

## ABSTRAK

Pelanggan merupakan aset utama dalam sebuah perusahaan, tidak terkecuali industri telekomunikasi. Pelanggan churn merupakan masalah utama yang ditemukan dalam industri telekomunikasi, karena sangat berpengaruh terhadap pendapatan perusahaan. Pada saat proses churn mulai terjadi, biasanya persentase data yang menggambarkan pelanggan churn tidak banyak. Kondisi ini menyebabkan diperlukannya model prediksi untuk dapat menentukan pelanggan yang berpotensi melakukan churn. Pendekatan Data mining dapat menghasilkan model prediksi dengan mempelajari data historis transaksi pelanggan. Minimnya data pelanggan churn di antara sejumlah data pelanggan yang dimiliki oleh perusahaan akan menimbulkan masalah data yang tidak seimbang. Data yang tidak seimbang akan menyebabkan kesulitan dalam pembuatan model prediksi sehingga hasil prediksi pelanggan churn menjadi tidak akurat. Sampel data yang digunakan dalam penelitian ini memiliki persentase churn 0,7 %. Penelitian ini menerapkan teknik kombinasi sampling dan *Weighted Random Forest* (WRF) untuk menghasilkan model prediksi pelanggan churn. WRF telah dikembangkan dari metode *Random Forest* (RF) untuk mengatasi masalah data yang tidak seimbang yang biasa terjadi pada prediksi churn. Metode ini diklaim dapat menghasilkan kinerja yang cukup baik pada data yang tidak seimbang. Namun, pada penelitian ini ditemukan kendala bahwa performansi yang dihasilkan masih rendah. Dengan menggunakan teknik *sampling*, permasalahan performansi yang masih rendah dapat diatasi. Adapun teknik *sampling* yang digunakan adalah *simple under sampling* dan *Synthetic Minority Over-sampling Technique* (SMOTE). Hasil penelitian ini menunjukkan bahwa kombinasi SMOTE dan *simple under sampling* terbukti dapat meningkatkan kinerja model prediksi yang dihasilkan oleh WRF.

Kata kunci: Churn, Prediksi, Metode Random Forest, Metode Weighted Random Forest, kombinasi sampling, SMOTE