Improved marketing decision making in a customer churn prediction context using generalized additive models

Kristof Coussement\textsuperscript{a,b,*}, Dries F. Benoit\textsuperscript{b}, Dirk Van den Poel\textsuperscript{b}

\textsuperscript{a}University College HUBrussel, Faculty of Economics and Management, Stormstraat 2, B-1000 Brussels, Belgium
\textsuperscript{b}Ghent University, Faculty of Economics and Business Administration, Department of Marketing, Tweekerkenstraat 2, B-9000 Ghent, Belgium

\textbf{A R T I C L E   I N F O}

Keywords: Customer Relationship Management (CRM) Churn modeling Marketing decision making Generalized Additive Models (GAM)

\textbf{A B S T R A C T}

Nowadays, companies are investing in a well-considered CRM strategy. One of the cornerstones in CRM is customer churn prediction, where one tries to predict whether or not a customer will leave the company. This study focuses on how to better support marketing decision makers in identifying risky customers by using Generalized Additive Models (GAM). Compared to Logistic Regression, GAM relaxes the linearity constraint which allows for complex non-linear fits to the data. The contributions to the literature are three-fold: (i) it is shown that GAM is able to improve marketing decision making by better identifying risky customers; (ii) it is shown that GAM increases the interpretability of the churn model by visualizing the non-linear relationships with customer churn identifying a quasi-exponential, a U, an inverted U or a complex trend and (iii) marketing managers are able to significantly increase business value by applying GAM in this churn prediction context.

\section{1. Introduction}

Today, the business environment is characterized by fierce competition and saturated markets. In this context, companies increasingly derive revenue from the creation and enhancement of long-term relationships with their customers. This move towards a customer-centric approach to marketing, coupled with the increasing availability of customer-transactional data has made Customer Relationship Management (CRM) the leading strategy for marketing decision makers and this is reflected in firms\textsuperscript{3} significant investments in CRM (Reinartz & Kumar, 2002; Teo, Devadoss, & Pan, 2006). Companies realize that their existing customer database is their most valuable asset (Athanassopoulos, 2000; Jones, Mothersbaugh, & Beatty, 2000; Thomas, 2001). It has been shown that it is more profitable to keep and satisfy existing customers than to constantly attract new customers who are characterized by a high attrition rate (Reinartz & Kumar, 2003). It is even suggested that it costs 12 times more to gain a new customer than to retain an existing one (Torkzadeh, Chang, & Hansen, 2006). Moreover, retained customers produce higher revenues and margin than new customers (Reichheld & Sasser, 1990).

It is clear that customer retention rates are important metrics in CRM (Hoekstra, Lee, & Wittink, 1999), resulting in an increasing number of customer churn related papers (e.g. Burez & Van den Poel, 2009; Coussement & Van den Poel, 2009). Predicting churn enables the elaboration of targeted retention strategies to limit the losses and to improve marketing decisions (Shaffer & Zhang, 2002). For example, specific incentives may be offered to the most risky customer segments, i.e. the most inclined to leave the company, with the hope that they remain loyal (Burez & Van den Poel, 2007). Two possible outcomes are observed: the customer churned or he/she stayed with the company. So, technically spoken, customer churn modeling is a binary classification problem. Moreover, for the prediction of customer churn using the information stored in the data warehouse one needs tools that can handle large amounts of data, i.e. data-mining tools (Shaw, Subramaniam, Tan, & Welge, 2001).

A broad range of data mining techniques have been used in the past, but one of the most popular models in the churn context remains the (binary) Logit model (Lemmens & Croux, 2006). This model has been used extensively in marketing decision making to model churn problems (e.g. Buckley & Van den Poel, 2005; Hwang & Euiiho Suh, 2004; Kim & Yoon, 2004). The popularity of Logistic Regression is not surprising since the model has some outspoken advantages. First of all, the method is able to combine relative simplicity and good performance. The parameter estimates of the Logit model are interpretable in terms of odds ratios which facilitates full understanding of the results. Moreover, the technique is relatively robust (Buckley & Van den Poel, 2005), while
its popularity is reflected in the availability in almost all statistical software packages. However, as often, one of the advantages can also be considered as a disadvantage. The simplicity of the model is achieved by assuming that the functional form of the dependence on the explanatory variables is known. However, the functional form is seldom known in practice. If the functional form is misspecified then the estimates of the coefficients and the inferences based on them are misleading (Horowitz & Savin, 2001).

