Industrial Marketing Management xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Industrial Marketing Management

Managing B2B customer churn, retention and profitability

Ali Tamaddoni Jahromi [a,1], Stanislav Stakhovych [a,2], Michael Ewing [b,*]

[a] Department of Marketing, Monash University, PO Box 197, Caulfield East, VIC 3145, Australia
[b] Faculty of Business and Law, Deakin University, 221 Burwood Highway, Burwood, VIC 3125, Australia

ARTICLE INFO

Article history:
Received 18 May 2013
Received in revised form 1 November 2013
Accepted 21 January 2014
Available online xxxx

Keywords:
B2B customer churn
Data mining
Non-contractual setting
Retention campaign
Profit

ABSTRACT

It is now widely accepted that firms should direct more effort into retaining existing customers than to attracting new ones. To achieve this, customers likely to defect need to be identified so that they can be approached with tailored incentives or other bespoke retention offers. Such strategies call for predictive models capable of identifying customers with higher probabilities of defecting in the relatively near future. A review of the extant literature on customer churn models reveals that although several predictive models have been developed to model churn in B2C contexts, the B2B context in general, and non-contractual settings in particular, have received less attention in this regard. Therefore, to address these gaps, this study proposes a data-mining approach to model non-contractual customer churn in B2B contexts. Several modeling techniques are compared in terms of their ability to predict true churners. The best performing data-mining technique (boosting) is then applied to develop a profit maximizing retention campaign. Results confirm that the model driven approach to churn prediction and developing retention strategies outperforms commonly used managerial heuristics.

1. Introduction

Customer churn remains a particularly salient concept in contemporary marketing and should not be ignored by B2B companies. Nowadays, due to improved access to information, customers are more transient and it is easier and less costly for them to switch between competitors (Wiersema, 2013). Firms recognize this and are interested in identifying potential churners in order to attempt to prevent defection by targeting such customers with incentives. Previous studies in B2B contexts have investigated churn from the point of view of resource allocation or customer profitability prediction (Rust, Kumar, & Venkatesan, 2011; Venkatesan & Kumar, 2004). However, the application of data mining techniques to predict churn and to develop efficient retention campaigns in B2B contexts does not appear to have received much attention in either literature or practice. Therefore, the current study aims to (1) introduce data mining techniques to model customer churn in a B2B setting and (2) develop a profit maximizing retention campaign based on customer heterogeneity.

Kamakura et al. (2005, p. 286) define churn as "the tendency for customers to defect or cease business with a company". Losing a customer can reduce sales revenues and increase acquisition costs (Athanassopoulos, 2000; Risselada, Verhoef, & Bijmolt, 2010). Since net return on investments (ROI) for retention strategies are generally higher than for acquisition, it is now widely accepted that companies should concentrate their marketing resources on customer retention which translates into increased revenue for the company because loyal customers continue transacting for relatively longer durations, while ignoring offers from competitors (Lam, Shankar, Erramilli, & Murthy, 2004). The importance of retention for suppliers becomes even clearer in the B2B context where customers make larger and more frequent purchases with far higher transactional values (Rauyruen & Miller, 2007). Moreover, business customers are fewer in numbers and more valuable which makes losing even one highly undesirable (Stevens, 2005). Add to that today's austere global economic outlook and increased web-based comparison shopping, and the importance of effective retention strategies becomes even more salient for B2B organizations (Waxer, 2011).

In tackling the ongoing challenge of customer churn, some firms favor untargeted approaches which rely on mass advertising and superior product (Burez & Van den Poel, 2007). For instance, a company that is dealing with high rates of churn might send incentives to all customers. Although such a 'shot-gun' strategy is easy to implement, it risks wasting resources by sending incentives to customers already intent on staying with the company regardless. Recent innovations in IT have led to increased availability of transactional customer data which companies have started utilizing in more targeted marketing approaches (Burez & Van den Poel, 2007). Adopting more targeted strategies enables firms to use customers' transactional data in conducting customer-base analyses and predicting customers' future behavior (Fader & Hardie, 2009). Over the past decade, the ability to collect and process large amounts of data and disseminate relevant knowledge has been considered to be a necessary core competency for successful companies (Cooper, Watson,

* Corresponding author. Tel.: +61 3 9244 6548.
E-mail addresses: Ali.Tamaddoni@monash.edu (A. Tamaddoni Jahromi), Stanislav.Stakhovych@monash.edu (S. Stakhovych), pvc-bl@deakin.edu.au (M. Ewing).
[1] Tel.: +61 3 990 34232.
[2] Tel.: +61 3 990 31455.

http://dx.doi.org/10.1016/j.indmarman.2014.06.016
0019-8501/

Wixom, & Goodhue, 2000; Sher & Lee, 2004). In response, astute companies have turned to analytical methods to better exploit the data collected from customers in their strategic and operational decision making process (Earl, 2001; Orriols-Puig, Martínez-López, Casillas, & Lee, 2013). In this regard, in an attempt to prevent customer churn, organizations are increasingly utilizing customers' transactional data to develop predictive models to identify customers who are more likely to defect. The identified churners are then targeted with tailored incentives to be encouraged to stay (Neslin, Gupta, Kamakura, Junxiang, & Mason, 2006).

Several studies have addressed the application of data mining techniques in identifying potential churners in B2C contexts — particularly in contractual settings (Burez & Van den Poel, 2007; Coussement & Van den Poel, 2008a; Neslin et al., 2006). However, the analysis of extant literature revealed less attention on modeling customer churn in B2B contexts. This paucity stems from the fact that 'big data' has not (yet) been touted in B2B contexts with the same enthusiasm as it has in B2C (Wiersema, 2013). Moreover, the issue of customer switching or defection in B2B contexts has been mostly studied from a resource allocation perspective, or in the context of predicting future customer profitability (Rust et al., 2011; Venkatesan & Kumar, 2004). The major focus of such studies has been on identifying and nurturing profitable customers rather than on pinpointing customers who are more likely to churn, or on the potential profitability of targeted retention campaigns derived from such predictive models.

Therefore, the current study contributes to the extant literature in two important ways. First, it introduces data mining modeling approaches for churn prediction to B2B contexts and evaluates their predictive power. Second, it proposes a profit maximizing retention campaign based on the output of data mining churn prediction.

In the first stage of our analysis, customers' transactional data from a company operating in a B2B context is utilized to construct a predictive model. To this end, Classification and Regression Tree (CART), as a single learner data mining technique (Breiman, Friedman, Olshen, & Stone, 1984), and boosting technique, as an ensemble learner (Breiman, 1996), have been employed to construct predictive models. Once constructed, all models have been benchmarked against logistic regression, as a popular binary classifier in choice modeling and customer churn literature (Ainslie & Pitt, 1992; Lemmens & Croux, 2006). Thereafter, the performance of the models is evaluated using an 'area under ROC (receiver operating characteristic) curve' and cumulative lift measures, to identify the best performing churn prediction model.

The second stage of analysis uses individual churn probabilities for each customer produced by the best performing technique in the first stage, to develop a profit maximizing retention campaign for the company in question.

The remainder of the paper is structured as follows. Section 2 highlights the importance of targeted retention approaches and reviews the extant literature on churn modeling and churn management profitability. The study methodology is presented in Section 3 followed by an empirical application in Section 4 to demonstrate how this retention approach works in practice. Finally, the conclusions and managerial recommendations are presented in Section 5.

2. Background

2.1. Retention approaches in customer relationship management

Though CRM is a frequently used term in contemporary marketing, it has its roots in the IT realm, only entering the scholarly marketing lexicon in the late 90s (Srivastava, Shervani, & Fahey, 1999; Wübben, 2008). While a variety of definitions exist for CRM (Kotler & Keller, 2006; Richards & Jones, 2008), and consensus has yet to be reached, the majority of definitions have 'acquisition' and 'retention' at their common core (D'Haen & Van den Poel, 2013; Jackson, 2005). An extensive body of literature now exists in both B2B and B2C contexts that emphasizes several benefits of retention over acquisition approaches (Ganesh, Arnold, & Reynolds, 2000; Lam et al., 2004; Rauyruen & Miller, 2007).

The inherent advantages of retention over acquisition strategies have resulted in many companies concentrating their efforts on the former (Bhattacharya, 1998; Colgate & Danaher, 2000). It has been established that a small change in retention rates can result in significant changes in contribution (Van den Poel & Larivière, 2004).

Specifically, in B2B contexts, customer retention is considered central to developing business relationships (Eriksson & Vaghult, 2000; Grönroos, 1990; Kalwani & Narayandas, 1995). Due to the large amounts of money that B2B customers typically spend, retention can be extremely financially rewarding for firms (Boles, Barksdale, & Johnson, 1997; Rauyruen & Miller, 2007).

