FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

Evaluating the Added Value of Pictorial Data for Customer Churn Prediction

Michel Ballings[1]

Dirk Van den Poel[2]

Emmanuel Verhagen[3]

December 2013

2013/869

[1] PhD candidate, Faculty of Economics and Business Administration, Ghent University, Tweekerkenstraat 2, B-9000 Ghent, Belgium, Michel.Ballings@UGent.be
[2] Professor of Marketing Analytics, Faculty of Economics and Business Administration, Ghent University, Tweekerkenstraat 2, B-9000 Ghent, Belgium, dirk.vandenpoel@UGent.be, Website: www.bigdata.UGent.be
[3] Psilogy, Leopold De Waelstraat 17a, 2000 Antwerp, Belgium

D/2013/7012/40

Evaluating the Added Value of Pictorial Data for Customer Churn Prediction

Michel Ballings[1], Dirk Van den Poel[1], Emmanuel Verhagen[2]

[1] Faculty of Economics and Business Administration, Department of Marketing, Ghent University, Tweekerkenstraat 2, B-9000 Ghent, Belgium, http://www.crm.ugent.be, Michel.Ballings@UGent.be, Dirk.VandenPoel@UGent.be

[2] Psilogy, Leopold De Waelstraat 17a, 2000 Antwerp, Belgium, Emmanuel.Verhagen@Psilogy.com

Abstract. The purpose of this paper is to evaluate whether pictorial data can improve customer churn prediction and, if so, which pictures are most important. We use Random Forest and five times twofold cross-validation to analyze how much pictorial stimulus-choice data increase the AUC and top decile lift of a churn model over and above administrative, operational, complaints and traditional survey data. We found that pictorial-stimulus choice data significantly increase models with administrative and operational data. The most important pictures are facial expressions, colors, and motivational scenes. Pictorial variables can reach importance of up to 21% of the importance of the best predictor included in the predictor set. Given this importance, managers could mine pictorial data from social media sites (e.g., Pinterest.com) in order to augment their internal customer database. To the best of our knowledge this study is the first that evaluates the added value of pictorial stimulus-choice data in predictive models. This is important because social media platforms are increasingly sharing their data and because of the recent rise of social media based on pictures. Pictorial data may soon become a viable option for data data-augmentation.

Keywords: CRM, Data Augmentation, Customer Retention, Customer Churn, Pictorial Stimulus-Choice Data

1 Introduction

Companies have come to recognize that in today's saturated marketplace their single most valuable asset is their customer base (Thomas, 2001; Athanassopoulos, 2000). Consequently, churn management has become the primary ingredient of companies' Customer Relationship Management (CRM) strategies. From an analytical viewpoint, churn management consists of (1) predicting which customers are most likely to leave the company and (2) assessing which action is most effective in keeping those customers (Hung, Yen, & Wang, 2006). This study focuses on the former component.

A customer's profitability improves over time. Hence even small changes in the churn rate can have considerable long term implications for a company's bottom line (Gupta, Lehmann, & Stuart, 2004). Van den Poel & Larivière (2004) show that even a decrease of one percentage point in the churn rate can have a substantial influence on the results. Companies adopt three main strategies in trying to improve customer churn prediction: (1) improving analytical techniques, (2) optimizing the time window, and (3) augmenting the customer database (Baecke, & Van den Poel 2011; Ballings & Van den Poel 2012). This study focuses on the latter strategy which consists in evaluating the added value, in terms of predictive performance, of new data over and above the internal database. Extant literature shows that finding new data that improves churn prediction over the baseline internal data is a particularly difficult task. This is because the internal database contains the top predictors in database marketing modeling (Cullinan, 1977; Coussement, & Van den Poel, 2008a): recency, frequency, monetary value (RFM) and length of relationship (LOR) (Baesens, et al., 2002; Van den Poel, 2003). Table 1 provides a literature review of studies that augment the internal database in order to improve predictive models in a CRM context.

Table 1. Literature review of data augmentation in CRM

Study	*Data-Variables*
Steenburgh, Ainsle, Engbretson (2003)	Geographical data (ZIP-codes)
Hu, & Zhong (2008)	Clickstream data
Hill, Provost, Volinsky (2006)	Consumer network data
Benoit, & Van den Poel (2012)	
Coussement, & Van den Poel (2008b)	Call center e-mails
Coussement, & Van den Poel (2009)	
Baecke, & Van den Poel (2011)	Commercially available survey data
D'Haen, Van den Poel, & Thorleuchter (2013)	
Baecke, & Van den Poel (2010)	Situational variables: weather, time, sales-person variables
Thorleuchter, Van den Poel & Prinzie (2012)	Customer websites
D'Haen, Van den Poel, & Thorleuchter (2013)	
Gilman, Narayanan & Paul (2004)	Call center dialog transcripts
Buckinx, Verstraeten, & Van den Poel (2007)	Company survey data: loyalty
Lariviere, & Van den Poel (2004)	Product features
Wong, Leung, Guo, Zeng, & Mok	RFID

(2012)

To the best of our knowledge no study has attempted to augment the internal database with pictorial data. As we will argue in the next section, social media platforms (e.g., Facebook, Pinterest) are the natural habitat of pictorial content. These platforms are increasingly sharing data through applications (Facebook 2013)(e.g., which user liked which picture) and hence mining pictorial data for data enhancement becomes a possibility. The recent rise of a social network in which pictures are the centerpiece (Pinterest.com) is also a promising evolution. This study aims to fill this gap in literature by answering the following two questions: (1) do pictorial stimulus-choice data have added value in customer churn prediction over and above the internal database and, (2) which pictures are most predictive?

The remainder of this paper is organized as follows. In the next section we provide a typology of different data types. In the methods section we discuss the data, time window, variables, estimation technique, cross-validation method, and assessment criteria. After discussing the results we conclude this study and cover managerial implications. The last section sets out directions for future research and addresses the limitations.

2 Data types

2.1 Traditional data types

We define four customer data types that companies can tap into for customer churn prediction. These definitions are based on the required level of investment, and on which customer behavior the data represents.

The first data type is administrative data. It represents customer identification and contract specifics acquired through the administrative process at the beginning of a subscription to a company's services. Alshawi, Missi, Irani (2011) point out that smaller or startup companies have less financial abilities to invest in information and data gathering. Therefore if they want to save money those companies will limit themselves to this type of data because they cannot (yet) stem the costs required to store, maintain and mine huge amounts of data. Administrative data is a requirement for operation in the market. A company needs to collect administrative data because it represents what currently needs to be delivered and invoiced.

The second customer data type is operational data. It comprises the entire history of subscriptions, contracts and operations (Larivière, & Van den Poel, 2005). Companies that do have the financial abilities to make the necessary investments in the extra storage, maintenance, software and skill requirements are probably more mature and bigger than companies that only use administrative data.

The third data type, complaints data (Coussement, & Van den Poel, 2008b), is related to the customer feedback process and requires a significant supplementary investment to analyze given its often unstructured nature (e.g., emails, call transcriptions). Only companies with mature customer intelligence departments will try to mine this data. According to the exit-voice theory (Hirschman, 1970), complaining behavior is conceived of as one of two options when a customer is dissatisfied, next to churning, and can therefore be a valuable data type for predictive models.

The fourth and final customer data type is surveys. Surveys are primarily aimed at uncovering insights (obtaining cross-sectional data for descriptive models), from a small number of representative customers. A typical example is a satisfaction survey. Survey data cannot be collected for all customers because of the following reasons. First of all, many customers simply don't want to invest time resulting in low response rates (Sheehan 2001). Second, in the case of paper based surveys, the cost is high both for administration and data input. Nevertheless, some companies try to employ surveys for predictive analytics. For example, Baecke and Van den Poel (2011) collected survey data from a limited amount of customers, added that data to the internal database, and imputed surveys responses for the other customers based on the predictors of the internal database. While this approach can mitigate the problem of not having data for all customers, its performance is dependent on the relationship between predictors of the internal data and predictors of the survey data. The added value of survey data is thus inherently limited. Companies that do invest in surveys for predictive analytics can be conceived of as having advanced customer intelligence departments.

2.2 Pictorial data

For all the aforementioned types of companies pictorial data can be a valuable candidate for data augmentation. Two reasons underlie this statement: availability and added value.

Pictorial data may soon be available on a large scale thanks to the increasing openness of social media platforms (Facebook 2013) and the recent rise of picture centered social networks (e.g., Pinterest.com). Social media are the natural habitat of pictorial content because pictures are (1) language independent (De Pelsmacker & Van Kenhove 2006) and, (2) an efficient and entertaining means of instantaneously conveying an enormous amount of information. Unlike words, pictures do not have to be translated to another language in order to be understood, hence paving the way for widespread dispersion. In sum, pictures are likely to become available for mining from social media.

