Customer event history for churn prediction : how long is long enough?

Michel Ballings, Dirk Van den Poel*

Faculty of Economics and Business Administration, Department of Marketing, Ghent University,
Tweekerkenstraat 2, B-9000 Ghent, Belgium, http://www.crm.ugent.be,
Michel.Ballings@UGent.be, Dirk.VandenPoel@UGent.be
*Corresponding author, Tel: + 32 9 264 89 80, Fax: +32 9 264 42 79

Abstract

The key question of this study is: How long should the length of customer event history be for customer churn prediction? While most studies in predictive churn modeling aim to improve models by data augmentation or algorithm improvement, this study focuses on a another dimension: time window optimization with respect to predictive performance. This paper first presents a formalization of the time window selection strategy, along with a literature review. Next, using logistic regression, classification trees and bagging in combination with classification trees, this study analyzes the improvement in churn-model performance by extending customer event history from 1 to 16 years. The results show that, after the 5th additional year, predictive performance is only marginally increased, meaning that the company in this study can discard 69% of its data with almost no decrease in predictive performance. The practical implication is that analysts can substantially decrease data-related burdens, such as data storage, preparation and analysis. This is particularly valuable in times of big data where decreasing computational complexity is paramount.

Keywords: Predictive Analytics, Time Window, Length of Customer Event History, Predictive Customer Churn Model

1   Introduction

To cope with the fierce competition in their business environments, companies are increasingly focusing on analytical Customer Relationship Management (CRM) (Reinartz & Kumar, 2002). Within CRM, churn management has received most of the attention because it has been shown that retaining customers is several times more profitable than acquiring new ones (Reinartz & Kumar, 2003; Reichheld & Sasser, 1990). As such, companies are realizing that their most valuable asset is their customer base (Athanassopoulos, 2000).

From an analytical viewpoint, churn management consists of (1) predicting which customers are going to churn and (2) evaluating which action is most effective in retaining these customers (Hung, Yen & Wang, 2006). Especially the former has attracted a lot of research. These churn-prediction studies typically use two strategies to improve model performance: an algorithm- based strategy and a data-based strategy (Baecke & Van den Poel, 2010). The former consists in evaluating multiple algorithms on given data and improving or inventing algorithms. The latter consists in augmenting the existing

database with new data sources. From an algorithm- based perspective, CRM has evolved from using RFM (Recency, Frequency and Monetary) models, over classifiers such as logistic regression and decision trees (McCarty & Hastak, 2007), to more recent advanced techniques such as random forests (Larivière & Van den Poel, 2005), neural networks (Zahavi & Levin, 1997) and support vector machines (Shin & Cho, 2006). From a data- based perspective, CRM has evolved from including geographical data (Steenburgh, Ainsle & Engbretson, 2003), over consumer network data (Hill, Provost & Volinsky, 2006), to clickstream (Van den Poel & Buckinx, 2005; Hu & Zhong, 2008) and survey data (Baecke & Van den Poel, 2011).

Although the data and algorithm dimensions are very important, they constitute only two out of three components of the modeling problem. The third component, the time window, remains thus far under-researched (see fig. 1).
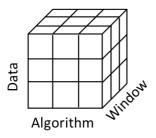

Fig. 1. Three dimensions of predictive modeling

While data and algorithm strategies are often, but not exclusively, focused on adding to the computational burden by feeding more data to the model or by estimating more parameters, time window- based strategies focuses on the opposite. If the analyses show that the added value of an extra year of data is marginal, this data could be discarded. Researchers have already noted that obtaining accurate predictions is a costly process of gathering, storing, cleaning, preparing and analyzing data (e.g. Malthouse & Derenthal, 2008). In that regard, Risselada, Verhoef & Bijmolt (2010) underline that striking a good balance between model building efficiency and model accuracy is desirable. Hence this study intends to fill this gap in extant literature by investigating how time window selection influences model performance.

The remainder of this paper is organized as follows. First we'll elaborate on time window configurations in order to formalize the time window- based strategy for model improvement and provide a literature review of studies that use this strategy. Second, in the methodology section, the data, time window, techniques, variables and model performance metrics will be discussed. Third, we provide a discussion of the results. Fourth, a conclusion will be formulated and practical implications will be elaborated on. Finally, limitations and directions for future research are given.

