

Abstrak

Kanker merupakan salah satu penyakit yang mematikan di dunia. Setiap tahunnya, penderita kanker terus meningkat dan banyak menelan korban jiwa. Hingga sampai saat ini, obat untuk penyakit yang mematikan ini masih sulit ditemukan. Dalam beberapa tahun terakhir, teknologi data *microarray* banyak digunakan untuk mendiagnosa kanker sejak dini, data DNA *microarray* adalah teknologi yang digunakan untuk melihat urutan sekuens asam nukleat yang berada pada lokasi tertentu pada struktur DNA yang dapat digunakan untuk menganalisa ribuan sampel pada waktu yang bersamaan sehingga nantinya dapat diklasifikasikan mana yang tergolong kanker dan bukan kanker. Oleh karena itu, data *microarray* adalah data yang memiliki ukuran dimensi data yang sangat besar. Data yang ukuran dimensinya sangat besar dapat mengakibatkan hasil perhitungan menjadi tidak optimal dan akurasi klasifikasi yang dihasilkan kecil. Untuk mengoptimalkan data dan meningkatkan nilai akurasi klasifikasi dari data yang dimensinya besar tersebut, dilakukan pengurangan dimensi dengan seleksi fitur *Genetic Algorithm* (GA). *Genetic Algorithm* biasanya mampu memberikan hasil yang baik dan tingkat akurasi yang cukup baik. Klasifikasi data *microarray* menggunakan metode *Naive Bayes* dan Regresi Logistik. Adapun akurasi terbaik dari *Genetic Algorithm* dan klasifikasi Regresi Logistik 100% pada data *colon tumor* dan *mll leukemia*. *Genetic Algorithm* dan klasifikasi *Naive Bayes* 57,7778% pada data MLL Leukemia. Dan Regresi Logistik 67% pada data *mll leukemia*.

Kata Kunci : kanker, seleksi fitur, klasifikasi, data *microarray*, *genetic algorithm*, *naive bayes*, regresi logistik