In addition, with Logistic Regression, a linear relationship is modeled on the data, which is an oversimplification of the real relationship in the data (Allison, 1999). It is possible to relax these restrictive assumptions by applying the Generalized Additive Models (GAM) approach.

The GAM approach has at least two distinct advantages compared to Logistic Regression. First, by relaxing the usual assumptions it becomes possible to uncover (non-linear) structures in the data that otherwise might be missed. These structures often give substantial new insights into the effects of the covariates. Furthermore, since the GAM approach allows for more complex relationships between the dependent variable and the covariates, more accurate predictions are likely (Hastie & Tibshirani, 1986).

The GAM approach has shown its relevance in a broad range of different research domains ranging from biology (e.g. Jowett, Parkyn, & Richardson, 2008) over cancer research (e.g. Lee et al., 2007) and supply-chain management (e.g. Cakir, 2009). However, only very few marketing-related studies have used a related non- or semi-parametric methodology (e.g. Kalyanam, 1998; Kumar, Scheer, & Steenkamp, 1998; Shively, 2000). Moreover, to the best of our knowledge this is the first study in a customer churn context that evaluates and uses GAM models.

This study contributes to the existing literature in several ways. Three distinct perspectives on GAM models in the current churn context are elaborated. First, a comparison in terms of predictive performance is made between the popular Logistic Regression model and the GAM approach. Second, it is shown that relaxing the linearity assumption results in a much richer insight into the effects of the covariates which helps decision makers in a thorough understanding of the churn problem of their company. Finally, it is shown how GAM models increase business value of the company. The added value of the GAM approach is therefore shown in monetary terms based on a real-life business setting.

The remainder of this paper is organized as follows. Section 2 focuses on the methodology behind both the logistic model and the GAM approach. The evaluation criteria used in the empirical part of this study are explained in the same section. Section 3 applies the proposed methodology to a real-world churn setting, i.e. the newspaper case. Section 4 treats the managerial implication of the proposed methodology and finally, Section 5 gives conclusions and directions for further research.

2. Methodology

2.1. Base learner

The methodological link between the Logistic Regression and the GAM approach is explained starting from the Generalized Linear Model (GLM) framework. The most common representation of the GLM is as follows (Tabachnick & Fidell, 1996):

$$E(Y) = \mu = g(X \beta)$$

(1)

with $E(Y)$ representing the expected value of $Y$, $X \beta$ is a linear combination of the data with unknown parameters $\beta$ and $g$ is the link function. As Eq. (1) shows, each outcome of the dependent variable $Y$ is assumed to be generated from a particular distribution function. Different choices of distribution function and link function result in different statistical techniques. For example, the normal distribution in combination with the identity link function results in the normal multiple regression model. When the responses $Y$ are binary, the distribution function is the binomial distribution and the interpretation of $\mu$ is then the probability, $p$, of $Y$ taking on the value one. A Logistic Regression model is then obtained by applying the logit link function by

$$\logit\{P(X)\} \equiv \log \left\{ \frac{P(X)}{1 - P(X)} \right\} = \alpha + \sum_{j=1}^{p} s_j(X_j)$$

(2)

with $P(X) = Pr(Y = 1|X)$. The parameter vector $\beta$ is estimated from observations $(y, x)$ and this can be done using maximum likelihood among others. More technical details on Logistic Regression are found in Anderson (1983).

2.2. Generalized additive models

An attractive alternative to Logistic Regression is Generalized Additive Models (GAM) (Hastie & Tibshirani, 1986; Hastie & Tibshirani, 1987; Hastie & Tibshirani, 1990). Additive Logistic Regression relaxes the linearity constraint and suggests a nonparametric fit to the data. In other words, the regression function is modeled in a nonparametric way and the data itself decides on the functional form. However, Hastie and Tibshirani (1990) argue that nonparametric methods perform worse when the number of explicative variables increases, because the sparseness of the data inflates the variance of the estimates. This is often referred to as the curse of dimensionality (Bellman, 1961). It was Stone (1985) who proposed additive models to approximate the multivariate regression function. Consequently, the curse of dimensionality is avoided since each individual additive term is estimated using a univariate smoother. Hastie and Tibshirani (1990) pick up the additive model principle by extending the GLM framework.

