

ABSTRACT

Solving imbalance problems is one of the challenging tasks in machine learning. When learning from highly imbalanced data, most classifiers are biased towards the majority class examples. In practice, churn prediction is considered as one of data mining application that reflects imbalance problems. This study investigates how to handle class imbalance in churn prediction using RUSBoost, a combination of random under-sampling and boosting algorithm, which is combined with feature selection for better performance result. The datasets used are broadband internet data collected from a telecommunication industry in Indonesia. The study firstly select the important features using Information Gain, and then building churn prediction model using RUSBoost with C4.5 as the weak learner. The result shows that the combination of feature selection and RUSBoost improve 16% of the performance of prediction and reduce 48% of the processing time.

Keyword: Information Gain, RUSBoost, Imbalance Problems, Churn Prediction