

Abstrak

Decision Tree yang merupakan salah satu cara dalam klasifikasi ini berguna untuk mendapatkan seperangkat *rule* yang tepat dari jumlah instance yang besar. Namun, ia memiliki kesulitan dalam memperoleh hubungan antara poin data yang memiliki nilai kontinu. Banyak dari algoritma *Decision Tree* dikembangkan untuk dapat menangani atribut kontinu dengan cara menerapkan konsep *pre-discretization*. Atribut discretization menjadi salah satu komponen penting dari tahap persiapan data untuk induksi *Decision Tree*. Atribut discretization, membuat partisi domain atribut menjadi set interval, melibatkan dua langkah utama: (1) penentuan jumlah interval terhadap atribut yang harus didiskretisasi; dan (2) penentuan batas setiap interval.

Beberapa algoritma *Decision Tree* yang menerapkan konsep *pre-discretization* ini adalah C4.5 dan NBTree yang merupakan algoritma *state-of-art* karena hasil akurasi klasifikasinya sudah sangat baik bahkan dalam ukuran data yang besar. Namun jika ditelusuri lebih lanjut dari sudut pandang teori informasi, hilangnya informasi akibat penerapan *pre-discretization* dapat menurunkan akurasi klasifikasi.

Dalam tugas akhir ini dibuat suatu perangkat lunak yang mengimplementasikan metode klasifikasi yaitu Self-adaptif NBTree, yang menginduksi secara *hybrid Decision Tree* dan Naive Bayes. *Bayes measure*, yang digunakan untuk membangun *Decision Tree* secara *post-discretization*, dapat langsung menangani atribut kontinu dan secara otomatis menemukan jumlah interval setabatas-batas yang paling tepat untuk diskretisasi. Simpul Naive Bayes yang berada pada *leaf* membantu untuk memecahkan masalah *overgeneralization* dan *overspecialization* yang sering terlihat pada *Decision Tree*.

Performansi algoritma Self-adaptive NBTree akan dibandingkan dengan NBTree, dan dua algoritma pendirinya yaitu C4.5(*Decision Tree*) serta Naive Bayes. Hasil pengujian menunjukkan Self-adaptive NBTree memiliki performansi akurasi yang lebih baik dan ukuran pohon lebih kecil dari NBTree, C4.5, dan Naive Bayes ketika menangani banyak nilai kontinu. Sedangkan NBTree dan C4.5 bekerja baik pada dataset dengan karakteristik tertentu. Naive Bayes memiliki waktu pembangunan model dan klasifikasi tercepat, namun akurasinya kurang baik pada dataset ukuran besar.

Kata kunci : *Decision tree, Naive Bayes, Discretization, Nbtree, Self-adaptive NBTree.*