

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Dalam penelitian ini dilakukan empat skenario pengujian. Berdasarkan hasil pengujian dapat disimpulkan bahwa:

1. Pada skenario pertama, hasil pengujian menggunakan metode *decision tree* ID3 memiliki nilai *accuracy* yang buruk pada semua *dataset* dikarenakan tidak mampu mengklasifikasikan *dataset* bertipe data *continue*. Sedangkan dengan menggunakan *Decision tree* C4.5 memiliki nilai *accuracy* yang pada *dataset glass* dan *lympografi* berturut-turut sebesar 68.22% dan 75.68%. Sedangkan *decision tree* C5.0 memiliki nilai *accuracy* yang baik pada *dataset vehicle*, *wine* dan *thyroid* berturut-turut sebesar 71.54%, 94.38% dan 94.42%.
2. Pada skenario kedua, hasil pengujian menggunakan *dataset* hasil seleksi PSO mampu meningkatkan performa *decision Tree* C4.5 pada *dataset vehicle* dan *lympografi* berturut-turut sebesar 72.34% dan 79.71%. Sedangkan pada *decision tree* C5.0 memiliki nilai *accuracy* yang lebih baik pada *dataset glass*, *thyroid* dan *wine* berturut-turut sebesar 70.09%, 94.38% dan 94.42%.
3. Pada skenario ketiga, hasil pengujian klasifikasi *dataset* menggunakan *cost sensitive decision tree* C4.5 memiliki nilai *accuracy* yang lebih baik pada *dataset vehicle* dan *wine* berturut-turut sebesar 76.86% dan 97.62%. Sedangkan klasifikasi *dataset* menggunakan *cost sensitive decision tree* C5.0 memiliki nilai *accuracy* yang lebih baik pada *dataset glass* dan *thyroid*, berturut-turut sebesar 75.27% dan 95.81%.
4. Pada skenario empat, Pengujian menggunakan metode *cost sensitive decision tree* C5.0 memiliki nilai *accuracy* yang lebih baik dari pada menggunakan metode *cost sensitive naïve bayes* pada *dataset glass*, *lympografi*, *vehicle* dan *wine* berturut-turut 76.17%, 83.33%, 75.27% dan 95.83%. Sedangkan dengan menggunakan metode *cost sensitive naïve bayes* memiliki nilai *accuracy* yang lebih baik pada *dataset thyroid* sebesar 97.67%.

Berdasarkan penelitian yang telah dilakukan, menunjukkan bahwa performa *accuracy* dipengaruhi oleh beberapa faktor antara lain jenis *dataset*, jumlah dan kualitas atribut, jumlah ratio distribusi data, jumlah label class, proses preprocessing dan algoritma.

6.2 Saran

Berdasarkan penelitian yang telah dilakukan masih memiliki beberapa kekurangan pada hasil klasifikasi dikarenakan oleh beberapa faktor antara lain jumlah label *class* dan jenis data yang digunakan. Oleh karena itu berikut saran yang bisa digunakan sebagai pengembangan penelitian selanjutnya.

1. Klasifikasi data *imbalanced multiclass* menggunakan metode *cost sensitive decision tree* dapat dikembangkan lebih lanjut dengan menggunakan algoritma *cost sensitive learning* lainnya seperti teknik *thresholding* (*cost sensitive classifier* dan *empirical thresholding*) dan teknik *sampling* (*coasting* dan *weighting*) untuk meminimalkan *cost* supaya mendapatkan performa *accuracy* yang optimal.
2. *Dataset* yang bersifat *continue* agar dapat diklasifikasikan menggunakan algoritma ID3 terlebih dahulu dilakukan proses konversi nilai atribut data menggunakan teknik logika fuzzy yaitu merubah data *continue* menjadi data *categorical*. Metode fuzzy yang dapat digunakan untuk mengelompokkan nilai berdasarkan derajat keanggotaan menggunakan metode mamdanni.