

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Kanker adalah salah satu penyakit yang menjadi penyebab utama kematian manusia diberbagai negara. Berdasarkan data GLOBOCAN, *International Agency for Research on Cancer* (IARC), diketahui bahwa pada tahun 2012 terdapat 14.1 juta kasus baru mengenai kanker, 8.2 juta kematian akibat kanker dan 32.6 juta orang hidup dengan kanker di seluruh dunia [1]. Karena itu, sangat diperlukan suatu teknologi yang mampu mendeteksi penyakit kanker lebih awal dengan analisis yang akurat, sehingga penyakit kanker dapat ditangani sejak dini.

Pada beberapa tahun terakhir, suatu analisis dan diagnosis kanker dilakukan dengan menggunakan data *microarray*. DNA *microarray* merupakan suatu teknologi yang digunakan untuk memonitor suatu ekspresi data (gen) dalam jumlah yang sangat besar dan berbeda secara bersamaan. Teknologi *microarray* telah digunakan untuk diagnosis medis dan analisis gen, terutama untuk memeriksa perubahan pola ekspresi gen. Dengan analisis data ekspresi gen berupa *microarray* tersebut, memungkinkan suatu medis untuk dapat mengetahui apakah seseorang terkena suatu kanker atau tidak. Karena itu, *microarray* sangat berperan penting dalam perihal diagnosis kanker.

Pada data *microarray*, adanya perbedaan antara jumlah gen yang diukur dengan jumlah individu sampel yang ada akan sangat berpengaruh dalam menganalisis ekspresi gen yang akan digunakan dalam mengidentifikasi ataupun mencari gen informatif [2]. Klasifikasi ekspresi gen ataupun klasifikasi kanker merupakan salah satu kajian dalam bidang bioinformatika yang digunakan untuk analisis suatu kanker. Proses klasifikasi dilakukan untuk mengidentifikasi gen informatif yang dapat digunakan untuk memprediksi kelas dalam pengujian sampel baru [3], untuk itu diperlukan suatu skema maupun metode klasifikasi yang efisien, berkinerja tinggi dan memiliki akurasi yang baik. Adapun skema klasifikasi yang diusulkan yaitu dengan melakukan tiga tahapan proses, diantaranya: preprocessing, reduksi dimensi, dan klasifikasi.

Para peneliti telah melakukan penelitian terhadap metode klasifikasi kanker berdasarkan *microarray* data. Berikut ini adalah beberapa literatur yang berkaitan dengan penelitian ini. Vanitha D.A. [2], membangun skema *Mutual Information-based Gene Selection* (MI) dengan *Support Vector Machines* (SVM) sebagai *feature selection* dan *classifier* pada data *microarray*, hasil akurasi diperoleh berdasarkan LOOCV *mean accuracy rate*. Adapun hasil akurasi yang diperoleh pada data kanker *colon* berdasarkan metode ANN (*Artificial Neural Network*) mendapatkan akurasi sebesar 0.5094, SVM dengan kernel *linear* memiliki akurasi sebesar 0.6774, SVM dengan kernel RBF memiliki akurasi sebesar 0.6051 dan SVM dengan kernel *polynomial* memiliki akurasi sebesar 0.4683. Terlihat bahwa pada [2], SVM dengan kernel *linear* dan RBF menghasilkan akurasi yang lebih baik dari ANN. Sedangkan pada paper Seeja.K.R. [4], membangun skema F-Score dengan SVM sebagai *feature selection* dan *classifier* data *microarray*, pada data leukemia skema tersebut mampu menghasilkan *running time* yang lebih cepat dari ANN. Sedangkan pada paper Nurfalah.A [5], skema yang dibangunnya yaitu dengan menggunakan PCA dan MBP (*Modified Backpropagation*) sebagai reduksi dimensi dan klasifikasi data *microarray*, dari skema tersebut menghasilkan akurasi sebesar 96% untuk kanker ovarian, 76.92% untuk kanker *colon*, dan 97.14% untuk data leukimia. Berdasarkan [5], Metode kombinasi PCA dan MBP menghasilkan waktu *training time* yang lebih cepat dari metode *Backpropagation* standar