This study gives a general overview on the GAM principle. For more details, we kindly refer to Hastie and Tibshirani (1986), Hastie and Tibshirani (1987, Hastie and Tibshirani (1990) or Hastie, Tibshirani, and Friedman (2001). Suppose that $Y$ is the response variable with binary target labels $X_1, X_2, \ldots, X_p$ is a set of independent variables, GAM generalize the Logistic Regression principle by replacing the linear predictor $\sum_{j=1}^{p} \beta_j X_j$ in Eq. (2) with an additive component where

$$\logit\{P(X)\} \equiv \log \left\{ \frac{P(X)}{1 - P(X)} \right\} = \alpha + \sum_{j=1}^{p} s_j(X_j)$$

(3)

with $s_1(\cdot), s_2(\cdot), \ldots, s_p(\cdot)$ as smooth functions. This study uses smoothing splines for $s_1(\cdot), s_2(\cdot), \ldots, s_p(\cdot)$ to estimate the nonparametric trend for the dependence of the logit on $X_1, X_2, \ldots, X_p$. A smoothing spline solves the following optimization problem: among all functions $\eta(x)$ with continuous second order derivatives, find the function that minimizes the penalized residual sum of squares

$$\sum_{i=1}^{n} (y_i - \eta(x_i))^2 + \lambda \int_{a}^{b} (\eta''(t))^2 dt$$

(4)

where $\lambda$ is a fixed constant and $a \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq b$. The goodness-of-fit is measured by the first part of Eq. (4), while the second term is a penalty term that penalizes curvature in the function by the smoothing parameter $\lambda$. The complexity of $\eta(x)$ is measured by $\lambda$ and it is inversely related to the degrees of freedom (df). If $\lambda$ is small (i.e. the df are large), $\eta(x)$ is any function that approaches an interpolation to the data, when $\lambda$ is large (i.e. the df are small), $\eta(x)$ is closely related to a simple least squares fit. It is shown that an explicit and unique minimizer for Eq. (4) exists, i.e. a natural
cubic spline with knots at the unique values of $x_i$ (Hastie & Tibshirani, 1990).

In order to optimize the GAM, the local scoring algorithm (Hastie & Tibshirani, 1986) is applied, while the backfitting algorithm of (Friedman & Stuetzle, 1981) is used to estimate the additive models. The current study follows the recommendation of Hastie et al. (2001) to use a $\lambda$ corresponding to four degrees of freedom, because it is shown that a small number of $df$ fits most data-sets very well.

2.3. Feature selection

Feature selection techniques are often used in the context of churn modeling (e.g. Coussement & Van den Poel, 2008). It is the process of choosing a subset of the original predictive variables by eliminating variables that are either redundant or possess little predictive information. This significantly improves the comprehensibility of resulting models and makes resulting models generalize better (Kim, 2006). In addition, selection techniques defy the curse of dimensionality in order to improve prediction performance (Guyon & Elisseeff, 2003). Moreover, extracting a comprehensive subset of relevant predictor variables is critical for decision makers who want to build marketing strategies based on key drivers of customer response. A forward selection is used within this research setting. The choice for this relatively standard selection technique makes it easy to replicate results with most statistical software packages (Kim, 2006).

2.4. Evaluation criteria

It is essential to evaluate the classifiers in terms of performance. A brief review of evaluation metrics for marketing is found in (Rosset, Neumann, Eick, Vatnik, & Idan, 2001). Two important and often used performance evaluation criteria in the context of churn prediction are the Area Under the receiver operating characteristic Curve (AUC) and the top-decile lift (e.g. Buckinx & Van den Poel, 2005; Coussement & Van den Poel, 2008; Lemmens & Croux, 2006).

The first measure used in this study is the AUC. In contrast to some other evaluation measurements (e.g. accuracy), the AUC is not influenced by any threshold value since it takes into account all possible thresholds on the predicted probabilities. For all these points, it considers the sensitivity (the number of true positives versus the total number of defectors) and one minus the specificity (the number of true negatives versus the total number of non-defectors) of the confusion matrix in a two-dimensional graph (Egan, 1975). Plotting the outcomes of these calculations, results in a receiver operating characteristics curve. The area under this curve (AUC) lies between 0.5 and 1 and is then used to evaluate the predictive performance of classification models. Again, the larger the AUC, the better is the performance of the model. Comparing and evaluating the predictive performance of two or more models using AUC is conducted using the non-parametric test proposed by DeLong, DeLong, and Clarke-Pearson (1988).