Several arguments can be made for the added value account. All arguments can be traced back to the underlying function of pictures as an enabling technique (De Pelsmacker & Van Kenhove 2006) to elicit certain responses from the beholders. Pictures can convey complex situations or concepts and hence can elicit a specific response that would otherwise not have been elicited. Rationality, psychological or memory-based barriers (De Pelsmacker & Van Kenhove, 2006) can be alleviated by decreasing the demand on the respondent's cognitive processing by facilitating memory retrieval (Hermans, Dehouwer & Eelen 1994). In fact, it has been empirically demonstrated that individuals' memory retrieval is more effective when the memory is based upon the same emotion as the emotional state at retrieval; this is called the mood congruency effect (Fiedler, Nickel, Muehlfriedel, & Unkelbach 2001). As such, inducing the emotion that was experienced during encoding can enhance memory (Fiedler, Nickel, Muehlfriedel, & Unkelbach 2001). Given that pictures are strong stimuli of emotion (Dehouwer & Hermans, 1994), pictures may enhance memory retrieval if one of the pictures contains the same emotional content as the memory to be retrieved. An example may clarify how pictures can aid in alleviating psychological or memory based barriers. When presented with a picture of a family or couple watching a movie at home, customers may remember a similar joyful moment that they have experienced themselves. If a customer of a movie subscription service has indicated that he or she likes the picture this may be valuable data for customer churn prediction.

In sum, a picture can be conceived of as being the source of multiple advantages and would therefore be an attractive means of data augmentation. In addition to being widely used and thus available on social media, pictures can induce feelings and subsequently help recall specific memories. In turn these memories will stimulate interaction with the picture, such as indicating that one likes the picture, which can subsequently be used as a predictor in customer churn prediction. If words or text would have been used these feelings would less likely be induced (Dehouwer & Hermans, 1994), resulting in lower recall of the specific memory, in turn resulting in less data that can be used for mining.

3 Methods

3.1 Data

To collect the pictorial data we sent an email to the customers of two Belgian newspaper brands with an invitation to click on a link to go to the pictures and some questions. While using email to send the link is an ad-hoc approach, companies that want to collect pictorial data and that have social media profiles (e.g., on Facebook) could put the invitation and link on their profile or include the pictures directly in their profile or custom application.

Both newspaper brands are similar, except for their geographical targeting at the state level. One of the two brands consisted of different editions. We sent 25,897 emails. To stimulate participation in the study we offered a prize. Out of all the customers that received the email 25.7% (6,661 customers) opened the message and of those customers 65.9% (4,360) clicked on the link (many dropped out after the first page of the website). Of that group 59.7% (2,605) completed the pictorial data collection and were subsequently used in the analysis. The pictorial data collection consisted in a question that customers had to answer by clicking on a picture that best represented their answer. In the rest of this study the term pictorial data is short for pictorial stimulus-choice data and is defined as variables that indicate if a customer has clicked on a picture as response to a question. Next to the pictorial questions we also included some traditional questions such as involvement (Zaichkowsky, 1994), satisfaction (Fornell, Johnson, Anderson, Cha, & Bryant 1996), calculative commitment (Gounaris, 2005), affective commitment (Gustafsson, Johnson & Roos 2005), normative commitment (Ros, Schwartz, & Surkiss, 1999), renewal intentions (Rossiter, 2002) and value instrumentality (Lindeman, & Verkasalo, 2005). Table 2 displays the data characteristics (see later for cross-validation).

Table 2. Sample characteristics (5 times 2 fold cross-validation)

Across folds	*Average Number of accounts*	*Average Relative percentage*	*Standard deviation*
Training/Test data	1302.5		
Non-Churners	1146.94	88.023%	0.306
Churners	156.073	11.978%	0.305

We merged the collected data with internal data from which we computed variables in the aforementioned categories. As such, we were able to assess the added value of pictorial stimulus-choice data.

3.2 Time window

Customers have to pay a certain price for their newspaper subscription depending on the length of the subscription and the promotional context. When they are approaching the end of the subscription the company sends them a letter with a invitation to renew, along with instructions on how to do that. Subscriptions cannot be cancelled, as such churn prediction involves predicting whether the customer will or will not renew his or her subscription in the four-week period following the end of the subscription. During the four-week grace period, customers keep on receiving newspapers, which they will have to pay if they renew.

In order to predictively discriminate churners from non-churners, the data needs to be prepared according to a specific time window: predictor variables are to be computed from data from the period preceding the period in which the churn variable is observed. After the model building process, both the predictor and response period can be shifted forward in time while respecting their relative positions in order to deploy the model. Figure 1 represents the time-window.

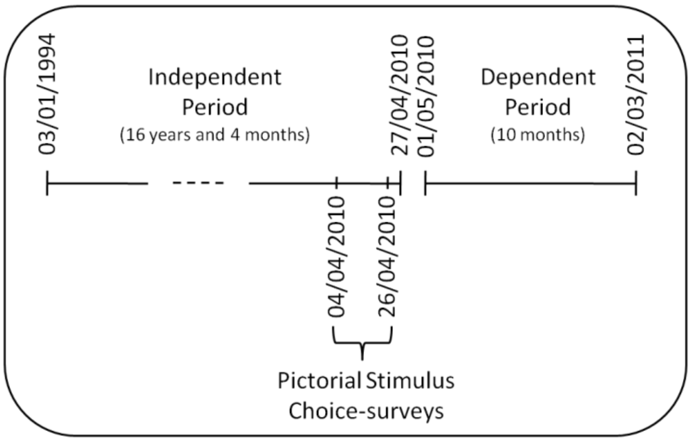


Figure 1. Time- window

3.3 Variables

A large number of predictor variables is used to predict churn. As discussed in section two, four plausible levels of data usage are administrative data only (level one), administrative and operational data (level two), administrative, operational, and complaints data (level three), and administrative, operational, complaints and survey data (level four). The number of variables included in the models for the four levels are respectively 129, 254 288 492. These four cases are the baseline. In order to determine the added value of pictorial data we will include the pictorial stimulus-choice data to each of these cases and compute the increase in predictive performance of the churn model. Since there are 48 pictorial variables (see later), this brings the total number of variables to respectively 177, 302, 336, and 540. Because of the large number of variables we will only provide some key examples and brief descriptions in the current section. The complete list of variables can be found in the appendix.

3.3.1 Administrative data

Administrative data contains all information regarding agreements made between the customer and the company at the time that the customer subscribed to the newspaper. Hence the data are acquired at the beginning of the relationship. This entails how much, where, when and to whom the newspaper needs to be delivered. It also comprises information about the price, payment method and possible promotions.

3.3.2 Operational data

Administrative data contains only data about the current subscription level. In contrast, operational data holds the entire customer history (at the subscriber's level). This means that variables are aggregated across subscriptions while they are per subscription in the administrative data. Operational data is acquired during the relationship, in contrast to at the beginning of the relationship. This data also holds socio-demographic data, as opposed to customer identification data, and data about suspensions, forward interruptions, credit handling, and marketing actions, as well as response to such action (e.g., participation in games). Much of the data is not merely contractual as in the administrative data but represents a more complex commercial policy (such as credit and suspension processes) and the subsequent usage of the product (subscription) by the customer.

3.3.3 Complaints data

This data type represents information about the number and topic of complaints and the company's solution and feedback to the customer. Examples of the topic of complaints are: non-delivery, incomplete newspaper, delivery is too late, wrong newspaper edition. Examples of solutions are: create credit, and redelivery. Examples of feedback are: force majeure, mailbox too small, mailman made a mistake, employee strike, and weather conditions. Feedback data and the Voice of Customers are acquired at customer initiated feedback moments and have been shown to be of value in predictive models (Coussement, & Van den Poel, 2008b).

3.3.4 Survey data

Survey data results from a company-initiated feedback process. Mindset variables (e.g., purchase intentions, commitment, product recommendations), customer life style information (e.g., interests and opinions) and product evaluation data (e.g., overall satisfaction, satisfaction drivers) are examples of this type of data and are impossible to collect from internal processes. Other examples are the frequency of reading specific sections (e.g., politics, economy, culture, science, TV guide) and whether the customer reads the newspaper online.

3.3.5 Pictorial stimulus- choice data

The final type of data was acquired through a choice process. Customers had to run through six picture sets (randomized between respondents) each containing nine pictures. The first five sets were accompanied with the following question. "Imagine a typical moment when you are reading [name of newspaper brand]. In general, how do you feel at that moment? Choose one of the nine pictures that best represents that feeling. Make your choice by clicking on the picture.". The sets respectively contained pictures of (1) motivational scenes, (2) a man's facial emotional expressions, (3) a woman's facial emotional expressions, (4) geometrical forms, and (5) colors. A sixth set contained pictures of couples in different relationship stages (such as being angry, an open relationship, engagement, marriage, having children, being grandparents) preceded with the question "Which of the following pictures best describes your relationship with [name of newspaper brand]?" For all pictures and sets, dummy-variables were then created indicating whether they were chosen by the respondent or not. This resulted in 48 dichotomous variables (6 sets times 9-1 pictures). Figure 2 represents an example such as the picture sets we used (a woman's facial emotional expressions). Unfortunately we cannot include the pictures that we have actually used in this study because of copy right issues. We purchased the pictures from a photo vending website. The pictures shown in this article are very close reproductions.

