

Klasifikasi data kanker otak menggunakan adaboost support vector machines dengan pemilihan fitur signal-to-noise ratio = Classification of brain cancer data using adaboost support vector machines with signal to noise ratio feature selection

Tommy Rachmansyah Adyalam, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20474308&lokasi=lokal>

Abstrak

Kanker otak adalah pertumbuhan sel-sel abnormal di organ otak yang bersifat ganas. Salah satu cara untuk mengurangi perkembangan penyakit ini adalah melakukan pendeteksian dini menggunakan machine learning. Metode machine learning yang digunakan adalah AdaBoost Support Vector Machines untuk klasifikasi. AdaBoost Support Vector Machines adalah metode ensemble antara AdaBoost dengan base classifier Support Vector Machines. Data kanker otak direpresentasikan dalam bentuk matriks berupa ekspresi gen yang disebut DNA microarray. Data DNA microarray yang berdimensi tinggi akan direduksi dengan pemilihan fitur Signal-to-noise Ratio.

Pemilihan fitur bekerja untuk menemukan fitur-fitur yang informatif dan membuang fitur-fitur yang tidak sesuai. Pertama, data diklasifikasi menggunakan AdaBoost Support Vector Machines tanpa pemilihan fitur, dilanjutkan klasifikasi menggunakan AdaBoost Support Vector Machines dengan pemilihan fitur.

Pendekatan one vs one digunakan untuk menyelesaikan masalah multi kelas. Setelah melakukan pengujian, hasil akurasi terbaik adalah 91,111 pada data training 90 dengan menggunakan pemilihan fitur sebanyak 60 fitur. Hasil tersebut lebih baik dibandingkan klasifikasi tanpa pemilihan fitur yaitu 86,667 pada data training 90.

.....Brain cancer is the growth of abnormal cells in the brain organ malignantly. One way to reduce the progression of this disease is to do early detection using machine learning. Machine learning method used is AdaBoost Support Vector Machines for classification. AdaBoost Support Vector Machines is an ensemble method between AdaBoost and base classifier Support Vector Machines. Brain cancer data is represented in the form of matrix of gene expression called DNA microarray. The high dimensional DNA microarray data will be reduced by Signal to noise Ratio feature selection.

Feature selection works to find informative features and discard irrelevant features. Firts, the data is classified using AdaBoost Support Vector Machines without feature selection, further classified using AdaBoost Support Vector Machines with feature selection. The one vs one approach is used to solve multi class problems. After testing, the best accuracy result is 91,111 in 90 training data by using feature selection of 60 features. The result is better than the classification without feature selection that is 86,667 in 90 data training.