

ABSTRACT

Loyal customers is valuable asset of the telecommunication industry. Therefore, the company will continue to strive to improve customer satisfaction in order to avoid losses due to churn. Of these issues required an accurate customer classification model to help companies reduce churn. Churn cases referred to as minor class, only 3.75% of overall dataset, that resulting imbalance data. Imbalance data will cause difficulty in making prediction models. One of the solution to solve the imbalance data by implementing undersampling techniques.

The data used in this final project is derived from customer's data in WITEL PT. Telekomunikasi. In this final project research conducted handling imbalanced data using Neighborhood Cleaning Rule (NCL) and Random Undersampling (RUS). After that, result from NCL and RUS used to build classification model using Weighted Random Forest (WRF) to classify class churn or not-churn. The result of this study proven can increase performance of predictive models generated by WRF with combination of NCL and RUS, that is accuracy model 93.799% with best F1 90.651% and oob error rate 7.36%.

Keyword : *imbalance data, churn, prediction, undersampling, weighted random forest, RStudio.*