Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 32 (2007) 277–288
Expert Systems with Applications
www.elsevier.com/locate/eswa

CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services

Jonathan Burez, Dirk Van den Poel *

Ghent University, Faculty of Economics and Business Administration, Department of Marketing, Hoveniersberg 24, B-9000 Ghent, Belgium

Abstract

The early detection of potential churners enables companies to target these customers using specific retention actions, and subsequently increase profits. This analytical CRM (Customer Relationship Management) approach is illustrated using real-life data of a European pay-TV company. Their very high churn rate has had a devastating effect on their customer base. This paper first develops different churn-prediction models: the introduction of Markov chains in churn prediction, and a random forest model are benchmarked to a basic logistic model.

The most appropriate model is subsequently used to target those customers with a high churn probability in a field experiment. Three alternative courses of marketing action are applied: giving free incentives, organizing special customer events, obtaining feedback on customer satisfaction through questionnaires. The results of this field experiment show that profits can be doubled using our churn-prediction model. Moreover, profits vary enormously with respect to the selected retention action, indicating that a customer satisfaction questionnaire yields the best results, a phenomenon known in the psychological literature as the 'mere-measurement effect'.

1. Introduction

The pay-TV company offers premium-channel content on a variety of topics and live broadcasts using an encrypted signal. The programming of the channel is mainly based on recent movies that have not been broadcasted yet (on free TV), as well as main sports events. In addition, it also offers a variety of information, home-made programs and series. During the early 1990s, the pay-TV company grew expansively and obtained a considerable customer base by 1996. After a few years of stagnation, 2001 marked the start of a constant decrease in membership (see Fig. 1). This significant decrease in membership urged management to enhance the relationship-building process with their customers. This move was motivated by the specific nature of pay-TV broadcasting, which is characterized by fairly high fixed costs due to high infrastructure investments (settop box, proprietary technology) as well as high broadcasting costs.

This paper shows how a company operating on a subscription basis can remedy high attrition rates. We apply different binary classification techniques on this customer-churn case: logistic regression, Markov chains and random forests. We choose the most appropriate model for our case, and justify that choice.

Next to developing a churn-prediction model, a company should also test its attrition prevention strategy. Any such strategy has to be implemented through specific marketing actions. This section describes three such courses of action, which have been implemented in the present case-study, but which can be applied to subscription services in general. The action types are (1) giving free incentives (enhancing the service), (2) organizing special events to pamper customers, and (3) obtaining feedback

* Corresponding author. Tel.: +32 9 264 89 80; fax: +32 9 264 42 79.
E-mail addresses: Jonathan.Burez@Ugent.Be (J. Burez), Dirk.VandenPoel@Ugent.Be (D. Van den Poel).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2005.11.037

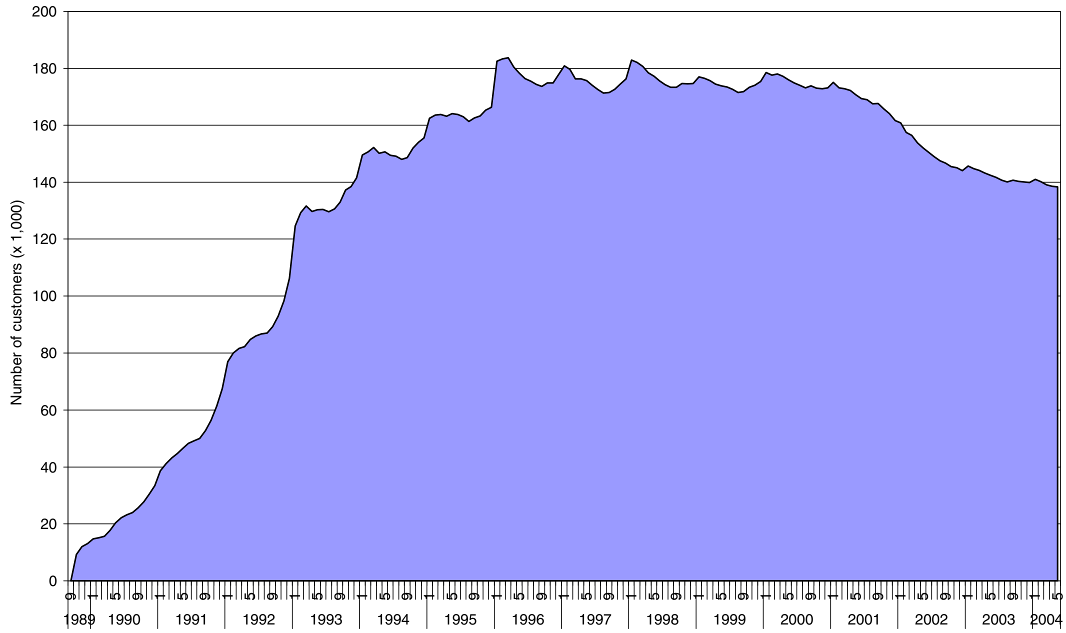

Fig. 1. The pay-TV company grew expansively during the early 1990s, but after a few years of stagnation, 2001 marked the start of a constant decrease in membership.

on customer satisfaction through questionnaires. The results of these strategies will be reviewed in the context of the pay-TV market, but we feel confident that this approach can be generalized to any professional, membership, or subscription services business; e.g. internet service providers, utility providers, telephone services (mobile, long distance...), newspapers, journals and magazines, web-based information delivery, health care, financial services, computing services, insurance.

After pointing out the importance of managing customer churn, this paper describes the model-building process and evaluates different proposed models. Next, the attrition-prevention strategies and the field experiment are described (see Fig. 2). The discussion of the results includes a quantification of profits for the pay-TV company. In addition, the transfer of the benefits of the model to other subscription-dependent industries is highlighted and directions for further research are provided.

2. Customer retention

An understanding of how to manage customer relationships effectively has become an important topic for both academics and practitioners in recent years. Although the concept of relationship marketing is not new, organizations have recently started to focus on identifying and retaining profitable long-term customers.

