

Klasifikasi data diabetes dengan learning vector quantization dan seleksi fitur chi-square = Learning vector quantization for diabetes data classification with chi-square feature selection

Nadisa Karina Putri, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20492723&lokasi=lokal>

Abstrak

Diabetes mellitus atau biasa disebut sebagai diabetes adalah penyakit metabolik yang disebabkan oleh penderita memiliki kadar gula darah yang tinggi dan organ pankreas tidak dapat memproduksi hormon insulin secara efektif. Diabetes dapat mengakibatkan penyakit yang lebih parah seperti kebutaan, gagal ginjal, dan penyakit jantung. Oleh karena itu, pendeteksian sejak dini dibutuhkan agar pasien dapat mencegah penyakitnya sebelum menjadi lebih parah. Karena data medis biasanya berukuran besar dan tidak berdistribusi normal, beberapa peneliti menggunakan metode klasifikasi untuk memprediksi gejala penyakit atau mendiagnosa penyakit. Pada penelitian ini, digunakan algoritma Learning Vector Quantization (LVQ) untuk klasifikasi data set diabetes dengan seleksi fitur Chi-Square. Pada penelitian ini digunakan dua data set diabetes yaitu data set I dengan 8 fitur dan data set II dengan 19 fitur. Hasil dari penelitian ini menunjukkan bahwa untuk data set dengan 8 fitur, akurasi dan performa model tertinggi diperoleh ketika data set mengandung hampir seluruh fiturnya yaitu 7 fitur dengan akurasi sebesar 76,55%. Sedangkan untuk data set dengan 19 fitur, akurasi dan performa model tertinggi diperoleh ketika data set telah melewati proses seleksi fitur dengan menggunakan metode Chi-Square yaitu pada model dengan 10 fitur dengan akurasi sebesar 78,96%.

.....Diabetes mellitus or commonly referred as diabetes is a metabolic disorder caused by high blood sugar level and the pancreas that does not produce insulin effectively. Diabetes can lead to more relentless disease such as blindness, kidney failure, and heart attacks. Therefore, early detection is needed in order for the patients to prevent the disease for being more severe. According to the non-normality and huge size of data in medical field, some researchers use classification methods to predict symptoms or diagnose patients. In this study, Learning Vector Quantization (LVQ) is used to classify the diabetes data set with Chi-Square Feature Selection. This study adopted two kinds of diabetes data set which are, data set I that contains 8 features and data set II that contains 19 features. The result of the experience shows that for data set I, the highest accuracy and model performance is achieved when the model contains most of its features which is the model that contains 7 features with 76,55% of accuracy. Moreover, for data set II, the highest accuracy and model performance is achieved when the model contains features that has been selected with the Chi-Square feature selection which is the model with 10 features and the accuracy achieved is 78,96%.