Abstrak

Churn prediction pada pelanggan telekomunikasi merupakan upaya memprediksi/mengklasifikasi pelanggan jasa telekomunikasi yang berhenti atau berpindah berlangganan dari suatu operator ke operator yang lain. Namun *dataset* pada kasus *churn* ini biasanya memiliki kelas yang *imbalance* dimana jumlah *instance* suatu kelas (kelas *active* atau tidak *churn* atau mayor atau negatif) jauh lebih besar dari jumlah kelas yang lain (kelas *churn* atau minor atau positif). Akibatnya, kebanyakan *classifier* cenderung memprediksi kelas mayor dan mengabaikan kelas minor sehingga akurasi kelas minor sangat kecil.

Salah satu pendekatan yang dilakukan untuk menangani permasalahan ini adalah dengan memodifikasi distribusi *instances* dari *dataset* yang digunakan atau yang lebih dikenal dengan pendekatan *sampling-based.* Teknik *resampling* ini meliputi *over-sampling*, *under-sampling*, dan *combine-sampling*.

Dengan metode *Ensemble of SVM* (EnSVM) diharapkan dapat meminimalisir kesalahan klasifikasi kelas mayor dan minor yang dihasilkan oleh *classifier* SVM tunggal. Metode ini menggunakan teknik *Combined Sampling* untuk menurunkan kekuatan dari kedua teknik *resampling* (SMOTE dan *under-samping*) sehingga meningkatkan akurasi kelas minor dan mayor.

Hasil penelitian menunjukkan bahwa metode *CombinedSampling*-EnSVM secara umum hanya mampu menghasilkan performansi *Gini Index* yang lebih baik daripada metode SMOTE-SVM dan tanpa *sampling* (*pure*-SVM) pada kasus Data Telekomunikasi.

Kata kunci: *churn prediction, imbalance, combined sampling, ensemble*, SVM, performansi.