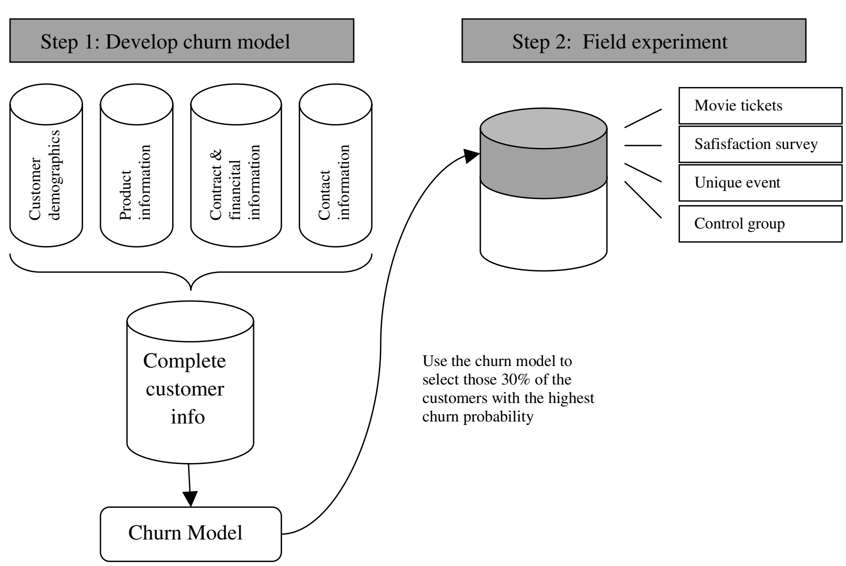

Fig. 2. The different steps in developing a churn prevention strategy.

It is indeed considerably more profitable to keep and satisfy existing customers than to constantly renew a customer base characterized by a high attrition rate (Reinartz & Kumar, 2003). Recently, Reinartz, Krafft, and Hoyer (2004) showed that the more firms engage in implementing CRM processes, especially at the initiation and maintenance stage, the better they perform.

A widespread concern exists about customer retention in service industries. The identification of those customers prone to switching carries a high priority (Keaveney & Parthasarathy, 2001).

3. Managing customer churn

There are two basic approaches to managing customer churn. *Untargeted* approaches rely on superior product and mass advertising to increase brand loyalty and retain customers. *Targeted* approaches rely on identifying customers who are likely to churn, and then either provide them with a direct incentive or customize a service plan to stay. There are two types of targeted approaches: reactive and proactive. Adopting a reactive approach, a company waits until customers contact the company to cancel their (service) relationship. The company then offers the customer an incentive, for example a rebate, to stay. Adopting a proactive approach, a company tries to identify customers who are likely to churn at some later date *in advance*. The *company* then targets these customers with special programs or incentives to keep the customer from churning. Targeted proactive programs have potential advantages of having lower incentive costs (because the incentive may not have to be as high as when the customer has to be "bribed" not to leave at the last minute) and because customers are not trained to negotiate for better deals under the threat of churning. However, these systems can be very wasteful if churn predictions are inaccurate, because then companies are wasting incentive money on customers who would have stayed anyway. That is why the goal is to predict customer churn as accurately as possible (Larivière & Van den Poel, 2004; Prinzie & Van den Poel, forthcoming).

4. Model development

Pay-TV is a subscription service requiring the customer to pay a fixed monthly contribution only; there is no extra per-minute-of-use charge (as is the case for instance in the mobile phone industry). Pay-per-view is considered as a separate branch of the market, and is not considered here. At the pay-TV company, all customers have a 12-month subscription. Canceling within that 12-month period is not allowed, as is prematurely reporting that a subscription will not be renewed: customers have to report that they will not renew their subscription during the last month of the 12-month contract. If nothing is reported, the subscription is automatically renewed for a period of 12 months.

We extracted data for this study from the data warehouse of the pay-TV company—a single integrated source of information combining data from all departments and services. It contains information about all customers that the pay-TV company has ever had. We used Oracle PL/SQL for data preparation and manipulation, and MATLAB to perform the statistical analyses.

We considered all active customers on February 28th 2002 (sampling date). Active customers are customers who had a subscription at that moment in time and were not suspended for defaulting on their payments or for another reason. The period between the startup and the sampling date is called the estimation period (see Fig. 3).

We constructed a dataset out of the database that contains information on active customers from their first subscription at the pay-TV company until the sampling date. This information is translated into explanatory variables. The following information is captured in those explanatory variables: all information about the customer's current and previous subscriptions, socio-demographics, information about reminders sent to the customer, and details on the contacts made with the customer (see Appendix A). Given the broadcasting technology used by the pay-TV company, they have no information on viewing patterns of individual viewers.

Those explanatory variables are both original (e.g. price of current subscription) and derived variables (e.g. a scaled indicator of bad payment behavior). Instead of performing a data imputation technique, we added a dummy variable for all variables having missing values. As shown by Buckinx and Van den Poel (2005) dummy variables, indicating that certain information is missing, can tell you a lot about a customer and can be very good predictors.

Previous studies confirm the importance of the product history of a customer when predicting future behavior (Van den Poel & Larivière, 2004). Therefore, we include all kinds

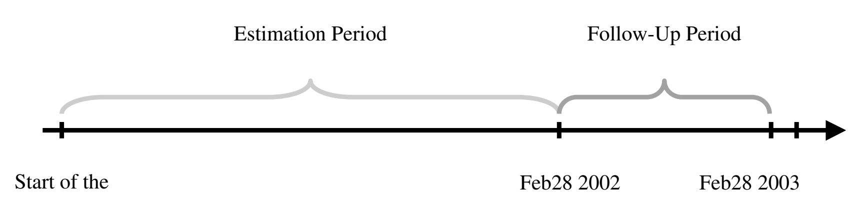

Fig. 3. For our customer retention model, we collected all information available about all customers with an active subscription on February 28 2002. In the follow-up period, we checked whether the customer churned or not.

of variables that describe the product history into the churn-prediction model: the type of product the customer started with, the type of product the customer currently holds, the number of upgrades/downgrades, etc. However, the problem with all those variables is that they do not take into account the sequence of events that took place. That is why the dataset also contains four variables based on Markov chains, which capture sequential information (see Modeling Techniques for more information).

