

Abstrak

Imbalance class adalah salah satu permasalahan yang ada dalam data mining. Permasalahan ini diakibatkan karena adanya jumlah data yang tidak seimbang pada kelas yang berbeda, dimana kelas yang satu memiliki jumlah data yang sangat banyak (mayoritas), sedangkan kelas lainnya memiliki jumlah data yang sangat sedikit (minoritas). Pada pengklasifikasian biasa, kelas minoritas tersebut tidak dapat terprediksi. Hal ini disebabkan karena bila jumlah data pada satu kelas sangat kecil, kelas tersebut akan salah diprediksi sebagai kelas mayoritas.

Salah satu metode klasifikasi yang biasa digunakan adalah metode *decision tree*. Metode *decision tree* ini akan membentuk model seperti *flow chart* dengan struktur *tree*. Pada Tugas Akhir ini diperbandingkan tiga algoritma pada *decision tree* yaitu C4.5, CART, dan HDDT. Setiap algoritma tersebut memiliki *splitting criterion* atau *attribute selection measure* yang berbeda dengan tingkat *sensitivitas* yang berbeda pula terhadap kecondongan data.

Analisis yang dilakukan pada Tugas Akhir ini adalah mengetahui bagaimana performansi dari algoritma C4.5, CART, dan HDDT yang memiliki *splitting criterion* yang berbeda pada permasalahan *imbalance class* serta bagaimana perbandingan performansi dari ketiga algoritma *decision tree* tersebut berdasarkan nilai *precision*, *recall*, dan *f-measure*.

Hasil yang didapat dari penelitian menunjukkan bahwa algoritma HDDT memiliki performansi yang lebih baik pada tingkat *imbalance* yang rendah dibandingkan algoritma CART dan C4.5. Hal ini disebabkan karena algoritma HDDT memiliki tingkat sensitivitas yang rendah terhadap kecondongan data yang menjadi masalah dalam *imbalance class*.

Kata kunci: *imbalance class*, algoritma C4.5, CART, HDDT, performansi.