

Abstrak

Churn adalah suatu fenomena dimana para pelanggan berpindah akibat tidak puas atas layanan yang diberikan, ataupun memutuskan untuk pindah dari *provider* satu ke *provider* yang lainnya karena tergiur penawaran fasilitas dan harga yang lebih menarik. Oleh karena itu, diperlukan prediksi awal untuk pengklasifikasian pelanggan yang diprediksi *churn* atau bukan, serta menganalisis atribut utama penentu prediksi awal *churn*. Namun, karakteristik dari data *churn* adalah data yang *imbalanced class*. Sehingga pada penelitian ini diimplementasikan metode *balanced random forest* untuk mengatasi data *imbalance class* serta dilakukan pengujian terhadap perbedaan jumlah pohon dan jumlah atribut pemecah. Dari hasil pengujian didapatkan bahwa penambahan jumlah pohon dan atribut pemecah meningkatkan nilai akurasi dan mengurangi nilai *OOB error rate* yang dihasilkan dengan nilai optimal jumlah pohon sebanyak 500 dan jumlah atributnya yaitu 9 dengan nilai rata-rata akurasi 79,05%. Serta atribut penentu utama prediksi awal yaitu jenis layanan yang dipakai pelanggan, disusul oleh kombinasi kategori penggunaan *voice* per hari selama seminggu, dan kombinasi dari variabel kategori panggilan.

Kata Kunci : *churn*, klasifikasi, *balanced random forest*, *imbalance class*.

Abstract

Churn is a phenomenon in which the switch due to unsatisfied customers for services rendered, or decide to switch from one provider to another provider because tempted facilities and offers a more attractive price. Therefore, it is necessary for the initial prediction pengkalsifikasian predictable customer churn or not, and to analyze the main attribute determinants of early churn prediction. Characteristics of the test dataset is imbalance class. So in this study implemented a balanced random forest method that solve the data imbalance class. In this research, testing of the difference in the number of trees and the number of attributes solver. From the test results showed that the addition of trees and increase the value attribute solver accuracy and reduce error rate oob value generated by the optimal value of the number of trees as much as 500 and the number of attributes that is 9 with an average value of 79.05% accuracy. As well as the attributes of the main determinant of early predictions that the type of service used by the customer, followed by a combination of voice usage category per day during the week, and the combination of variable call categories.

Keyword : *churn, classification, balanced random forest, imbalance class*