In addition, the dataset contains the binary variable that has to be predicted: does a subscriber churn. Different to what is done in churn-prediction modeling in the mobile phone industry, which predicts monthly churn, churn is considered here during the year after the sampling date. This period of 1 year, starting with the sampling date, is called the follow-up period. If a customer did not renew the subscription (commercial churn), or stopped paying (financial churn) or died, or moved (involuntary churn) the customer is considered a churner. Information on involuntary churn is not available and can consequently not be eliminated from the dataset in the investigation. The dataset from the pay-TV company exhibited a churn rate of 15%.

After building a churn-prediction model on an estimation sample, marketers may use it to predict the future churn behavior of their customers. To assess the precision of such predictions, one has to use a test sample. The principle is that this test set has not been used for constructing the classification model, and will therefore give reliable indicators of performance for marketers. Indeed, if one assesses a classifier on its respective estimation sample, the judgment could be biased because of overfitting: a classifier fits the idiosyncrasies of the estimation set too closely. It leads to lower error rates on the estimation sample, but at the same time to much higher error rates on the test set.

Hence the dataset, created as explained above, is split randomly into an estimation set (60%) and a test set (40%). We want a churn rate of 15% in both estimation and test sample, and thus the stratification is performed on the churn variable (Neslin, Gupta, Kamakura, Lu, & Mason, 2004).

5. Modeling techniques

5.1. Logistic regression

Logistic regression modeling is a well-known technique. It is very appealing for five reasons: (1) logit modeling is well known, conceptually simple and frequently used in marketing (Bucklin & Gupta, 1992), especially at the level of the individual consumer (Bucklin, Gupta, & Han, 1995; Neslin et al., 2004); (2) the ease of interpretation of logit is an important advantage over other methods (e.g. neural networks); (3) logit modeling has been shown to provide good and robust results in general comparison studies (Neslin et al., 2004) and (4) more specifically in database marketing, it has been shown by several authors (Levin & Zahavi, 1998) that logit modeling may even outperform more sophisticated methods; (5) whereas survival analysis would give us the estimated time to event (in this case churn), we are only interested in the estimated probability of the event within 1 year in a first phase.

5.2. Markov chains

Let's consider the following example: two customers, both having a length of relationship of 10 years. Assume there are two types of products (A and B), and the two customers both upgraded their product once (from type A to type B): the first one did so 8 years ago, the other one did that 1 year ago. The product both customers started with is the same, the current product of both customers is the same, and the number of product upgrades is the same for both customers; but the influence of those upgrades on their churn probability just isn't the same! The time aspect of those changes is important too. Moreover, recent events are more important when analyzing churn behavior. Therefore we used Markov chains.

A Markov chain (MC) is a probabilistic technique used to represent correlations between successive observations of a random variable (Berchtold, 2001). This sequence analysis technique is a form of time-series modeling, and was introduced at the beginning of the 20th century by Andrej Andreevic Markov. It is used in many disciplines, including meteorology, geography, biology, chemistry, physics, social sciences and music. In marketing, it has already been successfully applied in modeling purchases of financial services (Prinzie & Van den Poel, 2006), predicting website purchases using clickstream data (Montgomery, Li, Srinivasan, & Liechty, 2004) or predicting software performance (Bai, Hu, Xie, & Ng, 2005; Durand & Gaudoin, 2005). It has, to our knowledge, never been used for churn prediction, neither directly nor in an indirect way (as additional input variable to a classification model).

The product sequences of the example are shown in Table 1. A fourth-order MC (see Appendix A) then computes what is the probability that a customer will have

Table 1
Not only does it matter that a customer upgraded his product, but the moment in time when he did so is equally important

	$t-10$	$t-9$	$t-8$	$t-7$	$t-6$	$t-5$	$t-4$	$t-3$	$t-2$	$t-1$
Customer 1	A	A	B	B	B	B	B	B	B	B
Customer 2	A	A	A	A	A	A	A	A	A	B

A certain product type sequence can indicate an approaching churn event.

product type A this year (p_A), or product type B (p_B), or that customer will have no product at all this year (p_{Churn}), given his/her product history over the last 4 years. The two customers in our example have a different product history over the last 4 years, and thus have a different p_{Churn}. We include this MC probability estimate that the customer will churn as an additional predictor in the churn-prediction model.

5.3. Random forests

Decision trees have become very popular for solving classification tasks, because they can deal with predictors measured at different measurement levels (including nominal variables), and because of their ease of use, and interpretability (Duda, Hart, & Stork, 2001, Chapter 8). However, they also have their disadvantages such as lack of robustness and suboptimal performance (Dudoit, Fridlyand, & Speed, 2002). Recently, many of these disadvantages have been dealt with by creating a large number of trees and letting them vote for the most popular class, labeled forests (Breiman, 2001).

Several successful paths have been explored how to grow ensembles of trees: (1) bagging, where to grow each tree a random selection (without replacement) is made from the examples in the training set (Breiman, 1996); (2) random split selection, where at each node the split is selected at random from among the K best splits (Dietterich, 2000); (3) random subspace method, which does a random selection of a subset of predictors to grow each tree. For a comparison of those 3 methods, see Hamza and Larocque (2005). In this paper, we select the random forests as proposed in Breiman (2001) which uses the latter strategy.

Since random forests are a recent technique, we would like to highlight its attractiveness for its consistently high performance (Larivière & Van den Poel, 2005).

6. Model accuracy

It is important to know if the customers that would be targeted with special retention actions are indeed the most inclined to churn. The most commonly used way to quantify the accuracy of a predictive model is by calculating its lift, e.g. among say the 10% of customers predicted as most likely to churn, what percentage of them actually churn, relative to the percentage of all customers who churn (see Fig. 4). The higher the lift, the more accurate the model, and intuitively, the more profitable a targeted proactive churn management program will be.