Figure 2. A woman's facial emotional expressions

3.4 Classification algorithm

To model churn we used Random Forest (Breiman, 2001) because of multiple reasons. First, literature shows that the algorithm is one of the best performing classification techniques available (Luo, et al. 2004) and is very robust and consistent (Breiman, 2001). Random Forest copes with the limited robustness and suboptimal performance (Dudoit, Fridlyand, & Speed 2002) of decision trees by building a committee of trees (e.g., 1000 trees), deploying each tree to produce a vote, and subsequently choosing for the most popular class (Breiman, 2001). Each tree is grown on a bootstrap sample with a random subset of all available predictors. Second, the method does not overfit (Breiman, 2001), which is of particular importance for this study due to the relatively large number of predictors we want to test and the small sample size. Third, variable importance measures are available for all predictors (Ishwaran, 2004). Fourth, the algorithm has reasonable computing times (Buckinx, & Van den Poel, 2005). Fifth, the procedure is easy to implement: only two parameters are to be set (number of trees and number of predictors) (Larivière, & Van den Poel, 2005; Duda, Hart & Stork 2001). We follow the recommendation of (Breiman, 2001) by using a large number of trees (1000) and the square root of the total number of variables as the number of predictors. Random Forest is implemented using the R *randomForest* package of Liaw and Wiener (2002).

3.5 Cross -validation

Research has shown that the usual method to compare classification methods (t-tests to confirm significant differences in performance measures obtained from k-fold cross validation) results in increased type-I error (Dietterich, 1998). Dietterich (1998) and Alpaydin (1999) propose five times two-fold cross validation (5x2cv) to cope with the problem. Five times two-fold cross validation randomly partitions the sample (2,605 observations) in two parts (1,302 and 1,303) and repeats this process five times. Each time the first partition is used as training, the second as test sample, and vice versa. This process results in 10 performance measures per model (Dietterich, 1998). In order to determine whether models are significantly different we follow the recommendation of Demšar (2006) to use the Wilcoxon signed-ranks test (Wilcoxon, 1945).

The Wilcoxon signed-ranks test (Wilcoxon, 1945) is a non-parametric test, that ranks the differences in performance of two models, while ignoring the signs. Ranks are assigned from low to high absolute differences, and equal performances get the average rank. The ranks of both the positive and negative differences are summed and the minimum of those two is compared to a table of critical values. To be significant, the smallest sum of ranks should be smaller than the critical value. This means that if there are too many differences that are small, or if there are too few differences that are large, two models are insignificantly different. Greater differences count more (Demšar, 2006).

This test is safer than parametric tests such as the t-test because it does not assume normal distributions or homogeneity of variance (Demšar, 2006) and is less sensitive to outliers. When the assumptions of the t-test are met, the Wilcoxon signed ranks test has less power than the t-test. However, since in this case the sample size is only ten differences, verifying normality and homogeneity of variance is problematic and hence the Wilcoxon signed ranks test is strongly recommended over the paired t-test (Demšar, 2006).

3.6 Assessment criteria

The first research question was: do pictorial stimulus-choice data have added value in customer churn prediction over and above the internal database. To answer this question at all levels of data usage we will compare churn models without pictorial data with models that include pictorial data. To nuance the results we also provide variable importances of the pictorial variables in relation to all the other (non-pictorial) variables.

The second research question was: which pictures are most predictive? We identify the top five pictorial variables and have a closer look at their rank, percentile rank, importance and relative importance to the best predictor in the database. The following paragraphs will provide details on the model assessment criteria and the variable importance measure used in this study.

3.6.1 Models

The primary ingredient of a churn strategy consists in a prediction of which customers are about to churn (Hung, Yen, & Wang, 2006). Because marketing budgets for customer retention programs are often limited, only a fraction of the customer base can be targeted and the budget may not sustain action towards all customers predicted to churn (Coussement, Benoit, & Van den Poel 2010). Hence, customers that are most at risk for churning are targeted (Lemmens, & Croux 2006). For this purpose a ranking of the customers is required (Coussement, & Buckinx 2011) which can be produced by classifiers in the form of probabilities, as opposed to a class labels. In Random Forest such probabilities are computed by normalizing the votes. For example, if 1000 trees are generated, and 800 trees vote for class *churn* for a given customer then the probability for that customer to churn is 80%. By ranking the customers based on their probabilities, a budget constrained fraction of the most riskiest customers can be selected.

Two important measures to analyze predictive performance based on probabilities are the area under the receiver operating characteristic curve (AUC or AUROC) and the top decile lift (Benoit, & Van den Poel 2012; Coussement, Benoit, & Van den Poel 2010). AUC is argued to be an objective criterion for classifier performance by several authors (Provost, Fawcett, & Kohavi 1998; Langley, 2000; De Bock, Coussement, Van den Poel, 2010; Ballings & Van den Poel, 2013) and is extensively used in the field of CRM (e.g., Coussement, & Van den Poel, 2008a; Benoit, & Van den Poel 2012; Coussement, Benoit, & Van den Poel 2010). The receiver operating characteristic (ROC) curve is obtained from plotting sensitivity and 1-specificity considering all possible cut-off values (Hanley, & McNeil, 1982). AUC ranges from .5, if the predictions are not better than random, to 1, if the model predicts the behavior perfectly (Baecke, & Van den Poel 2011). AUC is defined as follows (Ballings & Van den Poel, 2013):

$$AUC = \int_0^1 \frac{TP}{(TP + FN)} \, d\frac{FP}{(FP + TN)} = \int_0^1 \frac{TP}{P} \, d\frac{FP}{N}$$

, with TP: True Positives, FN: False Negatives, FP: False Positives, TN: True Negatives, P: Positives (churn), N: Negatives (non-churn).

We use AUC instead of accuracy (Percentage of correctly classified, PCC) because AUC, in contrast to PCC, is the insensitive to the cut-off value of the 'a posteriori' probabilities (Baecke, & Van den Poel 2012; Thorleuchter & Van den Poel 2012). The AUC avoids the difficulty of choosing a cut-off value to discriminate between the predicted churners and non-churners (Benoit, & Van den Poel 2012). In addition, AUC is a more appropriate measure than accuracy in imbalanced data settings (Janitza, Strobl, & Boulesteix 2013). He and Garcia 2009 also state that accuracy can be deceiving when imbalanced data is used. The AUC puts equal weight on both classes, in contrast to the accuracy which gives more weight to the majority class (Calle, Urrea, Boulesteix, Malats, 2011). Consider a situation in which 12% of the samples is a churner and 88% is a non-churner. A naïve approach of classifying every observation to be a non-churner would yield an accuracy of 88%. At face value this seems to be a very good classifier, even though zero percent of the churners are identified. In sum, the accuracy measure does not provide an appropriate assessment of the predictive performance with respect to the type of classification required (He, & Garcia, 2009). Many other studies denote the ineffectiveness of accuracy in imbalanced settings (Ling, Huang, & Zhang 2003; Maloof, 2003; Sun, Kamel, Wong, & Wang, 2007; Provost, Fawcett, & Kohavi 1998) and AUC is deemed a more adequate performance measure (see Baesens, et al., 2002).

Lift is another very important measure of predictive performance in CRM (Lemmens, & Croux 2006; Miguéis, Camanho, Falcão e Cunha, 2013). Neslin, Gupta, Kamakura, Lu, and Mason (2006) even profess that lift is probably the most commonly used measure in predictive modeling and demonstrate that it is directly related to profitability. AUC and lift provide complementary information in that the former measures predictive performance for the whole sample and the latter measures predictive performance on the subset of riskiest customers (Piatetsky-Shapiro and Masand, 1999, Lemmens, & Croux 2006). Therefore it is perfectly possible to obtain a higher lift whilst not obtaining a higher AUC (Masand, & Piatetsky-Shapiro 1996; Piatetsky-Shapiro and Masand, 1999). Top decile lift is a popular measure because analysts are often interested in 10% of the customer base most likely to churn (Coussement, Benoit, & Van den Poel 2010; Coussement & Van den Poel, 2008a). In sum, top decile lift is a very appealing measure because it integrates the notion that marketing budgets are limited and that in turn actions to retain customers are limited to a subset of customers with high churn risk (Coussement, Benoit, & Van den Poel 2010). Top decile lift is defined as follows:

$$Top\ decile\ lift = \frac{\frac{P_{Top\ 10\%}}{P_{Top\ 10\%} + N_{Top\ 10\%}}}{\frac{P}{P + N}}$$

, where *top* refers to the highest probabilities, P: Positives (churn), N: Negatives (non-churn).

