Abstract

Technological developments are now increasing rapidly. The need for information and communication increases. Competition to earn revenue between telecommunication companies led to Churn. Churn is a move from one provider to another provider. The company prefers to keep the customer, because it costs less than adding a new customer. However, in this case churn has unbalanced and extreme data compared to non-churn, so there is a need for handling in the distribution of the majority class (non-churn) and minority (churn). In the method of approach with data mining, unbalanced data will lead to the classification process that tends to be mistaken in the minority class (churn). Therefore, it is necessary to handle unbalanced classes with sampling techniques.

The method used in this Final Project research is Underbagging method to handle imbalance data combined with classification method using Logistic Regression (LR). Testing is done by using the customer dataset WITEL PT. Telecommunication Regional 7 with 53 attributes. Total data churn 7,513 records and non-churn data 192.848 records. This research resulted the highest f1-measure value of 84.985% and increased the f1-measure more than 20% against the classification result without imbalance data handling.

Keywords: Classifications, Data Mining, Churn prediction, Logistic Regression, Imbalance Data, UnderBagging