Journal of Business Research 66 (2013) 1629–1636

Contents lists available at SciVerse ScienceDirect

Journal of Business Research

Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning

Kristof Coussement *, Koen W. De Bock

IESEG School of Management, Université Catholique de Lille (LEM, UMR CNRS 8179), Expertise Center for Database Marketing (ECDM), Department of Marketing, 3 Rue de la Digue, F-59000, Lille, France

A R T I C L E I N F O

Article history:
Received 1 January 2012
Received in revised form 1 September 2012
Accepted 1 October 2012
Available online 10 January 2013

Keywords:
Customer relationship management
Online gambling
Customer churn prediction
Ensemble algorithms
GAMens
Random forests

A B S T R A C T

The online gambling industry is one of the most revenue generating branches of the entertainment business, resulting in fierce competition and saturated markets. Therefore it is essential to efficiently retain gamblers. Churn prediction is a promising new alternative in customer relationship management (CRM) to analyze customer retention. It is the process of identifying gamblers with a high probability to leave the company based on their past behavior. This study investigates whether churn prediction is a valuable option in the CRM palette of the online gambling companies. Using real-life data of poker players at bwin, single algorithms, CART decision trees and generalized additive models are benchmarked to their ensemble counterparts, random forests and GAMens. The results show that churn prediction is a valuable strategy to identify and profile those customers at risk. Furthermore, the performance of the ensembles is more robust and better than the single models.

1. Introduction

Online gambling is an upcoming trend in today's society due to the digitalization and availability of new technologies, and its popularity is reflected by a tremendous amount of academic research studies (e.g. Smith, Levere, & Kurtzman, 2009). For instance, the online poker market has grown everywhere in the world, accounting for an estimated $3.6 billion in total revenues (Fiedler & Wilcke, 2011). Popular online gambling activities involve sports betting, virtual poker, and traditional casino games.

Previous research mainly focuses on the psychological facets of gambling behavior (Bleichrodt & Schmidt, 2002; Lam, 2006, 2007; McDaniel & Zuckerman, 2003; Mowen, Fang, & Scott, 2009), while other research papers investigate the determinants of individual differences amongst gamblers (e.g. Cronce and Corbin (2010)). Thus far, business and marketing literature approaches gambling from a purely exploratory insight perspective. In detail, former research gives insight into the individual betting behavior (Andrade & Iyer, 2009; Grant & Xie, 2007; Seybert & Bloomfield, 2009; Smith et al., 2009), the legitimation and development process of gambling and the impact on residents' perceptions (Humphreys, 2010; Roehl, 1999), the forecasting possibility of the outcomes of sporting events (Stekler, Sendor, & Verlander, 2010), the gambling agents' learning processes (Dana & Knetter, 1994), etc.

Although Oh and Hsu (2001) report that past gambling behavior impacts future gambling activities and that later on, several researchers tested the impact of past (aggregated) gambling behavior on future gambling behavior (Jolley, Mizerski, & Olaru, 2006; Lam, 2006; Lam & Mizerski, 2009; Mizerski, Miller, Mizerski, & Lam, 2004), no research study has explored the possibility of using real, individual past gamblers' behavior to predict future behavior to improve the customer relationship management (CRM) of the gambling company. However, gambling companies are searching new approaches to retain customers using their CRM strategy, because the rapid growth of the Internet does not facilitate the choice of the gambling website amongst the many options available (Jolley et al., 2006). Consequently, the CRM literature is expanding (Cooper, Gwin, & Wakefield, 2008; Ko, Kim, Kim, & Woo, 2008) and the fight to maintain the current players' base is a very important challenge within the gambling industry nowadays (Lal, Martone Carrolo, & Harrah's entertainment, 2001).

Since the concept of problem gambling is associated with the retention of gamblers, major gambling companies are actively approaching their players with targeted marketing actions to continue playing (Jolley et al., 2006; National Gambling Impact Study Commission, 1999; Productivity Commission, 1999). Indeed, in accordance with retailers and financial service providers, a lot of gambling companies realize that collecting customer data and storing the information into a customer database is an essential element in their retention strategy (Athanassopoulos, 2000).

Customer churn prediction is one of the key activities of a proactive retention strategy (Keaveney & Parthasarathy, 2001). Customer churn prediction is the process of assigning a probability of future churning

* Corresponding author.
E-mail addresses: k.coussement@ieseg.fr (K. Coussement), k.debock@ieseg.fr (K.W. De Bock).

0148-2963/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jbusres.2012.12.008

behavior to each gambler in the database by building a prediction model based on past information. Intuitively, customers' past behavior is in line with their future gambling behavior. A good churn prediction model is a model that is able to assign high churn probabilities to real churners, and low churn probabilities otherwise. Practically, the marketing analyst uses the churn probabilities of the prediction model to rank the customers from most likely to leave the company to least likely to leave the company. The top X% of customers, containing on average more churners than the random model, are used for the company's targeted retention campaign.

Neslin, Gupta, Kamakura, Lu, and Mason (2006) indicate that numerous steps in the customer churn prediction process have an impact on its success. However, they strongly suggest focusing on the prediction technique due to its huge impact on the return on investment of subsequent marketing actions. As a result, academic literature on the optimization of churn prediction algorithms has exploded in the last few years. Prediction algorithms that are used in customer churn prediction literature include *single algorithms* like decision trees (Lemmens & Croux, 2006), generalized additive models (Coussement, Benoit, & Van den Poel, 2010), logistic regression (Smith, Willis, & Brooks, 2000) or support vector machines (Coussement & Van den Poel, 2008), etc., and *ensemble learners* that combine the predictions of multiple single algorithms (e.g. random forests (Coussement & Van den Poel, 2009)).