Fig. 5 shows the cumulative gains chart. This tells us what cumulative percentage of churners is accounted for by the x% of customers predicted to be most likely to churn. Obviously, one would want cumulative lift to be as high as possible. With a churn rate of 15.13%, and in a perfect model the 15.13% of customers predicted most likely to churn would account for 100% of churners. Hence, the maximum lift is then 100/15.13 or 6.61. Note that the diagonally increasing line with unit slope in Fig. 5 depicts the random model. If predictions are random, the top 25% customers will account for 25% of churners. The rule previously used by the pay-TV company selects random customers out of all customers that have length of relationship shorter than one and shorter than 2 years; this gives a lift of 1.9 (<1 year) and 1.4 (<2 years), or by selecting all customers in their first 2 years of subscription (20.16% of the customer base) you reach almost 33.1% of all churners.

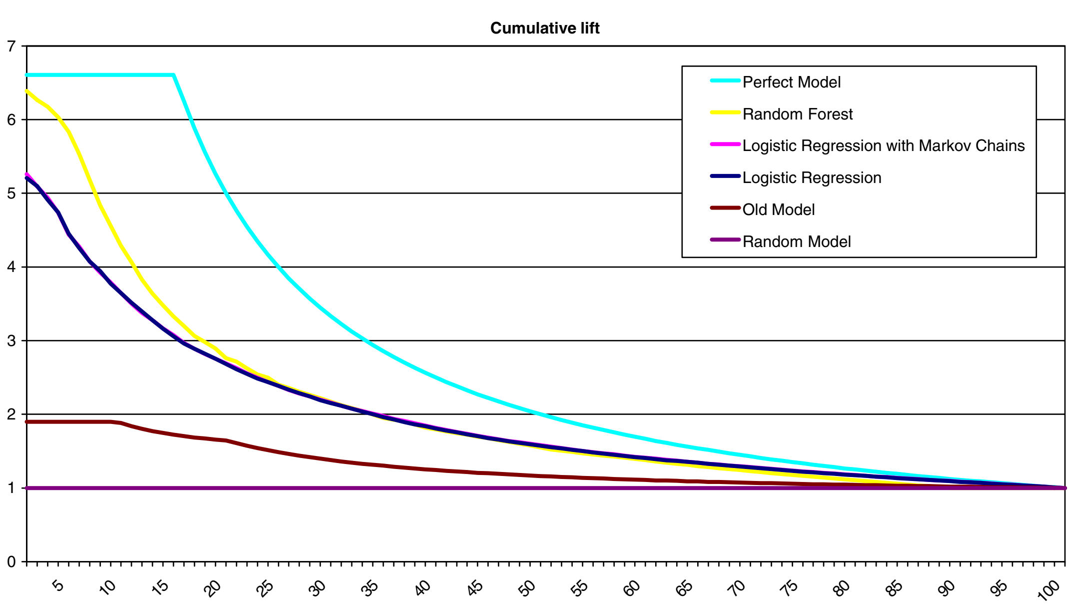

Fig. 4. Lift is the most commonly used performance measure of a churn prediction model. The cumulative lift curve shows the difference in effect between using a model and not using any model.

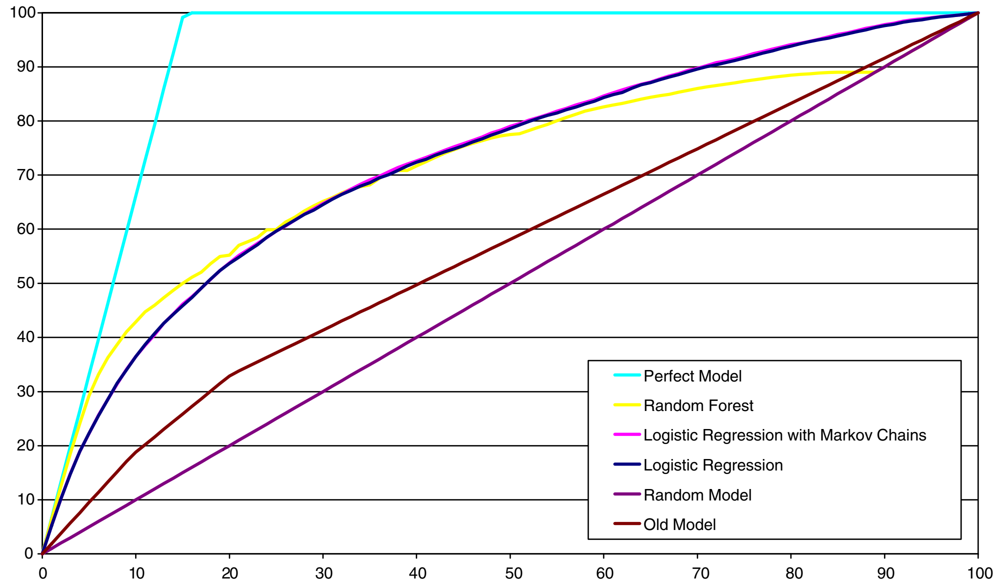

Fig. 5. Next to lift, the AUC coefficient gives a clear indication of the performance of a churn-prediction model. The bigger the surface between the cumulative gains curve of a model and the diagonal, the better that model can distinguish between churners and non-churners.

The goal is to generate a lift curve separated as far as possible from the random lift curve. This is measured by the AUC coefficient—the Area Under the receiver operating Curve—and is related to the area between an entry's cumulative gains curve and the random gains curve. The AUC coefficient can range from .50 to 1. A higher AUC value reflects better separation between the achieved gains curve and random gains. Note that top-decile lift focuses on lift among the top 10% of customers, whereas AUC measures lift along the entire continuum.

Percentage Correctly Classified (PCC), also known as accuracy is another measure of model accuracy. It computes the ratio of correctly classified customers to total number of customers to be classified.

