

INTI SARI

Perbandingan Performansi Metode C5.0 dan Metode CHAID dalam Mengklasifikasi Data Pendapatan Penduduk

by

Rizty Anggraini Anugrah

12/331565/PA/14767

Data mining merupakan sebuah alat yang sangat bermanfaat untuk memperoleh informasi-informasi dari data yang sangat besar. Salah satu bagian penting pada *data mining* adalah pengklasifikasian data. Klasifikasi digunakan untuk menggolongkan data berdasarkan pada sifat data yang sudah dikenali masing-masing kelasnya. Ada berbagai macam teknik yang digunakan untuk mengklasifikasi data, seperti algoritma C5.0 dan algoritma CHAID.

Pada skripsi ini, difokuskan pada metode klasifikasi C5.0 dan CHAID pada data pendapatan penduduk. Pendapatan penduduk per tahun ini dibagi kedalam dua jenis, yakni ≤ 50.000 dollar dan > 50.000 dollar. Dengan menggunakan kedua metode tersebut, dapat diperoleh perbandingan akurasi sehingga dapat disimpulkan metode yang paling cocok untuk mengklasifikasi data pendapatan penduduk serta dapat dilakukan prediksi berdasarkan variabel prediktornya.

Kata-kata kunci : *data mining*, klasifikasi, metode C5.0, metode CHAID, data pendapatan penduduk

ABSTRACT

Comparing the Performance of C5.0 Method and CHAID Method to Classify Income Dataset

by

Rizty Anggraini Anugrah

12/331565/PA/14767

Data mining is the useful tool to discovering knowledge from large data. One of the most important thing is classification of the data. Classification is used to classify the data based on their characteristics which is recognized each classes. There are some techniques to classify the data, C5.0 algorithm and CHAID algorithm.

In this minithesis, concerned to classify income dataset using C5.0 and CHAID. This income per year divided into $\leq \$50,000$ and $> \$50,000$. Using those methods, we can compare the accuracy so we can conclude the best method to classify the income dataset and predict it consider to the predictor variables.

Keywords : data mining, classification, C5.0 method, CHAID method, income dataset