Abstract

Churn prediction is an effort that is used to help telecommunications service providers in order not suffered heavy losses by way of predicting or classifying telecommunication customers who stop and switch to use other telecommunications. Dataset in this churn case typically has the majority of non-churn customer data rather than the minority of churn customer data, so that accuracy for minority data will have very little accuracy to cause imbalance data. Data that imbalance will cause difficulty in developing a good predictive model. One approach to addressing the imbalance data is sampling-based approaches. In this Final Project research using customer data from the WITEL PT. Telecommunications Regional 7. In the data used there are 53 attributes with 200,361 data records (192,848 data records for non-churn and 7,513 data records for churn). The amount of churn data is 3.7% of the overall data for handling imbalance data using Synthetic Minority for Oversampling Technique (SMOTE) combined with a method of Random Forest (RF) for the classification of data classes non-churn or churn. SMOTE for oversampling techniques is used to generate synthetic data from churn classes to increase chances of predicting churn. Study conducted proved to be able to increase the performance of model classification of RF with handling imbalance data using SMOTE that provides increased value of f-measure more than 20% with a value of the f-measure is the best 95.27% and 4.60% of error rate.

Keywords: *imbalanced data, churn prediction, smote, random forest.*