Abstract

Churn prediction is one of application in data mining task which has goal to predict customer who potentially to churn. Churn prediction is one of imbalance class case. Imbalance class problem in Clementine 9.0 use node balancing (use undersampling, oversampling or merging from each technique). The problem is model too overvitting. *DataboostIM* is method which developed from original *boosting* method which combine with *data generation* concept. Boosting is one of ensemble method which usually use to improve overall accuracies in the imbalance class. The *boosting* weakness is much focus to hard example in minor class, this is can be a reason for major class accuracies tend to decrease although minor class accuracies was increase. DataboostIM not only improve accuracies in minor class, but can hold out accuracies from major class. In this final project we make software which implemented DataboostIM method with base classifier integrated from Classification model in Clementine 9.0. Implementation had done which take the case study *churn prediction* on the one of Indonesian Telco company. The Result show that DataboostIM can improve minor class and keep accuracies in the major class, beside that Lift Curve from DataboostIM better than balancing node Clementine 9.0.

Keywords: *databoostIM, boosting, imbalance class, data generation, churn prediction.*