

Abstrak

Klasterisasi adalah proses mengelompokkan data ke dalam suatu kelompok (klaster) sehingga objek pada suatu klaster memiliki kemiripan yang sangat besar dengan objek lain pada klaster yang sama, tetapi memiliki ketidakmiripan yang besar dengan objek pada klaster lain.

Banyak algoritma klasterisasi ditujukan untuk data yang bersifat numerik. Salah satunya adalah algoritma *hierarchical clustering* yang mengelompokkan objek dengan membuat suatu hirarki dimana objek yang mirip akan ditempatkan pada hirarki yang berdekatan dan objek yang tidak mirip pada hirarki yang berjauhan. Namun, permasalahan timbul ketika algoritma tersebut diterapkan pada data yang mempunyai nilai atribut yang bersifat boolean atau kategorik. Seringkali objek data yang mempunyai nilai *similarity* yang kecil dikelompokkan menjadi satu klaster meskipun objek-objek tersebut tidak mempunyai kesamaan item.

Untuk menangani masalah data kategorik, dalam Tugas Akhir ini digunakan algoritma ROCK (*RObust Clustering using linKs*) yang melakukan klasterisasi dengan cara mengelompokkan data yang mempunyai *link* (jumlah item yang sama) paling banyak dengan tetangganya (*neighbor*) dengan parameter jumlah klaster (k) dan θ (*threshold*).

Ditunjukkan dalam Tugas Akhir ini bahwa algoritma ROCK menghasilkan *cohesion* dan *separation* rata-rata yang lebih baik dan dapat menangani *outlier* pada data dengan baik dibandingkan dengan algoritma *hierarchical clustering*.

Kata kunci: klasterisasi, klaster, data kategorik, *hierarchical clustering*, *robust clustering using links*