Considering the importance of retention, it is becoming a widespread belief across most industries that retaining existing customers and decreasing churn are both essential (Tsai & Lu, 2009). A primary way to retain customers is to offer retention incentives to reduce churn (Shaffer & Zhang, 2002). However, due to the costs associated with such retention strategies, coupled with the fact that not all customers are at risk of defecting, it is advisable not to focus churn management efforts across the entire customer base (Hadden, Tiwari, Roy, & Ruta, 2007). Rather, firms are encouraged to develop models to predict which customers are more likely to defect (Keaveney & Parthasarathy, 2001; Neslin et al., 2006). Once identified, these likely defectors should be targeted with appropriate incentives to convince them to stay (Hadden et al., 2007).

In this regard, churn management strategies are typically two-pronged: identifying would-be churners and then targeting them with retention incentives. While classic churn management strategies focus on targeting customers with the highest probability of defection (e.g. top 10%) (Lemmens & Croux, 2006), a more recent approach advocates selecting target customers based on the profit potential of each and to select the size of the target group to maximize the overall return on the retention campaign (Lemmens & Gupta, 2013; Verbeke, Dejaeger, Martens, Hur, & Baesens, 2012).

In this regard, the current study is inspired by Neslin et al. (2006) and comprises two distinct stages: predicting customer churn and maximizing retention campaign profitability.

2.2. Modeling customer churn

Over the past few decades the availability of various marketing-related data such as scanner data and internet data along with organizations' demand for new analytical methods, has spurned an increasing interest in Artificial Intelligence (AI)-based marketing problem solving (Orriols-Puig et al., 2013). Of all the roles that AI-based systems can play in marketing, predictive modeling and more specifically, churn modeling is one of the most promising (Wierenga, 2010).

Customer churn prediction is the process of calculating the probability of future churning behavior for each customer in the database, using a predictive model, based on past information/prior behavior (Coussement & De Bock, 2013). Thus, with the aim of developing an effective customer retention program, the utilized models should be as accurate as possible (Coussement & Van den Poel, 2008b), otherwise these systems would be very wasteful when spending incentive money on customers who will not churn (Tsai & Lu, 2009). In this regard, data mining techniques, with their roots in AI, have been widely favored to model churn (Risselada et al., 2010; Wierenga, 2010). The tendency to employ data mining techniques in customer churn prediction stems from the fact that churn is a rare event in a dataset and making an accurate forecast calls for techniques that emphasize predictive ability (Kamakura et al., 2005).

Literature on churn modeling confirms the simplicity and transparency of decision trees as popular single algorithm data mining techniques (Coussement & De Bock, 2013) for predictive modeling (Duda, Hart, & Stork, 2001). This algorithm has been utilized either as a part of hybrid approach in combination with other techniques (Jamal & Zhang, 2009) or as an individual, independent technique to predict customer churn (Hung, Yen, & Wang, 2006; Wei & Chiu, 2002). Furthermore, marketing scholars have recently introduced ensemble learners (Coussement & De Bock, 2013; Lemmens & Croux, 2006) such as Bagging and Boosting techniques to the customer churn prediction tool kit.

Nevertheless, studies reveal that, in general, research directions calling for applications of strategic intelligence (i.e. business intelligence, competitive intelligence, and knowledge management) in industrial marketing have received insufficient attention from both academics and practitioners (Martínez-López & Casillas, 2013; Wiersema, 2013). Accordingly, Martínez-López and Casillas (2013) argue that the application of intelligent systems in handling industrial marketing problems has been limited which has also been reflected in the paucity of academic studies in the strategic intelligence realm. In other words, this research theme has been underdeveloped in business and management journals. Digging deeper to uncover the underlying reasons for this gap, Wiersema (2013) points out that as opposed to the B2C field, the availability of B2B 'big data' is more limited. Thus, mining large datasets to extract knowledge about customers is not as common as in the B2C field. Furthermore, in cases where the data is available, the practices to exploit this data and transform it into information are still underdeveloped in B2B companies.

Similarly, by delving more deeply into the existing literature on churn modeling, one also notes that among all studies concentrating on predicting churn across different sectors such as telecommunications (Lemmens & Croux, 2006; Neslin et al., 2006), online retail (Yu, Guo, Guo, & Huang, 2011), finance (Risselada et al., 2010; Van den Poel & Larivière, 2004), and retail (Buckinx & Van den Poel, 2005), the majority are within B2C contexts and the application of data mining techniques in B2B churn prediction is still an underdeveloped area. This indicates the opportunity that exists to introduce different techniques and approaches of data mining modeling to churn prediction area in B2B contexts. The magnitude of the opportunity becomes even clearer when the nature of B2B contexts, with large purchases and transactions, is taken into account (Rauyruen & Miller, 2007).

2.3. Maximizing retention campaign profitability

A traditional churn management strategy typically involves ranking customers based on their estimated propensity to churn and sending out incentives to the subset of customers with the highest churn ranking (Lemmens & Gupta, 2013). The rationale behind this approach is that targeting the customers with higher propensity to defect and persuading them to stay with the company can increase firm's profitability. However, customers with the highest propensity to churn are not necessarily always the ones who can deliver the highest profit. In fact, as Lemmens and Gupta (2013) note, on the basis of Neslin et al. (2006) conceptual framework, at the customer level, the profit of targeting a customer by a retention incentive is influenced by several factors including, a customer's probability of churn, their probability of accepting an incentive offer, the cost of the incentive to the company, and the value of the customer.

According to Neslin et al. (2006) the profit of a churn management campaign at an aggregated level can be calculated by Eq. (1):

$$\Pi = N\alpha[\beta\gamma(CLV - c - \delta) + \beta(1-\gamma)(-c) + (1-\beta)(-c-\delta)] - A \qquad (1)$$

where, N is the total number of customers, α is the subset of the customer base which is targeted for the retention campaign, β is the fraction of potential churners within the targeted customers, δ is the cost of the incentive for the company, γ is the portion of would-be churners who decide to stay because of receiving the incentive, c is the cost of contacting a customer to offer the incentive, CLV is the lifetime value of an average customer, and A is the fixed cost of running the retention campaign.

Basically, when a company embarks on a retention campaign by implementing a churn prediction model, it correctly identifies a fraction of potential churners and sends retention promotions to them. From those real 'would-be churners' who receive the offer, a fraction of γ would accept it and stay loyal. Accordingly, the first term within the brackets in Eq. (1) represents the contribution of the retention campaign when potential churners are retained by incentives. The second term within the brackets reflects the money that the company loses due to contacting the fraction of would-be churners who do not accept the offer and ultimately defect. The last term within the brackets represents the loss of money when the company sends the retention incentive to a customer who would stay with the company anyway and the customer redeems the offer.

The framework proposed by Neslin et al. (2006) provides a metric to calculate the profit of a retention campaign at the aggregated level, when the company identifies the churner using a prediction model, ranks customers based on their propensity to churn and targets a subset of likely churners with incentives. However, one pitfall of this approach is that it does not maximize the profitability of the retention campaign. Thus, Verbeke et al. (2012) proposed an approach to profit maximization by evaluating the optimum target size. However, neither of these approaches account for heterogeneity in lifetime value of customers as well as the heterogeneity in probability of being targeted by incentives which comes from individual's propensity to churn. However, Lemmens and Gupta (2013) have addressed both of these limitations by replacing β with probability of being a would-be churner (extracted from a binary classifier) and by replacing lifetime value of an average customer with individual customer lifetime values. These changes allow the analyst to assess the individual contribution of each customer to the total profit of a targeted retention campaign. Instead of ranking customers based on their probability to churn, they can be ranked based on their individual profitability, allowing the firm to determine the optimum campaign target size to maximize profitability.

However, although the improved framework might work well in B2C contexts, it has to be modified to suit B2B markets. One of the modifications needed concerns the assumption of fixed incentives δ. While such assumptions might work well in B2C contexts (i.e. whereby all customers receive coupons of the same value), this does not necessarily hold in B2B contexts. As Mogelefsky (2000) note, B2B incentives, in order to be impactful, need to be personalized and customized. Moreover, it is more common in B2B relationships to offer percentage volume discounts rather than fixed discounts (Harrison-Walker & Neeley, 2004).

3. Methodology

3.1. Defining 'customer churn' in non-contractual settings

Defining customer churn in non-contractual settings is complex. In the absence of a contract(s) between the focal company and its customers to be renewed or terminated, it is difficult to estimate the exact time of defection or churn. Because of the inherent difficulty of specifying churn in non-contractual settings, most of the extant churn literature has focused on contractual settings. Analysts interested in non-contractual settings need to take into account exactly what is meant by "churn". In such settings, since it is probable that a customer returns after a period of inactivity (i.e. 'always-a-share' scenario) non-contractual churn has a different meaning (Rust, Lemon, & Zeithaml, 2004; Venkatesan & Kumar, 2004).

Thus, in the first stage of the current study, instead of attempting to predict churn as a 'permanent phenomenon', we have focused on predicting inactivity in the next time period. The length of time periods of interest can be defined arbitrarily as long as they follow what is practical and important for business decisions. Therefore, we choose half a year as a unit of measurement and define churn as being inactive in the second half of the year (prediction period) while being active in the first half of the year (calibration period). This definition is in line with Buckinx and Van den Poel (2005) and is based on the frequency of purchases.

