

Abstrak

Pesatnya pertumbuhan dan keanekaragaman informasi contohnya artikel berita, konsumen berita semakin sulit mencari informasi yang diinginkan. Untuk itu diperlukan pengelompokan artikel berita dengan menerapkan salah satu teknik yang terkait dengan Data Mining, yaitu “*clustering*”.

Metode *clustering* yang digunakan adalah *Ward's method*, metode pengelompokan objek yang menerapkan prinsip *agglomerative hierarchical clustering* yang bersifat *bottom up* dimana pengelompokan dibentuk hirarki dimulai dari *bottom* dengan menentukan tiap-tiap objek membentuk kluster masing-masing hingga ke *up* yaitu, mengelompokkan setiap pasangan kluster paling mirip tahap demi tahap sampai membentuk sebuah kluster tunggal. Pasangan yang mirip ini merupakan pasangan yang memiliki peningkatan *error sum of square* paling minimum.

Pengujian dilakukan untuk menganalisis hasil pengelompokan *Ward's method* berdasarkan parameter *error sum of square*(ESS), *Cophenetic Correlation Coefficient*(CPCC), *FMeasure*, dan waktu eksekusi. Penambahan jumlah dokumen uji dan topik pada koleksi dokumen uji menyebabkan perubahan nilai ESS yang cenderung semakin meningkat. Selain itu, penambahan jumlah dokumen uji juga mempengaruhi nilai CPCC, nilai tertinggi diperoleh koleksi dengan jumlah terkecil (terdiri dari 10 dokumen) yaitu 0,9519 (kualitas hirarki kluster sebesar 95,19%). Sedangkan berdasarkan *FMeasure* dan waktu eksekusi, rata-rata *FMeasure* tertinggi menunjukkan nilai keakurasian dalam pengelompokan diperoleh *complete linkage* yang merupakan metode pembandingan. Meskipun begitu dilihat dari waktu eksekusinya, *Ward's method* lebih cepat dalam melakukan pengelompokan.

Kata kunci : pengelompokan dokumen, *agglomerative hierarchical clustering*, *Ward's method*, *error sum of square*, *cophenetic correlation coefficient*, *FMeasure*.