2   Time window- based strategy for model improvement

Predictive modeling refers to learning the relationship between data that are observed in a period (or window) that ends before a certain point in time and data that is observed in a period that starts after the same point in time. The former period is called the independent, predictors or explanatory period, or customer event history, and the latter is called the dependent or response period. In between these two periods, there's often a gap, sometimes called the retention period, which is meant to serve as a

practical or operational period for organizing the actual CRM actions before customers exhibit the focal response behavior (Wei & Chiu, 2002).

For model estimation purposes, we provide a formalization of different types of windows. A window type can be defined by the variability of the length of the window: constant (C) or variable (V). Given the three time windows in one configuration, the predictors period, the operational period and the response period, there are theoretically 8 possible configurations (not all configurations may be practically useful). The configuration are displayed in fig. 2.

For model validation purposes, configurations do not change themselves, but their point in time. While fixed (F) point validation uses data from the same point in time, sliding (O) point validation uses data from a different point in time.
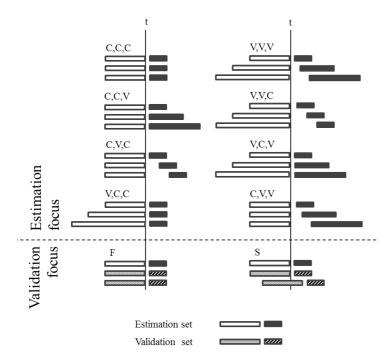

Fig.2. Window configurations

Literature on time windows is scarce but there are some notable studies. Wei and Chiu (2002) contribute to literature by estimation models with varying length of the operational period and validating it at the same point in time (CVCF). Several other authors opt for a pure validation focus in an S configuration: Zahavi & Levin (1997) and Neslin, Gupta, Kamakura, Lu & Mason (2006) build a predictive model with one period ahead, Wei and Chiu (2002) build a model with two periods ahead and Risselada, Verhoef & Bijmolt (2010) go up to three periods in order to assess the staying power of their models.

A pure F configuration can be considered the baseline validation configuration and is well represented in literature (see for example Coussement, Benoit & Van den Poel, 2010; Larivière & Van den Poel, 2005).

To the best of our knowledge, all other configurations are not represented in literature. The purpose of this paper is to investigate one configuration, the VCCF configuration. The reason for this choice is that investigating this configuration answers one of today's most thorny questions, namely how to

lower the computational burden in expert systems in a world where data is growing at an exponential rate (Manyika , Chui, Brown, Bughin, Dobbs, Roxburgh & Byers, 2011, p16).

More formally, the research question of this study is: Do different lengths of the customer event history result in different predictive model performance? In other words: How long is long enough in terms of the predictors period?

3 Methodology

3.1 Data and time windows

We analyze the entire customer database of a newspaper company. All the customers active at the end of the explanatory period are included in the analysis (n = 129,892 of which 75% for estimation and 25% for validation). The churn rate for the estimation and validation is respectively 11.15% and 11.47%. The price customers have to pay depends on the length of the subscription and the promotional context. The newspaper company sends a letter to remind them that they are approaching the end of the subscription and to ask them whether they want to renew their subscription, along with instructions on how to do that. Customers cannot cancel the subscription and they have a four-week grace period once their subscription has lapsed. As such churn prediction involves predicting whether the customer will or will not renew his or her subscription in the four-week period following the end of the subscription.

Predictor variables are computed from data starting at t and ending at 01/03/2010, where t takes on values from 03/01/2009 until 03/01/1994 by one year. Churn is computed from data from the period 01/08/2010 until 02/03/2011.

3.2 Techniques

In churn literature, the two most commonly used analytical techniques are logistic regression and classification trees (Neslin, Gupta, Kamakura, Lu & Mason, 2006; Risselada, Verhoef & Bijmolt, 2010). The performance of the two techniques is very similar (Levin and Zahavi, 2001; Neslin, Gupta, Kamakura, Lu & Mason, 2006), and depends on a multitude of factors such as the normality of the data, the number of categorical variables, the size of the training sample and the signal-to-noise ratio (Perlich, Provost & Simonoff, 2004; King, Feng & Sutherland, 1995).