There are numerous other ways to define churn, but they all have different drawbacks in the context of this paper. For instance, Miguéis, Van den Poel, Camanho, and Falcão e Cunha (2012) defined churn based on changes in customers spending from one period to another. We employ a 'change in monetary' variable as a predictor of a churn rather than the definition of the churn itself. Also there are several statistical approaches to calculate the probability for a customer to be 'alive' at a point in time (Kumar & Reinartz, 2012; Schmittlein & Morrison, 1985; Schmittlein, Morrison, & Colombo, 1987) but they have two weaknesses in the current application: first, we require the probability of being inactive for the whole time period and not at a point in time; and, second, in order to define the indicator of churn, the subjective cutoff point is required to convert the probability to a binary variable.

3.2. Classification techniques

3.2.1. Simple decision tree

Several classifiers including decision trees, support vector machines (SVM), Bayesian classifiers and artificial neural networks (ANN) have been identified by machine learning modelers (Han, Kamber, & Pei, 2011; Tan, Steinbach, & Kumar, 2006). Among all existing classification techniques, decision trees are the most popular in business since their underlying logic is typically more understandable to managers (Wei & Chiu, 2002; Xie, Li, Ngai, & Ying, 2009). One of the main reasons behind the popularity of decision trees is their transparency and interpretability (Olafsson, Li, & Wu, 2008).

A decision tree is a tree-shaped structure that represents sets of decisions and is able to generate rules for the classification of a data set (Lee & Siau, 2001). This technique is suitable for describing sequences of interrelated decisions or predicting future data trends (Chen, Hsu, & Chou, 2003; Kim, Song, Kim, & Kim, 2005) and is capable of classifying specific entities into specific classes based on feature of entities (Buckinx, Moons, Van Den Poel, & Wets, 2004; Chen et al., 2003). Decision trees are one of the top three most popular techniques of data mining in CRM (Xie et al., 2009), and have been included in our comparative study of churn prediction.

3.2.2. Decision tree with cost-sensitive learning

In modeling customer churn, class imbalance is a common challenge for model developers. In such cases, it is usually the rare class that is of primary interest (Burez & Van den Poel, 2009). In most churn data sets, the number of non-churners is greater than the churners. Thus, despite the fact that misclassifying the real churners might not have a great impact on model's accuracy, it can cause a costly loss for the companies.

Therefore, cost-sensitive learning methods have been utilized by academics to solve the problem of class imbalance in churn prediction (Burez & Van den Poel, 2009; Weiss, 2004). Basically, cost-sensitive learning methods consider the fact that the correct classification of churners has more value than correct classification of non-churners and this is done for a binary classification problem via assigning more cost to false negatives than the false positives (Burez & Van den Poel, 2009).

Thus, in order to handle the problem of class imbalance, in addition to simple decision trees, we compare the performance of cost-sensitive learning as well.

3.2.3. Boosting as an ensemble learner method

The idea of aggregating classifiers was initially proposed by Breiman (1996) who believed that the combination of several base classifiers can increase the overall accuracy of the aggregated model. In this regard, a class of ensemble learners such as random forests, bagging, and boosting have been introduced within the data mining stream of churn modeling (Breiman, 1996).

Among exiting ensemble learners, the boosting technique is popular due to its outstanding churn prediction capabilities (Burez & Van den Poel, 2009; Lemmens & Croux, 2006). Basically, the boosting technique manipulates the weight of misclassified instances by attributing more importance to them over multiple training iterations to help the classifier in the classification of instances which are difficult to classify correctly (De Bock & Van den Poel, 2011). Several versions of boosting exist, such as logitboost (Friedman, Hastie, & Tibshirani, 2000), adaptive boosting (Freund & Schapire, 1996), and brownboost (Freund, 2001). In this study we use adaptive boosting as it is one of the most well-known and capable boosting techniques (De Bock & Van den Poel, 2011; Han et al., 2011). For details of adaptive boosting algorithm please refer to Appendix A.

3.2.4. Logistic regression

In cases such as churn prediction, where the dependent variable is binary (e.g. churner as '1' vs. non-churner as '0'), the ordinary linear regression is not applicable as it allows the dependent variable to fall outside the range of 0–1. Thus, as a special case of general linear models – logistic regression – is favored.

Ease of use and robustness of results (Buckinx & Van den Poel, 2005) have made logistic regression a popular binary classifier among marketing academics as well as the first choice for customer churn modeling (Lemmens & Croux, 2006; Neslin et al., 2006). Therefore, in this study, logistic regression is used as a benchmark technique to compare against the performance of more sophisticated models. We are interested to see whether the added complexity of more sophisticated techniques pays off for predicting churn when simpler methods such as logistic regression can be utilized instead.

3.3. Assessment criteria

To evaluate the predictive performance of constructed models, two criteria have been employed: area under the receiver operating characteristic curve (AUC) and cumulative lift curve.

The simplest form of data produced by a classification model is a confusion matrix (Han et al., 2011) of counts of correctly classified and misclassified cases from each class. Consider the case of churn prediction. A binary classifier predicts the status of a customer as churner or non-churner. Such classifier might classify a real churner as 'churner' (true positive), a real churner as 'non-churner' (false negative), a real non-churner as 'non-churner' (true negative), and a real non-churner as 'churner' (false positive) as illustrated in Table 1.

Using the confusion matrix, one can extract *FP* rate (*FP*/*N*) and *TP* rate (*TP*/*P*). By plotting the *FP* and *TP* pairs for different decision thresholds, receiver operating characteristic (ROC) curve can be achieved (Bradley, 1997). In other words, a ROC curve plots the true positive rates vs. false positive rates, for a binary classifier system as its discrimination threshold is varied. However, as the ROC curve gives the models performance with two measures there might be cases where increasing the true positive rate causes an increase in false positive rate. In such cases the comparison of two models would become a challenging task. In order to tackle this problem, in this study, the performance of the models is compared based on the area under their ROC curve (AUC) (Bradley, 1997; Huang, Kechadi, & Buckley, 2012). While a random classifier possesses an AUC of 0.5, an ideal classifier has an AUC equal to 1. On this basis, all classifiers in practice have an AUC greater than 0.5 and smaller than 1. Area under ROC curve is computed as follows (Burez & Van den Poel, 2009):

$$\mathrm{AUC} = \int_0^1 \frac{\mathrm{TP}}{P} d\frac{\mathrm{FP}}{N} = \frac{1}{P \cdot N}\int_0^N \mathrm{TP} d\mathrm{FP}. \tag{2}$$

As a second evaluation criterion, the cumulative lift chart (Jamal & Bucklin, 2006; Risselada et al., 2010), the most popular prediction criterion in predictive modeling (Neslin et al., 2006), is employed. The focus of this measure is on customers with the highest probability to be 'positive' (i.e. churner in the case of churn modeling) and it is defined as the ratio of 'positives' in a segment, divided by the ratio of 'positives' in the whole test set (Berry & Linoff, 2004). For example, by using no model it is expected that by targeting *n*% of customer base with incentives *n*% of real churners would be targeted. However, by using a classification model with top *nth*

Please cite this article as: Tamaddoni Jahromi, A., et al., Managing B2B customer churn, retention and profitability, *Industrial Marketing Management* (2014), http://dx.doi.org/10.1016/j.indmarman.2014.06.016

percentile lift of c, it is expected that $c \times n$ % of all real churners would receive the incentive, when the top n% of customer are targeted. In other words, in order that a model performs better than the random classifier, its top-decile lift c should be greater than 1. Hence, high top-decile lift is an indicator of promising performance of the classifier (Lemmens & Croux, 2006).

3.4. Developing a profit maximizing retention campaign

Contrary to traditional retention approaches where customers are ranked on their likelihood to churn and the first *nth* percentile of customers is targeted, we propose a method that 1) maximizes the total profit of a retention campaign, and 2) determines the optimum target size. We follow Lemmens and Gupta (2013) and assume that individual profitability of targeting the customer i (π_i) can be computed. The overall profit of a retention campaign across all targeted customers is:

$$\Pi = \sum_{i \in target}^{N} \pi_i \tag{3}$$

The individual profit of targeting customer i depends on several parameters and can be written as follows, extending the work of Neslin et al. (2006) and Lemmens and Gupta (2013):

$$\pi_i = p_i[\gamma_i(V_{iP} - \delta_i)] + (1 - p_i)[-\varphi_i \delta_i] \tag{4}$$

where π_i is the profit of targeting customer i with a retention incentive, i.e. the extra revenue that company can generate by preventing customer i from becoming inactive in next period; p_i is the probability of being a would-be churner for customer i, extracted from the classification model; γ_i represents the probability of accepting an incentive offer by a would-be churner and staying active for the duration of the prediction period; δ_i is an incentive offered to customer i; φ_i represents the probability of accepting a retention incentive by a customer who would not become inactive in prediction period; and V_{iP} is an expected revenue for customer i in prediction period if customer is retained.

