

Abstrak

Cost Sensitive Learning merupakan pendekatan pada proses pembelajaran klasifikasi dimana dalam proses klasifikasi diperhitungkan juga *misclassification cost*. *Misclassification cost* merupakan kerugian (*cost*) yang timbul apabila suatu data salah diklasifikasikan. *Misclassification cost* merupakan hal yang sangat penting karena data yang salah diklasifikasikan dapat menimbulkan *cost* yang berbeda-beda. Untuk penerapan *cost sensitive learning* dalam klasifikasi *data imbalance* digunakan metode pembelajaran *AdaCost*, yaitu metode pembelajaran yang melibatkan perhitungan *misclassification cost*. Sedangkan untuk klasifikasi data digunakan metode klasifikasi *Naive Bayes*, yaitu metode klasifikasi yang menggunakan perhitungan probabilitas dalam memecahkan persoalan klasifikasi.

Pada Tugas Akhir ini telah dianalisis dan diimplementasikan klasifikasi data *imbalance* menggunakan *AdaCost* dan *Naive Bayes*. Pengujian dilakukan dengan beberapa skenario, antara lain pengujian dengan perbedaan jumlah ronde *boosting*, pengujian dengan membandingkan performansi *AdaCost* dengan metode klasifikasi dan metode *boosting* lain, dan pengujian menggunakan data yang terdapat *missing value*. Metode klasifikasi yang digunakan sebagai pembandingan dalam pengujian adalah *Decision Tree*, *OneR*, *Neural Network*, dan *Naive Bayes*. Sedangkan metode *boosting* yang digunakan sebagai pembandingan adalah *AdaBoost*, *AdaBoost M1*, *MultiBoost AB*, dan *LogitBoost*.

Dari hasil pengujian didapatkan kesimpulan bahwa algoritma *AdaCost* lebih memprioritaskan pencarian pola kelas *imbalance* sehingga dapat meningkatkan performansi *base classifiernya* dalam mengklasifikasikan kelas *imbalance* dan meminimalkan *misclassification cost*. *AdaCost* juga lebih baik dalam mengklasifikasikan kelas *imbalance* bila dibandingkan dengan metode *AdaBoost*, *AdaBoost M1*, *MultiBoost AB* dan *LogitBoost*. Namun, performansi dari *AdaCost* tergantung dari jumlah ronde *boostingnya*.

Keyword : *cost sensitive learning, misclassification cost, data imbalance, adacost, naive bayes*