Several more advanced techniques have been used in the marketing literature, such as random forests (Larivière & Van den Poel, 2005), rotation-based ensemble methods (De Bock & Van den Poel, 2011) and neural networks (Zahavi & Levin, 1997), but their application is still relatively limited due to the increase in complexity (for an extensive literature review we refer to Risselada, Verhoef & Bijmolt, 2010). In this regard, Neslin, Gupta, Kamakura, Lu & Mason (2006) found that 68% of techniques in a modeling contest for practitioners and academics were logistic regressions or classification trees.

More recently, classification trees have been shown to benefit from aggregation methods. Bagging, or bootstrap aggregating, consists in estimating a model on a multitude of bootstrap samples which

results in multiple predictions for each customer. The final prediction equals the mean of all predictions (Breiman, 1996; Lemmens & Croux, 2006). While bagging substantially increases the predictive performance of classification trees, Perlich, Provost & Simonoff (2004) show that this is not the case for logistic regression. It is argued that the reason for this difference is that logit has less sample sensitivity than classifications trees and hence averaging the predictions from the models on different bootstrap samples will have less of an effect (Risselada, Verhoef & Bijmolt, 2010). Van den Poel & Prinzie (2008) try to "solve" the latter issue for logit by also randomly selecting subsets of predictors.

In sum, logistic regression and classification trees are widely used by practitioners and academics and bagging is a simple approach to increase the predictive performance of classification techniques.

We used CHAID (Chi-squared automatic interaction detection) (Kass, 1980) and not CART (Breiman, Friedman, Olshen & Stone, 1984) as the method of classication tree construction. Haughton & Oulabi (1993) assert that the best solution to choose between CART and CHAID is to compare the results and choose the best. In some preliminary testing, CHAID outperformed CART so we continued our analyses with the former. To avoid overfitting in CHAID we tried multiple minimum node sizes required for splitting. To avoid overfitting in logistic regression we used stepwise variable selection.

As aforementioned, the bagging procedure consists in estimating models on a number of bootstrap samples of the original sample. We used 10 bootstrap samples to estimate 10 models (Perlich, Provost & Simonoff, 2004). We didn't consider more samples because of the large sample size. Although more bootstrap samples might yield better performance, the reason for this number is twofold. First of all, companies that have a lot of customers do not have the time to estimate a large number of models. Second, our analyses showed that even with 10 bootstrap samples, we realized a substantial increase in predictive performance in comparison to a single tree. Third, the focus of this study is not on computational techniques but the length of the independent period. Our intent is to find out if bagging shows the same pattern as the other classifiers across different lengths of independent period.

3.3   Variables

In the model based on an explanatory period of one year, 113 variables are tested for significance. When the model is provided an additional year of data the variables are recomputed and the variables from the previous year are kept. Hence, the longer the length of the predictors period, the more variables are tested in the analysis. Variables that are invariant to time (e.g. gender, relationship category) and variables that are perfectly related to time (e.g. age) are of course not recomputed. As such 108 variables are added with each additional year, amounting to 1733 variables that need to be tested when the length of the independent period is 16 years. The amount of variables the different models have selected, and that are actually included in the final models, ranges from 31 to 54.

Both customer characteristics and relationship characteristics are included in the analyses. Since their identification (Cullinan, 1977), many studies report three variables from the latter group as being the best predictors of customer behavior (e.g., Buckinx & Van den Poel, 2005; Bhattacharya, 1998; Keaveney & Parthasarathy, 2001): recency (and length of relationship), frequency and monetary value (Bauer, 1988; Baesens, Viaene, Van den Poel, Vanthienen & Dedene, 2002; Van den Poel, 2003). Moreover, the direction of the relation between these variables and repurchase behavior and churn has been shown to be consistent across studies. Hence, in what follows we'll review them briefly. More

frequent and heavier buyers (represented by the variable frequency) are more likely to display loyal behavior (Bauer, 1988; Van den Poel, 2003; Morrison, 1966; Lawrence, 1980). Similarly, the more money a customer spends with the company (captured by monetary value), the higher the repurchase likelihood (Ganesan, 1994; Levin & Zahavi, 1996). Adversely, the more time passes since the last purchase (recency), the lower the likelihood of repurchase (Van den Poel, 2003; Cullinan, 1977). In addition to the RFM variables, length of relationship (LOR) also is a top predictor (Van den Poel, 2003; Buckinx & Van den Poel, 2005). Simpson (1987) asserts that the duration of the relationship also indicates relationship stability, meaning that the longer the LOR is, the higher the repurchase probability. Table 1 provides an overview of the included variables in this study and a brief overview of the relevant literature. All categorical variables were dichotomized in the analysis.