We assume that the retained customer will spend in the next (prediction) period the same amount of money as in previous (calibration) period but adjusted by the overall churn rate r:

$$V_{iP} = V_{iC} \times (1 - r) \tag{5}$$

where r is the population churn rate and V_{iC} is the revenue that customer i generates during the calibration period. We adopt a conservative position in defining the future spending V_{iP} assuming that the retained customers will stay active for only one period (i.e. prediction period). By having this assumption π_i gives us the minimum profit that retaining customer i can generate for the company. If the customer stays active for longer than one period, the generated profit π_i would be even higher.

As for retention incentive δ_i, we adopt a pricing incentive in the form of a discount. As Harrison-Walker and Neeley (2004) note, in B2B contexts, pricing incentive is a primary tool to increase customer retention. Therefore, the cost of retaining different customers will be different and will depend on their future spending in prediction period V_{iP} and discount parameter λ_i:

$$\delta_i = \lambda_i V_{iP}. \tag{6}$$

It's important to note that π_i can have positive and negative values. Positive value means that the benefit of bringing back the would-be-churner to the customer base is larger than the loss obtained (discount given). These customers are more likely to have high probability to churn p_i and/or large previous spending V_{iC}. Negative values of π_i mean that the loss is higher than the gain. Such customers are likely to have lower churn probabilities and/or lower spending patterns.

Once π_i for all customers is calculated, customers are ranked based on their individual expected targeting profit. Doing so, the cumulative profit for each percentile from top to bottom of the customer base can be calculated, using Eq. (3); and the optimum target size and the maximum profit of a retention campaign can be determined.

4. Empirical analysis

4.1. Data

The data used in this study constitutes the transactional records of 11,021 business customers of a major Australian online Fast Moving Consumer Goods (FMCG) retailer. Compared to consumers of the organization, business customers can be characterized by higher spending in each transaction and having more regular purchasing behavior.

The process of placing a B2B order is fairly simple. All a firm needs to do is to register on the focal company's website. Once registered, orders can be placed by adding the required items to the shopping cart and paying the due amount electronically. Apart from the commitments that placing each individual order brings for each of the parties (i.e. paying the price of the order and delivering the order), no long term agreements or contracts exist between the focal company and its business customers. This means a business customer can stop being a 'customer' at any time without notifying the focal company (i.e. making this a true non-contractual setting).

The products involved in transactions can be considered as routine products (Kotler & Keller, 2006, p. 219). Such products generally have low value and cost to the customers which somehow justifies why companies in this database have tended to have rather a transactional and non-contractual relationship with the focal company.

To date, several approaches have been introduced by academics in terms of splitting the timeline of the study for prediction purposes. What the majority of the approaches have in common are a calibration period where the model is trained and a prediction period where the trained model is used to make predictions about individual behavior. However, in terms of treating and locating these two periods on the data timeline to make the predictions, academics have yet to reach consensus. For instance, Pancras (2009) splits the prediction time period into two sub-periods with equal lengths. Predictions are made for each sub-period and then the two predictions are aggregated to obtain the prediction for the whole prediction period. Berry and Linoff (2004) adopt a different approach where the prediction period is assumed to be one integrated period but is defined with a time gap from the calibration period. On the other hand, several studies exist where the prediction period has been considered as one integrated period, which comes immediately after the calibration period (Buckinx & Van den Poel, 2005; Coussement & De Bock, 2013).

For analytical purposes in the current study the latter approach has been adopted due to its popularity among modelers. To this end, transactional records of business customers for the period September 1, 2011 through September 1, 2012 (i.e. one year) have been utilized. This time window has been split into two periods with equal lengths:

- From September 1, 2011 to March 2, 2012 (183 days): Calibration Period
- From March 3, 2012 to September 1, 2012 (183 days): Prediction Period.

A business customer is considered to be included in the study only if it has made its first purchase in the first 12 weeks of the calibration period.

4.2. Stage 1: constructing churn prediction models

At the first stage of our analysis we construct several churn prediction models and compare them on their ability to identify true churners.

Please cite this article as: Tamaddoni Jahromi, A., et al., Managing B2B customer churn, retention and profitability, *Industrial Marketing Management* (2014), http://dx.doi.org/10.1016/j.indmarman.2014.06.016

Table 1
Confusion matrix for a binary classifier.

Actual	Predicted		
	0	1	
0	True negative (*TN*)	False positive (*FP*)	Number of real non-churner ($TN + FP = N$)
1	False negative (*FN*)	True positive (*TP*)	Number of real churners ($FN + TP = P$)

4.2.1. Variables operationalization

4.2.1.1. Predictor variables. The raw data available for the analysis consists of transactional information at the individual (B2B) customer level. In this study, in order to keep the models in their simplest form, we sought to employ as few predictors as possible, but seek to maximize predictive power. This more focused and pragmatic approach will enable academics and practitioners to apply our method to other cases with the most basic predictors available in any databases. On this basis, recency and frequency of purchases, as well as the magnitude of changes in total spending of customers during the calibration period, have been chosen as predictor variables to construct the models. Recency, frequency, and monetary variables have been proven to play an undeniable role in predicting customer churn (Buckinx & Van den Poel, 2005; Coussement & De Bock, 2013). As (Wu & Chen, 2000) noted, the more recent a customer's purchase is, the more likely that the customer is active. In addition, according to Reinartz and Kumar (2000) frequency of purchases made by a customer can be a measure of defection likelihood in future. Previous studies also suggest that the monetary value of past purchases of a given customer can be an indicator to predict the future behavior (Schmittlein & Peterson, 1994).

For the current study, recency, frequency, and monetary variables have been extracted as follows:

1 — x: number of transactions observed during the calibration period;
2 — t_x: time of the last transaction in observation period ($0 \leq t_x \leq T$);
3 — T: known as observation period which varies across the customers; it is the time between the first purchase and end of calibration period;
4 — Δm: relative change in total spending of a customer in the second half of the observation period (m_2) when compared with the first half of the observation period (m_1), i.e. $\Delta m = (m_2 - m_1)/m_1$.

Fig. 1 illustrates the predictor variables of this study on the data timeline, for two customers with different behaviors.

4.2.1.2. Target variable. The target variable in the current study is 'churn' which is defined based on business customers' transactional history in both calibration and prediction periods. Therefore, a customer is defined as a 'churner' (coded as 1) when the company has been active in the calibration period (i.e. has at least one transaction in the calibration period) but has no activity (i.e. purchase) in prediction period. On the other hand, a non-churner (coded as 0) is defined as a customer who has been active in the calibration period and has made at least one purchase in the prediction period.

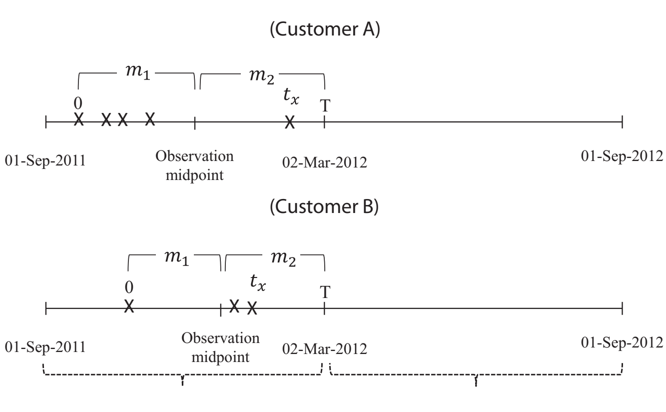

Fig. 1. Modeling timeline for two customers with different observation lengths.

4.2.2. Model construction

With the aim of constructing the classification models, the customer base is divided into a training set and a test set (70%/30%), such that 7714 customers are allocated to the training set and the remaining 3307 to the test set. Also, for the purpose of constructing the cost-sensitive model, the ratio of 1:3 (*FP*:*FN*) is employed.

Firstly, customers in the training set were used to construct the training model. Predictors and the target variable (churn) were extracted for all customers in the training set according to the time window presented in Fig. 1. By incorporating all of these variables into simple decision tree, cost-sensitive decision tree, logistic regression, and boosting technique, all four models were trained. Once the models were trained, their performance was assessed using the test set, such that the predictors and the actual churn status of the remaining 30% of customers in the dataset were extracted. At this point, and with the aim of testing the models' performance, the extracted predictor variables of customers in the test set should be incorporated in the models which have been constructed in the training phase. Using these predictors, the trained models would predict customer status (i.e. whether a customer is a churner or not) for customers in the test set. By comparing the customer status, predicted by models, with customers' actual status, extracted from prediction period, models' performance can be evaluated (see Fig. 2).

