

Klasifikasi multikelas data kanker leukemia menggunakan fuzzy support vector machines dengan pemilihan fitur principal component analysis = Multiclass classification of leukemia cancer data using fuzzy support vector machines with feature selection using principal component analysis

Inry Raudiatul Fauzi, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20466422&lokasi=lokal>

Abstrak

"ABSTRAK
"

Kanker merupakan penyakit penyebab kematian terbesar kedua di dunia. Menurut prediksi WHO 2015 kasus kematian akibat kanker akan meningkat menjadi 21,6 juta kasus pada tahun 2030. Salah satu usaha untuk mengurangi penyebaran kanker dengan menggunakan machine learning adalah melakukan pendeteksian jenis kanker dengan memanfaatkan microarray data. Pada umumnya, microarray data kanker terdiri dari banyak fitur. Namun, tidak semua fitur yang ada pada data kanker memiliki informasi penting. Oleh karena itu, fitur-fitur tersebut akan diekstraksi menggunakan metode Principal Component Analysis PCA . Kemudian dipilih fitur-fitur yang paling informatif dari data hasil ekstraksi PCA. Fitur-fitur terpilih dari data hasil ekstraksi akan dibentuk dalam data baru. Data sebelum dan data setelah dilakukan pemilihan fitur akan diklasifikasi menggunakan metode Fuzzy Support Vector Machines FSVM . Akurasi dari proses klasifikasi dua tahap tersebut akan dibandingkan. Pendekatan one versus one akan digunakan pada masalah klasifikasi multikelas data kanker leukemia. Dengan pendekatan tersebut akan terbentuk sebanyak $k-1/2$ masalah dua kelas, di mana k menunjukkan jumlah kelas. Hasilnya, tanpa melakukan pemilihan fitur, diperoleh akurasi tertinggi sebesar 87.69 . Setelah dilakukan pemilihan fitur, diperoleh akurasi terbaik dengan menggunakan 60 fitur dengan akurasi sebesar 96,92 .

<hr>

ABSTRACT

Cancer is the second leading cause of death globally. According to WHO prediction 2015 cases of cancer deaths will increase become 21.6 million cases by 2030. One of the effort to reduce the spread of cancer by using machine learning is to detect the types of cancer. We can use microarray data to detect the types of cancer. In general, microarray cancer data consist of many features. However, not all features in cancer data have important information. Therefore, these features will be extracted by using Principal Component Analysis PCA method. Then, we select the most features who have important information of data extraction. The selected features of extracted data will be formed in the new data. Data, before and after selection will be classified using Fuzzy Support Vector Machines FSVM method. The accuracy of the classification process will be compared. The one versus one approach will be used on multiclass leukemia cancer data. This approach will formed the multiclass problem into $k-1/2$ binary class problems, where k denotes the number of classes. The results, without doing feature selection, the highest accuracy is 87.69 . After doing feature selection, the best accuracy is obtained by using 60 features with the accuracy is 96.92 .