

Klasifikasi data diabetes dan kanker menggunakan entropi fuzzy C-means berbasis = Diabetes and cancer data classification using entropy fuzzy C-means based on kernel with chi-square feature selection

Nedya Shandri, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20494868&lokasi=lokal>

Abstrak

Penyakit kronis adalah penyakit yang diderita dalam waktu panjang dan dapat berkembang secara cepat, salah satunya adalah penyakit kanker dan diabetes. Oleh karena itu, dengan melakukan pendeteksian dini maka perkembangan penyakit kanker dan diabetes akan menurun. Salah satu cara pendeteksian dini dapat dilakukan oleh machine learning. Teknik machine learning banyak digunakan dalam berbagai bidang khususnya untuk analisa data medis. Clustering merupakan salah satu metode dari machine learning yang bertujuan untuk mengelompokkan suatu dataset ke dalam subset berdasarkan ukuran jarak. Salah satu contoh metode clustering adalah metode Entropi Fuzzy C-Means yang dapat mengidentifikasi entropi disetiap titik data dan memilih pusat kluster terdekat dengan entropi minimum. Pada penelitian akan digunakan data kanker dan diabetes dari UCI Repository dengan menggunakan metode Entropi Fuzzy C-Means yang akan dimodifikasi dengan kernel RBF. Sebelum dilakukan klasifikasi, dilakukan pemilihan fitur menggunakan Chi-Square. Tujuan dari penelitian ini adalah untuk mendapatkan fitur-fitur yang optimal dan mengetahui hasil akurasi menggunakan untuk klasifikasi data diabetes dan kanker. Diperoleh hasil akurasi tertinggi pada klasifikasi data medis menggunakan metode Entropi Fuzzy C-Means berbasis kernel dengan pemilihan fitur Chi-Square yaitu sebesar 83.33% untuk data diabetes dan 77.77-100% untuk data kanker.

<hr>

Chronic disease is a disease that occur for a long time and can develop quickly, one of them is cancer and diabetes. The early detection is very helpful to reduce the development of the disease. One of the ways to detect cancer and diabetes disease is using machine learning technic. Machine learning technic is widely use in many aspects especially in medical data analysis. Clustering is part of machine learning technic that is used to group a dataset into subset based on space size. Entropy Fuzzy C-Means is one of the methods which can identify entropy in every data and can choose the cluster center similar with minimum entropy. In this paper we will use cancer and diabetes medical data from UCI Repository using Entropy Fuzzy C-Means method which is modified by kernel RBF. Before classification, we will select the feature using Chi-Square to get the optimal subset feature. The purpose of this study was to obtain optimal features and find out the results of accuracy using for the classification of diabetes and cancer data. The medical data classification using Entropy Fuzzy C-Means based on kernel with Chi-Square feature selection gives the 100% highest accuration result for cancer data and 83,33% for diabetes data.