

ABSTRAK

Perkembangan industri telekomunikasi saat ini ,menghasilkan persaingan yang sangat ketat antara penyedia layanan (service provider). Persaingan ini mengakibatkan Customer Churn menjadi salah satu permasalahan yang sering dihadapi, karena dapat mempengaruhi pendapatan perusahaan, profitabilitas, kelangsungan hidup serta kualitas pelayanan perusahaan . Oleh karena itu, mengetahui pelanggan yang akan melakukan churn dike-mudian hari (Customer churn prediction) secara dini menjadi salah satu cara yang sering dilakukan, karena dapat membantu perusahaan dalam membuat rencana yang efektif un-tuk tetap mempertahankan pelanggannya.

Teknik klasifikasi data mining dapat digunakan untuk customer churn prediksi. Ran-dom forest merupakan salah satu teknik klasifikasi yang dikenal memiliki kinerja sangat baik jika dibandingkan dengan teknik klasifikasi lainnya, karena sangat mudah digunakan dan memberikan performansi akurasi yang lebih tinggi. Namun algoritma klasifikasi tidak dapat berjalan dengan baik jika dihadapkan pada data yang tidak seimbang karena dapat mempengaruhi kinerja teknik klasifikasi serta performansi yang dihasilkan.

Sedangkan data Customer Churn salah satu data yang memiliki karakteristik data yang tidak seimbang, dimana memiliki salah satu class data yang sedikit dari pada class data lainnya.

Tujuan dari penelitian ini adalah menangani data imbalance pada Customer Churn Prediction untuk meningkatkan efektivitas teknik klasifikasi dalam menghasilkan perfor-mansi prediksi yang lebih baik. Oleh karena itu dalam penelitian ini dilakukan klasi-fikasi pada data Customer churn PT Telekomunikasi Indonesia dengan mengusulkan se-buah metode yang dinamakan Modified Balance Random Forest (MBRF). Proses MBRF melakukan perubahan proses Balance Random Forest dengan menerapkan strategi under-sampling berdasarkan clustering untuk setiap bootstrap data yang akan dijadikan dalam pembentukan setiap pohon keputusan pada random forest, oleh sebab itu pendekatan MBRF disebut juga pendekatan handling imbalance data berdasarkan algoritma. Metode yang diusulkan pada penelitian ini (MBRF) memberikan hasil performansi yang lebih baik jika dibandingkan dengan metode Balance Random Forest (BRF) dan Random Forest (RF). MBRF memberikan nilai akurasi AUC terbaik (91.65%), Sensitivity atau True Positive Rate (TPR) terbaik (88%), Specificity atau True Negative Rate (TNR) terbaik (94%), dan G-Means terabaik (91%). Selain memberikan performansi yang lebih baik, MBRF juga memperbaiki jumlah waktu running time dengan menghasilkan time con-sumption proses yang lebih rendah.

Kata Kunci: Customer Churn, Imbalanced Data , Teknik Klasifikasi, Radom Forest, Undersampling , Clustering