

Abstrak

Churn prediction merupakan salah satu aplikasi data mining yang bertujuan untuk memprediksi para pelanggan yang berpotensi untuk *churn*. *Churn prediction* merupakan salah satu permasalahan kelas *imbalance* dan golongan *churn* merupakan kelas minor yang mengakibatkan kerugian bila tidak terdeteksi. Metode klasifikasi pada data mining secara umum tidak memperhitungkan kerugian akibat kesalahan prediksi suatu kelas ke dalam kelas lain, hal ini kurang sesuai dengan permasalahan *churn* yang mempertimbangkan faktor *benefit* sebagai pertimbangan utama dalam penanganan pelanggan.

Pada tugas akhir ini diteliti dua algoritma *cost-sensitive learning* berbasis *sampling* yaitu *Costing* dan *CSRoulette* serta 2 modifikasi algoritma tersebut yaitu *UnderCosting* dan *UnderCSRoulette*. Evaluasi dilakukan melalui penghitungan akurasi model *churn prediction* yang dinyatakan dalam bentuk *total benefit*, *top decile lift*, *top decile benefit*, *lift curve*, dan *gini coefficient*. Algoritma non-*cost-sensitive learning* yang digunakan sebagai pembandingan adalah *Boosting* dengan memakai data training yang diproses dengan teknik *balancing*.

Hasil yang didapat dari penelitian menunjukkan bahwa metode *cost-sensitive learning* berbasis *sampling* tidak selalu lebih bagus dari algoritma non-*cost-sensitive learning* pada semua parameter evaluasi. Algoritma yang berdasarkan *undersampling* seperti *UnderCosting*, *UnderCSRoulette*, dan *Boosting-UnderSampling* cenderung bagus pada nilai *cost* yang rendah dan buruk pada nilai *cost* yang tinggi karena menghasilkan *precision* yang rendah.

Kata kunci: *cost-sensitive learning*, *sampling*, *boosting*, *benefit*, *cost*.