

Klasifikasi data talasemia menggunakan K-nearest neighbor dan naive bayes = Classification of data thalassemia using K-nearest neighbor and naive bayes

Annisa Kamalia, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20485398&lokasi=lokal>

Abstrak

ABSTRACT

Talasemia adalah penyakit yang disebabkan oleh adanya kelainan dalam hemoglobin. Penyakit talasemia merupakan penyakit hereditas atau penyakit keturunan dimana pembawa gen talasemia adalah orang tua dari penderita. Di Indonesia, pada tahun 2015 diketahui jumlah kasus talasemia mencapai 7.029 kasus. Sampai saat ini talasemia belum dapat disembuhkan namun dapat dikenali sifat pembawanya dengan skrining. Dalam tugas akhir ini, akan dibandingkan performa dari dua metode yang digunakan untuk mengklasifikasikan data talasemia, yaitu K-Nearest Neighbor dan Naive Bayes. Data yang digunakan adalah 82 data pasien talasemia dan 68 data pasien non-talasemia dari Rumah Sakit Anak dan Bunda Harapan Kita, Jakarta Barat. Hasil akhir menunjukkan bahwa metode Naive Bayes memberikan nilai akurasi yang lebih besar dari K-Nearest Neighbor dalam mengklasifikasikan talasemia. Rata-rata akurasi Naive Bayes sebesar 99.775% dengan rata-rata waktu running 0.0554 detik dan rata-rata akurasi K-Nearest Neighbor adalah 97.142% dengan rata-rata waktu running 0.081 detik. Untuk nilai spesifikasi, keduanya memberikan performa yang sama, yaitu dari K-Nearest Neighbor diperoleh ketika $K=3$ yaitu sebesar 100% dan dari Naive Bayes sebesar 100%. Hasil rata-rata sensitivitas tertinggi diberikan oleh Naive Bayes yaitu sebesar 99.59%, sedangkan K-Nearest Neighbor sebesar 96.25% untuk $K=1$.

ABSTRACT

Thalassemia is a disease caused by abnormalities in the hemoglobin. Thalassemia is a hereditary disease which the thalassemia gene carriers are parents of sufferers. In Indonesia, in 2015 it was found that the number of thalassemia cases reached 7,029 cases. Until now thalassemia has not been cured, but it can be recognized the nature of its carrier by screening. In this final project, the performance of the two methods will be compared to classify thalassemia data, namely K-Nearest Neighbor and Naive Bayes. The data used were 82 data on thalassemia patients and 68 data on non-thalassemia patients from Harapan Kita Children and Womans Hospital, West Jakarta. The final results show that the Naive Bayes method provides greater accuracy value than K-Nearest Neighbor in classifying thalassemia. The average accuracy of Naive Bayes is 99.775% with an average running time of 0.0554 seconds and the average accuracy of K-Nearest Neighbor is 97.142% with an average running time of 0.081 seconds. For specification values, both give the same performance. The result of specification values using K-Nearest Neighbor yield when $K = 3$ that is 100% and from Naive Bayes that is 100%. The highest average sensitivity results are given by Naive Bayes is 99.59%, while K-Nearest Neighbor is 96.25% for $K = 1$.