ABSTRACT

The most valuable assets owned by a telecommunication companies are customers. Therefore a telecommunications company will always strive to maintain their assets by providing satisfying services. However, intense competition between telecommunication companies will provide many choices of potential customers to churn (stop serving). Churn greatly affects the loss of revenue of a telecommunications company, besides that the cost to attract new customers is greater than retaining existing customers. For this reason the churn case received serious attention from the telecommunications company to prevent it. A system is needed to predict the churn of customers.

In making a churn prediction system, a problem that is often found is unbalanced data between negative (not churn) samples and positive samples (churn). However, any classification algorithm and technique cannot run well if it is faced with unbalanced data because it can affect the performance of classification techniques and the resulting performance.

The purpose of this study is to handle imbalance data on customer churn prediction to improve the effectiveness of classification techniques in producing better performance. Therefore, in this study the classification of PT Telekomunikasi Indonesia customer churn data is proposed by a method called Modified Ensemble-Undersampling Boost (Modified EUS-Boost). Modified EUS-Boost process changing the balancing process by applying an undersampling strategy based on clustering for each boostrap of data that will be used in the formation of each subset of training data on Adaboost. Therefore the Modified EUS-Boost approach is also called an imbalance handling data approach based on algorithms level approach. The method proposed in this study provides better performance results when compared to other boosting based methods, namely Real Adaboost and Random Undersampling-Boost. Modified EUS-Boost provides the best accuracy value (95%), F-Score (49%), recall (82%), and the best precision (35%). In addition to providing better performance, Modified EUS-Boost also improves the amount of running time by producing a lower process time consumption.

Keywords: Customer Churn, Imbalanced Data , Classification, Undersampling, Clustering, Adaboost