This paper introduces the concept of customer churn prediction in the gambling literature, whereas its value as an additional instrument in the CRM toolbox is investigated. Section 2 describes the predictive modeling methodology. Moreover, Section 3 digs into the churn prediction algorithms. Section 4 describes the evaluation metrics used, while Section 5 gives more insight into the real-life gambling dataset. Furthermore, the results are described in Section 6. Finally, the managerial implications for customer churn prediction and the use of ensemble methods are discussed in Section 7, while Section 8 concludes this research paper.

2. Predictive modeling methodology

The purpose of predictive modeling is to predict what is likely to happen in the future based upon what happened in the past (Blattberg, Kim, & Neslin, 2008). In the context of CRM, predictive modeling uses historical transactions and characteristics of a customer to predict future customer behavior.

In predictive modeling, two distinct phases are identified, i.e. a training phase and a prediction or test phase. In the training phase, a model that links future customer behavior to historical customer information is created using a training dataset. This dataset consists of input variables and a target variable for a range of customers. The input variables describe the customers' profiles and their past behavior in a certain time period, and usually include demographic characteristics and historical transactions. The latter category includes the popular RFM variables (McCarty & Hastak, 2007), i.e. the recency of customer purchases (R), the purchase frequency (F) and the monetary value of the historical transactions (M). Several studies show that RFM variables are amongst the strongest performing variables in explaining future customer behavior (e.g. Bose & Chen, 2009). The target variable reflects the behavior of interest to be predicted in a subsequent period of time. In the case of customer churn prediction, the target variable is represented by a binary variable representing whether or not a customer churned during the observation period. The purpose of training a predictive model is trying to capture the relationship between the input variables and the target variable.

In the prediction or test phase, the trained model is deployed on data not used during model training, i.e. the test set. Indeed, a test set of customers that are not included in the training process is used to assess the quality of the predictions output by the model. Measuring the performance on unseen data better reflects the true prediction performance, because measuring the performance on the training dataset often results in finding idiosyncratic patterns in the training dataset that do not hold up in a real-life test setting. As such, the original dataset is randomly split into a training set and a test set. In order to disclose the true capabilities of a prediction algorithm, it is good practice to repeat the training phase multiple times in order to reduce the risk of erroneously identifying a particular model as a good model (Malthouse, 2001). This objective is pursued in the cross-validation methodology as suggested by Witten and Frank (2000) and already implemented in a variety of predictive modeling settings in marketing (e.g. Cui, Wong, & Lui, 2006). In this study, results are based upon a five times twofold cross-validation procedure (5×2-fold cv). This involves five replications of a twofold cross-validation. In each replication, customers of the original dataset are randomly assigned to one of two equally-sized cross-validation parts. One part is once used as training data to build the prediction model, while the performance is calculated for the other part acting as a test dataset. This process is then repeated, switching the roles of the two cross-validation parts. Furthermore before putting a prediction model into practice, it is important to compare its performance to alternative types of prediction models in order to find the best algorithm (Neslin et al., 2006). As such, Section 3 describes the different prediction algorithms used in this research study.

3. Churn prediction algorithms

3.1. Single algorithms

3.1.1. CART decision tree

Decision trees are popular techniques in predictive modeling due to their simplicity and transparency (Duda, Hart, & Stork, 2001). For instance, Murthy (1998) found more than 300 academic references that use decision trees in a variety of settings. A decision tree is composed of a set of rules that divide a dataset that is heterogeneous in terms of the target variable into smaller, more homogeneous sets. The technique is best explained by referring to its prediction phase. To determine the churn probability of a customer in the test dataset, the customer enters the decision tree at the start or root node at the top of the decision tree. This node represents a test that attributes the customer to one of the lower-level or child nodes. The test is a logical question formulated in terms of the input variable, chosen in such a way that it discriminates maximally between churners and non-churners. The result of the test determines which child node is chosen for the customer. Tests at subsequent child nodes redirect the customer through the decision tree until a terminal or leaf node is reached. The collection of unique paths between the root node and the leaf nodes makes up the rules used to generate predictions. All customers from the test set reaching the same leaf node receive the same prediction probability, calculated as the percentage of churners from all training customers in that leaf node.

This study employs the CART algorithm, acronym for Classification and Regression Trees, because it is a popular decision-tree algorithm using the *Gini coefficient* as splitting criterion (Breiman, Friedman, Olsen, & Stone, 1984). It implements binary recursive partitioning; it uses *two-way splits* that split a node into exactly two child nodes, while its growing process continues recursively until splits are no longer possible because all customers in a node are identical in terms of the target variable or the input variables. Finally, the decision-tree pruning that aims at reducing the decision-tree complexity is necessary, because decision trees are susceptible to overfitting. In CART, pruning is done via the *maximum performance objective criterion* (Breiman et al., 1984; Steinberg & Colla, 1998).

3.1.2. Generalized additive model

The most popular technique among predictive modeling techniques that have been applied to the prediction of customer churn