

## Introduction

### Spring 2018 Sports Marketing Analytics Contest: Season ticket churn

Customer churn is a classic customer relationship management (CRM) problem for marketers in general, referring to the attrition of customers in renewable contractual relationships with the firm. An application in sports marketing can be found when considering the renewal of season tickets for a sports team, with non-renewal equivalent to churn. The context for this semester's contest is to build a predictive churn model for a team.

The overriding objective of this contest is to encourage the students of our college to explore data analytics and get first-hand experience in seeing how data can be used to produce better informed decisions. No particular coursework, experience, aptitude, or anything else is required to participate in the contest – submissions can be simple or quite sophisticated. Hopefully, each student that participates will benefit from enhancing their analytical skills.

The Spring 2018 Sports Marketing Analytics Contest focuses on the prediction of season ticket holders that will not renew, i.e. “churn” or “defect”. In contemporary sports organizations, as in businesses in general, predictions like these are produced by models developed through the analysis of customer level data. Typically, a churn model will produce a score (often times in the form of a probability) for each customer, designed in this context to correspond with the likelihood of that customer defecting. The scores will be a function of the individual-level data. To make a binary prediction, 1=not renew and 0=renew, a threshold score must be identified, with customer scores that exceed that threshold indicating churn, and those below the threshold indicating renewal. So essentially, this is a binary classification model.

#### A simple example

A rudimentary scoring model, based on heuristics, might be as follows:

- fans that never used their tickets the previous year, or only used them once, are more likely to defect, so fans never using their tickets get 2 points, and those only using their tickets once get 1 point; and
- newer fans are more likely to defect than those that have held season tickets for a long time, so those fans that have held tickets for three seasons or less get 1 point.

An arbitrary threshold value might be chosen, say 2 points, with those fans with a score of 2 or more points predicted to defect, and those with less than 2 points predicted to renew.

Given this scoring system, fans predicted to churn would be:

- established fans that never used their tickets the previous year (2 points),
- new fans that only used their tickets once ( $1 + 1 = 2$  points), and
- new fans that never used their tickets last season ( $2 + 1 = 3$  points).

All fans with scores of 0 or 1 point would be predicted to renew.

The challenge is to develop a scoring system and identify a threshold that most effectively classifies the actual 0's and 1's, renewals and non-renewals, respectively. In a practical sense, an effective classification model allows for a more efficient allocation of marketing resources, e.g. retention efforts can be focused on those season ticket holders that are predicted to be most at risk to defect.

We acknowledge The Association for Computing Machinery's Special Interest Group on Knowledge Discovery and Data Mining and their KDD Cup Competitions. The set-up for this contest has borrowed from the general organization, rules, and evaluation approach of the 2009 KDD Cup. <http://www.kdd.org/kdd-cup/view/kdd-cup-2009>