The second performance measure used in this study is the top-decile lift. The top-decile lift focuses exclusively on the most critical group of customers and their churn risk. The top 10% riskiest customers (i.e. the group of customers with the highest predicted churn probabilities) represents an ideal segment for targeting in a retention-marketing campaign (Lemmens & Croux, 2006). This performance measure is very appealing because it incorporates somewhat the notion that marketing budgets are limited and actions to reduce churn are therefore limited to a segment of customers that is at high risk. In practice one will order the customers on decreasing predicted churn probability. Next, the proportion of real churners in the top 10% is compared with the proportion of churners in the total dataset. It is clear that the higher the top-decile lift, the better is the model. For example, a top-decile lift of 2 means that the model under investigation identifies twice as many churners in the top 10% as a random assignment would do.

It is important to see that the AUC and the top-decile lift are measuring different aspects of the predictive accuracy of the models. Both evaluation criteria provide complementary information. A model can be good at identifying the most risky segment but less effective at recognizing less risky customers. Combining the two metrics provides a thorough evaluation of the performance of the models under investigation.

3. Empirical validation

This Section applies the proposed methodology to a real-world churn setting. In a first subsection, an overview of the research setting is given. Section 3.2 defines the best subset of marketing variables by means of a Logistic Regression and a Generalized Additive Model using a forward selection procedure, while both models are compared in terms of the predictive performance. Finally, Section 3.3 clarifies the interpretability of both prediction models.

3.1. Data

The current study is performed on data supplied by the largest Belgian newspaper publishing company. Churn prevention is an important issue for the newspaper business (Lemmens & Croux, 2006), which makes this setting an ideal testbed for validating the proposed methodology. Customers subscribe during a certain period for a specific newspaper. This customer churn prediction context comes down to predicting whether or not a subscriber will renew his/her subscription during the 30-day period after the expiry date. The company still delivers the newspaper to the subscribers during this 30-day period in order to allow some flexibility in the renewal procedure. The churn response variable is coded as ‘1’ when the subscription is not renewed within a 30-day period after the renewal date, and ‘0’ otherwise. The explanatory variables contain information covering a 30-month period returning from every individual renewal point. The predictors include client/company interaction variables (e.g. the number of responses on direct marketing actions…), renewal-related variables (e.g. the number of times the customer churned in the past…), subscriber-describing variables (e.g. age…) and subscription-describing variables (e.g. the length of the current subscription…).

Appendix A gives a detailed overview of the independent variables used within this research setting.

This database contains all renewals from the first of March 2005 till the end of June 2005. In order to make a correct prediction on whether or not a customer will leave the company, the total dataset is divided into a training and test set. The training set is used for model building, while the test set allows for valid assessment of performance (Lemmens & Croux, 2006) and as such, one is sure that the trained model generalizes well. Practically, we randomly assign 70% of the data to the training set in order to build the prediction model, while the other 30% of the data is used to validate the proposed models to unseen data. The training set contains 93,859 subscriptions with a churn incidence of 11.95%, while the test set contains 40,225 subscriptions of which an equal proportion of 11.95% are churners. More information on the data characteristics is found in Table 1.

In order to answer the research question whether or not it is beneficial to allow more flexibility in the prediction model by using nonparametric trends, two churn-prediction models are built: (i) MOD_LOG containing the parametric terms of a forward Logistic Regression and (ii) MOD_GAM which models customer churn via a forward GAM model which allows for nonlinear relationships.
Best subset of predictors.

Table 2
Overview of the data characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Number of subscriptions</th>
<th>Relative percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subscriptions not renewed</td>
<td>11,218</td>
<td>11.95</td>
</tr>
<tr>
<td>Subscriptions renewed</td>
<td>82,677</td>
<td>88.05</td>
</tr>
<tr>
<td>Total</td>
<td>93,895</td>
<td>100</td>
</tr>
<tr>
<td>Test set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subscriptions not renewed</td>
<td>4808</td>
<td>11.95</td>
</tr>
<tr>
<td>Subscriptions renewed</td>
<td>35,417</td>
<td>88.05</td>
</tr>
<tr>
<td>Total</td>
<td>40,225</td>
<td>100</td>
</tr>
</tbody>
</table>

3.2. Predictive performance

By employing MOD_LOG and MOD_GAM, one is able to obtain the best subset of predictors differentiating between churners and non-churners. The variable selection procedure results in a subset of 18 predictors for MOD_LOG and 27 variables for MOD_GAM, both at a significance level of $p < 0.05$ (see Table 2).

Table 2 clearly indicates that MOD_GAM selects more significant variables describing the churn behavior than MOD_LOG due to added flexibility in the prediction model.

Table 2
Best subset of predictors.