4.2.3. Results of models' comparison

All constructed churn prediction models, including simple and cost-sensitive decision tree, boosting model, and logistic regression are applied on the test set. Tables 2 to 5 illustrate the confusion matrix of

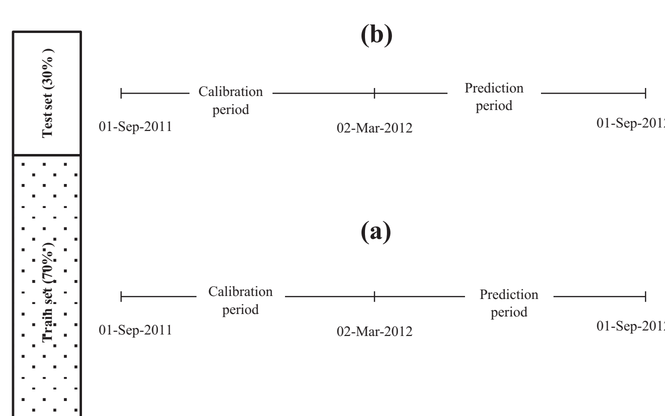

Fig. 2. (a) RFM variables from calibration period and 'churn' variable from prediction period are incorporated in the model to build the training model. (b) Model constructed on train set is utilized to predict the 'churn' variable in prediction period using the RFM variables from calibration period.

Please cite this article as: Tamaddoni Jahromi, A., et al., Managing B2B customer churn, retention and profitability, *Industrial Marketing Management* (2014), http://dx.doi.org/10.1016/j.indmarman.2014.06.016

each of constructed models. The actual fraction of churners in the data is 28% and non-churners 72%.

The low misclassification errors in confusion matrices indicate that all four models are sufficiently accurate. Although the accuracy measure extracted from confusion matrix has been traditionally used to evaluate the performance of classifiers, experts argue that for rare events such as customer churn, it might not be an optimal choice (Morrison, 1969). To elaborate more, consider a dataset with 10% churners. If a model classifies all customers as non-churners, the error rate of this model would be equal to 10% which is considered quite low. However, the reality is that the model has completely failed to classify the churners correctly. Another weakness of error rate measure is that this criterion does not take into account the value of the churn scores of individual customers, thereby overlooking potentially valuable information (Lemmens & Croux, 2006).

We observe from Tables 2 to 5 that in terms of error rates, the models are very similar, indicating that the error rate measure was not able to capture the differences between models' performance. The weak applicability of error rates is consistent with the findings of Lemmens and Croux (2006). Therefore, additional measures of models' performance such as AUC and cumulative lift are required to identify the better performing churn prediction model.

The results based on AUC measurement show more differences in models' performance by having the area under the ROC curve equal to 0.83, 0.85, 0.91, and 0.92 for cost-sensitive decision tree, simple decision tree, logistic regression, and boosting models, respectively. Therefore, the boosting and logistic regression models outperform both simple and cost-sensitive decision tree models based on AUC; while the boosting model performs slightly better than the logistic regression model. Note, that all four models considerably outperform the null model (random classifier) which by definition has AUC equal to 0.5.

A closer examination of the ROC curve in Fig. 3 also reveals that for both 'conservative' and 'liberal' discrimination thresholds (below and above 0.2 false positive rate respectively) the boosting and logistic regression model demonstrate higher *TP*/*FP* rates in comparison with both decision trees. It is also evident from Fig. 3 that while for more conservative discrimination thresholds simple decision tree performs better than the cost sensitive one, for more liberal thresholds it is the cost sensitive model that dominates (here the terminology has been adopted from Fawcett (2006)). It should also be mentioned that for both 'conservative' and 'liberal' discrimination thresholds the boosting model demonstrates the best performance followed by the logistic regression model.

In addition to the AUC measure, the models constructed in this study have also been evaluated using the cumulative lift chart. As depicted in Fig. 4, all four models perform considerably better than the random classifier. Similar to area under ROC curve, the cumulative lift chart in Fig. 4 also demonstrates the superior performance of boosting and logistic regression techniques in comparison to simple and cost-sensitive decision tree models.

Table 6 summarizes the performance of the constructed models based on cumulative lift measure for the top four deciles. For instance, if the company aims to target 30% of its customer base by using simple decision tree, cost-sensitive decision tree, logistic regression, or boosting model, it can capture 78%, 67%, 80%, and 81% of real churners, respectively; while by using no model (i.e. random classifier) only 30% of real churners can be captured. All in all, taking into account the results from the confusion matrices, AUC, and lift measures, we can conclude that the boosting technique performs the best in identifying churners. Therefore, we use the output of this technique to design the retention campaign in the second stage of analysis.

Table 3
Confusion matrix (cost-sensitive decision tree).

Actual	Predicted	
	0	1
0	57%	15%
1	3%	25%

We now consider which predictors contribute more to predicting churn using the boosting technique. Across 200 iterations, the time of last transaction t_x and number of transactions x were used the most often in predicting customer churn, namely 157 and 152 times respectively. This finding confirms that recency and frequency are indeed powerful predictors of churn. Specifically, higher number of purchases (frequency) and longer time since last order (recency) positively influence the likelihood of the customer to be a churner. Other two variables, duration of activity in observation period T and relative change in total spending Δm have less impact on predicting churn appearing only 122 and 120 times across 200 iterations.

4.3. Stage 2: developing profit maximizing retention campaigns

Being able to precisely identify churners is important but potentially worthless unless the results can guide managerial actions. Therefore, the results from the boosting technique from the first stage of churn management are used to compute expected individual profitability of targeting as in Eq. (4).

In order to compute π_i we need to define some parameters. The retention incentive $\delta_i = \lambda_i V_{iP}$ depends on the discount parameter λ_i. Without loss of generality we assume that all targeted customers receive 5% discount on their future purchases in the next period, i.e. $\lambda_i = \lambda = 0.05$. It is assumed that when an offer is made to would-be churners, the probability of accepting the offer by them and staying active for another period (i.e. prediction period) is the same across all customers and equals to 0.3 ($\gamma_i = \gamma = 0.3$). It is also assumed that if an offer is made to non-churners, the response probability of them is equal to 1 (i.e. $\varphi_i = \varphi = 1$). The fact that $\varphi = 1$ is justified by the nature of retention action when the discount is applied automatically to B2B customers. This is different from B2C contexts where the incentive is usually in the form of coupon which should be redeemed and, therefore, parameter φ would be less than 1. Besides, fixing φ to 1 allows us to be on conservative side in deriving retention campaign's profit. If by any reason in practice parameter φ becomes less than 1, the campaign profit will be increased.

After that the customers are ranked based on their individual retention campaign profits π_i and the cumulative profit as in Eq. (3) is computed.

Fig. 5 illustrates the predicted cumulative profit of a retention campaign (dashed line) and actual profit of the campaign (solid line) calculated using known status of the customer in the prediction period plotted against the percentiles of targeted customers ranked on their individual profits π_i. The predicted and actual profits are remarkably close, indicating

Table 2
Confusion matrix (simple decision tree).

Actual	Predicted	
	0	1
0	64%	8%
1	5%	23%

Table 4
Confusion matrix (logistic regression).

Actual	Predicted	
	0	1
0	65%	7%
1	5%	23%

Please cite this article as: Tamaddoni Jahromi, A., et al., Managing B2B customer churn, retention and profitability, *Industrial Marketing Management* (2014), http://dx.doi.org/10.1016/j.indmarman.2014.06.016

Table 5
Confusion matrix (boosting).

Actual	Predicted	
	0	1
0	65%	7%
1	7%	21%

excellent predictive power of the proposed approach. Even though the predicted profit is slightly overestimated, this does not detract from our case as we are interested more in the customers who should be targeted and who will maximize the total profit. Therefore, when $\gamma = 0.3$, the expected maximum profit of the campaign is equal to \$94,437 and is achieved when 43% of the customer base is targeted with the incentive. This is close to the realized profit of \$84,832 when the actual churn status of customers in the prediction period is used instead of probabilities. As was mentioned before, we have chosen quite conservative values of the parameter $\gamma = 0.3$. When 40% or 50% of would-be-churners are retained (i.e. $\gamma = 0.4$ and $\gamma = 0.5$), the expected profit increases to \$131,236 and \$168,673 respectively.

We compare these results with alternative methods used in practice when the customers are ranked based on probability to churn rather than individual profitability and fixed decile of customers are targeted. For instance, when customers are ranked based on probability to churn and 10% of such customers are targeted the actual profit is \$25,894 and when 20% is targeted the profit is \$41,155, which is considerably lower than what the proposed method is able to generate. Also, note that targeting all customers with incentives may even be unprofitable for the company.

