

Abstrak

Data mining merupakan proses semi-otomatis untuk mengeksplorasi dan menganalisis data yang berjumlah besar gunanya untuk mendapatkan pola yang berguna. Data mining ini merupakan proses gabungan antar bidang-bidang terutama adalah *machine learning*, analisis statistik dan basis data. Data mining berusaha untuk menemukan kaidah dan pola dari data.

Salah satu *task* yang penting dalam data mining adalah *classification* (klasifikasi). Klasifikasi ini dapat dideskripsikan sebagai berikut: terdiri dari data input yang disebut juga sebagai *training set* terdiri dari sejumlah *examples (record)* yang masing-masing memiliki sejumlah atribut atau disebut juga fitur. Adapun tujuan klasifikasi ini adalah untuk menganalisis data input dan mengembangkan sebuah model yang akurat untuk setiap kelas berdasarkan beberapa variabel prediktor.

Algoritma *Soft Decision Tree - ID3* (SDT - ID3) adalah salah satu metode pengklasifikasian data. Algoritma SDT - ID3 mencari solusi dengan cara membentuk *decision tree* berdasarkan teori *fuzzy set* dimana memiliki rangkaian kemungkinan (*possibility*) untuk beberapa atau semua kelas pada data yang akan diuji.

Pada tugas akhir ini menganalisis performansi algoritma SDT - ID3 untuk klasifikasi data *breast cancer*, *wine* dan *glass*, analisis pengaruh jumlah variabel linguistik dan fungsi keanggotaan terhadap nilai performansi, dan analisis kelebihan algoritma SDT - ID3 pada kasus klasifikasi keempat *dataset (breast cancer, wine dan glass)*.

Dari hasil pengujian diperoleh kesimpulan bahwa algoritma SDT - ID3 lebih akurat dibandingkan dengan algoritma J48 dikarenakan algoritma *Soft Decision Tree - ID3* menggunakan *soft discretization* berupa *fuzzy set* dimana dapat menangani masalah ketidakpastian yaitu keraguan, ketidaktepatan, dan ketidaklengkapan informasi.

Kata kunci : *decision tree, fuzzy set, soft discretization*, akurasi