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable name</th>
<th>Variable description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOD_LOG</td>
<td>1</td>
<td>mon_val Monetary value</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>rec_complaint Elapsed time since the last complaint</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>avg_complaint_pos the average positioning of the complaints in the current subscription</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>nbr_renew_point The number of renewal points</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>term Elapsed time since last renewal point</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>age Age of the subscriber</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>days_renew_after How many days after the expiry date, the previous subscriptions were renewed</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>nbr_suspension The number of suspensions</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>nbr_change_edition The number of conversions made in edition</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>rec_change_pay Elapsed time since last conversion in payment method</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>lor The length of relationship</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>length_current_subsc The length of the current subscription</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>avg_renewaltime The average of the subscription renewal time</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>nbr_complaints The number of complaints</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>avg_suspension time Average suspension time</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>nbr_prior_churn The number of prior churning events</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>nbr_dm_action The number of responses on direct marketing actions</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>timelag_response Elapsed time since response on direct mailing campaign</td>
</tr>
<tr>
<td>MOD_GAM</td>
<td>1</td>
<td>mon_val Monetary value</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>rec_complaint Elapsed time since the last complaint</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>rec_change_pay Elapsed time since last conversion in payment method</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>days_renew_before How many days before the expiry date, the previous subscriptions were renewed</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>term Elapsed time since last renewal point</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>nbr_renew_point The number of renewal points</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>rec_suspension Elapsed time since last suspension</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>timelag_response Elapsed time since response on direct mailing campaign</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>days_renew_after How many days after the expiry date, the previous subscriptions were renewed</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>age Age of the subscriber</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>avg_complaint_pos the average positioning of the complaints in the current subscription</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>nbr_change_pay The number of conversions made in payment method</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>avg_suspension time Average suspension time</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>lor The length of relationship</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>rec_dm_action Elapsed time since last response on direct marketing action</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>nbr_complaints The number of complaints</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>nbr_suspension The number of suspensions</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>length_current_subsc The length of the current subscription</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>nbr_change_channel The number of conversions made in distribution channel</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>rec_change_channel Elapsed time since last conversion in distribution channel</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>avg_renewaltime The average of the subscription renewal time</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>rec_change_edition Elapsed time since last conversion in edition</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>nbr_dm_action The number of responses on direct marketing actions</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>nbr_free_copy The number of free newspapers</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>var_renewaltime The variance on the subscription renewal time</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>nbr_compensation The number of compensation newsletters</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>nbr_change_edition The number of conversions made in edition</td>
</tr>
</tbody>
</table>

This finding results in a better predictive performance of MOD_GAM compared to MOD_LOG. Table 3 shows the predictive performance of MOD_LOG and MOD_GAM. It is clear that MOD_GAM outperforms MOD_LOG in terms of AUC. The AUC increases from 0.8298 to 0.8452 on the test set. This improvement is significant ($\chi^2 = 187.40, df = 1, p < 0.001$) (Delong et al., 1988). Insightfully, MOD_GAM is able to better distinguish churners from non-churners compared to MOD_LOG.

Moreover, the beneficial effect of using an Additive Logistic Regression contrary to a traditional Logistic Regression is confirmed when having a look at the top-decile lift. MOD_GAM is able to increase the lift in the first decile from 4.08 to 4.26. In other words, MOD_LOG is able to increase the base line churn rate of 11.95% in the first decile to 48.75%, while an additional increase is obtained by MOD_GAM to 50.91% churners.

These results suggest that applying an Additive Logistic Model instead of a traditional Logistic Regression is a valuable strategy to improve the prediction power of a customer churn model.

3.3. Interpretability of the proposed models

This subsection explores the impact of the best subset variables on customer attrition. Table 4 shows the univariate parameter
estimates from a traditional Logistic Regression model. As such, the marketing analyst is able to measure the true impact of a single variable on the churn behavior.

The relationship of the variables in Table 4 is linear to the logit of churning. However, this assumption can be relaxed by taking into account nonlinear effects between the independent and the logit. This is done by applying the GAM approach. Appendix B visualizes the effect of every single variable in the analysis on the churn probability. On all sub panels, the horizontal axis represents the fitted churn probabilities. The orange line represents the non-montonic relationship between the independent variable and the churn probability as estimated using a GAM. Appendix B clearly shows that more flexible solutions are obtained when one relaxes the linearity constraint of the Logistic Regression. It gives the marketing analyst a much better insight into the way a certain variable affects churn behavior. From this graph we can learn that both high and low values lead to smaller churn probabilities, while more central values push the churn probability curve is striving to a maximum, but when the independent variable gets even bigger, the churn probability becomes smaller. This is the case for the variable ‘term’. Looking only at the linear effect from the Logistic Regression model (see Table 4), one concludes that higher values for this variable result in lower churn probabilities. Again, the added value of the GAM approach is shown in Appendix B. The graph shows a much richer insight into the effect of the covariate. From this graph we can learn that both high and low values lead to smaller churn probabilities, while more central values push the churn probability up.

