

ABSTRACT

Customers are the key asset in an industry, the telecommunications industry is no exception. Customer churn is a major problem that is found in the telecommunications industry, because it affects the company's revenue. At the time of the customer churn is taking place, the percentage of data that describes the customer churn is usually not much, unfortunately the churn data is the data which have to predict earlier. This condition causes the need for predictive models in order to determine the potential customers do churn. Data mining approach can produce prediction models by studying the historical data of customer transactions. The lack of data on customer churn among a number of customer data held by the company will lead to the problem of imbalanced data. Data that is not balanced will cause difficulty in making a prediction model so that the results of customer churn prediction become inaccurate. The sample data used in this study has a percentage of 0.7 % churn. This research applies a combination of sampling techniques and Weighted Random Forest (WRF) to produce the customer churn prediction model . WRF has been developed from the method Random Forest (RF) to overcome the problem of unbalanced data which is common in churn prediction. This method is claimed to produce a reasonably good performance on the imbalanced data. However, this study found that the performance of the results is still low. By using sampling technique, the low performance problems can be overcome. The sampling technique used is simple under sampling and Synthetic Minority Over-sampling Technique (SMOTE). The results of this study indicate that the combination of SMOTE and under- sampling simple proven to increase performance of predictive models generated by WRF.

Keywords: Churn, Prediction, Random Forest, Weighted Random Forest, Combined-sampling, SMOTE