

BAB I

PENDAHULUAN

Statistika merupakan ilmu yang berhubungan dengan analisa data. Salah satu permasalahan di dalam analisa data multivariat adalah permasalahan analisa cluster. Permasalahan analisa cluster merupakan suatu metode untuk mengelompokkan bilangan dalam variabel-variabel atau unit-unit data kedalam bilangan dari group variabel atau group unit data.

Metode clustering dalam analisa cluster terdapat dua metoda yaitu metoda clustering hirarki dan metoda clustering non hirarki. Metoda clustering hirarki digunakan pada matriks similar atau matriks disimilar. Dalam pembahasan tugas akhir ini digunakan metode clustering hirarki. Diberikan data dalam bentuk $x_i = (x_{1i}, x_{2i}, \dots, x_{zi}) \in R^z$; $i \in N$, selanjutnya didefinisikan jarak antara dua elemen dari N , yang dinyatakan dengan matriks d_{ij}^2 yang elemen di dalamnya adalah jarak antara i dan j untuk $i, j \in N$ merupakan matriks mahalnobis (disimilaritas).

Permasalahan yang akan dijabarkan dalam tugas akhir ini adalah meminimalkan :

$$C'(J) = \sum_{i=1}^m \tau(j_i)$$

dari fungsi diatas akan diperoleh cluster yang memuat elemen-elemen a_j dalam d_{ij}^2 sehingga akan diperoleh $C'(J)$ merupakan batas bawah. Dengan pendekatan

pencabangan dan pembatasan akan diperoleh dua cluster yang mempunyai batas bawah optimal.

Sistematika penulisan tugas akhir ini meliputi Bab I yang merupakan bab pendahuluan, Bab II merupakan analisa cluster tentang jarak antara dua elemen, metode clustering, metode clustering hierarchichal dan skema hierarchichal, Bab III menguraikan isi tentang batas bawah dari partial cluster, langkah-langkah pencabangan dan pembatasan, aturan skema pencabangan, dan sebagai penutup Bab IV yang merupakan kesimpulan dari optimalitas complete cluster dengan pendekatan pencabangan dan pembatasan.

