

ABSTRAK

Microarray data mengambil bagian penting di dalam mendiagnosis dan mendeteksi kanker karena analisis microarray dapat digunakan untuk melihat tingkat ekspresi gen dalam sampel sel tertentu yang berfungsi untuk menganalisis ribuan gen secara bersamaan. Namun, microarray data memiliki sangat sedikit data sample dan memiliki dimensi data yang tinggi. Sehingga untuk melakukan klasifikasi pada microarray data membutuhkan proses reduksi dimensi. Reduksi dimensi dapat menghilangkan redundancy pada data sehingga fitur yang digunakan pada klasifikasi adalah fitur yang memiliki correlation yang tinggi terhadap kelasnya. Ada 2 jenis reduksi dimensi yaitu seleksi fitur dan ekstraksi fitur. Pada penelitian ini akan menggunakan seleksi fitur, dengan menggunakan algoritma clustering k-means untuk mengelompokkan fitur yang memiliki similarity yang tingkat pada 1 cluster, kemudian untuk setiap cluster dilakukan proses perankingan menggunakan metode Relief. Setelah itu, fitur yang memiliki skor yang tinggi akan dipilih sebagai subset fitur untuk proses klasifikasi. Tujuannya adalah untuk menghapus redundancy pada data yang dapat menurunkan akurasi pada klasifikasi. Selanjutnya pada proses klasifikasi akan menggunakan algoritma Random Forest. Dari hasil penelitian ini diperoleh hasil akurasi untuk setiap dataset, yaitu Colon 85.87%, Lung Cancer 98.9%, dan Tumor Prostate 89%. Hasil akurasi yang diperoleh lebih tinggi daripada penelitian sebelumnya, yang hanya menggunakan algoritma Random Forest sebagai seleksi gen dan klasifikasi, sehingga dapat disimpulkan pendekatan clustering untuk menghapus redundancy dimensi dapat digunakan dan diterapkan pada klasifikasi menggunakan microarray data.

Kata kunci: Microarray, high dimensional, reduksi dimensi, seleksi fitur, ekstraksi fitur, clustering, k-means, information entropy, klasifikasi, random forest.