

Abstrak

Seiring dengan perkembangan teknologi, semakin banyak informasi yang diberikan dalam bentuk dokumen teks digital. Dokumen teks memiliki informasi yang beraneka ragam, sehingga untuk memudahkan dalam pengambilan informasi agar sesuai dengan keinginan perlu adanya pengelompokan dokumen. *Clustering* adalah proses untuk mengelompokkan data ke dalam suatu klaster, sehingga objek pada suatu klaster memiliki kemiripan yang sangat besar dengan objek lain pada klaster yang sama, tetapi sangat tidak mirip dengan objek pada klaster yang lain. *Clustering* yang dilakukan terhadap dokumen-dokumen disebut *document clustering*.

Pada tugas akhir ini diimplementasikan suatu algoritma *clustering* yaitu algoritma *cure*. *Algoritma cure* merupakan algoritma yang bekerja dengan cara mengukur jarak antar dokumen dengan list poin perwakilan *cluster* yang sudah dipilih sebelumnya.

Pengujian yang dilakukan dengan menghitung jumlah *cluster* dan menghitung nilai cohesi dan *separation cluster* yang dihasilkan menggunakan algoritma *cure* . Berdasarkan nilai cohesi yang dihasilkan pengujian ini dalam membentuk *cluster* yang sesuai dengan kategori yang terdapat dari dataset diperoleh bahwa kualitas yang dihasilkan cukup bagus sekitar 0.0855. Sedangkan berdasarkan nilai *separation* yang dihasilkan pengujian ini yaitu 0.927039 meskipun membentuk *cluster* yang tidak sesuai dengan kategori dataset. Akan tetapi kualitas *cluster* yang dihasilkan cukup bagus, karena kualitas *clustering* baik jika semakin kecil nilai cohesi dan semakin besar nilai *separation*.

Kata kunci: algoritma, *cure*, *clustering*, dokumen, cohesi, *separation*