Table 1
Tested predictors

Variable	Variable Type	Study
	Customer characteristics	
• Language • Gender • Age, Whether the age is known • Relationship type: person, company, advertiser • Whether information is available: telephone number, fax, email, bank account number	Socio demographics	Larivière & Van den Poel, 2005
	Relationship characteristics	
• Frequency: Number of subscriptions • Number of newspapers in last subscription • Sum number of newspapers across subscriptions • Number of changes in newspaper editions: sum, mean • Subscription type: Renewal or not • Newspaper edition	Frequency related	Cullinan, 1977; Bauer, 1988; Van den Poel, 2003; Morrison, 1966; Lawrence, 1980

• Payment method: direct debit
• Changes in payment method: sum, mean
• Price single newspaper:
  -Last formula: Net
  -All formulas: Sum net, mean gross
• Price product formula (Monetary value):
  -Last formula: Net, Gross
  -All formulas: Sum net, mean gross, sum gross, mean gross
	Monetary value related	Ganesan 1994; Levin & Zahavi 1996; Cullinan 1977

• Length of last subscription
• Which month and season the end of subscription falls in
• Length of relationship with interruptions
• Length of relationship without interruptions
• Mean length of subscription
• Length of last subscription divided by mean length of subscription
• Recency: elapsed time since start date last subscription
• Elapsed time since last change in edition
• Prior churn: not having a subscription in between subscriptions
  -Yes/no: there is a start date 1, 10, 20, >30 day(s) than the expiry date of the previous subscription (both for last and all subscriptions)
  -Count
  -Duration: sum, mean
	Recency and duration related	Van den Poel, 2003; Cullinan, 1977; Buckinx & Van den Poel, 2005; Simpson, 1987

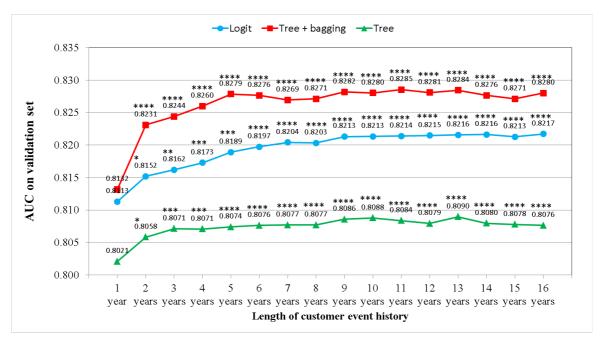|		
• Sum of length last 50%(20%) of subscriptions divided by sum of length first 50%(20%) of subscriptions	Combinations	This study
• Monetary value of last 50%(20%) of all subscriptions divided by monetary value of first 50%(20%) of all subscriptions		
• Interactions: Recency x Frequency x Monetary x LOR, Recency x Frequency x Monetary, Recency x Frequency x LOR, Recency x Monetary x LOR, Frequency x Monetary x LOR, Recency x Frequency, Recency x Monetary, Recency x LOR, Monetary x LOR, Frequency x Monetary, Frequency x LOR,		
• Divisions: Monetary / LOR, Frequency / LOR, Monetary / Frequency, LOR / Frequency, Recency / length of last subscription, Recency / LOR		

3.4 Model performance

To evaluate the classification models' performance we'll use the area under the receiver operating characteristic curve (AUC or AUROC). AUC is argued to be an objective criterion for classifier performance by several authors (Provost, Fawcett & Kohavi, 1998; Langley, 2000). The receiver operating characteristic (ROC) curve is obtained from plotting sensitivity and 1-specificity considering all possible cut-off values (Hanley & McNeil, 1982). AUC ranges from .5, if the predictions are no better than random, to 1, if the model predicts the behavior perfectly (Baecke & Van den Poel 2011). We use AUC instead of accuracy (Percentage of correctly classified, PCC) because AUC, in contrast to PCC, is not sensitive to the cut-off value of the 'a posteriori' probabilities. Hence, AUC is a more adequate performance measure (see Baesens, Viaene, Van den Poel, Vanthienen & Dedene, 2002).

