Expert Systems with Applications 40 (2013) 6225–6232

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# Customer attrition in retailing: An application of Multivariate Adaptive Regression Splines

V.L. Miguéis *, Ana Camanho, João Falcão e Cunha

*Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal*

ARTICLE INFO

*Keywords:*
Marketing
Customer relationship management
Churn analysis
Retailing
Classification
Logistic regression
Multivariate Adaptive Regression Splines

ABSTRACT

The profit resulting from customer relationship is essential to ensure companies viability, so an improvement in customer retention is crucial for competitiveness. As such, companies have recognized the importance of customer centered strategies and consequently customer relationship management (CRM) is often at the core of their strategic plans. In this context, a priori knowledge about the risk of a given customer to mitigate or even end the relationship with the provider is valuable information that allows companies to take preventive measures to avoid defection. This paper proposes a model to predict partial defection, using two classification techniques: Logistic regression and Multivariate Adaptive Regression Splines (MARS). The main objective is to compare the performance of MARS with Logistic regression in modeling customer attrition. This paper considers the general form of Logistic regression and Logistic regression combined with a wrapper feature selection approach, such as stepwise approach. The empirical results showed that MARS performs better than Logistic regression when variable selection procedures are not used. However, MARS loses its superiority when Logistic regression is conducted with stepwise feature selection.



## 1. Introduction

Due to globalization, the retail sector is becoming more competitive, triggering significant changes in the way companies do business. The relationship between companies and customers has become a critical factor of companies' strategy. In sectors characterized by high competition, it is difficult to attract new customers, and consequently companies have to intensify efforts to retain current consumers. Therefore, companies are moving away from past strategies centered on selling products and services without giving due attention to the customers who bought these products and services. An increasing number of companies are shifting from a transactional to a relational approach (Eriksson & Vaghult, 2000), and adopting customer relationship management practices to promote closer relationships with customers (Ranjan & Bhatnagar, 2008; Greenberg, 2001). Companies wishing to be at the leading edge have to continuously guarantee a good business relationship with customers.

CRM can be viewed as the development of a customer-oriented culture by which a strategy is created for acquiring, retaining and enhancing the profitability of customers, enabled by the use of IT. Hence, a successful implementation of CRM results in an increased customer retention and loyalty, higher customer profitability, creation of value for the customer, customization of products and services, and higher quality products and services. Therefore, it results in mutual benefits for both the organization and the customers (Rababah, Mohd, & Ibrahim, 2010).

The benefits of CRM, combined with the explosion of data collection and storage procedures observed in recent decades, promoted an increase in the adoption of customer relationship management practices. In recent decades, companies generate and handle increasingly larger amounts of data. Companies' databases are able to store a big amount of customer-related data. For each customer many data objects are collected, allowing the analysis of the complete customers purchasing history. Retailers state that one of the basis of CRM strategies is the existence of a loyalty program, supported by a loyalty card, that allows to gather customer data.

Following Swift (2000); Parvatiyar and Sheth (2002); Kracklauer, Mills, and Seifert (2004); Ngai, Xiu, and Chau (2009) analytical CRM can be categorized in four dimensions: customer identification, customer attraction, customer development and customer retention. These four dimensions can be seen as a closed cycle of the customers management system (Au, Chan, & Yao, 2003; Kracklauer et al., 2004). In today's competitive environment, customer retention is receiving particular attention from companies, as customer life cycles are becoming shorter than in the past. Some customers reveal switching behavior in their purchases (Peterson, 1995) and others split their purchases between several

* Corresponding author. Tel.: +351 225082132; fax: +351 225081538.
*E-mail address:* vera.migueis@fe.up.pt (V.L. Miguéis).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.05.069

competitors (Dwyer, 1989). Particularly in non-contractual settings, such as retail sector, this tendency is of utmost relevance, since customers do not have to inform companies about their intention to leave the company, and experience very little switching cost. According to Kracklauer et al. (2004), customer satisfaction is the main issue regarding customers retention. Customer satisfaction can be defined as the comparison of customers' expectations (resulting from personal standard, image of the company, knowledge of alternatives, etc.) with the perceptions (resulting from actual experience, subjective impression of product performance, appropriateness of the product or service, etc). The customer's perception of the value offered by the company leads to sustained customer retention. Moreover, a high quality shopping experience leads to a positive emotional feeling, which enables the company to achieve the desired customer loyalty.

Several authors have highlighted the advantages of customer retention (e.g., Larivire & Van den Poel, 2005; East, Hammond, & Gendall, 2006; Martin, Clark, Peck, & Payne, 1995; Strandvik & Liljander, 1994). Reichheld and Sasser (1990) provides evidence about those advantages, that are based on the strong connection between customer retention and companies' profit. This study acknowledges that long time customers spend more over time, become more loyal and then promote the word-of-mouth. They are also less price-sensitive and the operating cost to serve them declines over time. Furthermore, the literature suggests that retaining customers costs less than attracting new customers (Dick & Basu, 1994; Saren & Tzokas, 1998). In this context, a small improvement in customer retention can mean a significant increase in profit (Reichheld & Sasser, 1990; Larivire & Van den Poel, 2005).