Reported performance measures are all averages on the test samples obtained through five times two-fold cross validation (Dietterich 1998; Alpaydin 1999) as explained in the previous section.

3.6.2 Variables

An often used variable importance measure in Random Forest is the decrease in accuracy if variables are permuted. As discussed in the previous subsection, accuracy is not a good performance measure in imbalanced settings (as is the case in this study). Janitza, Strobl, and Boulesteix (2013) have compared permutation based decrease in accuracy and AUC. They conclude that the AUC based importance measure is preferred to the standard accuracy based measure whenever the two response classes have different class sizes. As this is the case in our study (see the data section), we will follow their recommendation and use the mean decrease in AUC as the variable importance measure in this study.

Janitza, Strobl, and Boulesteix (2013) work with out of bag data, which is different for all trees. Hence they average the AUC over all trees per ensemble. Since we work with test data in order to compare the models on the same observations, we compute the measure for the ensemble. This is essentially the same approach. The importance measures are averaged across the ten folds obtained through five times twofold cross-validation. The mean decrease in AUC for a predictor *j* is defined as follows:

$$Mean\ Decrease\ AUC = \frac{1}{n\ folds} \sum_{f=1}^{n\ folds} (AUC_{fj} - AUC_{fj_p})$$

with n folds the number of folds equal (ten in this case) and j_p the variable j in permuted form.

4 Results

To answer the first research question we look at whether the models including pictorial data are significantly more predictive of churn than models without pictorial data. All the measures that are displayed in this article are averages across the test folds obtained from 5x2cv. Figure 3 presents the AUC per data type. The lower line is the baseline representing the data types without pictorial data. The upper line represents the pictorial stimulus choice data in addition to the baseline.

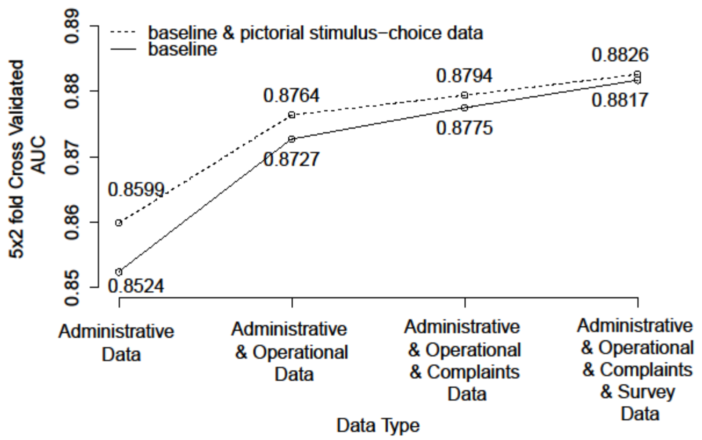


Figure 3. The added value per data type in terms of AUC

There are three important conclusions. First, the results indicate that the addition of the operational data accounts for the biggest increase in AUC. Second pictorial stimulus-choice data improves predictive performance. Third, the performance improvement of pictorial stimulus-choice data decreases marginally when more data types are considered.

When only administrative data is analyzed pictorials add 0.9% to the predictive performance. In the case of administrative and operational data, pictorials add 0.4%. If administrative, operational and complaints data are considered then pictorial data adds 0.2% to predictive performance. Finally, when all data are modeled (administrative, operational, complaints, and survey data) pictorial stimulus-choice data adds 0.1%. Wilcoxon signed-ranks tests of the null hypothesis that both models (baseline versus baseline plus pictorial stimulus-choice data) perform equally well are rejected for models based on administrative data (V=3, p < .01), and administrative and operational data (V=1, p < .01). Models based on administrative, operational & complaints data are not significantly different (V=11, p >.10). The same conclusion can be drawn for models based on administrative, operational, complaints and survey data (V=13, p > .10).

Figure 4 displays the top decile lift. The conclusions are different. Because this measure only takes into consideration the top ten percent of the predictions, it is perfectly possible to obtain a higher lift whilst not obtaining a higher AUC (Masand, & Piatetsky-Shapiro 1996; Piatetsky-Shapiro and Masand, 1999).

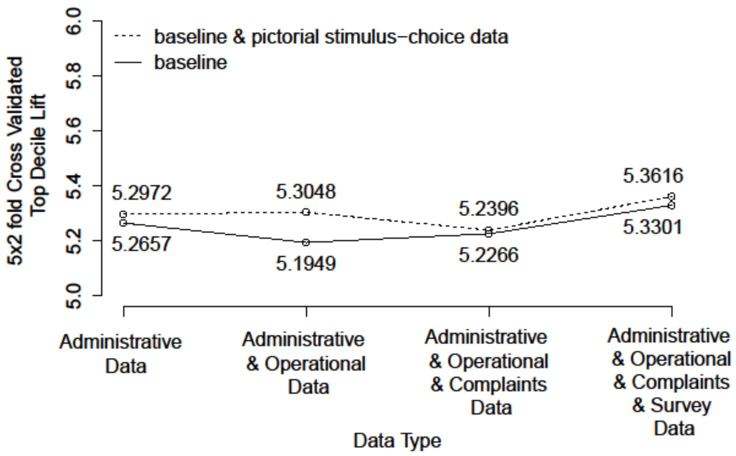


Figure 4. The added value per data type in terms of top decile lift

The main conclusion from Figure 4 is that the biggest increase in top decile lift due to pictorial data is obtained in the case of administrative and operational data. Although one could expect that the lift would monotonically increase when more data is considered this is not necessarily the case because only ten percent of the data is considered in the top decile lift (Masand, & Piatetsky-Shapiro 1996). Measures that consider all the data (i.e., AUC, Figure 3) do increase monotonically with more data.

Pictorials add 0.6% to the predictive performance when only administrative data are considered and 2.1% in the case of administrative and operational data. When administrative, operational and complaints data are analyzed pictorial data adds 0.2% to predictive performance. When all data are modeled pictorial data adds 0.6%. Wilcoxon signed-ranks tests confirm that models are only significantly different in the second case: administrative data (V=12, p > .10), administrative and operational data (V=0, p < .05), administrative, operational & complaints data (V=15, p >.10), administrative, operational, complaints and survey data (V=8, p > .10).

To investigate which pictorial predictors are driving these results we made scree plots per data level (Figure 5 to 8). In the scree plots the variables are sorted in descending order by the five times twofold cross-validated mean decrease in AUC. Only the variables that had an importance higher than zero are included. The pictorial variables are indicated by a black dot. We can see from the plots that the pictorial variables are well dispersed (smaller ranks are better): they are ranked in all quartiles. Some pictures are ranked very high (e.g, the first ranked pictorial variable obtains the eighth position out of a total of 90 variables in Figure 5). The scree plots indicate that there is potential for improving the models if only the right pictorials can be found.

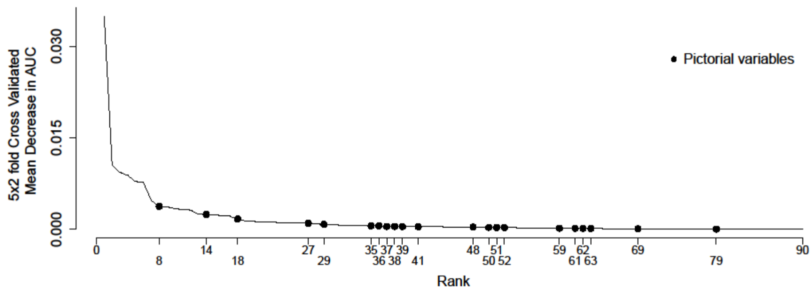


Figure 5. Scree plot of cross-validated variable importance for administrative data

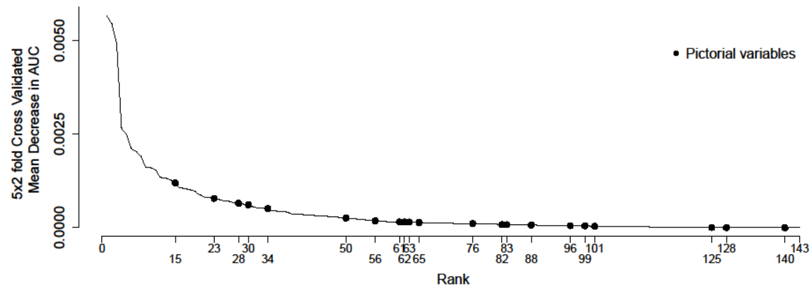


Figure 6. Scree plot of cross-validated variable importance for administrative and operational data

