

Abstrak

Dengan semakin berkembangnya teknologi informasi, jumlah dokumen teks berupa digital semakin berkembang dengan pesat. Untuk melakukan pencarian informasi tentang kejadian-kejadian baru yang berkaitan dengan topik tertentu dalam kumpulan dokumen akan lebih sulit. *Clustering* merupakan salah satu metode data mining yang bersifat *unsupervised learning* untuk mengelompokkan dokumen berdasarkan kemiripannya. Untuk melakukan pengelompokan tersebut, digunakan salah satu algoritma *clustering* yaitu *single pass clustering*. *Single pass clustering* merupakan suatu tipe *clustering* yang berusaha melakukan pengelompokan data satu demi satu dan pembentukan kelompok dilakukan seiring dengan pengevaluasian setiap data yang dimasukkan ke dalam proses *cluster*. Jumlah *cluster* yang dihasilkan sangat tergantung pada nilai *threshold*. Kemiripan antar dokumen yang digunakan adalah *standard cosine similarity*. [4]

Dalam Tugas Akhir ini kualitas *cluster* yang dihasilkan diukur dengan dua parameter yaitu *cohesion* dan *separation*. Nilai *threshold* 0.03 menghasilkan nilai rata-rata *cohesion* sebesar 0.13863266 dan rata-rata *separation* sebesar 0.009379814.

Kata kunci: *single pass clustering, unsupervised learning, cosine similarity, threshold, cohesion, separation.*