Abstract

Churn prediction is one of the application in data mining which has purpose to predict potential customer to churn. Churn prediction is imbalance class problem with churn as minor class. There are several ways to solve any imbalance class problem especially churn case. For example is balancing data training or using special method to solve that.

Analysis in this final assignment is understanding whether *Bagging* and Boosting methods can solve to classify data churn. To support that, software implementing Bagging, SMOTEBoost, and Lazy Bagging methods is made. Evaluation uses two datasets. There are data telecommunication company and data tournament. The methods for that evaluation are Bagging, SMOTEBoost, Lazy Bagging, Boosting Clementine 10.1 and C 5.0 clementine 10.1. After that, calculate the accurateness of imbalance data in churn prediction model which is implemented as lift curve, top decile, gini coefficient and f measure.

Bagging and Boosting Clementine methods can predict data churn after balancing data training. SMOTEBoost method is used to predict imbalance data for data telecommunication company and data tournament.

Keywords : bagging, boosting, imbalance data, churn prediction, accurateness