The last category of variables is characterized by a complex trend. The effects on the churn probability are Gordian, what makes interpretation burdensome. One can say that the variables in this category enhance the predictive power of the model more than its interpretability. Moreover comparing Table 2 and Appendix B, this type of variables is often not selected in the MOD_LOG model.

In the end, it should be clear that relaxing the linearity constraint by using GAMs gives data analysts a more profound insight into the explicative variables.

4. Managerial implications

As the difference in predictive performance between MOD_LOG and MOD_GAM is significant, the impact on the company’s marginal profitability will not fail to appear. Indeed, several studies summarized by four main categories: a quasi-exponential, a U, an inverted U and a complex trend.

A first category of variables follows a quasi-exponential trend. Hereby the effect on the churn probability is very pronounced for small values of the independent variable, while it decreases rapidly for middle to higher values. For example, in Appendix B, the variable ‘length_current_subsc’ clearly shows a strong effect for small values, while for values larger than 100, the effect is minimal and starts to stabilize. A much richer insight into the effect of ‘length_current_subsc’ on the churn probability is obtained compared to what is learned from the Logistic Regression coefficients (see Table 4), i.e. increasing values of the explicative variable leads to decreasing churn probabilities.

For variables following the U-trend, the churn probability decreases till it reaches a minimum. When the values of the independent variable become larger, the chance of not renewing the subscription gets bigger. The ‘nbrSuspension’ exerts such an effect. Extreme values for this variable lead to higher churn probabilities, while more central values result in a lower churn probability. Interpretation based only on the Logistic Regression coefficient (see Table 4) overlook the insight of low values for ‘nbrSuspension’ leading to higher churn probabilities.

The inverse happens when one has an inverted U trend. The churn probability curve is striving to a maximum, but when the independent variable gets even bigger, the churn probability becomes smaller. This is the case for the variable ‘term’. Looking only at the linear effect from the Logistic Regression model (see Table 4), one concludes that higher values for this variable result in lower churn probabilities. Again, the added value of the GAM approach is shown in Appendix B. The graph shows a much richer insight into the effect of the covariate. From this graph we can learn that both high and low values lead to smaller churn probabilities, while more central values push the churn probability up.

The last category of variables is characterized by a complex trend. The effects on the churn probability are Gordian, what makes interpretation burdensome. One can say that the variables in this category enhance the predictive power of the model more than its interpretability. Moreover comparing Table 2 and Appendix B, this type of variables is often not selected in the MOD_LOG model.

In the end, it should be clear that relaxing the linearity constraint by using GAMs gives data analysts a more profound insight into the explicative variables.

Table 3 Test set performance in terms of AUC and top-decile lift.

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>Top-decile lift</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOD_LOG</td>
<td>0.8298*</td>
<td>4.08</td>
</tr>
<tr>
<td>MOD_GAM</td>
<td>0.8452*</td>
<td>4.26</td>
</tr>
</tbody>
</table>

* Significant with \( p < 0.01 \).

Table 4 Overview of the univariate standardized parameter estimates and odds ratios.