Berdasarkan beberapa penelitian sebelumnya, maka pada penelitian ini skema yang diusulkan untuk deteksi kanker berdasarkan klasifikasi data *microarray* adalah kombinasi metode PCA dan SVM. Pada tahap pengurangan dimensi akan dilakukan dengan menggunakan metode PCA, PCA telah banyak digunakan pada beberapa penelitian sebelumnya dan terbukti sangat efektif dalam pengurangan dimensi. Sedangkan pada tahap klasifikasi akan menggunakan metode *Support Vector Machines* (SVM), dimana SVM dapat memecahkan klasifikasi yang kompleks. Dalam metode klasifikasi gen yang diuraikan oleh Furey et.al [6], SVM memiliki kinerja yang baik, dengan kernel sederhana dalam analisis ataupun klasifikasi data *microarray*. Adapun data kanker yang akan digunakan pada penelitian ini diantaranya: data kanker *colon, central nervous, lung, prostate, breast, ovarian* dan leukemia yang berupa data *microarray*.

1.2. Perumusan Masalah

Adapun rumusan masalah dari penelitian ini, diantaranya:

1. Bagaimana menerapkan skema reduksi dimensi pada data *microarray*?
2. Bagaimana klasifikasi kelas kanker berdasarkan data *microarray* hasil reduksi dimensi?
3. Bagaimana menganalisis performansi hasil klasifikasi kanker?

1.3. Batasan Masalah

Dalam penelitian ini terdapat beberapa batasan masalah yang akan dilakukan, diantaranya:

1. *Software* yang digunakan yaitu Matlab R2013a.
2. Jumlah kernel *option* yang digunakan yaitu 9, 18, dan 27. Kernel *option* merupakan parameter masukan (*input*) pada fungsi kernel SVM.
3. Selain kernel *option*, parameter SVM lainnya yang digunakan yaitu $C=1000$.

1.4. Tujuan

Adapun tujuan dari penelitian ini, diantaranya:

1. Menerapkan *Principal Component Analysis* (PCA) sebagai metode reduksi dimensi pada data *microarray*.
2. Menerapkan *Support Vector Machines* (SVM) untuk klasifikasi kelas kanker.
3. Mengetahui performansi dari algoritma *Support Vector Machines* (SVM) dengan *Principal Component Analysis* (PCA).

1.5. Metodologi Penyelesaian Masalah

Berikut ini merupakan tahapan-tahapan yang dilakukan:

1. Penelusuran Pustaka
Mencari referensi yang dibutuhkan berdasarkan topik untuk menunjang penyelesaian penelitian terhadap topik yang diajukan.
2. Analisis Data
Data yang digunakan merupakan data *microarray*. Data *microarray*, merupakan data yang berdimensi tinggi dimana jumlah *feature* yang ada sangat banyak namun jumlah *sampelnya* sedikit.
3. Analisis dan Perancangan Sistem
Menganalisis suatu skema yang mampu mengimplementasikan penelitian serta membuat skema sistem dengan algoritma PCA dan SVM (*Support Vector Machines*), yang kemudian diimplementasikan pada suatu *tools* ataupun *software* tertentu.

4. Pengujian dan Analisis

Menguji performansi algoritma SVM (*Support Vector Machines*) dan PCA pada sistem, dari skema yang telah diterapkan. Melakukan analisis akurasi terhadap klasifikasi pada data kanker yang telah diuji.

5. Mengambil Kesimpulan

Mengambil kesimpulan dari hasil penelitian yang sudah didapatkan.

1.6. Sistematika Penulisan

Adapun sistematika penulisan pada penyusunan Tugas Akhir ini, diantaranya:

BAB 1 PENDAHULUAN

Pada bab ini, membahas mengenai latar belakang masalah, perumusan masalah, batasan masalah, tujuan, dan sistematika penulisan.

BAB 2 STUDI LITERATUR

Pada bab ini, membahas mengenai dasar teori yang menunjang untuk penelitian yang dilakukan.

BAB 3 PERANCANGAN SISTEM

Pada bab ini, membahas mengenai proses perancangan skema yang akan digunakan dalam pelaksanaan penelitian.

BAB 4 PENGUJIAN DAN ANALISIS

Pada bab ini, membahas mengenai implementasi beserta hasil akhir yang diperoleh dari implementasi.

BAB 5 KESIMPULAN DAN SARAN

Pada bab ini, berisi tentang kesimpulan dari semua proses dan hasil yang diperoleh, beserta saran untuk pengembangan penelitian selanjutnya.