

Abstract

Maintaining customers is important for telecommunication companies, because customers are a source of revenue for the company. Telecommunication companies that implement churn prediction model can detect the possibility of transferring customers from one provider to another provider. However, because customers who churn far less than non-churn, the distribution between majority and minority classes is not balanced, it is called imbalanced data. With an unbalanced distribution, approach methods to data mining tend to misclassify minority classes. Therefore, it is necessary to handle imbalanced data with sampling technique.

The method used in this Final Project research is Underbagging method to handle imbalanced data combined with classification method using Random Forest (RF). System testing performed using one of PT. WITEL Telekomunikasi Regional 7 customer datasets which has 53 attributes with a total churn data of 7,513 records and non-churn data of 192,848 records. This research produces the highest f-measure performance of 92.58% and increased the f-measure performance of 22.32% to the classification result without handling imbalanced data.

Keywords: *churn prediction, imbalanced data, Underbagging, Random Forest.*