

## 1. Pendahuluan

### Latar Belakang

Menurut data dari WHO pada tahun 2015 terdapat sekitar 8.8 juta orang meninggal disebabkan oleh penyakit kanker [3]. Untuk tahun 2018 diperkirakan sekitar 9.6 juta orang meninggal disebabkan oleh kanker [15]. Terdapat 2.09 juta kasus kanker paru-paru dan 1.76 juta orang yang meninggal disebabkan oleh kanker paru-paru pada tahun 2018 [15]. Sebanyak 2.09 juta orang terkena kanker payudara dan 627 ribu diantaranya meninggal dikarenakan kanker payudara pada tahun 2018 [15]. Dari data tersebut, terdapat peningkatan angka kematian yang disebabkan oleh kanker setiap tahunnya. Selama ini masyarakat masih mendeteksi penyakit kanker dengan cara yang tradisional yaitu dengan melihat ciri-ciri gejala terjangkit penyakit kanker. Teknik *microarray* merupakan salah satu teknik yang dapat digunakan untuk mendeteksi kanker dengan akurasinya yang dapat dipertimbangkan.

Teknik *microarray* merupakan salah satu teknik yang dapat membantu kita dalam mendeteksi adanya penyakit kanker. Di dalam *microarray* terdapat banyak *gene expressions* yang dapat mendeskripsikan sel-sel DNA yang menjadi penyebab kanker. *Microarray* memiliki data dengan dimensi yang besar. Dengan menganalisis menggunakan *gene expression* dapat mempersingkat waktu yang dibutuhkan untuk menyimpulkan penyakit kanker dibandingkan dengan menyimpulkan menggunakan metode tradisional. Dan hasil yang didapatkan bisa dipertanggung jawabkan berdasarkan aturan analisa yang terdapat di ahli medis.

Salah satu masalah yang sering terjadi dalam melakukan deteksi kanker adalah teknik klasifikasi data *microarray*. Hal ini bisa terjadi dikarenakan data yang digunakan memiliki dimensi yang besar. Namun data *microarray* dapat dimaksimalkan dengan melihat data yang berpengaruh untuk menentukan penderita penyakit kanker.

Pada tahun 2018 terdapat penelitian tentang klasifikasi data *microarray* menggunakan metode klasifikasi C4.5 *Decision Tree* dan BPSO, dengan hasil akurasi 99% [16]. Sedangkan pada penelitian deteksi kanker menggunakan metode reduksi dimensi *Minimum Redundancy Maximum Relevance* dan *Support Vector Machine* menghasilkan akurasi 100% [2]. Sehingga dalam penelitian tugas akhir ini, metode yang digunakan adalah *Minimum Redundancy Maximum Relevance* dan C4.5 *Decision Tree*.

Alur pengerjaan pada penelitian ini, dilakukan dengan 3 tahapan, yaitu *preprocessing*, *dimensional reduction*, dan *gene classification*. Pada proses *preprocessing* menggunakan metode persamaan normalisasi data. Untuk seleksi fitur pada proses *dimensional reduction* menggunakan *Minimum Redundancy Maximum Relevance (MRMR)* dan *Genetic algorithm*. Sedangkan pada proses *gene classification* menggunakan metode C4.5 *decision tree*.

### Topik dan Bahasan

Berdasarkan latar belakang dari penelitian terkait, maka rumusan masalah pada tugas akhir ini adalah bagaimana membangun sistem yang dapat mereduksi dimensi pada data *microarray* menggunakan metode MRMR GA dan bagaimana klasifikasi data *microarray* menggunakan data C4.5 dengan parameter yang digunakan berupa akurasi.

Adapun batasan masalah pada penelitian kali ini adalah, dataset yang digunakan merupakan data *microarray* cancer yang berasal dari Kent-Ridge Repository (<http://leo.ugr.es/elvira/DBCRepository/>). Terdapat 5 jenis data yang diolah pada penelitian kali ini, yaitu data *Colon Tumor*, *Lung Cancer*, *Leukimia*, *Ovarian Cancer* dan *Breast Cancer*.

### Tujuan

Tujuan dari tugas akhir ini adalah membangun sebuah sistem untuk mendeteksi kanker pada dataset *microarray* menggunakan metode reduksi dimensi MRMR GA dan menggunakan metode klasifikasi C4.5. Hasil akurasi yang didapatkan akan menjadi parameter keberhasilan dari suatu sistem dalam mendeteksi kanker dengan data *microarray*. Sistem diimplementasikan menggunakan perangkat lunak Spyder.