

Perbandingan Implementasi Algoritma KC-Means Clustering dan Kernel KC-Means Clustering untuk Memprediksi Penyakit Talasemia = The Comparison Between the Implementation of KC-Means Clustering and Kernel KC-Means Clustering Algorithm in Predicting Thalassemia Disease

Sri Hartini, supervisor

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20509530&lokasi=lokal>

Abstrak

Talasemia merupakan penyakit genetik dimana terjadi kelainan pada sel darah merah akibat kesalahan produksi hemoglobin. Perkawinan antara sesama pembawa gen talasemia dapat menyebabkan munculnya keturunan yang menderita talasemia mayor. Talasemia mayor merupakan jenis talasemia yang dapat menyebabkan kematian. Oleh karena itu, pendeteksian dini terhadap penyakit talasemia merupakan hal yang penting. Salah satu cara untuk melakukan pendeteksian talasemia adalah dengan menggunakan machine learning. Pada skripsi ini, algoritma KC-Means (KCM) clustering digunakan untuk memprediksi talasemia. Kemudian performa algoritma tersebut dibandingkan dengan algoritma Kernel KC-Means (KKCM) clustering, dimana fungsi kernel yang digunakan adalah kernel Gaussian Radial Basis Function (RBF) dan polinomial. Pengujian kedua algoritma tersebut dilakukan dengan menggunakan hold-out evaluation dan 5-fold cross validation. Data yang digunakan adalah data talasemia yang berasal dari Rumah Sakit Harapan Kita, Jakarta, Indonesia. Data talasemia tersebut terdiri dari 150 sampel dengan komposisi 82 sampel talasemia dan 68 sampel non-talasemia, yang tiap sampelnya memiliki 11 fitur. Berdasarkan percobaan yang dilakukan, algoritma KCM dan KKCM menghasilkan prediksi dengan akurasi, sensitivitas, presisi, spesifisitas, dan F1-Score yang sama, yaitu di atas 96 persen. Namun, waktu komputasi yang dibutuhkan oleh KKCM dapat mencapai hingga 10.25 kali lebih cepat dari waktu komputasi KCM.

Thalassemia is a genetic disease in which there are abnormalities in red blood cells due to an error in the production of hemoglobin. Marriage between thalassemia carriers can cause the birth of a child suffering from thalassemia major. Thalassemia major is a type of thalassemia that can cause death. Therefore, early detection of thalassemia is essential. One of the techniques to detect thalassemia is to use machine learning. In this thesis, the KC-Means (KCM) clustering algorithm is used to predict thalassemia. Then the performance of the algorithm is compared with the KC-Means Kernel (KKCM) clustering algorithm, where the kernel functions used are the Gaussian Radial Base Function (RBF) and polynomial kernel. Evaluation of those algorithms is carried out using hold-out evaluation and 5-fold cross-validation. The data used is thalassemia data from Harapan Kita Hospital, Jakarta, Indonesia. The data consists of 150 samples with a composition of 82 thalassemia samples and 68 non-thalassemia samples; each has 11 features. Based on the experiments, the KCM and KKCM algorithms make predictions with the same accuracy, sensitivity, precision, specificity, and F1-Score, which is above 96 percent. However, the computing time needed by KKCM can reach up to 10.25 times faster than the KCM computing time.