

Perbandingan pemilihan fitur genetic algorithm dan laplacian score menggunakan klasifikasi support vector machines pada data kanker = Comparison of feature selection of genetic algorithm and laplacian score using classification of support vector machines on cancer data

Intan Primasari, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20458403&lokasi=lokal>

Abstrak

ABSTRAK

Kanker merupakan salah satu penyakit yang paling mematikan bagi manusia. Menurut WHO 2015, kanker adalah penyebab kematian nomor 2 di dunia sebesar 13 setelah penyakit kardiovaskular. Salah satu hal yang dapat dilakukan untuk penelitian kanker menggunakan machine learning adalah melakukan pendeteksian jenis kanker dengan memanfaatkan microarray data. Microarray data yang memiliki banyak fitur. Itu merupakan salah satu kendala dalam penerapan teknik machine learning. Hal ini akan mempengaruhi performa atau keakuratan dari hasil klasifikasi pada data kanker. Oleh karena itu, metode pemilihan fitur diperlukan untuk meningkatkan performa dalam pendeteksian kanker. Dalam tugas akhir ini dilakukan perbandingan pemilihan fitur menggunakan Genetic Algorithm dan Laplacian Score. Fitur-fitur yang sudah terpilih pada data kanker kemudian digunakan dalam proses klasifikasi Support Vector Machines. Hasilnya, didapatkan akurasi terbaik saat dengan metode pemilihan fitur menggunakan Genetic Algorithm yaitu 98,69 dengan penggunaan 40 fitur untuk data kanker prostat dan 98,97 dengan penggunaan 30 fitur untuk data kanker kolon.

ABSTRACT

Cancer is one of the most deadly diseases for humans. According to the WHO 2015, cancer is the causes of the death number two in the world by 13 after cardiovascular disease. Taking advantage from microarray data, machine learning methods can be applied to help cancer prediction according to its types. Microarray data has many features. It is one of the obstacles in the machine learning techniques. This will affect the performance or accuracy of the classification results on cancer data. Therefore, feature selection methods are required to increase performance in cancer prediction. This research proposed comparison of feature selection using Genetic Algorithm and Laplacian Score. Features that are already selected in the cancer data then used in the Support Vector Machines classification. The results show that the best accuracy obtained when using Genetic Algorithm with percentage of 98,69 by using 40 features for prostate cancer data and 98,97 by using 30 features for colon cancer data.