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Standardized parameter estimates</th>
<th>Odds ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>mon_val</td>
<td>-0.72</td>
<td>0.994</td>
</tr>
<tr>
<td>rec_complaint</td>
<td>0.26</td>
<td>1.046</td>
</tr>
<tr>
<td>avg_complaint_pos</td>
<td>-0.29</td>
<td>0.100</td>
</tr>
<tr>
<td>nbr_renew_point</td>
<td>-1.90</td>
<td>0.642</td>
</tr>
<tr>
<td>term</td>
<td>-0.03</td>
<td>0.988</td>
</tr>
<tr>
<td>age</td>
<td>-0.19</td>
<td>0.976</td>
</tr>
<tr>
<td>days_renew_after</td>
<td>0.05</td>
<td>1.010</td>
</tr>
<tr>
<td>nbr_suspension</td>
<td>-0.14</td>
<td>0.819</td>
</tr>
<tr>
<td>nbr_change_edition</td>
<td>-0.26</td>
<td>0.314</td>
</tr>
<tr>
<td>rec_change_pay</td>
<td>-0.09</td>
<td>0.960</td>
</tr>
<tr>
<td>lor</td>
<td>-0.75</td>
<td>0.986</td>
</tr>
<tr>
<td>length_current_subsc</td>
<td>-0.87</td>
<td>0.984</td>
</tr>
<tr>
<td>avg_renewtime</td>
<td>-0.18</td>
<td>0.285</td>
</tr>
<tr>
<td>nbr_complaints</td>
<td>-0.42</td>
<td>0.815</td>
</tr>
<tr>
<td>avg_suspension time</td>
<td>-0.08</td>
<td>0.632</td>
</tr>
<tr>
<td>nbr_prior_churn</td>
<td>0.07</td>
<td>1.381</td>
</tr>
<tr>
<td>nbr_dm_action</td>
<td>0.11</td>
<td>1.229</td>
</tr>
<tr>
<td>timelag_response</td>
<td>0.12</td>
<td>1.951</td>
</tr>
</tbody>
</table>

* Significant with \( p < 0.05 \).
showed that minor increases in customer attrition rates result in significant changes in contribution (e.g. Reichheld & Sasser, 1990; Van den Poel & Larivière, 2004). Table 5 gives an overview of the increase in profitability between the base model (i.e. random customer selection or RANDOM), MOD_LOG and MOD_GAM performance. Table 5 focuses on the top-decile performance to calculate the additional profits, because marketing analysts are often interested in just 10% of the customer base who are most likely to churn (Coussement & Van den Poel, 2008). The reason is two-fold: marketing budgets for customer retention campaigns are often limited and actions to reduce churn typically involve only 10% of the entire list of customers in order to maintain a feasible operational solution.

Table 5 contains the profit implications for the random model, MOD_LOG and MOD_GAM. RANDOM has always a top-decile lift of 1, because theoretically selecting X random customers from the total customer base in the current case results in the selection of 11.95% churners, i.e. the base churn rate. A sensitivity analysis on the profitability is done based on the number of churners in the top-decile of RANDOM. The discounted profit is defined as today's value of the future one-year profit gained by the retained churners from the top-decile. The discounted profit for model X is then calculated by

\[
\text{discounted profit}_X = \frac{\varphi \cdot P}{1 + \frac{d}{1 + \frac{d}{1 + d}}}
\]

(5)

with \( P \) equal to the average yearly profit per subscription, \( d \) equal to the discount rate and \( \varphi \) equal to the number of retained churners. \( \varphi \) is calculated by

\[
\varphi = (K \ast \tau) \ast \eta
\]

(6)

with \( K \) equal to the number of churners in the top-decile from RANDOM, \( \tau \) equal to the top-decile lift of model X and \( \eta \) equal to the conversion rate of the churners.

Within this research paper, the marginal (discounted) profit of model X over model Y is defined as the difference between the discounted profit of model X and the discounted profit of model Y or

\[
\text{marginal profit}_X - Y = \text{discounted profit}_X - \text{discounted profit}_Y
\]

For this real-life example, the company's average yearly profit per subscription \( P \) is close to 200 Euro, the discount rate \( d \) is set equal to 4%, the number of churners in the top-decile from RANDOM \( K \) ranges arbitrarily from 1000 to 5000 by steps of 1000 or \( K = \{1000; 2000; 3000; 4000; 5000\} \), the churn conversion rate \( \eta \) is set to a realistic 10% meaning that one out of ten churners is retained by the company's retention program and the top-decile lift \( \tau \) is obtained from the results of MOD_LOG and MOD_GAM, while \( \tau = 1 \) for RANDOM meaning that the relative number of churners in the top-decile equals the base churn rate of 11.95%.

