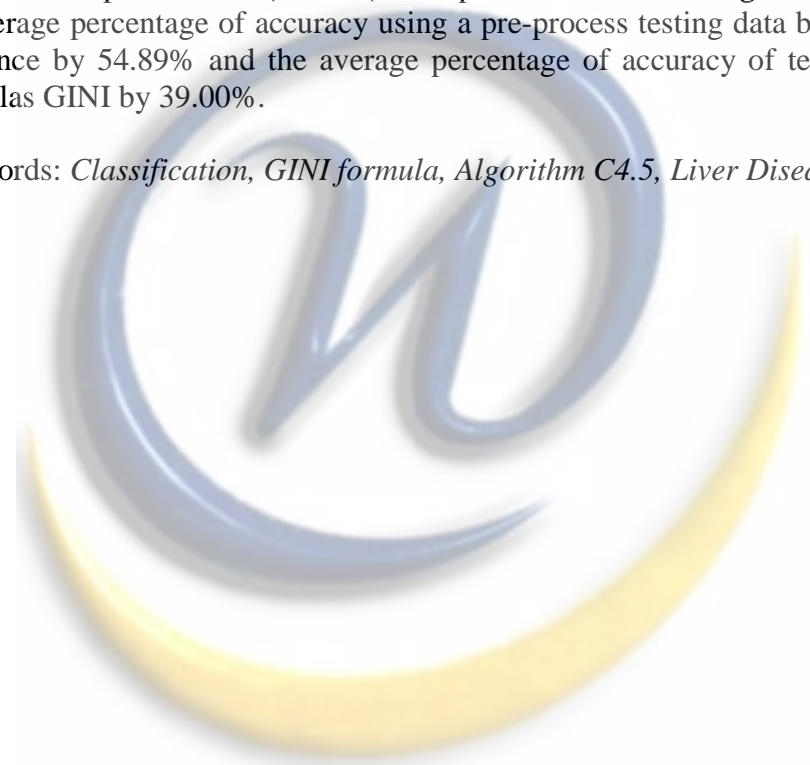


ABSTRACT

The process of detection of conventional liver disease requires a long process and takes a long time. Data mining techniques to classify liver patient data resulted in the decision whether patients include those with liver or not. The data set used is taken from WEKA. The data were analyzed using the algorithm C4.5. This algorithm has input training data in the form of data and test data are selected by determining the percentage of the total data. Pre-process data can be selected by two ways, namely by reference to health or use formula GINI. Algorithms and case studies implemented in software. Software development method is a method of structured approach to using tools Data Flow Diagrams (DFD), data dictionary, and Process Specification (PSPEC). The performance of the algorithm C4.5 with an average percentage of accuracy using a pre-process testing data based medical reference by 54.89% and the average percentage of accuracy of test data using formulas GINI by 39.00%.

Keywords: *Classification, GINI formula, Algorithm C4.5, Liver Disease*



ABSTRAKSI

Proses pendeteksian penyakit liver secara konvensional memerlukan proses yang panjang dan membutuhkan waktu yang cukup lama. Teknik *data mining* dapat mengklasifikasikan data pasien liver yang menghasilkan keputusan apakah penderita termasuk penderita liver atau bukan. Data set yang digunakan diambil dari WEKA. Data tersebut dianalisis menggunakan algoritma C4.5. Algoritma ini memiliki *input* data berupa data latih dan data uji yang dipilih dengan menentukan persentase dari total data. Pra proses data dapat dipilih dengan melakukan dua cara, yaitu berdasarkan referensi kesehatan atau menggunakan rumus GINI. Algoritma dan studi kasus diimplementasikan dalam bentuk perangkat lunak. Metode pengembangan perangkat lunak adalah metode pendekatan terstruktur dengan menggunakan *tools Data Flow Diagram (DFD)*, kamus data, dan *Process Specification (PSPEC)*. Adapun kinerja dari algoritma C4.5 dengan rata-rata persentase akurasi pengujian data menggunakan pra proses berdasarkan referensi kesehatan sebesar 54,89% dan rata-rata persentase akurasi pengujian data menggunakan rumus GINI sebesar 39,00%.

Kata kunci : *Klasifikasi, Rumus GINI, Algoritma C4.5, Penyakit Liver*