Results presented in Table 2 lead us to conclude that predicting customer churn is a viable strategy: first, PCC performance of 0.879 for random forests on a test sample (i.e., on cases not used during estimation) should be benchmarked to Morrison's (1969) proportional chance criterion of 0.7432 ($=0.1512^2 + (1 - 0.1512)^2$) or the majority prediction rule of 0.8487 ($=1 - 0.1512$); second, AUC performance of 0.7693 (again for random forests on the test sample) exceeds the 0.5 benchmark of the null model.

When comparing the different classification techniques, random forests clearly outperform the other techniques for cut off values that select less then 25% of the customers as churners (see Fig. 5). For a higher cut off value, the logistic regression models perform as good, if not better than the random forest model. As the experiment obliged us to select 30% of the customers, this was a first reason to opt for the logistic regression model with Markov chains. The second reason comes from the fact that random forests tend to give small groups of customers the same churn probability. This results into bumpy lift curves (see e.g. percentile 20 in the Fig. 4). The experiment used the predictions of the logistic regression model with Markov chains for the selection of defection-prone customers.

To investigate whether our model is robust and stable, we tested it on a more recent dataset we received from the pay-TV company. For this test, we scored all the active customers of the company on February 28, 2003. Since we only had a follow-up period of 3 months in this test (March, April and May 2003), we only selected those

Table 3
A comparison of the test results shows that the out-of-period performance of the churn-prediction model is as good, if not better than that on the holdout sample

Data	AUC
Holdout sample (same-period)	0.7673
Out-of-period sample	0.7709

Table 2
The new churn-prediction model clearly outperforms the old heuristic

Cut off	10%		30%		Overall
Performance measure	PCC	Cumulative lift	PCC	Cumulative lift	AUC
Random forest	0.879	4.29	0.743	2.17	0.7693
Logistic regression	0.859	3.65	0.744	2.15	0.7651
Logistic regression with Markov	0.859	3.64	0.745	2.16	0.7673

customers who had to renew their subscription in those months. The following numbers (see Table 3) show that the model performed equally well on the new data set.

7. Customer retention program

Having a churn prediction score is one thing, using it to effectively reduce churn is something else. The pay-TV company is using the churn predictions proactively both continuously and on a monthly basis. Continuously in two ways: first of all, the predictions determine the order of answering incoming phone calls. Customers with the highest churn probability will be answered first, while the others remain on hold. Secondly, the score is one of the parameters that appear on the screen of the telemarketer of the contact center when a customer calls in with e.g. a complaint or a question. This enables the operator to take the appropriate action: a customer with a high churn probability, who calls in very unsatisfied, can be offered one month for free, or something similar.

On a monthly basis, the company uses the churn predictions to target customers with customer-retention actions. As most customers have a 12-month subscription, the best strategy is to target (a part of) all customers that have to renew in a certain month, and approach them a few months before their expected renewal. Possible customer-retention actions range from free movie theater tickets to free tickets to a sporting event, from invitations to an avant-première to giving the customer one month subscription for free (in case of subscription renewal), from giving some gadgets to calling the customer for a satisfaction questionnaire.

8. Field experiment to reduce churn

Until September 2003, the only incentive that was used for proactive targeting, was a small satisfaction questionnaire. With a better churn prediction score available at that time, the pay-TV company wanted to test if other incentives would work better to woo back potential churners to the company. The incentives were chosen by availability (e.g. there were too few tickets for sporting events to use them as an incentive) and by cost to the company (the concept of one month for free pay-TV at renewal was rejected by management because it was considered too expensive). Three incentives remained: free movie tickets, an invitation to a unique event, and the satisfaction questionnaire mentioned earlier.

Table 4
The field experiment proved that our churn-prediction model obtains the desired results

	Total	Churn rate (%)
Top 30%[a]	7326	14.02
Bottom 70%	16859	5.46
Total	24185	8.05

[a] Including those customers that were targeted with a retention action.

A large-scale field experiment was conducted to test the influence of those three customer retention actions on relational behavior. All pay-TV customers who had to renew their subscription in December 2003, and of whom the company had a telephone number, were candidates for one of the actions.

24,185 customers satisfied this condition. We selected 7350 customers with the highest churn probability (labeled top 30%). Table 4 shows that our churn-prediction model clearly was able to identify customers with a much higher attrition rate. Note that the overall churn rate of these 24,185 subscribers (8.05%) differs substantially from the ones for all customers. This is partially due to a seasonal effect (fewer customers churn during the winter months), next to the influence of the retention actions (see further).

Those customers were divided into four groups via systematic sampling (Churchill & Iacobucci, 2002, p. 484): one group of 1050 customers would receive two free movie tickets, a second group of 2100 customers would receive an invitation for two to a unique event, a third group of 2100 customers was asked to respond to a satisfaction questionnaire by telephone; the fourth group (2100 customers) served as a control group. The differences between those figures (column 'A priori selected' in Table 5) and the number of customers actually contacted (column 'Contacted' in Table 5) are due to several reasons: the address or telephone number available in the database was incorrect, the customer left the company (died, moved, stopped paying...) between the calculation of the churn scores and the selection (a one-month period).

There were 650 two-person tickets for the unique event. The invitation to the unique event was a two-step process: 2087 customers were asked whether they were interested,

Table 5
This table summarizes the results of the field experiment

	Action	A priori selected	Contacted	Participated	Churn rate (%)	Δ (%)	Prob(χ^2)
Top 30%	Control	2100	2087	N/A	15.67		
	Unique event	2100	1841	354	12.87	2.79	0.0127
	Tickets	1050	998	N/A	12.12	3.54	0.0090
	Satisfaction questionnaire	2100	1992	627	10.94	4.72	<.0001
Bottom 70%			15679		5.40		

The influence of the retention actions is visible through the lower churn rate compared to the control group. Δ = Absolute difference in churn rate with the control group. Prob(χ^2) = Probability.

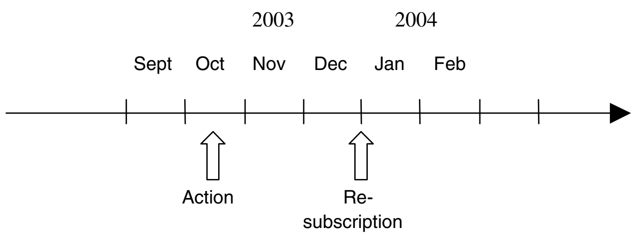

Fig. 6. For our field experiment, we used the above timeline.

and had to reply if they were; they then obtained two free passes. The final turnout at the event was low, as can be seen in the column labeled 'Participated' in Table 5.

