

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kanker merupakan penyakit mematikan yang telah menelan banyak korban. Setiap tahun, angka kematian yang diakibatkan kanker terus mengalami kenaikan. Dalam banyak kasus, sel kanker baru ditemukan pada pertumbuhan tumor. Sel kanker tersebut dapat menyusup ke jaringan sekitar tumor sehingga tumor tersebut sudah diklasifikasikan kedalam kanker [11]. Dibandingkan dengan mendeteksi kanker secara tradisional yaitu berdasarkan analisis kemunculan tumor, deteksi kanker melalui ekspresi gen akan jauh lebih terpercaya dan akurat [3]. Mendeteksi kanker melalui ekspresi gen akan sangat membantu para ahli medis dalam penanganan selanjutnya bagi pasien yang menderita kanker sehingga dapat menekan angka kematian yang kian bertambah.

DNA mengandung sifat dan informasi suatu makhluk hidup, sehingga suatu penyakit dapat diprediksi dari ekspresi DNA [1]. Perkembangan terakhir dalam diagnosis menunjukkan DNA *microarray* dapat menggolongkan kanker pada tingkat gen [1]. DNA *microarray* memiliki kemampuan memantau ribuan ekspresi gen secara bersamaan dalam satu kali percobaan. Teknologi ini membantu para peneliti dalam mempelajari berbagai penyakit, terutama kanker. Oleh karena itu, teknologi *microarray* dapat membantu melakukan diagnosis, memonitor dan memprediksi suatu penyakit [2].

Beberapa tahun terakhir, DNA *microarray* telah menunjukkan pengaruh besar dalam menentukan gen yang menjadi penyebab kanker. Kekurangan utama dalam DNA *microarray* yaitu masalah dimensi (*curse of dimensionality*) [2]. Data DNA *microarray* ini mengandung jumlah gen yang melebihi jumlah sampel, sehingga diperlukan metode seleksi fitur untuk menentukan gen informatif [5]. Gen informatif yang dipilih, akan digunakan untuk melatih *classifier*. Kemudian *classifier* ini akan menggolongkan sampel data *microarray* kedalam kelasnya masing-masing berdasarkan model klasifikasi yang telah dibuat. Dengan begitu, *dataset* kanker yang dimiliki dapat diklasifikasikan termasuk kedalam kanker atau *non-kanker*.

Dalam pengerjaan tugas akhir ini, Support Vector Machine (SVM) bersama tiga kernel dasar yaitu Linear, Polynomial dan Radial Basis Function (RBF) akan diterapkan untuk membagi *dataset microarray* menggunakan Analysis of Variance (ANOVA) sebagai metode seleksi fitur. ANOVA dipilih sebagai metode seleksi fitur karena menjadi pasangan terbaik bersama *classifier* SVM dibandingkan dengan pasangan lainnya [1]. Juga dengan memainkan kernel pada SVM, dapat diketahui kernel mana yang menghasilkan akurasi paling besar untuk klasifikasi. Berdasarkan kajian yang telah dilakukan oleh Bharathi dan Natarajam, akurasi terbaik yang dihasilkan SVM dan ANOVA yaitu sebesar 97,91%, lebih besar bila dibandingkan dengan T-Test dan BPN [1], dan berdasarkan kajian yang dilakukan oleh Devi Arockia Vanitha dan tim, akurasi terbesar dihasilkan oleh SVM dengan kernel linear yaitu sebesar 97,7% [3].

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dibuat, berikut akan dipaparkan perumusan masalah yang dapat dirumuskan.

1. Bagaimana cara mendesain sistem pengklasifikasian kanker dengan menggunakan *dataset* hasil dari teknologi DNA *microarray*.
2. Bagaimana cara melakukan seleksi fitur agar masalah dimensi yang tinggi dari *dataset* dapat teratasi hingga menghasilkan akurasi yang baik.
3. Model seperti apa yang dihasilkan *classifier* agar *dataset* DNA *microarray* dapat diklasifikasikan dengan benar.
4. Kernel manakah yang memiliki akurasi paling tinggi dalam mengklasifikasikan *dataset* kanker.

1.3 Batasan Masalah

Berikut di bawah ini akan disebutkan hal-hal yang menjadi batasan masalah dalam pengerjaan tugas akhir.

1. *Dataset microarray* yang digunakan dalam tugas akhir ini terdiri dari empat *dataset* yaitu leukimia, *colon tumor*, *lung cancer* dan *ovarian cancer*. Data tersebut diambil dari Kent Ridge Bio-medical *Dataset* Repository [20].
2. Membahas terbentuknya *dataset*, namun tidak menjelaskan secara jelas atribut apa saja yang ada didalamnya.
3. Hanya membahas penyebab kanker secara umum, tidak membahas penyebab terjadinya kanker secara *detail* pada keempat jenis kanker tersebut.
4. Menunjukkan pemenuhan asumsi - asumsi pada ANOVA, tanpa menjelaskan *detail* mengenai metode yang digunakan dalam pemenuhan asumsi.
5. Hasil dari klasifikasi terbagi menjadi dua *class*.

1.4 Tujuan

Dibawah ini akan dijelaskan tujuan dari pembuatan tugas akhir ini.

1. Dapat membuat sistem pengklasifikasian kanker dengan menggunakan *dataset* yang dihasilkan dari teknologi DNA *microarray*.
2. Mampu mengurangi dimensi yang tinggi dengan menyeleksi gen, yaitu melakukan perhitungan korelasi dan menghasilkan pasangan gen informatif dengan menggunakan seleksi fitur ANOVA, sehingga menghasilkan akurasi klasifikasi yang tinggi.
3. Menghasilkan model klasifikasi yang cocok dengan menggunakan SVM sehingga *dataset* dapat diklasifikasikan dengan benar.
4. Menemukan salah satu kernel dengan akurasi terbaik dalam pengklasifikasian setiap *dataset* DNA *microarray* dari ketiga kernel yang digunakan.

1.5 Metodologi Penyelesaian

Metodologi penyelesaian dalam tugas akhir ini yaitu sebagai berikut:

1. Studi Literatur

Bahan – bahan pembelajaran untuk merancang tugas akhir ini yaitu berdasarkan referensi buku, data kesehatan pemerintah, artikel- artikel terkait dan jurnal – jurnal yang diunduh di internet. Pencarian referensi ini khususnya yang berkaitan dengan teknologi DNA *microarray*, penyakit kanker, Analysis of Variance (ANOVA) serta Support Vector Machine.

2. Cara Pengumpulan Data

Data yang digunakan merupakan hasil dari teknologi DNA *microarray*. Data tersebut diambil dari Kent Ridge Bio-medical Data Set Repository.

3. Rancangan Penelitian

Pada bagian ini akan dijelaskan *dataset* kanker dengan teknologi *microarray* berasal. Kemudian cara pemilihan gen informatif yang ada pada *dataset* dengan metode seleksi fitur ANOVA. Gen informatif yang dihasilkan, akan dilatih menggunakan *classifier* SVM. Dengan memainkan empat fungsi dasar kernel SVM, maka dapat diketahui kernel mana yang menghasilkan akurasi terbaik.

4. Pengujian dan Analisis

Pada tahap ini akan dilakukan pengujian terhadap model yang telah dibuat, serta membuat analisis terhadap hasil pengujian tersebut dengan mempertimbangkan beberapa faktor.

5. Penulisan laporan tugas akhir

Laporan tugas akhir ini terdiri dari lima bab, yaitu pendahuluan, tinjauan pustaka, perancangan sistem, implementasi dan penutup, serta lampiran-lampiran yang diperlukan.