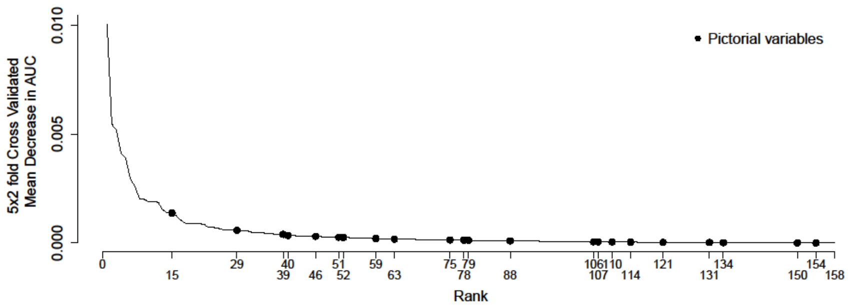


Figure 7. Scree plot of cross-validated variable importance for administrative, operational, and complaints data

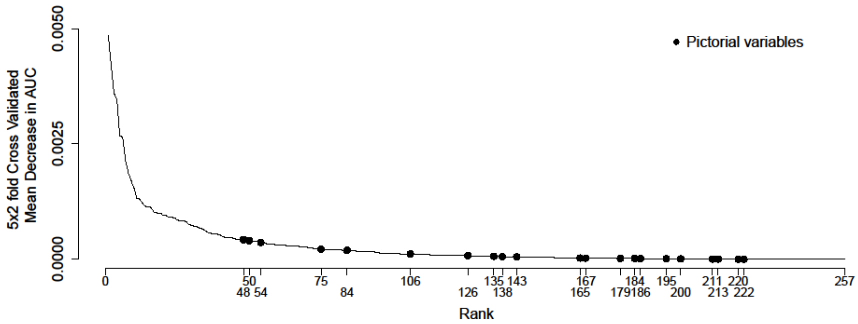
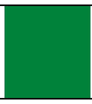
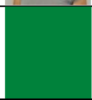
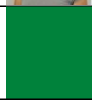
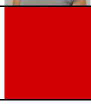
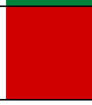


Figure 8. Scree plot of cross-validated variable importance for administrative, operational, complaints, and survey data

To answer the second research question we identify the top five pictorial variables (Table 3) per data level. Pictures of four pictorial sets are present: (1) a woman's facial emotional expressions, (2) a man's facial emotional expressions, (3) colors and (4) motivational scenes. Note that the pictures are close copies of the ones we used in this research. We cannot include the actual ones because of copyright issues.

Table 3. Top five most important pictures along with their rank among all variables

Administrative data (Rank)	Administrative and operational data (Rank)	Administrative, operational and complaints data (Rank)	Administrative, operational, complaints, and survey data (Rank)
8 	15 	15 	48 

14		23		29		50	
18		28		39		54	
27		30		40		75	
29		34		46		84	

Because we did not test the objective meaning of the pictures (in a descriptive context this would be required, in contrast to a predictive context) we cannot make any statements about what the pictures actually mean. The intended meanings of the displayed pictures are ‘astonishment’, ‘interest’, and ‘surprise’. The motivational pictures was intended to mean ‘journey’.

Table 4 contains the rank and percentile rank of the top five predictors based on the cross-validated performance. The percentile rank is computed by dividing the rank by the total number of contributing (variables with importance greater than zero) variables. Except for the last data level, the best pictorial predictor is ranked in the top decile. In the highest data level the top predictor is ranked in the 19th percentile (lower is better: most important variables get lowest ranks).

Table 4. Rank and percentile rank of the top five predictors

	Rank (percentile rank) based on cross-validated importance			
	Administrative data	Administrative and operational data	Administrative, operational and complaints data	Administrative, operational, complaints, and survey data
1st picture	8 (9%)	15 (10%)	15 (9%)	48 (19%)
2nd picture	14 (16%)	23 (16%)	29 (18%)	50 (19%)
3rd picture	18 (20%)	28 (20%)	39 (25%)	54 (21%)
4th picture	27 (30%)	30 (21%)	40 (25%)	75 (29%)
5th picture	29 (32%)	34 (24%)	46 (29%)	84 (33%)
Total number of contributing variables	90	143	158	257

More information can be obtained by looking directly at the variable importances (Table 5). Because it makes little sense to look at the absolute values of the variable importances of models with many variables (from 90 to 257) we computed the relative importance of the five best pictorial predictors relative to the best of all predictors. Depending on the data level the best pictorial predictor added between 9% and 21% to the AUC of what the best predictor in the model explained. Because the ranked variable importances (i.e., a scree plot) follow a negative exponential distribution (the decrease in importance from the top to the first quartile is very steep, while the decrease from the first quartile to the bottom is very low) we consider this to be a strong indication that the pictorial variables are a valuable addition to the model.

Table 5. Cross-validated importance and importance relative to best of all predictors.

	Cross-validated Importance (importance relative to best of all predictors)			
	Administrative data	Administrative and operational data	Administrative, operational and complaints data	Administrative, operational, complaints, and survey data
1st picture	0.003772 (11%)	0.001195 (21%)	0.001377 (14%)	0.000428 (9%)
2nd picture	0.002470 (7%)	0.000785 (14%)	0.000585 (6%)	0.000403 (8%)
3rd picture	0.001710 (5%)	0.000654 (12%)	0.000394 (4%)	0.000368 (8%)
4th picture	0.001003 (3%)	0.000613 (11%)	0.000344 (3%)	0.000221 (5%)
5th picture	0.000831 (2%)	0.000508 (9%)	0.000306 (3%)	0.000196 (4%)
Best predictor	0.034955 (Length last subscription)	0.005671 (Length last subscription)	0.010060 (Length of relationship)	0.004857 (Length of relationship)

In the next section we conclude this article and discuss the managerial implications. The last section will discuss the limitations of this study along with directions for future research.

5 Conclusions and managerial implications

One of the main strategies for companies to improve their predictive churn models is data augmentation. An increasing number of data sources are accessible and managers' choice-options for data augmentation are constantly growing. Hence, managers are in need of information about the added value of these options. One of those options is pictorial data from for example social media. Social media are the natural habitat of pictorial content because pictures are language independent. In contrast to words, pictures do not have to be translated so that they can be understood. In addition they can have a high entertainment value and can convey information in an instantaneous fashion. Because social media platforms are increasingly sharing their data (see for example Facebook 2013) and because of the recent rise of social media based on pictures (Pinterest.com), pictorial data may soon become a viable option for data data-augmentation.

In this study we have analyzed if pictorial data provide added value in the prediction of customer churn behavior. We show that the augmentation of companies' databases with pictorial data can significantly improve customer churn prediction, even over and above a large number of traditional internal predictors. We also show that, as expected, companies with less data to deploy (administrative and administrative & operational data) in their customer intelligence projects would benefit most from this kind of data. Figures 5-8 indicate that the importance of the pictorial predictors varies considerably. This means that there is a lot of potential to increase the predictive performance of models if only the right pictures can be found.

In addition we have taken a closer look at which pictorial variables are most predictive. The most important pictures are specific facial emotional expressions, specific colors, and motivational scenes. The top pictorial predictors have an importance of up to 21% of the importance of the best predictors in the entire predictor set.

Managers can easily augment the customer database by adding a link to an online pictorial survey on the social media profiles of their products or companies. They can also include picture sets directly on the social media platforms or embed them in a platform specific application. The pictures could be presented with a question such as 'Which picture do you like best' or 'Which picture represents how you feel about product x'. A reactive approach could also be meaningful: users will indicate that they like a picture, even without asking for it. Hence it might prove valuable to simply mine all pictures that are liked and search for relationships between these pictures and churn behavior. On a more practical note; because it is often the case that social media (e.g., Facebook) force their users to use their real names, the company can easily link the data to the internal data. Other information such as the date of birth is also available from the social media platforms if the name would not be enough for a full match with the internal data.

6 Future research and limitations

This study intends to gauge the added value of pictures in predictive models. Since a picture contains a huge amount of information, user actions (e.g., 'liking' a picture on a social media platform) could possibly say a great deal about future behavior. Whether this is the case depends on the informational content of the picture. A limitation in this study is then that we tested only six pictorial sets. A possible direction for future research is to test more pictorial sets. The results indicate that there is a large amount of variation in the importance of the pictorial variables. Some perform well, others perform poorly. Hence the construction of optimal pictorial sets could be an interesting path to explore.

A second idea for future research is to model pictorial stimulus-choice data with other dependent behavior as the criterion variable. For example, because in a customer acquisition case the available data to be incorporated in the model is far more limited than in a churn case, the added value of pictorial data could be much more pronounced.

A third direction is to use a more structural approach for data collection. In this study we used an ad-hoc approach (email). Future research could use a systematic approach by mining online social networks and including picture sets directly in the social media profiles of products, brands or companies.

