

ABSTRAK

Data mining adalah proses mengekstrak dan menganalisa data dalam jumlah besar untuk mendapatkan pengetahuan dari data yang ada. Perbedaan antara data mining dan ekstraksi informasi lainnya adalah menganalisa jumlah data yang besar, mengekstrak pengetahuan yang potensial serta tugas yang dilakukan serta tugas yang dilakukan cukup kompleks sehingga sulit dilakukan secara manual. Klasterisasi sebagai salah satu fungsionalitas dari data mining adalah proses mengelompokkan sekumpulan objek ke dalam klaster-klaster berdasarkan keterhubungan antar atribut data.

Tugas akhir ini menerapkan algoritma pCluster sebagai salah satu algoritma klasterisasi untuk mencari klaster pada data multidimensi. Algoritma pCluster merupakan algoritma untuk klasterisasi dengan penekanan pada kesamaan pola yang dibentuk pada analisis data *microarray*. Dalam analisis DNA *microarray*, level-level ekspresi dari dua gen mungkin muncul dan hilang dalam respon pada sekumpulan kondisi. Dalam pCluster, dua objek akan serupa jika diperlihatkan pada pola yang bertalian dalam sekumpulan dimensi. Walaupun besarnya level ekspresinya mungkin tidak bisa hilang, pola-pola diperlihatkan sedapat mungkin sama.

Pemodelan pCluster walaupun dibangun dalam studi kasus untuk analisis klasterisasi data *microarray*, tetapi dapat juga diaplikasikan untuk aplikasi-aplikasi lain yang membutuhkan penemuan pola-pola yang serupa atau sejenis di dalam sekumpulan data multidimensional *numeric*.

Algoritma pCluster mempunyai kemampuan yang baik dalam menemukan klaster beserta kumpulan dimensi yang berkorelasi pada data berdimensi tinggi. Skalabilitas pCluster terhadap penambahan jumlah objek dan atribut data bersifat linier terhadap waktu respon. Untuk jumlah klaster yang ditemukan dan waktu respon yang diperlukan berbanding terbalik terhadap penambahan jumlah minimal baris atau jumlah minimal kolom, dan berbanding lurus terhadap penambahan batas ambang.

Kata kunci : *data mining, data multidimensi, klasterisasi, frequent pattern, subspace clustering, algoritma pCluster.*