

1. Pendahuluan

Pada bagian ini berisi empat sub-bagian, yaitu Latar Belakang, Topik dan Batasannya, Tujuan, dan Organisasi Tulisan. Berikut adalah mengenai penjelasannya.

Latar Belakang

Kanker merupakan salah satu penyebab utama kematian di dunia, menurut World Health Organization (WHO). Kanker menduduki posisi ke dua dari 10 data penyakit penyebab kematian secara global, dengan estimasi angka mencapai 9.6 juta kematian di tahun 2018, [1].

Sementara itu seiring berkembangnya ilmu pengetahuan, telah ditemukan cara untuk mendeteksi kanker melalui pemrosesan ekspresi genetik dengan mengekspresikan data gen manusia ke dalam suatu media penampung yang disebut *microarray*. *Microarray* dapat digunakan sebagai media untuk mendeteksi gejala kanker dari suatu ekspresi genetik seseorang [7].

Untuk mendeteksi kanker dalam data *microarray*, diperlukan upaya pengolahan dalam proses *data mining* yang disebut klasifikasi. Namun di sisi lain data DNA *microarray* memiliki jumlah dimensi yang besar dan dapat berdampak buruk pada performansi klasifikasi [2]. Salah satu upaya yang sangat memungkinkan dapat dilakukan untuk mengatasi hal ini adalah dengan menyeleksi fitur yang dianggap kurang *relevant* sehingga menghasilkan performansi klasifikasi yang optimal [4]. Hal ini menunjukkan bahwa sebuah metode seleksi fitur di anggap perlu untuk di lakukan sebelum menerapkan proses klasifikasi, agar pendektisan kanker dapat dilakukan dengan lebih optimal.

Namun tidak semua metode seleksi fitur dan skema penerapannya dapat mengoptimalkan performansi klasifikasi, diperlukan banyak percobaan dan analisa perbandingan metode fitur seleksi serta skema penerapannya terhadap proses klasifikasi untuk menemukan hasil yang paling optimal. Berdasarkan hal tersebut, dalam penilitan ini penulis mengimplementasikan sebuah skema perbandingan dua metode seleksi fitur dengan jenis *filter* untuk mengoptimasi metode klasifikasi *Support Vector Machine* (SVM). Metode seleksi fitur *filter* dipilih karena memiliki penilaian *independent* terhadap kualitas fitur yang baik, dan memiliki komputasi yang sederhana dan cepat dibandingkan metode lain[16], maka dari itu metode filter akan tepat jika digunakan pada data *microarray* yang berdimensi banyak.

Dua metode seleksi fitur *filter* yang dipilih adalah ReliefF yang merupakan metode dengan pendekatan *ranking* dan *Correlation – based feature selection* (CFS) dengan pendekatan *subset*. ReliefF memberikan nilai *ranking* pada setiap fitur terhadap atribut kelas, yang mana nantinya fitur dengan nilai *weight* tertinggi memungkinkan memiliki pengaruh yang baik dalam proses klasifikasi. Sementara *Correlation – based feature selection* (CFS) akan menilai bagus tidak nya suatu *subset* fitur menggunakan perhitungan *Merits*, berdasarkan korelasi antar fitur dengan kelas, dan korelasi antar satu fitur dengan fitur – fitur yang lain, semakin besar nilai *Merits* yang didapatkan oleh sebuah subset maka akan berdampak lebih bagus untuk proses klasifikasi. Sedangkan metode klasifikasi SVM *Support Vector Machine* (SVM) dipilih karena dapat menghasilkan akurasi yang lebih baik dalam kasus *microarray*, dibandingkan dengan beberapa metode klasifikasi yang lain[15].

Topik dan Batasannya

Sistem klasifikasi data *microarray* ini dibuat menggunakan metode *Support Vector Machine* (SVM) dengan metode seleksi fitur ReliefF dan seleksi fitur CFS. Dimana kedua metode seleksi fitur tersebut akan dibandingkan performansinya dalam meningkatkan kinerja metode klasifikasi SVM. Dataset yang digunakan dalam tugas akhir ini diambil dari Kent Ridge Bio-medical Data Set Repository (<http://leo.ugr.es/elvira/DBCRepository/>) yang terdiri dari *breast cancer*, *colon tumor*, *ovarian*, *lung cancer*, *prostate*, *central nervous system*, dan *MLL leukemia*.

Tujuan

Dengan dibuatnya penelitian ini dapat mengetahui bagaimana pengimplementasian dan hasil perfromansi dari metode seleksi fitur ReliefF dan *Correlation based Feature Selection* (CFS) dalam meningkatkan kinerja klasifikasi dari metode *Support Vector Machine* (SVM) pada data *microarray*.

Organisasi Tulisan

Laporan ini dibagi menjadi beberapa bagian untuk menjelaskan hasil dari pengujian. Pada bagian 1, pendahuluan menjelaskan tentang latar belakang, batasan dan tujuan penelitian ini dibuat. Pada bagian 2, studi terkait menjelaskan tentang metode yang digunakan dan bagian 3 sistem yang dibangun menjelaskan bagaimana penerapan pada sistem. Bagian 4, Evaluasi menjelaskan tentang hasil pengujian yang dilakukan dan kesimpulan dituliskan pada bagian 5.