

ABSTRAK

Microarray merupakan salah satu teknologi bioinformatika yang dapat mengetahui profil dari ekspresi gen secara paralel dalam jumlah dimensi yang besar. Microarray digunakan untuk membantu peneliti dalam melakukan diagnosis terhadap penyakit. Proses kerja microarray adalah mengelompokkan gen berdasarkan rasio yang ditunjukkan dari keseluruhan gen tersebut. Dalam Data Mining, proses pengelompokkan ini disebut dengan klasterisasi.

Secara umum, data microarray dikumpulkan dalam jumlah sampel dan dimensi yang besar. Pada analisis penelitian, data yang memiliki jumlah dimensi besar sangat sulit untuk diteliti. Oleh karena itu, dibutuhkan klasterisasi untuk memperoleh kluster sampel sehingga menghasilkan informasi baru dari data tersebut.

Metode yang digunakan adalah clique partition yang didasari oleh prosedur branch and bound dan DFS untuk menelusuri setiap titik dalam graf. Algoritma clique bekerja dengan melakukan pencarian di setiap titik secara rekursif dan backtracking hingga semua titik dilewati. Proses menemukan kluster diawali dengan mentransformasikan data microarray ke dalam graf yang dibentuk menjadi matriks adjacency. Matriks ini merupakan matriks keterhubungan antar titik yang hanya terdiri dari dua nilai biner, 1 jika terdapat korelasi antar titik dan 0 jika sebaliknya. Dalam penelitian Tugas Akhir ini, penentuan korelasi ditentukan berdasarkan nilai threshold. Mencari kluster menggunakan clique partition berarti mencari maximal clique. Hasil yang diperoleh menunjukkan perubahan threshold mempengaruhi jumlah kluster yang diperoleh. Analisis hasil kluster untuk data microarray yang digunakan menunjukkan bahwa pemilihan threshold yang lebih kecil memeberikan nilai error SSE yang lebih sedikit.

Kata kunci: *clique partition, klastering, microarray, PCA, threshold*