

1. Pendahuluan

1.1. Latar Belakang

Menurut data dari *World Health Organization* (WHO), kanker dianggap sebagai penyebab utama kematian di dunia. Sekitar 8,2 juta orang meninggal karena kanker dan jumlah ini diperkirakan akan terus bertambah pada setiap tahunnya. Dikarenakan berbahayanya penyakit ini maka perlu dilakukan pendeteksian secara dini. Cara melakukan pendeteksian dapat dengan melalui pemeriksaan secara fisik, pemeriksaan darah, *rontgen* maupun *X-ray*. Namun pendeteksian kanker yang ada dapat memakan waktu yang cukup lama, serta sangat rentan terjadi berbagai kesalahan yang dapat dilakukan oleh manusia. Seperti yang dimuat pada jurnal kesehatan *British Medical Journal* (BMJ), kesalahan medis menyebabkan 251.454 kematian setiap tahunnya di Amerika Serikat. Oleh sebab itu diperlukan suatu teknologi yang dapat membantu dalam menyelesaikan permasalahan tersebut, salah satunya adalah teknologi *gene expression*. Teknologi *gene expression* merupakan salah satu metode pendeteksian kanker dengan bantuan teknologi *microarray* [1].

Teknologi *microarray* adalah salah satu teknologi yang dibangun untuk mempelajari ekspresi dari banyak sifat yang dibawa oleh gen dalam satu waktu [2], alasan ini telah menjadi dasar beberapa penelitian dalam mendeteksi kanker. Namun masalah utama dalam penyelesaian *microarray* adalah mengolah data dalam *microarray*, hal ini dikarenakan dimensi yang dimiliki *microarray* sangat besar [2].

Masalah pengklasifikasian data *microarray* telah banyak mengundang peneliti melakukan penelitian terkait hal ini, seperti Luis M. de Campos [3], Devi Arockia [4], dan Liwei Fan [5]. Dalam [3], dijelaskan secara rinci yang dimaksud dengan *microarray*, detail distribusi data, dan pendekatan untuk memecahkan kasus pengklasifikasian data *microarray*. Banyak pendekatan berbeda yang telah diteliti, seperti [4] yang mengusulkan pendekatan *feature selection* menggunakan *Mutual Information* (MI), pada penelitian tersebut kemampuan MI memiliki performa yang bagus dalam hal mereduksi dimensi. Namun, dalam penelitian ini memiliki kekurangan dalam hal metode pengklasifikasian yaitu dalam hal penyelesaian data

berskala besar. Sedangkan dalam [5], untuk mengklasifikasikan data *microarray*, digunakan *Naïve Bayes* (NB), yaitu dengan mengasumsikan ke independenan pada setiap atribut terhadap *class*.

Dalam penelitian Tugas Akhir ini, dihasilkan sistem yang mampu pengklasifikasian data *microarray*. Pendekatan ini diharapkan bisa mendapatkan informasi terpenting dari pola *feature selection* yang ada. Untuk *feature selection* yang digunakan difokuskan dengan menggunakan *Mutual Information*. Dengan demikian, kasus pengklasifikasian data *microarray* dapat ditangani. Sedangkan *machine learning* yang digunakan adalah *Naïve Bayes* (NB), NB digunakan karena mampu mencari pola dan keterhubungan atribut dengan *class* dengan pendekatan probabilitas. Dengan pendekatan ini, sistem yang dibangun ternyata mampu mengklasifikasikan data *microarray* dengan sangat baik.

1.2. Perumusan Masalah

Pada tugas akhir ini memiliki beberapa batasan masalah seperti;:

- a) Bagaimana cara mereduksi dimensi *feature* agar *learning time* lebih cepat?
- b) Bagaimana membangun sistem yang dapat mendeteksi kanker dengan cara mengklasifikasikan dari data *microarray*?
- c) Bagaimana menganalisis performa sistem yang dibangun dalam mendeteksi kanker dari data *microarray*?

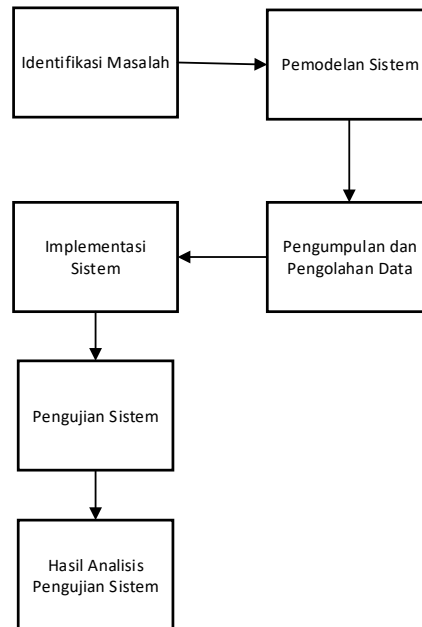
1.3. Tujuan

Berdasarkan rumusan masalah yang sudah dideskripsikan di atas dapat disimpulkan bahwa tugas akhir ini memiliki beberapa tujuan, sebagai berikut :

- a) Mereduksi dimensi dengan pendekatan *feature selection* menggunakan *Mutual Information*.
- b) Menganalisis dan mengimplementasikan sistem pengklasifikasian data *microarray* dengan metode *Naive Bayes*.
- c) Menganalisis performansi algoritma klasifikasi yang dibangun menggunakan *cross validation* dan *F1 Measure*.

