

INTISARI

PENERAPAN DISKRITISASI CHIMERGE DAN SELEKSI FITUR INFORMATION GAIN PADA PENGKLASIFIKASI NAÏVE BAYES

Oleh

Muhamad Emir Sultana Nur Akbar

20/462317/PA/20289

Analisis klasifikasi merupakan salah satu metode *machine learning* yang termasuk pada kategori *supervised learning* dan berguna untuk memprediksi kelas variabel dependen bertipe kategorik berdasarkan variabel independen (atribut atau fitur) yang dipunyai. Salah satu metode analisis klasifikasi yang populer adalah Naïve Bayes. Naïve Bayes melakukan prediksi kelas variabel dependen berdasarkan pada probabilitas dan memiliki asumsi bahwa antar atribut, untuk mengklasifikasikan kelas, adalah independen satu dengan lainnya. Pada sejumlah penelitian, penerapan diskritisasi maupun seleksi fitur mampu meningkatkan performa hasil klasifikasi Naïve Bayes. Diskritisasi merupakan proses untuk mengubah data bertipe numerik menjadi bertipe kategorik dan seleksi fitur adalah proses untuk mendapatkan fitur yang relevan terhadap variabel dependen. Pada penelitian ini dilakukan proses diskritisasi ChiMerge dan seleksi fitur *information gain* pada pengklasifikasi Naïve Bayes. Diskritisasi ChiMerge merupakan proses diskritisasi berdasarkan nilai χ^2 dan seleksi fitur *information gain* merupakan seleksi fitur metode *filter* yang didasarkan pada perhitungan *information gain* antara fitur dengan variabel dependen. Penerapan kedua metode tersebut pada pengklasifikasi Naïve Bayes akan dibandingkan hasilnya dengan pengklasifikasi tanpa proses tambahan. Penerapan diskritisasi ChiMerge dan seleksi fitur *information gain* mampu meningkatkan akurasi pengklasifikasi Naïve Bayes dalam mendiagnosis penyakit gagal ginjal kronis, memprediksi pasien yang bertahan hidup dengan gagal jantung, dan mendiagnosis penyakit liver.

Kata kunci: analisis klasifikasi, diskritisasi ChiMerge, seleksi fitur *information gain*, Naïve Bayes, diagnosis

ABSTRACT

APPLICATION OF CHIMERGE DISCRETIZATION AND INFORMATION GAIN FEATURE SELECTION TO THE NAÏVE BAYES CLASSIFIER

By

Muhamad Emir Sultana Nur Akbar

20/462317/PA/20289

Classification analysis is one of the machine learning methods that belongs to the supervised learning category and is useful for predicting the class of a dependent variable of categorical type based on the independent variables (attributes or features) it has. One of the popular classification analysis methods is Naïve Bayes. Naïve Bayes performs classification prediction based on probability and assumes that the attributes, to classify the class, are independent of each other. In a number of studies, the application of discretization and feature selection can improve the performance of Naïve Bayes classification results. Discretization is the process of converting numeric data into categorical data and feature selection is the process of obtaining features that are relevant to the dependent variable. In this research, the ChiMerge discretization process and information gain feature selection are carried out on the Naïve Bayes classifier. ChiMerge discretization is a discretization process based on the χ^2 value and information gain feature selection is a filter method feature selection based on the calculation of information gain between features and dependent variables. The application of these two methods to the Naïve Bayes classifier will be compared with the results of the classifier without additional processes. The application of ChiMerge discretization and information gain feature selection can improve the accuracy of the Naïve Bayes classifier in diagnosing chronic kidney failure, predicting the survival of patients with heart failure, and diagnosing liver disease.

Keywords: classification analysis, ChiMerge discretization, information gain feature selection, Naïve Bayes, diagnosis