

BAB I

PENDAHULUAN

1.1. Latar Belakang

Inovasi teknologi dari waktu ke waktu kian berkembang pesat. Tidak hanya dalam bidang komunikasi namun juga dalam bioteknologi. Dalam bioteknologi, terdapat salah satu teknologi khusus untuk meneliti gen atau DNA yang dikenal dengan sebutan microarray. Teknologi microarray memungkinkan para peneliti untuk menggali informasi ratusan bahkan ribuan data biologis berdimensi tinggi secara paralel dalam waktu yang bersamaan. Microarray terdiri dari rangkaian representatif ekspresi gen manusia yang dianalisis dalam reaksi tunggal. Analisis microarray hingga saat ini digunakan untuk melakukan diagnosis penyakit, misalnya kanker. Dewasa ini, penggunaan microarray telah semakin berkembang. Selain bioteknologi, microarray juga digunakan dalam bidang pertanian, makanan, kosmetik dan komputer.

Masalah yang dihadapi ketika menganalisis data microarray adalah jumlah dimensi yang sangat besar dan kompleks sehingga hal ini menyulitkan peneliti untuk memperoleh informasi dari data tersebut. Masalah seperti ini sering dikenal dengan istilah *Curse of Dimensionality*. Oleh karena itu, dibutuhkan analisis klasterisasi untuk memperoleh klaster sampel sehingga menghasilkan informasi baru yang cepat dan efektif.

Dalam Tugas Akhir ini akan dijelaskan cara klasterisasi untuk data microarray menggunakan metode clique partition. Clique partition adalah teknik partisi dalam graf yang menggunakan metode branch and bound untuk menelusuri setiap titik dalam graf yang dibentuk. Penelusuran titik digunakan untuk mencari titik-titik yang saling terhubung satu sama lain, atau istilahnya disebut dengan subgraf lengkap. Penelusuran titik akan dijalankan hingga syarat batasnya dipenuhi yaitu semua titik

dalam graf telah dilewati. Algoritma clique yang digunakan dalam penelitian Tugas Akhir ini adalah Bron-Kerbosch.

Klasterisasi data microarray berarti menemukan klaster dalam data microarray. Pada umumnya data microarray memiliki jumlah dimensi dan sampel yang besar. Hal ini tentunya akan menyulitkan selama proses penelitian berlangsung. Oleh karena itu, dibutuhkan teknik klasterisasi untuk memudahkan peneliti dalam melakukan penelitian sehingga menghasilkan informasi baru yang diinginkan dari data tersebut secara cepat dan efektif. Klasterisasi merupakan proses membagi data ke dalam kelas atau kelompok berdasarkan kemiripan yang dimiliki oleh titik data dengan titik data lainnya. Sebelum mencari klaster, data microarray dipresentasikan ke dalam graf terlebih dahulu. Dengan data yang telah diubah bentuk ke dalam graf, pencarian klaster dengan clique akan menjadi lebih mudah.

Graf yang terbentuk merupakan himpunan titik dan sisi dimana titiknya adalah record sampel pasien dan sisinya adalah hubungan kedekatan antar titik dalam data. Untuk mengetahui kedekatan antar titik dibutuhkan perhitungan korelasi. Pada penelitian ini akan digunakan threshold untuk melihat pengaruh korelasi terhadap hasil klasterisasi yang didapatkan. Untuk nilai korelasi yang lebih besar dari threshold maka diberi nilai 1 dan sebaliknya, untuk nilai korelasi yang lebih kecil dari threshold maka diberi nilai 0. Setelah nilai korelasi diperoleh maka langkah selanjutnya adalah membentuk matriks adjacency M_{ij} dimana $i, j = \{1, 2, 3, \dots, n\}$. Baris dan kolom M merepresentasikan titik dalam graf dan setiap elemen M merepresentasikan hubungan antara dua buah titik. Matriks adjacency adalah matrik simetri yang hanya memiliki 2 buah nilai, yaitu 1 jika terdapat korelasi antara dua buah titik dan 0 sebaliknya. Matriks adjacency merupakan representasi dari graf tak langsung yang menyatakan hubungan antar titik dalam data.

1.2. Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian Tugas Akhir yaitu:

1. Menganalisis dan mengimplementasikan clique partition untuk klasterisasi data microarray.
2. Menganalisis performansi sistem klasterisasi yang telah dibangun.

1.3. Rumusan Masalah

Dalam penelitian Tugas Akhir ini dirumuskan permasalahan sebagai berikut:

1. Bagaimana cara mengklasterisasi data microarray?
2. Bagaimana mengukur performansi hasil dari klasterisasi data microarray?

1.4. Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini antara lain:

1. Dataset yang digunakan adalah data microarray kanker prostat yang diperoleh dari *Kent Ridge Bio-medical Data Set Repository*[14].
2. Teknik reduksi dimensi data menggunakan PCA.
3. Algoritma yang digunakan untuk clique partition adalah algoritma Bron-Kerbosch.

1.5. Metodologi Penelitian

Metode yang digunakan dalam penelitian yaitu:

1. Studi literatur
Mengumpulkan dan membaca literatur dari buku/website/jurnal/paper mengenai klasterisasi data microarray, clique partition, serta algoritma Bron-Kerbosch.
2. Perolehan data

- Mengumpulkan data microarray yang akan diolah selama penelitian berlangsung.
3. Perancangan sistem
Merancang diagram alur serta menganalisis sistem klasterisasi yang akan digunakan saat penelitian.
 4. Implementasi
Mengimplementasikan sistem yang telah dibangun dengan algoritma Bron-Kerbosch.
 5. Analisis
Melakukan analisis terhadap hasil implementasi dan menarik kesimpulan dari hasil analisis tersebut.
 6. Penyusunan laporan
Menyusun laporan sebagai akhir dari penelitian.

1.6. Sistematika Penulisan

Penulisan Tugas Akhir mengikuti sistematika berikut:

Bab I Pendahuluan

Bab I memuat penjelasan mengenai latar belakang masalah, rumusan masalah, tujuan, batasan masalah yang digunakan, metodologi penelitian, serta sistematika penulisan.

Bab II Dasar Teori

Bab II memuat materi-materi yang mendukung pemahaman penelitian, contohnya definisi, teori, algoritma, serta materi pendukung tambahan.

Bab III Analisis dan Rancangan Sistem

Bab III memuat deskripsi sistem yang dibangun dan diagram alur perancangan sistem serta prosesnya menggunakan clique partitioning dengan algoritma Bron-Kerbosch.

Bab IV Implementasi dan Hasil Pengujian

Bab IV memuat implementasi dari rancangan sistem yang telah dibangun disertai hasil dan pembahasan penelitian Tugas Akhir.

Bab V Penutup

Bab V memuat kesimpulan dari hasil penelitian serta saran yang diberikan untuk perbaikan dan penyempurnaan penelitian di masa akan datang.

Daftar Pustaka

Daftar pustaka memuat referensi utama dan pendukung yang digunakan sebagai acuan selama pengerjaan penelitian Tugas Akhir.