1.4. Metodologi Penyelesaian Masalah

Pada bagian ini akan menggambarkan bagan alir beserta penjelasan dari metodologi penyelesaian masalah tugas akhir ini. Bagan alir dapat dilihat pada gambar 1-1 :



Gambar 1-1 Alur Bagan Metodologi Penelitian.

Gambar 1-1 menggambarkan alur proses penelitian pada tugas akhir ini. Dibawah ini terdapat penjabaran dari alur bagan metodologi penelitian tersebut, seperti :

1.4.1. Identifikasi Masalah

Berdasarkan permasalahan yang sudah dirumuskan pada sub bab rumusan masalah di atas. Maka pada tugas akhir ini bertujuan untuk meninjau pengaruh dari metode tersebut. Informasi dikutip dari *paper*, *e-book*, *slide*, buku dan bahan lain yang dapat menunjang pembuatan tugas akhir ini

1.4.2. Pemodelan Sistem

Pada tugas akhir ini dilakukan pemodelan sistem yang bertujuan untuk memberikan gambaran sistem yang dibangun. Reduksi dimensi yang akan diterapkan pada tugas akhir ini diterapkan metode *Mutual Information*. Kemudian pada proses klasifikasi diterapkan metode *Naive Bayes*.

1.4.3. Pengumpulan dan Pengolahan Data

Tahap pengumpulan *resource* adalah tahap yang harus dilakukan agar tercapainya suatu penelitian ilmiah tersebut. Tugas akhir ini memanfaatkan lima dataset yang berasal dari Kent Ridge Bio-medical [6]. Pada tabel 1-1 terdapat informasi mengenai dataset yang digunakan.

Tabel 1-1 Informasi Dataset.

No	Dataset	Number of Instances	Number of Genes	Number of Classes
1	<i>Colon Cancer</i>	62	2000	2
2	<i>Ovarian Cancer</i>	253	15154	2
3	<i>Leukimia ALL-AML</i>	72	7129	2
4	<i>Breast Cancer</i>	97	24481	2
5	<i>Lung Cancer</i>	181	12533	2

1.4.4. Implementasi Sistem

Pada tahap ini, menerapkan metode pada sistem yang akan dibangun menggunakan bahasa pemrograman java berdasarkan permasalahan pada penelitian sebelumnya yang sudah dijelaskan di sub bab rumusan masalah. Untuk penjelasan proses implementasi sistem pada tugas akhir ini dapat dilihat pada bab 3 selanjutnya.

1.4.5. Pengujian Sistem

Pada tahap ini akan dilakukan pengujian sistem terhadap lima dataset yang berasal dari Kent Ridge Bio-medical [6]. Pengukuran pada pengujian ini ditinjau dari hasil *Recall*, *Precision*, dan *F1-Score* yang dihasilkan untuk setiap metode yang diterapkan.

1.5. Sistematika Penulisan

Untuk lebih jelas memahami laporan tugas akhir ini, berikut terdapat gambaran singkat mengenai informasi setiap bab dalam laporan tugas akhir ini.

BAB I Pendahuluan

Pada bagian ini berisi tentang latar belakang, perumusan masalah, tujuan, metodologi penyelesaian masalah, dan sistematika penulisan.

BAB II Kajian Pustaka

Pada bagian ini berisikan tentang teori-teori yang digunakan pada analisis ini, berupa pengertian, definisi yang diambil dari kutipan buku, jurnal, *e-book*, serta beberapa literatur *review* yang berhubungan dengan penelitian.

BAB III Perancangan Sistem

Pada bagian ini menjelaskan tentang proses sistem yang diusulkan, dan detail tentang proses serta hasil kerja dari sistem yang di implementasikan. Kemudian dijabarkan secara satu persatu dengan menerapkan konsep sesudah adanya sistem yang diusulkan.

BAB IV Pengujian dan Analisis

Pada bagian ini berisikan hasil pengujian dan analisis dari pengujian yang sudah dilakukan. Dijelaskan secara detail satu persatu berdasarkan proses yang telah diuji.

BAB V Kesimpulan dan Saran

Pada bagian ini berisi kesimpulan dan saran yang berkaitan dengan analisis berdasarkan yang telah diuraikan pada bab-bab sebelumnya.