The questionnaire was conducted by telephone by a specialized firm, two and a half months before the moment in time on which the customers had to renew their subscription (October 17, 2003). The interviewer requested participation from the "person in your household that was the main promoter of subscribing to pay-TV". The survey took about 7 min to complete on average, and inquired into a number of motivations, usage and satisfaction questions eliciting respondents' evaluation of specific features (e.g. price, programming, movies broadcasted, sports programs...) followed by the general question, "Overall, how satisfied are you with the pay-TV company?" The 'Participated' column in Table 5 means that the customer was reached, and participated in the questionnaire.

Churn is defined in this case as a customer that did not renew his subscription as expected at the end of December, two and a half months after the action (see Fig. 6).

9. Discussion of the field experiment

All three different action types reduce the churn rate significantly. Table 5 reveals that the churn rate for the three action groups is significantly lower than the churn rate for the control group. For the satisfaction questionnaire, the reduction amounts to almost 5% (10.94% opposed to 15.67% for the control group). Giving tickets or inviting customers to a unique event also reduces churn, but to a lesser degree (decreases by 3.54% and 2.79% respectively). While those differences between the action types themselves seem to be substantial as well, none of them differ significantly.

10. Mere-measurement effect

The influence of the satisfaction questionnaire might be somewhat surprising. This phenomenon, however, is known in literature as the mere-measurement effect. Firms routinely collect data by conducting customer surveys to determine how satisfied their customers are with the firm and its offerings, and to evaluate different aspects of the firm's marketing mix. An important assumption underlying such surveys is that existing opinions are elicited from customers and these do not influence their subsequent behaviors. However, when responding to surveys, respondents are often induced by the measurement process to form judgments that would otherwise not be formed, which in turn influences subsequent responses and behaviors, making them more consistent with the expressed judgments (Feldman & Lynch, 1988). Moreover, such measurement-induced judgments are especially likely in contexts such as satisfaction or purchase intention surveys, in which most respondents are unlikely to have formed these judgments spontaneously beforehand or, indeed, given the issue much prior thought (Kardes, 1988).

A related stream of research has shown that measurement-induced judgments, specifically the elicitation of behavioral intentions, can change respondents' subsequent actions. In the marketing literature, several studies have shown that the process of measuring purchase intentions changes subsequent behaviors (see Morwitz & Fitzsimons (2004) for a recent review), a phenomenon that has been called the mere-measurement effect (Morwitz, Johnson, & Schmittlein, 1993) and the self-prophecy effect (Spangenberg & Greenwald, 1999). Recent research (Chandon, Morwitz, & Reinartz, 2004; Dholakia & Morwitz, 2002) has shown that the effect of measurement-induced judgments is persistent over time and broad in scope.

11. Profit quantification

Now that we know how much of the would-be churners are wooed back to the company by a specific action, we want to know what the profit would be of the model, compared to the rule that is now used, given a certain action. We use Neslin's approach (2004) to calculate the profit contribution, defined as a function of lift. We attach some plausible Euro values to the gains from higher predictive accuracy by filling in actual values for the parameters, and calculating the profit contributed by the churn management program, given the success rate of the retention actions tested.

We first define the following terms:

N	total number of customers
α	fraction of customers who are targeted for the churn-prevention program
β	fraction of targeted customers who are in fact potential churners
δ	cost of the customer incentive to the firm. For example, if the company offers customers a €50 rebate, the cost is €50
γ	fraction of targeted potential churners who are wooed back to the company by an incentive: the success rate of an incentive
c	cost of contacting customers to offer them an incentive
CLV	customer Lifetime Value, the present value of all future profits generated from a customer. The common approach, used here, is to assume that we know how long a customer will be with the

company and then generate a discounted cash flow for that time period (Gupta & Lehmann, 2003). $\mathrm{CLV} = \sum_{t=1}^{n} \frac{m_t}{(1+i)^t}$ where m is the margin of contribution for each customer in a given time period t (e.g. year), i is the discount rate, and n is the period over which the customer is assumed to remain active

$\prod$ profit contributed by the churn-prevention program

Given these definitions, the profit contributed by the churn-prevention program is

$$\prod = N\alpha[\beta\gamma(\mathrm{CLV} - c - \delta) + \beta(1-\gamma)(-c) + (1-\beta)(-c-\delta)] \quad (1)$$

The term β reflects the accuracy of the model, and is related to the concept of lift as follows. Let:

β_0 The fraction of all the company's customers who will churn

λ "Lift" from the predictive model

Then, we can express β as

$$\beta = \lambda\beta_0 \quad (2)$$

Substituting Eq. (2) into Eq. (1) and re-arranging terms, we get

$$\prod = N\alpha\{(\gamma\mathrm{CLV} + \delta(1-\gamma))\beta_0\lambda - \delta - c\} \quad (3)$$

The incremental gain in profit from a unit increase in predictive accuracy "λ" is the slope of Eq. (3), namely

$$\mathrm{GAIN} = N\alpha\{[\gamma\mathrm{CLV} + \delta(1-\gamma)]\beta_0\}$$

We attach some plausible Euro values to these terms:

N 150,000. The pay-TV company has 150,000 subscribers.