A decisive aspect for improving companies' retention involves the prediction of customer attrition (or churn). A customer attrition event expresses the customer decision to terminate the relationship with a provider, either because the customer does not need its products or services anymore or because the customer wants to switch to another product/service provider. Despite the relevance of this topic, churn analysis in retailing can still be considered incipient.

Effective customer churn management practices involve the construction of accurate churn prediction models. Among previous studies in the literature, statistical and data mining techniques have been used to build such models. Data mining is defined as the process of exploring and analyzing huge datasets, in order to find patterns and rules that can be important to solve a problem (Berry & Linoff, 1999). Following the framework proposed by Ngai et al. (2009), classification techniques are those more frequently used to assist customers churn prediction. These techniques include Logistic regression (e.g., Huang, Kechadi, & Buckley, 2012; Migueis, Van den Poel, Camanho, & Falcao e Cunha, 2012a), Decision tress (e.g. Nie, Rowe, Zhang, Tian, & Shi, 2011; Li & Deng, 2012), Random forests (e.g., Ballings & Van den Poel, 2012; Migueis, Van den Poel, Camanho, & Falcao e Cunha, 2012b), Neural networks (e.g., Tsai & Lu, 2009; Song, Yang, Wu, Wang, & Tang, 2006), support vector machines (e.g., Yu, Guo, Guo, & Huang, 2011; Chen, Fan, & Sun, 2012) and MARS (e.g., Lee, Chiu, Chou, & Lu, 2006b; Chen, Ma, & Ma, 2009).

This paper aims at investigating the prediction performance of MARS, compared with Logistic regression (typically considered a benchmarking technique) in the context of customer attrition prediction. Additionally, this paper aims to evaluate the performance of MARS in comparison with Logistic regression combined with stepwise feature selection.

The remainder of this paper is organized as follows. Section 2 introduces the literature on customer attrition prediction models. Section 3 introduces the methodology followed in this paper, namely the explanatory variables, the classification techniques and the performance evaluation criteria used. Section 4 presents the application, i.e., the company used as case study and the discussion of the results obtained. The paper finishes with the conclusion.

## 2. Related work

The literature in churn prediction is flourishing in several domains, such as retail (e.g., Buckinx & Van den Poel, 2005; Baesens et al., 2004), telecommunications (e.g., Verbeke, Dejaeger, Martens, Hur, & Baesens, 2012; Idris, Rizwan, & Khan, 2012) and banking (e.g., Nie et al., 2011; Lin, Tzeng, & Chin, 2011). Despite the large amount of research done on this topic, only a few studies focused on retail sector. To the best of our knowledge, only Buckinx and Van den Poel (2005), Migueis et al. (2012a, 2012b) have predicted customer churn in this context. Buckinx and Van den Poel (2005) used three classification techniques: Logistic regression, automatic relevance determination and random forests, to predict partial attrition by behaviorally loyal customers. Migueis et al. (2012a) proposed an attrition model that includes as predictors variables that represent the succession of the categories of the first products purchased. These are a proxy of the state of trust and demand maturity of a customer towards a company in grocery retailing. This study used Logistic regression as the classification technique. Migueis et al. (2012b) constructed a model to predict attrition that includes as a predictor the similarity of the sequences of the first products purchased with the sequences observed for those customers who partially left and who did not leave the company. The sequence of first purchase events is modeled by means of markov-for-discrimination. Two classification techniques were used in this empirical study: Logistic regression and Random forests.

The literature suggests several classifications of the attrition event. Depending on the agent who cancels the relationship, customer attrition can be involuntary and voluntary (Desai, 2007; Hadden, Tiwari, Roy, & Ruta, 2007). Involuntary attrition happens when the provider decides to terminate the relationship, due to missed payments, bad debts, or others. Voluntary attrition happens when customers choose to switch the provider or terminate their use of the service or product. Furthermore, Berry and Linoff (2004) introduces another class of attrition, i.e., expected attrition. Expected attrition occurs when the customer is no longer in the target market for a product or service. In a financial context, Burez and Van den Poel (2008) proposes three different concepts of attrition. Involuntary attrition concerns customers who died or moved abroad during the period of analysis, financial attrition corresponds to customers who stopped paying the service due to financial concerns, and commercial attrition corresponds to customers who canceled the service because they do not want it anymore. Pettersson (2004) introduces another classification of attrition. A total churn happens when a customer completely stops buying the service or goods provided, whereas partial churn happens when a customer cancels at least one product/service, but still buys other products/services from the company. This is the approach followed by Buckinx and Van den Poel (2005) and Migueis et al. (2012a, 2012b) to analyze customer attrition in retailing. Moreover, Pettersson (2004) distinguishes open attrition from hidden attrition. Open attrition occurs when the agreement between the provider and the customer is canceled and hence it is possible to detect the attrition in the customer database. Conversely, hidden attrition occurs when there is no specific event that enables to identify in the database the customers who left.

The distinction between open and hidden attrition is linked to the differences between contractual and non-contractual settings. Usually in contractual settings, a customer leaves the company after having contacted the service or product provider in order to cancel the contract. Therefore, companies are able to distinguish between active and inactive customers. In opposition, in non-contratual settings, companies do not know which customers are