5. Discussion

5.1. Conclusions

The application of intelligent systems to marketing issues has recently received a great deal of attention from academics (see the 2013 *Industrial Marketing Management* special issue on intelligent systems in industrial marketing). No doubt, these and related topics will continue to attract

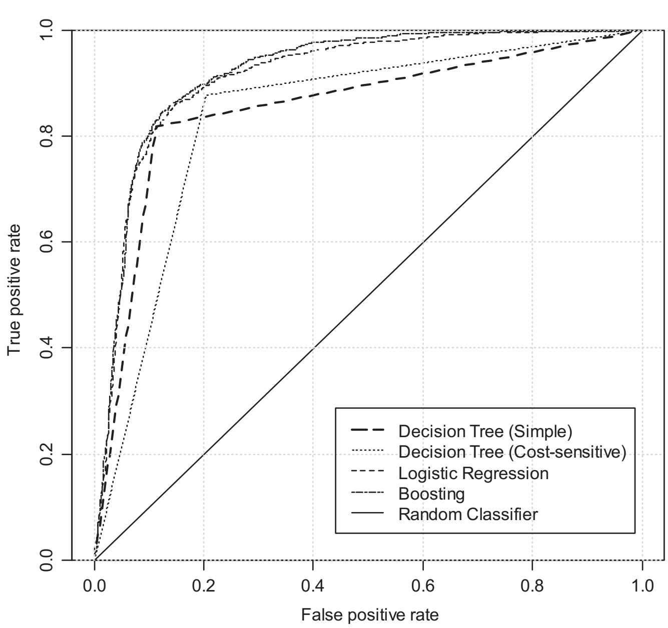

Fig. 3. ROC curve of the constructed models.

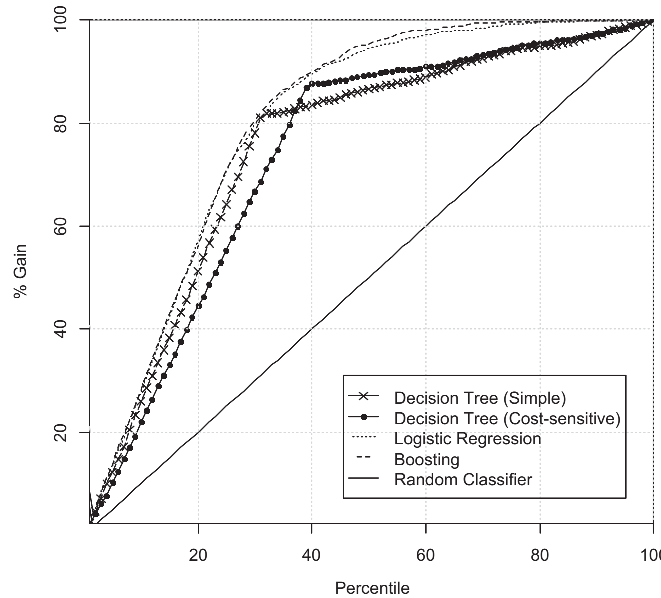

Fig. 4. Cumulative lift charts of constructed models against random sampling.

attention in coming years due to the inherent potential for information-aided possesses to drive change in B2B-marketing (Wiersema, 2013).

According to Martínez-López and Casillas (2013), applications of intelligent systems in B2B contexts exist on a broad continuum from pricing strategies to communication decisions and product development. Of all the roles that such systems can play in solving industrial marketing issues, managing customers' relationship would surely be a significant one. It has been well established in the marketing literature that, as a more profitable marketing strategy, firms should focus on establishing long term relationships with their customers by adopting appropriate retention approaches, rather than seeking to acquire new customers. By adopting this strategy, companies can benefit from lower servicing costs coupled with higher revenues.

In this regard, the current study highlights the importance of churn prediction models in retention approaches and applies different data mining classification techniques to identify customers who are more likely to churn in the future. Once the models are constructed their performance is evaluated based on AUC and cumulative lift measures.

Additionally, although several studies in both B2C and B2B contexts directly or indirectly investigate churn, only a few have quantitatively considered the profitability of retention campaigns. Therefore, we extend the work of Neslin et al. (2006) and Lemmens and Gupta (2013) to develop profit maximizing B2B retention strategies.

For this purpose, the transactional data of business customers from a major Australian FMCG retailer has been utilized to construct churn prediction models. Due to the nature of non-contractual B2B data, where 'churn' cannot be considered as a permanent phenomenon, this study focuses on two half year periods (calibration and prediction) and aims to predict (temporary) inactivity of customers in the prediction period.

Table 6
Lift measures of constructed models for the top four deciles.

Model	% Gain for percentile			
	10	20	30	40
Decision tree (simple)	26	51	78	84
Decision tree (cost-sensitive)	22	44	67	88
Logistic regression	28	57	80	90
Boosting	28	57	81	91

Please cite this article as: Tamaddoni Jahromi, A., et al., Managing B2B customer churn, retention and profitability, *Industrial Marketing Management* (2014), http://dx.doi.org/10.1016/j.indmarman.2014.06.016

We compare both simple and cost-sensitive decision trees as single learner data mining techniques, boosting as an ensemble learner, and logistic regression as a benchmark technique to predict churn. The results of our analysis show that on the basis of both AUC and cumulative lift measures, boosting technique, as an ensemble classifier, considerably outperforms both simple and cost-sensitive decision tree models in terms of churn prediction accuracy while only marginally outperforming logistic regression. This can stem from boosting's sensitivity to noise in the dataset (Freund & Schapire, 1996). Schapire, Freund, Bartlett, and Lee (1998) argue that margin (i.e. the difference between number of correct and incorrect predictions for an observation) of observations can be a measure for the effectiveness of a classification method (e.g. boosting). As Optiz and Maclin (1999) note in the case of boosting, the focus of later classifiers is on increasing the margins for the observations with poor current margins. As long as the overall accuracy of the classifier is not affected negatively, this can be an effective approach. However, when the data is noisy, overemphasizing on misclassified observations may boost the margins of noisy observations which can mislead the overall classification (for more details on how boosting algorithms works please refer to Appendix A).

Overall, the models' outcomes indicate that the application of data mining-based models such as those proposed in this paper to predict churn holds considerable promise in the identification of true churners.

Furthermore, according to our framework for calculating retention profit, the company can add $84,832 when implementing a retention campaign. It should also be noted that due to conservative assumptions made about the probability of accepting an offer by would-be churners/non-churners, as well as our assumption about the duration for which a retained customer stays active, our calculated profit is conservative and may in fact be even higher.

We would like to highlight that the nature of the products involved (i.e. routine products) as well as the nature of the focal company's business (i.e. FMCG retailer) has enabled the focal company to handle over 11,000 B2B customers which is not a common case in B2B markets. Nevertheless, the proposed approach offers valuable insights for the cases where information is available in both quantity and quality.

5.2. Managerial implications

Adopting a targeted approach to handle customer churn in B2B contexts, companies should identify customers who are at churn risk and then target such potential churners with tailored incentives to persuade them to stay. However, such strategies can incur several losses if firms are not able to accurately distinguish churners from non-churners. Basically, when an incentive which can be sent to a real churner, is sent to a non-churner, it means either that a real churner has not received an incentive to be persuaded to stay (increase in his/her risk of churn) or a non-churner has received an incentive which s/he was not supposed to receive (loss of marketing budget for the company). Thus, companies seeking to adopt such retention strategy should be equipped with models that can accurately identify customers who are likely to churn in a given future time period. This becomes even more critical in B2B contexts where the average value of customers is higher than for customers in B2C contexts.

In this study, a data mining-based model building approach has been adopted to develop churn prediction models. Such models can help B2B companies to more accurately identify churners and thereby develop more efficient, effective and targeted retention campaigns. In addition, a novel framework has been proposed to compute the profit of targeting customers with incentives at the individual level, while accounting for heterogeneity of incentives offered, value of customers, and the probability of being a would-be churner. The framework can give a preview of the possible benefits or losses that the campaign can deliver to the company. Likewise, it can help the company to send out the optimum number of incentive offers and reap maximum profit from each campaign.

Overall, the approach proposed here to predict customer churn in a B2B context and to maximize profitability of retention campaigns, can help companies not only to reduce costs by sending fewer incentives to non-churners, but also to increase revenues by contacting potential churners with the highest profit potential. This means that more valuable churners are more likely to change their mind and stay loyal to the focal company, leading to an increase in retention rate and delivering more revenue/profit. Needless to say, retained customers are less costly to serve and, more importantly, a company with a high retention rate will spend less on customer acquisition.

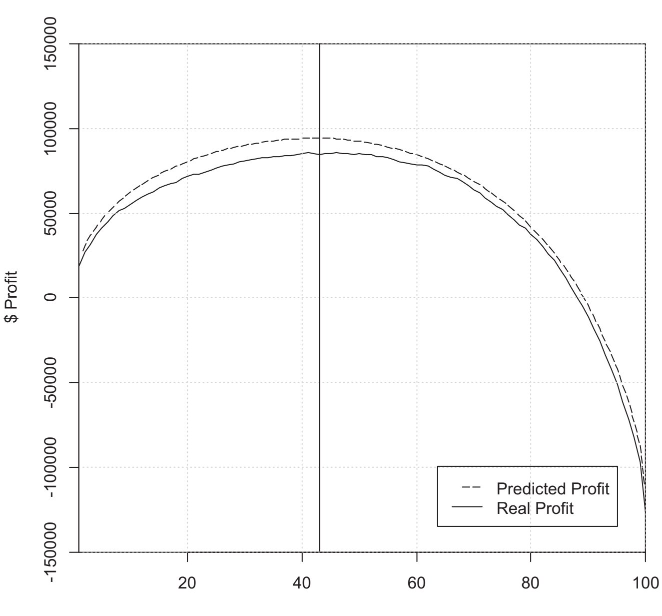

Fig. 5. Predicted vs. actual profit of the retention campaign for different percentiles.

