

ABSTRAKSI

Masalah ketidakseimbangan data adalah salah satu tugas yang menantang dalam pembelajaran mesin. Ketika belajar dari data yang sangat imbalance, kebanyakan klasifikasi algoritma condong terhadap kelas mayoritas. Dalam aplikasi dunia nyata, churn prediction dianggap sebagai salah satu aplikasi data mining yang mencerminkan masalah ketidakseimbangan data. Dalam penelitian ini, kami menyelidiki bagaimana kita bisa lebih baik menangani ketidakseimbangan kelas di prediksi churn. Menggunakan RUSBoost, kombinasi undersampling acak, dikombinasikan dengan pilihan fitur untuk hasil kinerja yang lebih baik. Dataset yang digunakan adalah data internet broadband yang didapat dari PT.Telekomunikasi Indonesia. Langkah pertama yang dilakukan dalam studi ini adalah pemilihan fitur dengan menggunakan Information Gain, dan kemudian membangun model prediksi churn menggunakan RUSBoost dengan C4.5 sebagai metode klasifikasinya. Hasil menunjukkan bahwa kombinasi dari fitur seleksi dan RUSBoost memiliki dampak dalam meningkatkan nilai prediksi sebesar 16% dan mengurangi waktu eksekusi hingga 48%.

Kata kunci: Information Gain, RUSBoost, Masalah ketidakseimbangan, Prediksi Churn