α 0.10. We assume the company will contact 10% of its customers in the churn management campaign

δ €17.5, €5 and €0. In the experiment, the pay-TV company offered an invitation to a unique event worth €17.5, movie tickets worth €5 or nothing (just asking about satisfaction)

γ 0.18,0.23, and 0.30. There are no published statistics on how many would-be churners accept this or that offer and stay with the company. Our experiment pointed out that 18% of would-be churners was converted by inviting them to a unique event, 23% by giving them free movie tickets, and 30% by interviewing them about their satisfaction

c €1, €1 and €4.55. It costs €1 to contact somebody to invite them to the unique event or to offer them free ticket. This is done via a separate mailing piece. Interviewing a customer costs €4.50 per interview, plus a fixed monthly reporting cost

Table 6
The higher the lift of a churn-prediction model, and the better a certain incentive woos back customers, the higher profits of the retention program will be

Model	Lift	Customer retention action		
		Unique event	Tickets	Satisfaction questionnaire
Churn-prediction model	3.65	2,105,180	2,809,590	3,752,464
Old model	1.88	951,416	1,403,487	1,899,679
Random model	1	377,794	704,408	978,521

β_0 0.1513. The average yearly churn rate for the customers in our data is 15,13%, so we use this as the baseline churn level

CLV €1528. Monthly revenues per customer for the pay-TV company are €38. If we use a 5% annual discount rate, assume 10% variable costs and a median survival time of 4 years, the lifetime value of the customer is an estimated €1528. Note that this calculation does not include the fixed cost of providing service. We assume the goal of the company is revenue growth to cover these costs

Using the above assumptions, Table 6 displays the profit contributed by the churn management program, given the success rate of the retention actions tested. The table shows that even without a churn-prediction model, customer-retention actions (in that case untargeted) pay off. The rule currently used by the pay-TV company (target those customers in their first or second year of subscription) almost doubles the company's profits of its current churn-prevention program. Compared to that rule, our model almost doubles the profits of the current churn-prevention program after submitting a satisfaction questionnaire to the most fragile 10% of its customers.

12. Transferability of the approach

The results indicated above can be transferred to other types of companies which provide a continuous service in exchange for a subscription fee. For such subscription service companies that allocate fixed costs across large numbers of customers and that depend on receipt of fixed membership or access fees on a continuous basis, customer switching behavior can have a particularly devastating effect on the bottom line (Keaveney & Parthasarathy, 2001).

When depending on usage fees, a company wants to retain the most profitable customers; if not, a company wants to retain all customers. This has to be taken into account in the customer relationship management of a company.

An indispensable condition for customer-retention models is that the company possesses an information-dense database. This can be the company database, with information on customer demographics, customer history and/or usage information (conditional upon whether the information is available), possibly complemented with purchased external data (e.g. neighborhood information).

13. Directions for further research

The model-selection procedure could be improved by using a leap and bound procedure similar to the one proposed by Furnival and Wilson (1974) to reduce the number of variables, and identify that set of variables that gives the best explanatory power.

There might be considerable heterogeneity in the churn-prediction model with respect to the magnitude and direction of the response parameters. As length of relationship appears to be the best predictor variable, it might be a good idea to split up the population (based on length of relationship) and define more than one model.

One could enquire the effect of other retention actions, next to the three actions mentioned in this article. Moreover, the long-term effect of the mere-measurement effect should be investigated further. The satisfaction questionnaire could be introduced as a normal, standard procedure (interview all new customers after 9 months—so 3 months before the first time they have to renew their subscription—instead of only the "at risk" ones). This would also enable an enhanced churn-prediction model. After all, the questionnaire provides the company with general viewing information and satisfaction figures, information that is not available in the database. First tests indicate that the answers to the questionnaire, or the knowledge that a certain customers refused the interview, will enhance the accuracy of the model.

With regard to the retention actions, it would be very interesting to investigate the influence of retention actions on all customers, and not only on defection-prone customer. Furthermore, knowing which defection-prone customer we have to target with which action would be an enormous advantage for the company.

14. Conclusions

The pay-TV company encountered huge churn rates at the beginning of the twenty-first century. We have shown in this article, what to do when a company is in such a situation. We developed a churn-prediction model, and targeted potential churners with three different churn-prevention actions, using that model. The empirical results of three alternative courses of action reveal that all three are equally effective to reduce customer attrition: (1) giving free incentives (enhancing the service), (2) organizing special events to pamper customers and (3) obtaining feedback on customer satisfaction through questionnaires. Given the fact that the latter has the added benefit of increasing 'knowledge' about the individual customer, it seems to be the most attractive one. In summary, our results show that, using the full potential of our churn-prediction model and the incentives available, the pay-TV company's profits of the churn prevention program would double when compared to its current model. We believe that a similar approach is applicable to most subscription or membership services.

Appendix A. Independent variables included in the churn-prediction model

Current subscription: the variables describing the active subscription on February 28, 2002:

- The number of months and the monthly payment of this subscription,
- The month of contract expiration (for seasonal effects),
- What product(s) a customer has,
- What option(s) a customer has (like sports, movies, porn, etc.),
- What type of decoder a customer has (digital or analogue),
- How a customer pays his bills (domiciling).

Socio-demographic variables: a group of variables that describe a subscriber:

- Age,
- Gender,
- Telephone (mobile, fixed phone, none),
- Region,
- Business (is the subscriber a company or a private person),
- Additional information.

Financial information: a group of variables that describe a customer's history of payments:

- The number of reminders,
- The type of reminders,
- The number of overdue accounts on February 28 2002 and its value,
- Elapsed time since the last reminder.

Previous subscriptions: information about the subscriber's past behavior, after excluding all the variables about bad payments and contacts between a subscriber and the pay-TV company:

- Length of relationship and actual length of subscription,
- Number of contracts a customer had over his relationship with the pay-TV company,
- Monetary value,
- Total discounts,
- Member-gets-member information,
- Recruitment and other campaigns,
- Cancellations,
- Subscription renewals,

- Place where first subscription was bought (display store, specialized store…),
- Probability of going to another product based on the past behavior of all customers (Markov chains, see B).

Contact variables: all direct contact between a customer and the pay-TV company:

- Surveys,
- Letters by type (about actions, about broadcasting information, etc.; personal as well as generic; incoming as well as outgoing),
- Calls by type (e.g. call concerning subscription, contract, additional information, games, technical problems).