5.3. Limitations and directions for future research

The authors aimed to limit the number of factors used to predict customer churn (recency, frequency, and monetary value). The motivation was to promote the use of data mining techniques among academics and practitioners even with few predictors in the database. However, further studies may consider the predictive power of other factors such as variables related to product type and/or customer profiles.

In addition, the authors use quite conservative levels of probability of accepting an offer by would-be churners (γ) and non-churners (φ) leading to equally conservative estimates of retention profitability. Knowing or being able to estimate the redemption rates would produce more objective estimates of future retention campaign profits.

Appendix A

In this section a brief explanation of adaptive boosting algorithm is presented. The following discussion is based on Tan et al. (2006) and Han et al. (2011). Keen readers are encouraged to refer to the above sources for deeper understanding.

As was mentioned in Section 3.2, boosting is a newly introduced classification technique from the family of ensemble methods. In boosting a series of k classifiers are learned in an iterative procedure. This is basically to update the weights of training observations so that the base classifiers can pay more attention to the observations which are hard to classify. The extracted weights then can be used (1) to draw a set of bootstrap samples from the original data, and (2) to train a model which tends towards observations with higher weights.

What happens in the algorithm is that, in each iteration, the algorithm draws a sample with replacement from all observations to train the model. On this basis, each observation's chance of being selected is determined by its weight. Initially, since all N observations are assigned equal weights (i.e. $1/N$), they are equally likely to be chosen in the training sample.

Once a classifier is constructed, its error rate is calculated based on the number of observations which have been misclassified by the classifier. To compute the error rate of a classifier we sum the weights of each of the observations in the sample which has been misclassified by the classifier (see Eq. (A.1)).

$$\epsilon_i = \sum_{j=1}^{N} w_j \times err\left(X_j\right) \quad \text{(A.1)}$$

here $err(X_j)$ is a binary variable to represent the misclassification error of observation X_j. If the observation was classified correctly then $err(X_j) = 0$, and if the observation is misclassified $err(X_j) = 1$.

The computed error rate is then employed to calculate the measure of importance of classifier (α_i) as a part of the ensemble (see Eq. (A.2)). Accordingly, the importance of a single classifier increases as its error rate decreases.

$$\alpha_i = \frac{1}{2} \ln\left(\frac{1-\epsilon_i}{\epsilon_i}\right). \quad \text{(A.2)}$$

Once the α_i and the error rate are obtained they can be used to calculate the updated weight of the observations according to Eq. (A.3) while $w_i^{(j)}$ is the weight assigned to an observation during the jth boosting round. In Eq. (A.3), Z_j is the normalization factor to make sure that their sum remains the same as it was before (i.e. 1).

$$w_i^{(j+1)} = \frac{w_i^{(j)}}{Z_j} \times \begin{cases} exp^{-\alpha_j} & \text{if the observation is classified correctly} \\ exp^{\alpha_j} & \text{if the observation is classified incorrectly} \end{cases}. \quad \text{(A.3)}$$

What in fact Eq. (A.3) does is that it increases the weight of the misclassified observations and decreases the weight of those that have been correctly classified. After updating the observations' weight, the classifier of the next boosting round is constructed noting that weight of some observations (the one which have been misclassified by the previous classifier) is multiplied to have more representation in the sample and in the algorithm. Therefore, the new classifier is more likely to correctly classify observations with higher weights.

Once all k classifiers are constructed, the ensemble of classifiers is used to predict the class label of a given observation. Doing so, the prediction of each classifier C_j is weighted according to α_j. The lower a classifier's error rate, the higher power its prediction has. In other words, this approach diminishes the effect of models with poor accuracy. For each observation X and for each class, c, we sum the weights of each classifier that assigned class c to the observation X. Regarding this, the class prediction for observation X is the class with the highest sum.

References

Ainslie, A., & Pitt, L. (1992). Customer retention analyses: An application of descriptive and inferential statistics in database marketing. *Journal of Direct Marketing*, *6*(3), 31–43.

Athanassopoulos, A.D. (2000). Customer satisfaction cues to support market segmentation and explain switching behavior. *Journal of Business Research*, *47*(3), 191–207.

Berry, M. J. A., & Linoff, G. S. (2004). *Data mining techniques for marketing, sales, and customer relationship management* (2nd ed.). Indianapolis, Indiana: Wiley Publishing, Inc.

Bhattacharya, C. B. (1998). When customers are members: Customer retention in paid membership contexts. *Journal of the Academy of Marketing Science*, *26*(1), 31–44.

Boles, J. S., Barksdale, H. C., & Johnson, J. T. (1997). Business relationships: An examination of the effects of buyer–salesperson relationships on customer retention and willingness to refer and recommend. *Journal of Business & Industrial Marketing*, *12*(3/4), 253–264.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145–1159.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. New York: Chapman and Hall.

Buckinx, W., Moons, E., Van Den Poel, D., & Wets, G. (2004). Customer-adapted coupon targeting using feature selection. *Expert Systems with Applications*, *26*(4), 509–518.

Buckinx, W., & Van den Poel, D. (2005). Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, *164*(1), 252–268.

Burez, J., & Van den Poel, D. (2007). CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, *32*(2), 277.

Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, *36*(3), 4626–4636.

Chen, Y. -L., Hsu, C. -L., & Chou, S. -C. (2003). Constructing a multi-valued and multi-labeled decision tree. *Expert Systems with Applications*, *25*(2), 199–209.

Colgate, M. R., & Danaher, P. J. (2000). Implementing a customer relationship strategy: The asymmetric impact of poor versus excellent execution. *Journal of the Academy of Marketing Science*, *28*(3), 375–387.

Cooper, B.L., Watson, H. J., Wixom, B. H., & Goodhue, D. L. (2000). Data warehousing supports corporate strategy at first American corporation. *MIS Quarterly*, *24*(4), 547–567.

Coussement, K., & De Bock, K. W. (2013). Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. *Journal of Business Research*, *66*(9), 1629–1636.

Coussement, K., & Van den Poel, D. (2008a). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, *34*(1), 313–327.

Coussement, K., & Van den Poel, D. (2008b). Integrating the voice of customers through call center emails into a decision support system for churn prediction. *Information & Management*, *45*(3), 164–174.

D'Haen, J., & Van den Poel, D. (2013). Model-supported business-to-business prospect prediction based on an iterative customer acquisition framework. *Industrial Marketing Management*, *42*(4), 544–551.

De Bock, K. W., & Van den Poel, D. (2011). An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications*, *38*(10), 12293–12301.

Duda, R. O., Hart, P. E., & Stork, D.G. (2001). *Pattern classification* (2nd ed.). New York, NY: John Wiley & Sons.

Earl, M. (2001). Knowledge management strategies: Toward a taxonomy. *Journal of Management Information Systems*, *18*(1), 215–233.

Eriksson, K., & Vaghult, A. L. (2000). Customer retention, purchasing behavior and relationship substance in professional services. *Industrial Marketing Management*, *29*(4), 363–372.

Fader, P.S., & Hardie, B. G. S. (2009). Probability models for customer-base analysis. *Journal of Interactive Marketing*, *23*(1), 61–69.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874.

Freund, Y. (2001). An adaptive version of the boost by majority algorithm. *Machine Learning*, *43*(3), 293–318.

Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. *Paper presented at the 13th International Conference on Machine Learning, Bari, Italy*.

Friedman, J. H., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, *28*(2), 337–374.

Ganesh, J., Arnold, M. J., & Reynolds, K. E. (2000). Understanding the customer base of service providers: An examination of the differences between switchers and stayers. *Journal of Marketing*, *64*(3), 65–87.

Grönroos, C. (1990). *Service management and marketing: Managing the moments of truth in service competition*. Lexington, MA: Lexington Books.

Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, *34*(10), 2902–2917.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

Harrison-Walker, L. J., & Neeley, S. E. (2004). Customer relationship building on the internet in B2B marketing: A proposed typology. *Journal of Marketing Theory and Practice*, *12*(1), 19–35.

Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. *Expert Systems with Applications*, *39*(1), 1414–1425.

Hung, S. Y., Yen, D. C., & Wang, H. Y. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications*, *31*, 515–524.

Jackson, T. W. (2005). CRM: From 'art to science'. *Journal of Database Marketing & Customer Strategy Management*, *13*(1), 76–92.

Jamal, Z., & Bucklin, R. E. (2006). Improving the diagnosis and prediction of customer churn: A heterogeneous hazard modeling approach. *Journal of Interactive Marketing*, *20*(3–4), 16–29.