A fourth direction for future research is to evaluate different introductory questions to the pictorial sets. In this study we asked to associate a picture with the product experience, which can be conceived of as a product attribute. These attributes in turn affect the overall attitude towards the target (e.g., overall satisfaction with the product) (Guo, Xiao & Tang, 2009). In turn, the attitude towards the target affects attitude towards behavior (e.g., commitment to staying in the relationship) (Bansal, Irving & Taylor, 2004). This attitude will subsequently result in a conscious plan to engage in the focal behavior (behavioral intentions) (Solinger, van Olffen & Roe, 2008). In sum, attitudes have an impact on observed behavior, from most proximal to most distal, through behavioral intentions, attitude toward behavior (commitment), attitude toward target (satisfaction) and drivers of attitude toward target (drivers of satisfaction) (Fishbein and Ajzen 1975; Solinger, van Olffen & Roe, 2008; Eagly & Chaiden's, 1993; Bansal, Irving & Taylor, 2004; Guo, Xiao & Tang, 2009). It might prove valuable to adapt the question to gauge a more proximal attitude to behavior (e.g., Which picture best represents your commitment to stay with the newspaper?).

A fifth and final avenue would be to compare traditional verbal scales to pictorial answer scales. In this study we didn't have commensurate variables that can be compared. A particularly interesting direction could be to ask how the customer feels about a certain brand, let him or her answer the question by clicking on pictorial facial emotional expressions and verbal (binary) scales and evaluate which type, pictorial or verbal, is most predictive of customer churn.

7 Acknowledgements

We would like to thank strategic marketing bureau Psilogy for their ideas on the topic of this paper and Katrien Meert for posing for the pictures presented in this study. We would also like to thank the three anonymous reviewers for their fruitful comments. Finally we acknowledge the IAP research network grant No. P7/06 of the Belgian government (Belgian Science Policy).

8 Appendix A: Variables included in the model

8.1 Administrative data

Frequency related variables
- Number of paid and free newspapers
- Number of newspaper deliveries per week

Time related variables
- Whether the start of subscription to renewal is more than 30 days
- Duration of change in contract specifics of the subscription
- The contracted length of the subscription
- Length of final subscription
- Which month and season the end of subscription falls in
- Recency: elapsed time since last subscription start date
- Which days the newspaper is to be delivered

Monetary value related variables
- Product formula: Net price, Gross price, Price Reduction
- One newspaper: Net price
- Title: Price, Price reduction
- Total: Price reduction, Monetary value (Total amount paid), Amount of credit
- Whether it is a regular subscription
- Whether the subscription is paid for
- Payment method: direct debit
- Whether this is a free subscription

Customer related variables
- Relationship category
- Address of delivery: Province, Postal code dummy variables

Product related variables
- Newspaper edition
- Whether the brand is brand 1
- Marketing action
- Product formula
- Whether the contract specifics of the subscription have changed

8.2 Operational data

Frequency related variables

- Number of subscriptions, Number of regular subscriptions, Mean Number of regular subscriptions
- Number of newspapers, Number of free newspapers
- Amount of forwarding interruptions, Amount of suspensions
- Total number of times reason of operational interruption: Non-payment, Vacation, Other
- Number of registrations in marketing actions
- Whether this is the subscriber's first subscription, Whether the subscription is a renewal
- Number of changes in newspaper editions
- Amount of newspaper edition changes divided by number of subscriptions
- Last operational interruption is a forwarding, Last operational interruption is a suspension
- Reason operational interruption: Non-payment, Vacation, Other

Time related variables

- Length of accumulated subscriptions (LOAS) last 50%(20%) of subscriptions divided by LOAS first 50%(80%) of subscriptions
- Division of length of accumulated subscriptions (LOAS) last 50%(20%) of subscriptions by number of subscriptions, divided by division of LOAS first 50%(80%) of subscriptions by number of subscriptions
- Elapsed time since last registration of participation in marketing actions, Elapsed time since average registration date of participation in marketing actions
- Elapsed time since last operational interruption (suspension or forwarding), Elapsed time since mean date of operational interruptions (suspension or forwarding)
- Elapsed time since mean of date credit settlements, Elapsed time since last credit settlement
- Elapsed time since change in newspaper edition
- Length of relationship (LOR): elapsed time since start of first subscription
- Average length of subscription, Mean length of subscription
- Number of days until subscription expiry date
- Duration final subscription divided by mean duration of all subscriptions
- End date current subscription minus renewal date new subscription, Renewal date current subscription minus end date minus theoretical time of reminder sent

- Sum across all subscriptions of renewal date current subscription minus end date minus theoretical time of reminder sent
- Sum across all subscriptions of end date current subscription minus renewal date new subscription, Mean across all subscriptions of renewal date current subscription minus end date minus theoretical time of reminder sent, Mean across all subscriptions of end date current subscription minus renewal date new subscription
- Elapsed time since mean renewal date across all subscriptions
- The variance in the number of days the previous subscriptions are renewed before expiry date
- Days to previous subscription / average days to renewal across all other subscriptions
- Elapsed time since change in payment mode
- Sum of subscription lengths, Sum of all contracted subscription lengths
- Prior churn: Number of times there is a start date in the entire purchase history that is greater than the expiry date of the previous subscription
- Total prior churn duration: Total duration of not having a subscription in between subscriptions
- Prior churn 10, Prior churn 20, Prior churn >30: Number of times there is a start date in the entire purchase history that is 10 (20,>30) days greater than the expiry date of the previous subscription
- Prior churn: Whether the current subscription's start date is greater than the expiry date of the previous subscription
- Prior churn 10, Prior churn >30: Whether the current subscription's start date is 10(>30) days greater than the expiry date of the previous subscription
- Prior churn duration: duration of not having a subscription in between the current and the previous subscription
- Mean prior churn duration: Mean duration of not having a subscription in between subscriptions
- Prior churn: Whether there is a start date in the entire purchase history that is greater than the expiry date of the previous subscription
- Prior churn 10, Prior churn 20, Prior churn >30 . Whether there is a start date in the entire purchase history that is 10,20,>30 days greater than the expiry date of the previous subscription
- Duration of last operational interruption (suspension or forwarding), Duration of all operational interruption: suspension or forwarding, Mean duration of all operational interruption (suspension or forwarding)

Monetary value related

- Division of monetary value of accumulated subscriptions (MOAS) last 50%(20%) of subscriptions by number of subscriptions, divided by division of MOAS first 50%(80%) of subscriptions by number of subscriptions

- Monetary value of accumulated subscriptions (MOAS) last 50%(20%) of subscriptions divided by MOAS first 50%(80%) if subscriptions
- Product formula: Net price, Gross price, Price Reduction
- Title: Price, Mean Title price
- Monetary value: Total amount paid, Mean Monetary value: Total amount paid
- Total: price reduction, amount of credit
- One newspaper: Mean Net price, Net price
- Product formula: Mean Net price, Mean Gross price, Mean Price Reduction
- Mean Total amount of credit, Mean Total price reduction
- Source of credit: subscription, complaint
- Credit code: to be confirmed, complaint, other reason
- Number of credit settlement by increasing amount of newspapers, Number of newspapers received from credit settlement, How credit is settled: increasing amount of newspapers
- Amount of last credit transaction, Total amount of credit transactions, Number of credit transactions

Customer related variables

- Relationship type: company, VAT N.A., limited liability company, N.A., Advertiser
- Language: Dutch
- Gender
- Age (relations-data)
- Whether contact information is available: phone number, email
- Whether the age is known

Multiples of recency, frequency, monetary and length of relationship

- Recency x Frequency x Monetary x LOR
- Recency x Frequency x Monetary
- Recency x Frequency x LOR
- Recency x Monetary x LOR
- Frequency x Monetary x LOR
- Recency x Frequency
- Recency x Monetary
- Recency x LOR
- Monetary x LOR
- Frequency x Monetary
- Frequency x LOR
- Monetary / LOR
- Frequency / LOR
- Monetary / Frequency

- LOR / Frequency
- Recency / length of last subscription
- Recency / LOR

8.3 Complaints data

Frequency related variables

- Number of complaints
- Number of times the topic of complaint is non post-delivery, delivery, incomplete newspaper, delivery too late, wrong newspaper edition, other
- Number of times the solution to complaint is no solution needed, create credit, post-delivery, take measures
- Number of times the feedback to complaint is force majeure, other, mailbox too small, mailman made mistake, pre-delivery too late, pre-delivery of incomplete newspapers, strike, weather conditions
- Whether last topic of complaint is other, non delivery, incomplete newspaper, is too late
- Whether last solution to complaint is no solution needed, create credit, post-delivery
- Whether last feedback to complaint is other, mailman made mistake, predelivery is too late, strike weather conditions

