

ABSTRAK

Dengan semakin luasnya penggunaan komputer di dalam bisnis, pemerintahan dan ilmu pengetahuan, penemuan pola-pola yang menarik dari basisdata berukuran besar menjadi sangat penting. Data mining muncul sebagai solusi bagi masalah analisis data yang dihadapi oleh banyak organisasi. Salah satu fungsionalitas dalam data mining adalah *clustering* yang bertujuan untuk mengelompokkan data ke dalam suatu *cluster* berdasarkan kemiripan karakteristiknya.

Subspace clustering merupakan pengembangan dari metode *clustering*, yaitu membentuk kumpulan *cluster* pada dataset dengan menentukan dimensi yang paling relevan untuk setiap *cluster*. FINDIT melakukan pendekatan perhitungan *dimension-oriented distance* dan *dimension voting* untuk membentuk suatu *cluster*. Pada tugas akhir ini telah diimplementasikan algoritma FINDIT dan juga dianalisis performansi algoritma berdasarkan jumlah data, dimensi dataset terhadap waktu, serta akurasi *cluster* yang dihasilkan berdasarkan parameter *Dmindist*.

Dmindist sebagai salah satu user parameter dapat mempengaruhi kinerja perangkat lunak. Jika semakin kecil maupun terlalu besar nilai *Dmindist*, akurasi *cluster* yang dihasilkan menjadi kurang baik, ditunjukkan dengan hilangnya satu atau lebih *subspace* pada *original cluster*. Peningkatan jumlah data mempengaruhi waktu untuk menemukan *cluster*, semakin banyak jumlah data maka semakin lama waktu yang dibutuhkan. Begitu pula untuk peningkatan jumlah dimensi data, akan menambah waktu untuk menemukan *cluster*.

Kata kunci : data mining, *subspace clustering*, algoritma FINDIT, *dimension oriented distance*, *dimension voting*, *Dmindist*.