

ABSTRACT

Customer churn has become a significant problem and also a challenge for Telecommunication company such as PT. Telkom Indonesia. It is necessary to evaluate the problems that make customer churn; so that, the company's managements will be able to make appropriate strategies to minimize the churn and try to retain the existing customers. The number of churns, which are categorized as churn "Atas Permintaan Sendiri" (APS), in PT. Telkom's customer data is very small. Per month is about 1 %. The lack of this churn data leads to the problem of the imbalanced data, and it causes the difficulties in developing a good prediction model. This issue then becomes one of the challenging tasks in machine learning.

This study investigates how to handle imbalanced data in churn prediction using data preprocessing, the combined sampling between Synthetic Minority Over-Sampling (SMOTE) and Random Under-Sampling (RUS). Those proposed combined techniques tried to overcome the above problem so it can provide a better performance of churn prediction's model. SMOTE tries to multiply the minority data (churn data) and avoid the overfitting data minor. RUS tries to reduce the majority data (non churn data) and keep from losing information too much from reducing process. The combined techniques have been proven to provide a better performance than other single technique. The First technique was data preprocessing and the second was classification. The datasets used in this study are Broadband Internet data collected from Telkom Regional 6 Kalimantan.

This study also tries to build churn prediction model using one of ensemble approaches by Bagging method, combines with C4.5 as the single classifier. Because of its simplicity, the Bagging method is worthwhile improves the result of the usage of data preprocessing and training a single classifier. The results of this research show that the combination of sampling SMOTE and RUS is able to rebalance the dataset and increase the performance of churn prediction to 571% compared to that of without sampling. In addition, the use of Bagging methods gives a positives synergy to improve the performance to 56 %.

Keyword: SMOTE, Random Under Sampling, Bagging Methods, Imbalance Problems, Churn Prediction