Jamal, Z., & Zhang, A. (2009). 2008 DMEF customer lifetime value modeling (Task 2). *Journal of Interactive Marketing*, *23*(3), 279–283.

Kalwani, M. U., & Narayandas, N. (1995). Long-term manufacturer–supplier relationships: Do they pay off for supplier firms? *Journal of Marketing*, *59*(1), 1–16.

Kamakura, W., Mela, C., Ansari, A., Bodapati, A., Fader, P., Iyengar, R., et al. (2005). Choice models and customer relationship management. *Marketing Letters*, *16*(3), 279–291.

Keaveney, S. M., & Parthasarathy, M. (2001). Customer switching behavior in online services: An exploratory study of the role of selected attitudinal, behavioral, and demographic factors. *Journal of the Academy of Marketing Science*, *29*(4), 374–390.

Kim, J. K., Song, H. S., Kim, T. S., & Kim, H. K. (2005). Detecting the change of customer behavior based on decision tree analysis. *Expert Systems with Applications*, *22*(4), 193–205.

Kotler, P., & Keller, K. (2006). *Marketing management* (12th ed.). New Jersey: Pearson Prentice Hall.

Kumar, V., & Reinartz, W. (2012). *Customer relationship management: Concept, strategy, and tools*. Berlin Heidelberg: Springer.

Please cite this article as: Tamaddoni Jahromi, A., et al., Managing B2B customer churn, retention and profitability, *Industrial Marketing Management* (2014), http://dx.doi.org/10.1016/j.indmarman.2014.06.016

Lam, S. Y., Shankar, V., Erramilli, M. K., & Murthy, B. (2004). Customer value, satisfaction, loyalty, and switching costs: An illustration from a business-to-business service context. *Journal of the Academy of Marketing Science*, *32*(3), 293–311.

Lee, S. J., & Siau, K. (2001). A review of data mining techniques. *Industrial Management and Data Systems*, *101*(1), 41–46.

Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, *43*(2), 276–286.

Lemmens, A., & Gupta, S. (2013). Managing churn to maximize profits. *Working Paper* (doi: http://www.hbs.edu/faculty/Publication%20Files/14-020_3553a2f4-8c7b-44e6-9711-f75dd56f624e.pdf).

Martínez-López, F. J., & Casillas, J. (2013). Artificial intelligence-based systems applied in industrial marketing: An historical overview, current and future insights. *Industrial Marketing Management*, *42*, 489–495.

Miguéis, V. L., Van den Poel, D., Camanho, A. S., & Falcão e Cunha, J. (2012). Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert Systems with Applications*, *39*(12), 11250–11256.

Mogelefsky, D. (2000). Smooth operators. *Incentive*, *174*(5), 104.

Morrison, D.G. (1969). On the interpretation of discriminant analysis. *Journal of Marketing Research*, *6*(2), 156–163.

Neslin, S. A., Gupta, S., Kamakura, W., Junxiang, L., & Mason, C. H. (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, *43*(2), 204–211.

Olafsson, S., Li, X., & Wu, S. (2008). Operations research and data mining. *European Journal of Operational Research*, *187*(3), 1429–1448.

Optiz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, *11*, 169–198.

Orriols-Puig, A., Martínez-López, F. J., Casillas, J., & Lee, N. (2013). Unsupervised KDD to creatively support managers' decision making with fuzzy association rules: A distribution channel application. *Industrial Marketing Management*, *42*(4), 532–543.

Pancras, J. (2009). A method to predict aggregate customer equity using activeness thresholds — 2008 DMEF customer lifetime value modeling competition (Task 1). *Journal of Interactive Marketing*, *23*(3), 276–278.

Rauyruen, P., & Miller, K. E. (2007). Relationship quality as a predictor of B2B customer loyalty. *Journal of Business Research*, *60*(1), 21–31.

Reinartz, W. J., & Kumar, V. (2000). On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *Journal of Marketing*, *64*(4), 17–35.

Richards, K. A., & Jones, E. (2008). Customer relationship management: Finding value drivers. *Industrial Marketing Management*, *37*, 120–130.

Risselada, H., Verhoef, P. C., & Bijmolt, T. H. A. (2010). Staying power of churn prediction models. *Journal of Interactive Marketing*, *24*(3), 198–208.

Rust, R. T., Kumar, V., & Venkatesan, R. (2011). Will the frog change into a prince? Predicting future customer profitability. *International Journal of Research in Marketing*, *28*(4), 281–294.

Rust, R. T., Lemon, K. N., & Zeithaml, V. A. (2004). Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing*, *68*(1), 109–127.

Schapire, R. E., Freund, Y., Bartlett, P., & Lee, W. S. (1998). Boosting the margin: A new explanation for the effectiveness of voting methods. *The Annals of Statistics*, *26*(5), 1651–1686.

Schmittlein, D. C., & Morrison, D.G. (1985). Is the customer still active? *The American Statistician*, *39*(4), 291–295.

Schmittlein, D. C., Morrison, D.G., & Colombo, R. (1987). Counting your customers: Who are they and what will they do next? *Management Science*, *33*(1), 1–24.

Schmittlein, D. C., & Peterson, R. A. (1994). Customer base analysis: An industrial purchase process application. *Marketing Science*, *13*(1), 41–67.

Shaffer, G., & Zhang, Z. J. (2002). Competitive one-to-one promotions. *Management Science*, *48*(9), 1143–1160.

Sher, P. J., & Lee, V. C. (2004). Information technology as a facilitator for enhancing dynamic capabilities through knowledge management. *Information & Management*, *41*(8), 933–945.

Srivastava, R. K., Shervani, T. A., & Fahey, L. (1999). Marketing, business processes, and shareholder value: An organizationally embedded view of marketing activities and the discipline of marketing. *Journal of Marketing*, *63*, 168–179 (Special Issue).

Stevens, R. P. (2005). *B-to-B customer retention: Seven strategies for keeping your customers.*

Tan, P. -N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining.* Boston, MA: Pearson Education Inc.

Tsai, C. -F., & Lu, Y. -H. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, *36*(10), 12547–12553.

Van den Poel, D., & Larivière, B. (2004). Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, *157*(1), 196–217.

Venkatesan, R., & Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing*, *68*(4), 106–125.

Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, *218*(1), 211–229.

Waxer, C. (2011). *Loyalty and the B2B customer.* Loyalty360 — The Loyalty Marketer's Association.

Wei, C. -P., & Chiu, I. -T. (2002). Turning telecommunications call details to churn prediction: A data mining approach. *Expert Systems with Applications*, *23*(2), 103–112.

Weiss, G. M. (2004). Mining with rarity: A unifying framework. *Sigkdd Explorations*, *6*(1), 7–19.

Wierenga, B. (2010). Marketing and artificial intelligence: Great opportunities, reluctant partners. In J. Casillas, & F. Martínez-López (Eds.), *Marketing intelligent systems using soft computing, Vol. 258.* (pp. 1–8). Berlin Heidelberg: Springer.

Wiersema, F. (2013). The B2B Agenda: The current state of B2B marketing and a look ahead. *Industrial Marketing Management*, *42*(4), 470–488.

Wu, C., & Chen, H. -L. (2000). Counting your customers: Compounding customer's in-store decisions, interpurchase time and repurchasing behavior. *European Journal of Operational Research*, *127*(1), 109–119.

Wübben, M. (2008). *Analytical CRM: Developing and maintaining profitable customer relationships in non-contractual settings.* Springer.

Xie, Y., Li, X., Ngai, E., & Ying, W. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, *36*(3), 5445–5449.

Yu, X., Guo, S., Guo, J., & Huang, X. (2011). An extended support vector machine forecasting framework for customer churn in e-commerce. *Expert Systems with Applications*, *38*(3), 1425–1430.

Ali Tamaddoni Jahromi is a PhD candidate in marketing modeling in the Department of Marketing at Monash University, Australia. Tamaddoni Jahromi's areas of research interest include analytical CRM, churn management, stochastic buyer behavior models, and data mining. His current research focuses on maximizing accuracy and profitability of customer churn models.

Stanislav Stakhovych is a Lecturer (Assistant Professor) in the Department of Marketing at Monash University. His research interests are in the area of retailing, customer base analysis, spatial modeling, and Bayesian methods. His work has been published in *Papers in Regional Science* and *Multivariate Behavioral Research*.

Mike Ewing is Alfred Deakin Professor and Dean of Business and Law at Deakin University. His research interests include marketing communications, marketing strategy and services marketing/CRM. Mike has published more than 115 journal articles, including in the *Journal of the Academy of Marketing Science*, *Information Systems Research*, the *International Journal of Research in Marketing* and *Social Science & Medicine*.

Please cite this article as: Tamaddoni Jahromi, A., et al., Managing B2B customer churn, retention and profitability, *Industrial Marketing Management* (2014), http://dx.doi.org/10.1016/j.indmarman.2014.06.016