Table 5 indicates that churn prediction modeling is highly beneficial. The marginal profits of MOD_LOG and MOD_GAM over RANDOM are high. For instance, per 1000 churners in the top-decile of the base model, the company is able to make an additional profit of 59,230.77 Euro when predicting churn based on Logistic regression (MOD_LOG) and 62,692.31 Euro when a Additive Logistic Model (MOD_GAM) is used. Predictive models are able to better distinguish churners from non-churners. In other words, the density of churners in the top-decile is drastically increased which results in additional profits. Moreover, the impact of distinguishing churners from non-churner using MOD_GAM instead of MOD_LOG is highly significant. The increase in top-decile from 4.08 to 4.26 results in an additional increase of 3,461.54 Euro per 1000 churners in the top-decile. Extending these results on a yearly basis to this real-life customer churn context has the following impact. The company has approximately 16,000 churners during a 4-month period (see Table 1), which results in about 48,000 subscriptions that are not renewed on a yearly basis. The real-world impact on profits for RANDOM, MOD_LOG and MOD_GAM for this setting is shown in bold and italic in Table 5. For instance, it is shown that the company is able to save a substantial amount of 16,538.46 Euro on a yearly basis by relaxing the linearity constraint in the customer churn prediction model. In summary choosing MOD_GAM instead of MOD_LOG for distinguishing churners from non-churners is a valuable strategy to improve the prediction power of a customer churn prediction model resulting in an additional increase in company's profit.

5. Conclusion and directions for further research

Nowadays, a lot of marketing analysts focus on predicting whether or not a customer will leave the company. Indeed, customer churn prediction is one of the facets in a well-considered CRM strategy within a lot of companies. Using the customer-transactional database as a starting point for their churn analyses, marketing analysts are using data mining techniques to set up a churn prediction system.

This study contributes to the current churn literature by applying Generalized Additive Models. GAM relaxes the linearity constraint in contrast to a traditional Logistic Regression model. As such, it is possible to allow for complex non-linear relationships between the explanatory variables and the churn probability. First, it is shown that one is able to significantly increase the prediction performance of the churn model by applying GAM. Allowing for more complex relationships between the covariates and the logit, results in better identification of the customers at risk. Second, one gains better insights into the effects of the covariates which are modeled using a non-parametric relationship. Visually, one is able to distinguish four different patterns between the covariates and the logit: a quasi-exponential, a U, an inverted U or a complex trend. Finally, from a managerial point of view it is highly beneficial in monetary terms to use GAM for churn prediction in this research context.

While we strongly believe that this research paper fills a large gap in today's literature, there are still some directions for future research. This study introduces the use of the GAM approach in a churn prediction context. In order to externally validate the obtained results, one can apply GAM in other than customer churn contexts, e.g. response modeling, cross- and up-sell applications... or by exploiting this principle in other than non-binary classification domains of CRM like for instance customer lifetime value modeling. Moreover, this study compares Logistic Regression with a relaxed version of it, namely Additive Logistic Regression which allows for nonlinear fits to the independent variables. Future research could be done in comparing other nonlinear prediction techniques like decision trees, neural networks, random forests...

Acknowledgments

We would like to thank the anonymous Belgian Publishing Company for making available their data. Next, we also like to thank Ilse Bellinck and Weijie Cai for their fruitful comments on this research paper.

Appendix A. Overview of the explicative variables

Client/company-interaction variables
- The number of complaints,
- Elapsed time since the last complaint,
- The average cost of a complaint (in terms of compensation newspapers),
- The average positioning of the complaints in the current subscription,
- The conversions made in distribution channel, payment method & edition,
- Elapsed time since last conversion in distribution channel, payment method & edition,
- The number of responses on direct marketing actions,
- The number of suspensions,
- The average suspension length (in number of days),
- Elapsed time since last suspension,
- Elapsed time since last response on a direct marketing action,
- The number of free newspapers.

Renewal-related variable
- How many days before the expiry date, the previous subscription was renewed,
- The average number of days the previous subscriptions are renewed before expiry date,
- The variance in the number of days the previous subscriptions are renewed before expiry date,
- Elapsed time since last step in renewal procedure,
- The number of times the churner did not renew a subscription.

Subscriber-describing variables
- Age,
- The length of relationship.

Subscription-describing variables
- Elapsed time since last renewal,
- Monetary value,
- The number of renewal points,
- The length of the current subscription,
- The number of days a week the newspaper is delivered (intensity indication).

Appendix B. Visualization of MOD_GAM

Quasi exponential trend

(continued on next page)
Appendix B (continued)

![Graphs showing churn probability vs. various metrics such as mon_val, nbr_renew_point, nbr_change_channel, and avg_renewaltime.](image)

**U-trend**

![Graphs showing churn probability vs. avg_complaints and other metrics.](image)
Appendix B (continued)

Inverted U-trend

(continued on next page)
Appendix B (continued)

Complex trend

- Age
- SBE_line_copy
- SBE_change_edition
- SBE_complaint
- SBE_channels
- SBE_complaint

Churn probability

Appendix B (continued)

References


