

Abstrak

Churn prediction merupakan upaya yang digunakan untuk membantu penyedia layanan telekomunikasi agar tidak mengalami kerugian yang besar dengan cara memprediksi atau mengklasifikasi pelanggan jasa telekomunikasi yang berhenti dan berpindah menggunakan jasa telekomunikasi yang lain. Dataset pada kasus *churn* ini biasanya memiliki data pelanggan *non-churn* yang mayor daripada data pelanggan *churn* yang minor sehingga akurasi untuk data minor akan memiliki akurasi yang sangat kecil sehingga menyebabkan *imbalance data*. Data yang *imbalance* akan menyebabkan kesulitan dalam mengembangkan model prediksi yang baik. Salah satu pendekatan untuk menangani *imbalance data* adalah pendekatan *sampling-based*. Pada penelitian Tugas Akhir ini menggunakan data pelanggan dari WITEL PT. Telekomunikasi Regional 7. Pada data yang digunakan memiliki 53 atribut dengan 200,361 *record* data (192,848 *record* data *non-churn* dan 7,513 *record* data *churn*). Tingkat data kelas *churn* pada dataset ini adalah 3.7%. Untuk menangani *imbalance data* menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) yang dikombinasikan dengan metode *Random Forest* (RF) untuk klasifikasi data kelas *churn* atau *non-churn*. SMOTE merupakan teknik *oversampling* yang digunakan untuk menghasilkan data sintetis dari kelas *churn* untuk meningkatkan prediksi peluang *churn* yang terjadi. Penelitian yang dilakukan terbukti dapat meningkatkan performansi model klasifikasi RF dengan penanganan *imbalance data* menggunakan SMOTE yang memberikan peningkatan nilai *f-measure* lebih dari 20% dengan nilai *f-measure* terbaik yaitu 95,27% dan *error rate* 4,60%.

Kata kunci: *imbalanced data, churn prediction, smote, random forest.*