

ABSTRAK

Penyebaran berita saat ini semakin tersebar luas semenjak perkembangan dunia internet semakin pesat. Perkembangan dunia internet membuat berita yang tersebar semakin beragam dan berjumlah sangat besar. Pembaca berita akan kesulitan untuk memperoleh berita yang diinginkan jika berita tersebut tidak terkelompok dengan baik. Dan jika harus dikelompokkan secara manual membutuhkan waktu yang sangat lama. Oleh sebab itu, *Clustering* menjadi solusi untuk mengatasi masalah tersebut. *Clustering* akan mengelompokkan dokumen berita berdasarkan tingkat kemiripan dari dokumen tersebut.

Metode *Single Linkage* merupakan metode pengelompokan *hierarchical clustering*. Metode *Single Linkage* mengelompokkan dokumen didasarkan dengan jarak terdekat antar dokumen. Variasi kelompoknya dari data sebagai satu kelompok sampai semua data bergabung menjadi kelompok tunggal. Komputasi *Single Linkage* merupakan komputasi yang mahal dan kompleks. Sedangkan metode *K-means* merupakan metode pengelompokan *partitioned clustering*. Metode *K-means* mengelompokkan dokumen didasarkan dengan jarak terdekat dengan *centroid*-nya. *K-Means* merupakan metode pengelompokan yang sederhana dan dapat digunakan dengan mudah. Pada jenis data tertentu, *K-means* tidak dapat memberikan segmentasi data dengan baik sehingga kelompok yang terbentuk tidak murni data yang sama.

Metode pengujian yang digunakan untuk mengukur kualitas *cluster* adalah *Silhouette Coefficient* dan *Purity*. Berdasarkan hasil pengujiannya metode *Single Linkage* memiliki performansi yang lebih baik dibandingkan dengan metode *K-means*. Nilai *silhouette coefficient* *Single Linkage* selalu lebih unggul dibandingkan dengan *K-Means*. Pertambahan jumlah dokumen membuat nilai *silhouette coefecient single linkage* semakin kecil sedangkan *K-means* terkadang menghasilkan nilai yang negatif. Untuk nilai *purity*, *Single Linkage* selalu bernilai 1 sedangkan *K-Means* tidak pernah bernilai 1. Hasil pertambahan jumlah cluster dan jumlah dokumen memberikan pengaruh terhadap nilai *silhouette coefficient* dan *purity*. Hal ini berarti *single linkage* selalu menghasilkan dokumen yang sama sedangkan *K-means* masih bercampur dengan dokumen yang lain.

Kata kunci : *Clustering*, HAC, *Partitioned*, *Single Linkage*, *K-Means*, *Silhouettte Coefficient*, dan *purity*.