Time related variables

- Elapsed time since last complaint, Elapsed time since mean complaint date

Monetary related variables

- Whether credit is settled

8.4 Survey data

Frequency related variables

- Reading frequency of brand, Buying frequency of brand
- Frequency of internet surfing

- Frequency of surfing on the brand's website
- Frequency of reading about politics, economy, money, accidents and casualties, crime and justice, national news, foreign news, regional news, sports, personalities and television, culture, science, leisure time- traveling – cooking, TV guide

Time related variables

- Recency of filling in the survey

Mind set related variables

- Overall brand quality
- Single item overall commitment
- Whether the respondent has reasons to stay and none to leave, Whether the respondent has reasons to stay but also to leave
- Overall satisfaction: 3 items: Overall, how satisfied are you with [brand]? (1:very dissatisfied, 7: very satisfied), How well does [brand] meet your expectations? (1:falls short of expectations, 7:exceeds expectations), How close is [brand] to your ideal newspaper? (1: not very close to ideal newspaper, 7: very close to ideal newspaper) (Fornell, Johnson, Anderson, Cha, & Bryant 1996)
- Affective commitment (4 items): I take pleasure in being a customer of [brand]., [brand] is the newspaper company that takes the best care of their customers., There is a presence of reciprocity in my relationship with [brand]., I have feelings of trust toward [brand]. (1:Strongly disagree, 7:Strongly Agree) (Gustafsson, Johnson & Roos 2005)
- Calculative commitment (3 items): It is hard to break my relationship with [brand], There are worthwhile alternatives to [brand], There are high costs to breaking my relationship with [brand] (1:Strongly disagree, 7:Strongly Agree) (Gounaris 2005)
- Normative commitment (4 items): Even if it were to my advantage, I do not feel it would be right to stop being a customer of [brand]. [brand] deserves my loyalty., I would feel guilty if I stopped being a customer of [brand]., I would continue being a customer of [brand] because I have a sense of obligation to them. (1:Strongly disagree, 7:Strongly Agree) (Bansal, Irving & Taylor 2004)
- Situational trigger: There has been a recent change in your working conditions, family situation, or living conditions that has caused you to consider switching to another operator. (1:Strongly disagree, 7:Strongly Agree) (Gustafsson, Johnson & Roos 2005)
- Recommendation to friends or colleagues: Would you recommend [brand] to a friend or colleague? (1: Absolutely not, 7: Absolutely yes) (Reichheld, 2003)
- Renewal intentions: Are you going to renew your subscription for [brand] when it expires? (1: very unlikely, 7: very likely) (Rossiter 2002)

- Product involvement: 10 seven-point semantic differentials: How would you describe a newspaper subscription (for any newspaper brand)? Important-unimportant, Boring-interesting, Relevant-irrelevant, Exciting-unexciting, Means nothing- means a lot to me, Appealing- unappealing, Fascinating-mundane, Worthless- valuable, Involving-uninvolving, Not needed-needed (Zaichkowsky, 1994).
- Value instrumentality: Imagine a typical moment at which you are reading [newspaper brand].In general, how do you feel at that moment? Indicate to which degree the following concepts are consistent with that feeling. 10 items: Power (Social power, authority, wealth, Preserving Public Image, Social Recognition), Achievement (Successful, capable, ambitious, influential, Intelligent), Hedonism (Pleasure, enjoying life, self indulgence), Stimulation (Daring, a varied life, an exciting life), Self-direction (Creativity, freedom, independent, curious, choosing own goals, Self- Respect), Universalism (Broad-minded, wisdom, social justice, equality, a world at peace, a world of beauty, unity with nature, protecting the environment, inner harmony), Benevolence (Helpful, honest, forgiving, loyal, responsible, meaning in life, a spiritual life, True Friendship, Mature Love), Tradition (Humble, accepting own portion of life, devout, respect for tradition, moderate, detachment from world concerns), Conformity (Politeness, obedient, self-discipline, honoring of parents and elders), Security (Family security, national security, social order, clean, reciprocation of favors, Sense of Belonging, Healthy) (1: Absolutely not, 7: Absolutely yes) (Lindeman & Verkasalo 2005)
- Which is the respondent's preferred brand

Customer related variables

- Whether the respondent is interested in books, movies, museums and exhibitions, music, concerts and theatres, cars and motorcycles, food and cooking, health, fashion, parenting, gardening, traveling, houses and interior design, sports, athletics, basketball, formula 1 and racing, golf, motocross, tennis, mountain bike, soccer, volleyball, cycling, other sports
- Whether the respondent practices sports, basketball, cycling, fitness-going to the gym, golf, jogging, tennis, soccer, volleyball, walking, maintain biking, swimming, other sports
- Whether the respondent is married or living together
- Whether the respondent has children
- Number of children the respondent has
- Whether and which child is still living with the respondent
- Age (company survey), Partner's age, Children's age
- Whether the respondent is the sole or joint decision maker for buying the newspaper

- Occupation : Independent, white collar, blue collar, civil servant, middle management, retired, Housewife/househusband, unemployed- looking for a job, student, not working
- How the respondent acquires the newspaper: subscription/delivery at home, buy at the shop- newspaper per newspaper, family member buys at the shop- newspaper per newspaper, via colleague or employer
- Browser used for online survey Chrome, Internet Explorer 6, Internet Explorer 7, Internet Explorer 8, Firefox 3, Firefox 3.5, Firefox 3.6, Safari 3.1, Safari 3.2, Safari 4
- Operating system used for online survey: Mac OS, Windows
- Whether the respondent reads online

Product related variables

- Which newspaper the respondent reads
- Which competitor brand the respondent reads
- Which sections the respondent reads: personalities and television, foreign news, culture, economy, crime and justice, accidents and casualties, politics, regional news, sports
- Whether the respondent reads online about personalities and television, foreign news, culture, economy, crime and justice, accidents and casualties, politics, regional news

9 References

Alpaydin, E. (1999). Combined 5 x 2cv F test for comparing supervised classification learning algorithms. Neural Computation. 11(8), 1885–1892.

Alshawi, S., Missi, F., Irani, Z. (2011). Organisational, technical and data quality factors in CRM adoption - SMEs perspective. Industrial Marketing Management. 40 (3), 376-383.

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

Baecke, P., & Van den Poel, D. (2012). Including spatial interdependence in customer acquisition models: a cross-category comparison. Expert Systems with Applications, 39(15), 12105-12113.

Baecke, P., Van den Poel, D. (2010). Improving purchasing behavior predictions by Data Augmentation with situational variables. International Journal of Information Technology & Decision Making. 36 (3), 367-383.

Baecke, P., Van den Poel, D. (2011): Data Augmentation by Predicting Spending Pleasure Using Commercially Available External Data. Journal of Intelligent Information Systems. 36 (3), 367-383.

Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J., Dedene, G. (2002). Bayesian neural network learning for repeat purchase modelling in direct marketing. European Journal of Operational Research. 138 (1), 191–211.

Ballings, M., & Van den Poel (2012). Customer event history for churn prediction: How long is long enough? Expert Systems with Applications, 39 (18) , 13517-13522

Ballings, M., & Van den Poel, D. (2013), Kernel Factory: An ensemble of kernel machines. Expert Systems with Applications. 40 , 8, 2904-2913

Bansal H. S., Irving P. G., & S. F. Taylor (2004). A Three-Component Model of Cus-tomer to Service Providers, Journal of the Academy of Marketing Science, 32, 3, 234-250.

Benoit, D.F., & Van den Poel, D. (2012). Improving customer retention in financial services using kinship network information. Expert Systems with Applications. 39 (13), 11435-11442.

Breiman, L. (2001). Random forests. Machine Learning. 45 (1), 5–32

Buckinx W., Verstraeten G., Van den Poel D. (2007); Predicting customer loyalty using the internal transactional database. Expert Systems with Applications. 32 (1), 125-134.

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;

Calle M., Urrea V., Boulesteix AL., Malats N. (2011). AUC-RF: A new strategy for genomic profiling with random forest. Human Heredity. 14 (2),121–132.

Coussement, K., & Buckinx, W. (2011). A probability-mapping algorithm for calibrating the posterior probabilities: A direct marketing application. European Journal of Operational Research. 214 (3), 732-738.

Coussement, K., Benoit, D. F., & Van den Poel, D. (2010). Improved marketing decision making in a customer churn prediction context using generalized additive models. Expert Systems with Applications. 37 (3), 2132-2143.

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.

Coussement, K., Van den Poel, D. (2009). Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers. Expert Systems with Applications. 36, 6127–6134.

Cullinan, G.J. (1977). Picking them by their Batting Averages' Recency – Frequency – Monetary Method of Controlling Circulation. Manual Release 2103, Direct Mail/Marketing Association, N.Y.

De Bock, K. W., Coussement, K., Van den Poel, D. (2010). Ensemble classification based on generalized additive models. Computational Statistical Data Analysis. 54 (6), 1535-1546.

