

Abstrak

Setiap lembaga mempunyai satu atau lebih jenis data yang berkaitan dengan lembaga yang bersangkutan. Salah satu jenis data yang sering ditemukan yaitu data dua dimensi dalam bentuk teks yang mempunyai struktur baris dan kolom untuk mendeskripsikan suatu bagian atau elemen data. Salah satu jenis pengolahan data yaitu pengelompokan data.

Similarity-Driven Cluster Merging merupakan salah satu metode pengelompokan data yang mengedepankan kemiripan antar data dalam membentuk suatu kelompok data atau *cluster*. Metode *Similarity-Driven Cluster Merging* dapat diterapkan untuk menentukan *cluster* suatu jenis data baik pada data yang sudah terklasifikasi maupun pada data yang belum terklasifikasi.

Dalam tugas akhir ini, metode *Similarity-Driven Cluster Merging* digunakan pada jenis data yang belum mempunyai struktur klasifikasi (*unsupervised*). Matrik *similarity* digunakan untuk menghitung dan menampung nilai-nilai kemiripan antar cluster, sehingga dapat ditentukan batasan nilai atau *threshold cluster merging* sebagai patokan untuk menentukan mirip tidaknya data-data. *Generalized Objective Function* untuk meng-*update* nilai keanggotaan data terhadap suatu *cluster*.

Berdasarkan hasil pengujian dan analisis hasil, metode *Similarity-Driven Cluster Merging* menghasilkan jumlah cluster dengan kualitas cluster yang paling baik yaitu pada nilai error epsilon 50 dan threshold cluster merging 1.2. Kualitas cluster yang baik ini ditunjukkan dengan nilai score function yang mendekati 1.0 yaitu 0.914686.

Kata kunci: *Similarity-Driven Cluster Merging, cluster, threshold cluster merging, error epsilon.*