4 Results and discussion

Figure 3 shows the predictive performance, in terms of AUC, across the different lengths of the customer event history. Significance is reported for the differences between respective lengths of the independent period and one year. To determine the significance of the differences between the ROC curves we used the nonparametric test of DeLong, DeLong & Clarke-Pearson (1988).

A first observation is that logistic regression outperforms the single classification tree. The performance of logistic regression and classification trees strongly depends on sample factors, such as the normality of the data, the amount of categorical variables, and the signal-to-noise ratio (Perlich, Provost & Simonoff, 2004; King, Feng & Sutherland, 1995). Hence depending on the sample either logistic regression or classification trees can perform better. A second observation is that bagging in combination with classification trees outperforms logistic regression. This is consistent with literature in that the combination of bagging and trees often improves classification performance substantially (Perlich, Provost & Simonoff, 2004) and it has been shown to outperform logistic regression (e.g., Risselada, Verhoef & Bijmolt, 2010). In sum, the performance of the classifiers we used is consistent with literature.

*p<.10, **p<.05, ***p<.01, ****p<.001

Fig. 3. Predictive performance across different lengths of the independent period

Next we'll discuss the relationship between the length of the predictors period on predictive performance. Interestingly, all three classifiers show a logarithmic increase in performance when the length of the independent period increases. After the fifth year, the increase in predictive performance seems to level off for two out of three classifiers (trees and trees + bagging). Moreover the significance of the difference with the minimum length of one year is not improving. Hence if the company in this study would take five years as a cut-off, it can discard almost 69% of its data and subsequently make substantial savings in data storage, data preparation and data analysis, with a minimum of loss in predictive performance. Other cut-offs can be considered too. For example, at a predictors period of 9 years, the increase in predictive performance of the third technique, logistic regression, also levels off. This still comes down to almost 44% percent of savings.

5   Conclusion and practical implications

In extant literature there are two main strategies to improve predictive performance: an algorithm-based strategy and a data- based strategy (Baecke & Van den Poel, 2010). The former revolves around the evaluation of multiple algorithms on given data and the improvement or invention of algorithms. The latter focuses on augmenting the existing database with new data sources. As we mention in the introduction, the third dimension of predictive modeling, the time window, remains under researched. Hence, in the second section of this study we provided a framework for strategies of classification improvement involving the time window. Next, we analyzed one of those strategies in an empirical study.

We have used logistic regression, classification trees and classification trees in combination with bagging to study the relation between the length of customer event history and classification performance. We can conclude that the length of the predictors period is logarithmically related to classification performance.

The practical implication of our findings is that companies can generate efficiency gains in the modeling process. In a world where data is growing at an exponential rate (Manyika , Chui, Brown, Bughin, Dobbs, Roxburgh & Byers, 2011, p16) companies are especially looking for efficiency. To the best of our knowledge, our study is the first to answer this call using a time window strategy that allows companies to fine-tune the costly process of gathering, storing, cleaning, preparing and analyzing data (e.g. Malthouse & Derenthal, 2008; Risselada, Verhoef & Bijmolt, 2010).

More specifically, the results show that the company in our analysis can trade a decrease of almost 69% of its data (i.e., increase in computational efficiency) for a small decrease in predictive performance (ranging from 0.0028 to 0.0001 depending on the classification technique) by reducing the length of the independent period from 16 years to 5 years.

6   Limitations and directions for future research

Although we are confident about our results, it is unclear whether they can be generalized over a wider range of subscription services. Hence, it would be interesting to validate our findings on customer databases in other industries. This would indicate whether the logarithmic relation between length of customer event history and predictive performance is typical for the newspaper industry or whether it holds for other services as well. Unfortunately, we only have limited access to such proprietary databases.

A second avenue for future research is to try additional analytical techniques. Although the techniques we used are the ones that are most commonly used in academics and business, it could prove valuable to investigate whether the same patterns can be observed with more recent advanced techniques such as random forests (Larivière & Van den Poel, 2005), neural networks (Zahavi & Levin, 1997), support vector machines (Shin & Cho, 2006) and random (multinomial) logit (Prinzie & Van den Poel, 2008).

