

Klasifikasi data kanker menggunakan support vector machines dengan pemilihan fitur menggunakan correlated based support vector machines-recursive feature elimination = Cancer classification using support vector machines with feature selection using correlated based support vector machines recursive feature elimination

Nurul Maghfirah, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20459050&lokasi=lokal>

Abstrak

Kematian yang disebabkan oleh kanker diperkirakan akan terus meningkat, padahal jumlah kematian ini dapat dikurangi dengan adanya deteksi dini. Salah satunya adalah dengan klasifikasi data kanker. Data kanker yang digunakan merupakan data kanker berdimensi tinggi dengan ribuan fitur, tetapi tidak semua fitur yang ada merupakan fitur yang relevan. Oleh karena itu, perlu adanya proses seleksi fitur. Untuk meningkatkan tingkat akurasi yang dihasilkan, digunakan sebuah metode seleksi fitur yang meninjau adanya korelasi antar gen, yaitu CSVM-RFE. Pada metode tersebut, data yang ada diproyeksikan dan diubah menjadi sebuah data baru dengan ekstraksi fitur, dan kemudian dilakukan proses seleksi fitur. Penggunaan dua metode tersebut pada klasifikasi tiga data kanker yang ada terbukti menghasilkan tingkat akurasi yang tinggi, pada data kanker kolon tingkat akurasi yang didapatkan adalah sebesar 96,6, pada kanker prostat sebesar 98,9, dan pada kanker lymphoma sebesar 98,6.

.....The number of death caused by cancer expected to rise over two decades, whereas the number of death can be reduced by early detection. One of them is cancer classification. Cancer dataset is a high dimensional dataset that consist of thousands of features, but not all of these features are relevant. Therefore, it is necessary to remove the redundant features using feature selection. Feature selection can also improve the accuracy of classification. Many feature selection methods do not consider the correlated genes, so we need a new feature selection method that consider the correlated genes. It is CSVM RFE, in this method the existing data is projected and converted into a new data with feature extraction. These two methods are applied to the cancer datasets, and produce the accuracy of 96.6 using colon cancer dataset, 98.9 using prostate cancer dataset, and 98.6 using lymphoma cancer dataset.