

Abstrak

Pada era sekarang dunia komputer dan informasi sudah berkembang sangat pesat. Terutama di bidang pengolahan data. Seiring dengan perkembangan zaman, data yang terbentuk di zaman sekarang sudah semakin banyak. Oleh karena itu dibutuhkan suatu solusi supaya data tersebut bisa diolah menjadi informasi yang berguna. Data mining merupakan suatu solusi untuk menangani banyaknya data. Data *mining* adalah sebuah alat yang memiliki peranan penting untuk mengatur dan mengolah banyaknya data.

Pada penelitian ini penulis membahas tentang *clustering* data untuk data teks, yaitu menggunakan data berita yang diambil dari situs *online* dan dipindah ke *notepad* dengan format *.txt*. Pada penelitian ini, penulis menggunakan metode *single linkage clustering*. Metode *single linkage clustering* merupakan bagian dari *agglomerative hierarchical clustering*. Metode *single linkage* mengelompokkan data berdasarkan pada data atau objek terdekat. Metode yang digunakan untuk mengetahui kualitas *cluster* adalah dengan validitas berorientasi kemiripan.

Hasil yang sudah tercapai dari sistem *summarize* dan *single linkage clustering* cukup baik. *Summarize* dengan persentase 50% adalah yang terbaik, yaitu memiliki tingkat akurasi 74.10%. Sedangkan, *Single linkage clustering* dengan 12 *cluster* memiliki tingkat akurasi yang terbaik dibandingkan dengan *single linkage clustering* 6 *cluster* dan akurasi yang dihasilkan oleh jumlah *cluster* sebanyak 12 *cluster* adalah 79.8%.

Kata kunci : *Data mining, text mining, clustering, agglomerative hierarchical clustering, single linkage, summarize*