ABSTRACT

The growth of telecommunication industry today, makes tight competition among service providers. This competition gives rise (Customer Churn) an important problems for companies, because it affects the company's revenue, profitability, survival and service quality. Therefore, conducting early customer churn prediction is important, because it can help companies make practical plans to keep their customers.

Classification techniques data mining can be used for customer churn prediction. Random forest is one of the most well-known classification techniques and it can perform very well compared to many other classification techniques because it is very easy to use and gives a higher accuracy performance. However, the classification algorithm cannot run properly if it is faced with imbalanced data because it can affect the performance of classification techniques and the resulting performance.

While the data "Customer Churn" are one of the data that has the characteristics of imbalanced data, which has one class of data slightly from the other data classes.

The purpose of this study is to handle imbalance data on "Customer Churn Prediction" to improve the effectiveness of classification techniques in producing better prediction performance. Therefore, in this study classification is done on the data churn Customer PT Telekomunikasi Indonesia by proposing a method called Modified Balance Random Forest (MBRF). The MBRF process changes the Balance Random Forest process by implementing an undersampling clustering strategy for each bootstrap of data to be created in the formation of each decision tree in the random forest. Accordingly the MBRF approach is also called the algorithmic data handling approach. The method proposed (MBRF) in this study provides better performance results when compared to the Balance Random Forest (BRF) and Random Forest (RF) method. MBRF gives the best accuracy AUC (91.65%), Best Sensitivity or True Positive Rate (88%), best Specificity or True Negative Rate (TNR) (94%), and best G-Means (91%). In addition to delivering better performance, MBRF also improves the amount of running time by producing a lower processing time consumption.

Keywords: Customer Churn, Imbalanced Data, Classification Technique, Random Forest, Undersampling, Clustering.