Appendix B. Markov chains

A Markov chain (MC) is a model in which the current value (time t) of a variable Y taking values in $\{1,\ldots,K\}$ is fully explained by the knowledge of the value taken by the same variable at time $t-1$ (Berchtold, 2001). This model is summarized in a transition matrix C_1 giving the probability distribution of Y_t given any possible value of Y_{t-1}

$$C_1 = [P(Y_t = j \mid Y_{t-1} = i)] = [p_{ij}] = \begin{array}{c} \\ t-1 \\ 1 \\ \vdots \\ K \end{array} \begin{array}{c} t \\ \begin{array}{ccc} 1 & \cdots & K \end{array} \\ \left| \begin{array}{ccc} p_{11} & \cdots & p_{1K} \\ \vdots & \ddots & \vdots \\ p_{K1} & \cdots & p_{KK} \end{array} \right| \end{array}$$

Each row of C is a probability distribution summing to one. Since the current value is fully determined by the knowledge of only one past period, this model is said to be of order 1. More generally, a Markov chain of order f, $f >= 0$, is a model in which the current value is explained by all lags up to $t-f$. The transition matrix is then of a larger size. For instance, for $f = 2$ (second-order MC) and $K = 3$, we have (in the reduced form defined by Pegram (1980)):

$$R_2 = \begin{array}{cc} t-2 & t-1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 1 & 2 \\ 2 & 2 \\ 3 & 2 \\ 1 & 3 \\ 2 & 3 \\ 3 & 3 \end{array} \begin{array}{c} t \\ \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \left| \begin{array}{ccc} p_{111} & p_{112} & p_{113} \\ p_{211} & p_{212} & p_{213} \\ p_{311} & p_{312} & p_{313} \\ p_{121} & p_{122} & p_{123} \\ p_{221} & p_{222} & p_{223} \\ p_{321} & p_{322} & p_{323} \\ p_{131} & p_{132} & p_{133} \\ p_{231} & p_{232} & p_{233} \\ p_{331} & p_{332} & p_{333} \end{array} \right| \end{array}$$

References

Bai, C. G., Hu, Q. P., Xie, M., & Ng, S. H. (2005). Software failure prediction based on a Markov Bayesian network model. *Journal of Systems & Software, 74*(3), 275–282.

Berchtold, A. (2001). Estimation in the mixture transition distribution model. *Journal of Time Series Analysis, 22*(4), 379–397.

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

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

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

Bucklin, R. E., & Gupta, S. (1992). Brand choice, purchase incidence and segmentation: an integrated modeling approach. *Journal of Marketing Research, 29*(2), 201–215.

Bucklin, R. E., Gupta, S., & Han, S. (1995). A brand's eye view of response segmentation in consumer brand choice behavior. *Journal of Marketing Research, 32*(1), 66–74.

Chandon, P., Morwitz, V., & Reinartz, W. (2004). The short- and long-term effects of measuring intent to repurchase. *Journal of Consumer Research, 31*(3), 566–572.

Churchill, G. A., Jr., & Iacobucci, D. (2002). *Marketing research methodological foundations* (8th ed.). Mason, Ohio: South-Western Thompson Learning.

Dholakia, U. M., & Morwitz, V. G. (2002). The scope and persistence of mere-measurement effects: Evidence from a field study of customer satisfaction measurement. *Journal of Consumer Research, 29*(2), 159–167.

Dietterich, T. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization. *Machine Learning, 40*(2), 139–157.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. NY: Wiley.

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.

Durand, J., & Gaudoin, O. (2005). Software reliability modelling and prediction with hidden Markov chains. *Statistical Modelling: An International Journal, 5*(1), 75–93.

Feldman, J. M., & Lynch, J. G. Jr., (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behaviour. *Journal of Applied Psychology, 73*(3), 421–435.

Furnival, G. M., & Wilson, R. W. (1974). Regressions by leaps and bounds. *Technometrics, 16*, 499–512.

Gupta, S., & Lehmann, D. R. (2003). Customers as assets. *Journal of Interactive Marketing, 17*(1), 9–24.

Hamza, M., & Larocque, D. (2005). An empirical comparison of ensemble methods based on classification trees. *Journal of Statistical Computation & Simulation, 75*(8), 629–643.

Kardes, F. R. (1988). Spontaneous inference processes in advertising: the effects of conclusion omission and involvement on persuasion. *Journal of Consumer Research, 15*(2), 225–233.

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

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 forest and regression forest techniques. *Expert Systems with Applications, 29*(2), 472–484.

Levin, N., & Zahavi, J. (1998). Continuous predictive modeling: A comparative analysis. *Journal of Interactive Marketing, 12*(2), 5–22.

Montgomery, A. L., Li, S., Srinivasan, K., & Liechty, J. C. (2004). Modeling online browsing and path analysis using clickstream data. *Marketing Science, 23*(4), 579–595.

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

Morwitz, V. G., & Fitzsimons, G. J. (2004). The mere-measurement effect: Why does measuring purchase intentions change actual purchase behaviour? *Journal of Consumer Psychology, 14*(1&2), 64–73.

Morwitz, V. G., Johnson, E., & Schmittlein, D. (1993). Does measuring intent change behaviour? *Journal of Consumer Research, 20*(1), 46–61.

Neslin, S.A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. (2004). Defection detection: Improving predictive accuracy of customer churn models. Working paper.

Pegram, G. G. S. (1980). An autoregressive model for multilag Markov chains. *Journal of Applied Probability, 17*, 350–362.

Prinzie, A., & Van den Poel, D. (2006). Investigating purchasing patterns for financial services using Markov, MTD and MTDg models. *European Journal of Operational Research, 170*(3), 710–734.

Prinzie, A., & Van den Poel, D. (forthcoming). Incorporating sequential information into traditional classification models by using an element/position-sensitive SAM. *Decision Support Systems*.

Reinartz, W., & Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing, 67*(1), 77–99.

Reinartz, W., Krafft, M., & Hoyer, W. (2004). The CRM process: Its measurement and impact on performance. *Journal of Marketing Research, 41*(3), 293–305.

Spangenberg, E. R., & Greenwald, A. G. (1999). Social influence by requesting self-prophecy. *Journal of Consumer Psychology, 8*(1), 61–89.

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