRACT

Clustering is a tool that use for analyze data, which solve the case problems of classification. The object is the case of distribution (people, objects, resources, events, etc) into groups, so the degree of connectedness betwen members of the same cluster is strong and weak betwen members of different cluster.

In this way each of cluster describe,in terms of aggregate data collection, which belong to class member. The result from cluster analysis may contribute to the defenition of a formal classification plan, such as a taxonomy for related animals, insects or plants,or suggest a statistical model that describe the population or mark the rules for asigning new cases to classes for identification and diagnostic purpose or provides a measure of the defenition, size and changes in the previous consept that not only lagre.

This thesis is focused on identifying the differences and analyze the impact on the application of clustering techniques to databases. This thesis also introduces the methodology, techniques, and tools that will be used for implementing the distribution database (clustering).

Keywords : Clustering, Data Analysis, Cluster Analysis