De Pelsmacker P., & Van Kenhove P. (2006). Marktonderzoek, methoden en toepassingen (2de editie), Pearson Education Benelux, Prentice Hall, 2006, 515.

Dehouwer J. & Hermans D. (1994). Differences in the Affective Processing of Words and Pictures, Cognition & Emotion, 8, 1, 1-20.

Demšar, J. (2006). Statistical Comparisons of Classifiers over Multiple Data Sets. Journal of Machine Learning Research. 7, 1-30.

D'Haen J., Van den Poel D., & Thorleuchter D. (2013). Predicting Customer Profitability During Acquisition: Finding the Optimal Combination of Data Source and Data Mining Technique. Expert Systems with Applications, 40(6), 2007-2012.

Dietterich, T.G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. Neural Comput. 10, 1895–1924.

Duda, R.O., Hart, P.E., & Stork, D.G. (2001). Pattern Classification. Wiley, NY, Chapter 8.

Dudoit, S., Fridlyand, J., Speed, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. Journal of the American Statistical Association. 97 (457), 77–87.

Eagly A, Chaiken S. (1993). The psychology of attitudes. Fort Worth, TX: Harcourt Brace Jovanovich.

Facebook (2013). Request for permission for sharing data. Facebook, Menlo Park, CA, https://www.facebook.com/about/privacy/your-info-on-other (accessed 17.06.2013).

Fiedler K, Nickel S, Muehlfriedel T, & Unkelbach C. (2001). Is mood congruency an effect of genuine memory or response bias? Journal of Experimental Social Psychology, 37, 3, 201-214.

Fishbein M. & I. Ajzen (1975). Belief, attitude, intention and behavior: An introduction to theory and research. Reading, Mass: Addison-Wesley.

Fornell, C., Johnson, M.D., Anderson, E.W., Cha, J.S., Bryant, B.E. (1996). The American customer satisfaction index: nature, purpose, and findings. Journal of Marketing. 60 (4), 7-18.

Gilman A., Narayanan B. & Paul S. (2004). Mining call center dialog data. In Zanasi A., Ebecken N.F.F. and Brebbia C.A. (Eds.) Data Mining V. WIT Press.

Gounaris, S.P. (2005). Trust and commitment influences on customer retention: insights from business-to-business services. Journal of Business Research. 58 (2) 126-140.

Guo L., Xiao JJ. & Tang CY. (2009). Understanding the psychological process underlying customer satisfaction and retention in a relational service, Journal of Business Research, 62, 11, 1152-1159.

Gupta, S., Lehmann, D. R., Stuart, J. A. (2004). Valuing customers. Journal of Marketing Research. 41, 7–19.

Gustafsson, A., Johnson, M.D., Roos, I. (2005). The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention. Journal of Marketing. 69 (4), 210-218.

Hanley, J.A., McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 143 (1), 29–36.

He, H. & Garcia, E. A. (2009). Learning from Imbalanced Data, IEEE Transactions on Knowledge and Data Engineering, 21(9).

Hermans D., Dehouwer J. & Eelen P. (1994). The Affective Priming Effect, Automatic Activation of Evaluative Information in Memory, Cognition & Emotion, 8, 6, 515-533.

Hill, S., Provost, F., Volinsky, C. (2006). Network-based marketing: Identifying likely adopters via consumer networks. Statistical Science. 21, 256–276.

Hirschman, A.O. (1970). Exit. Voice, and Loyalty—Responses to Decline in Firms. Organizations, and States. Cambridge, Harvard University Press.

Hu, J., Zhong, N. (2008). Web farming with clickstream. International Journal of Information Technology & Decision Making. 7, 291–308.

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

Ishwaran, H., Blackstone, E. H., Pothier, C. E., Lauer, M. S. (2004). Relative risk forests for exercise heart rate recovery as a predictor of mortality. Journal of the American Statistical Association. 99 (467), 591–600.

Janitza, S., Strobl, C., Boulesteix, A.-L. (2013). An AUC-based permutation variable importance measure for random forests. BMC Bioinformatics, 14.

Langley, P. (2000). Crafting papers on machine learning. In: Langley, P. (Ed.), Proc. of 17th International Conference on Machine Learning. ICML-2000. Stanford University, Stanford.

Larivière B., Van den Poel D. (2004). Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services. Expert Systems with Applications. 27 (2), 277-285.

Larivière, B., Van den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. Expert Systems with Applications. 29 (2), 472-484.

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

Liaw A. & Wiener M. (2002). Classification and Regression by randomForest. R News 2(3), 18--22.

Lindeman, M., Verkasalo, M. (2005). Measuring values with the short Schwartz's value survey. Journal of Personality Assessment. 85 (2), 170-178.

Ling, CX; Huang, J; Zhang, H. (2003). AUC: A better measure than accuracy in comparing learning algorithms. In Xiang, Y; ChaibDraa, B, Lecture Notes in Artificial Intelligence, Advances in Artificial Intelligence. 16th Conference of the Canadian-Society-for-Computational-Studies-of-Intelligence, Halifax, Canada, Jun 11-13, 2003, Volume 2671, pp 329-341.

Luo, T., Kramer, K., Goldgof, D.B., Hall, L.O., Samson, S., Remsen, A., et al. (2004). Recognizing plankton images from the shadow image particle profiling evaluation recorder. IEEE Transactions on Systems, Man and Cybernetics Part B-Cybernetics, 34 (4), 1753–1762.

Maloof, M.A. (2003). Learning When Data Sets Are Imbalanced and When Costs Are Unequal and Unknown, Proceedings of the International Conference on Machine Learning, Workshop Learning from Imbalanced Data Sets II.

Masand, B., Piatetsky-Shapiro G. (1996). A comparison of different approaches for maximizing business payoff of prediction models. In proceedings of KDD-96 (Portland, OR, Aug 1996), AAAI/MIT press, 195-201.

Miguéis, V.L., Camanho, A., Falcão e Cunha, J. (2013). Customer attrition in retailing: An application of Multivariate Adaptive Regression Splines. Expert Systems with Applications. In Press.

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

Piatetsky-Shapiro G., & Masand, B. (1999). Estimating Campaign Benefits and Modeling Lift. In Proceedings of KDD-99, San Diego, CA, 185-193.

Provost, F., Fawcett, T., & Kohavi, R. (1998). The case against accuracy estimation for comparing induction algorithms. In: J., Shavlik (Ed.), Proc. of 15th International Conference on Machine Learning. ICML-1998. Morgan Kaufman, San Francisco, CA.

Reichheld, F.F. (2003). The one number you need to grow. Harvard Business Review, 81(12), 46.

Ros, M., Schwartz, S.H., & Surkiss, S. (1999). Basic individual values, work values, and the meaning of work. Applied Psychology-An International Review. 48 (1), 49-71.

Rossiter, JR. (2002). The C-OAR-SE procedure for scale development in marketing. International Journal of Research in Marketing. 19 (4), 305-335.

Sheehan, K.B. (2001). E-mail Survey Response Rates: A Review. Journal of Computer-Mediated Communication, 6(2), 0.

Solinger, O.N., van Olffen W. & R. A. Roe (2008). Beyond the three-component model of organizational commitment, The Journal of applied psychology, 93, 1, 70-83.

Steenburgh, T. J., Ainsle, A., Engbretson, P. H. (2003). Massively categorical variables, revealing the information in ZIP codes. Marketing Science. 22, 40–57.

Sun, Y., Kamel, M.S., Wong, A.K.C., and Wang Y. (2007). Cost-Sensitive Boosting for Classification of Imbalanced Data. Pattern Recognition, 40 (12), 3358-3378.

Thomas, J.S. (2001). A methodology for linking customer acquisition to customer retention. Journal of Marketing Research. 38 (2), 262–268.

Thorleuchter D., Van den Poel D. & Prinzie A. (2012). Analyzing existing customers' websites to improve the customer acquisition process as well as the profitability prediction in B-to-B marketing. Expert Systems with Applications. 39 (3), 2597-2605.

Thorleuchter, D., & Van den Poel, D. (2012). Predicting e-commerce company success by mining the text of its publicly-accessible website. Expert Systems with Applications, 39(17), 13026-13034.

Van den Poel, D. (2003). Predicting mail-order repeat buying: Which variables matter? Tijdschrift Voor Economie En Management. 48 (3), 371–403.

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

Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics. 1, 80–83.

Wong W. K., Leung S. Y. S., Guo Z. X., Zeng XH., & Mok, P. Y. (2012). Intelligent product cross-selling system with radio frequency identification technology for retailing. International Journal of Production Economics. 135(1), 308-319.

Zaichkowsky, JL. (1994). The Personal Involvement Inventory- Reduction, Revision, and Application to Advertising. Journal of Advertising. 23 (4), 59-70.