Finding the high-retaining customer

A study of boosted decision trees to classify user retention on an audio streaming service

Carl Samuelsson*

June 2020

Abstract

In an audio streaming service where its success is greatly dependent on the degree to which its users retain on the service, it is of evident interest to predict who will stay and not for proactive measures. Moreover, to optimize the service towards the preferences of its users, inferential work is needed towards obtaining a better understanding of the underlying processes of why certain users decide to stay on the service or not. This work centers around the study and evaluation of boosted tree models on predictive and inferential grounds. Empirical results shows however only a marginal predictive benefit of using these rather complex models in comparison with a baseline logistic regression. However, more exhaustive future work which utilizes a larger degree of the vast amount of information available could prove the boosted tree models justice for the task of interest.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden.
E-mail: c.samuelsson94@gmail.com . Supervisor: Chun-Biu Li.