Finally, in the second section of our paper we provided a framework for time window- based strategies for model improvement. This paper studies one of those strategies. It could be especially valuable to explore the insights that can be obtained from the analysis of other time window configurations.

7   References

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

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

Baecke, P., & Van den Poel, D. (2010). Improving purchasing behavior predictions by Data Augmentation with situational variables. *International Journal of Information Technology & Decision Making, 9*(6), 853-872.

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

Bauer, C. L. (1988). A Direct Mail Customer Purchase Model. *Journal of Direct Marketing, 2*(3), 16-24.

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

Breiman, L. (1996). Bagging Predictors. *Machine Learning, 24(*2), 123– 40.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Belmont: Wadsworth.

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.

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

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

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

DeLong, E.R., DeLong, D.M., & Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics, 44*(3)*, 837-845.

Ganesan, S. (1994). Determinants of long-term orientation in buyer–seller relationships. *Journal of Marketing, 58*(2), 1–19.

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

Haughton, D., & Oulabi, S. (1993). Direct Marketing Modeling with CART and CHAID. *Journal of Direct Marketing, 11*(4), 42–52.

Hill, S., Provost, F., & Volinsky, C. (2006). Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science, 21*(2), 256–276.

Hu, J., & Zhong, N. (2008). Web farming with clickstream. *International Journal of Information Technology & Decision Making, 7*(2), 291–308.

Hung, S.-Y., Yen, D.C., & Wang, H.-Y. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications*, 31(3), 515–524.

Kass, G.V. (1980). *An Exploratory Technique for Investigating Large Quantities of Categorical Data. Applied Statistics, 29(*2), 119–127.

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

King, R. D., Feng, C., & Sutherland A. (1995). Statlog — Comparison of Classification Algorithms on Large Real-world Problems. *Applied Artificial Intelligence, 9(*3), 289–333.

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

Larivière, B., & Van den Poel, D. (2005). Predicting Customer Retention and Profitability by Using Random Forests and Regression Forests Techniques. *Expert Systems with Applications, 29(*2), 472 – 84.

Lawrence, R. J. (1980). The Lognormal Distribution of Buying Frequency Rates. *Journal of Marketing Research, 17*(2), 212-220.

Lemmens, A., & Croux, C. (2006). Bagging and Boosting Classification Trees to Predict Churn. *Journal of Marketing Research , 43(*2), 276 –286.

Levin, N., & Zahavi, J. (2001). Predictive Modeling using Segmentation. *Journal of Interactive Marketing, 15(*2), 2 –22.

Levin, N.,& Zahavi, J. (1996). Segmentation Analysis with Managerial Judgment. *Journal of Direct Marketing, 10*(3), 28-47.

Malthouse, E.C., & Derenthal, K. M. (2008). Improving Predictive Scoring Models through Model Aggregation. *Journal of Interactive Marketing, 22(*3), 51–68 .

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A.H. (2011). Big data: The next frontier for innovation, competition and productivity. McKinsey Global Institute.

McCarty, J. A., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of Business Research, 60*(6), 656–662.

Morrison, D. G. (1966). Interpurchase Time and Brand Loyalty. *Journal of Marketing Research, 3(*3), 281-291.

Neslin, S.A., Gupta, S., Wagner, K., Lu, J., & Mason C.H. (2006). Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models. *Journal of Marketing Research , 43(*2), 204 –211.

Perlich, C., Provost, F., & Simonoff, J. S. (2004). Tree Induction vs. Logistic Regression: A Learning-curve Analysis. *Journal of Machine Learning Research, 4(*2), 211–255.
Prinzie, A., & Van den Poel, D. (2008). Random forests for multiclass classification: Random MultiNomial Logit. *Expert Systems with Applications, (34)*3, 1721-1732.

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

Reichheld, F.F., & Sasser, W.E. Jr. (1990). Zero defections: quality comes to service. *Harvard Business Review, 68*(5). 105-111.

Reinartz, W.J., & Kumar, V. (2002). The mismanagement of customer loyalty. *Harvard Business Review, 80*(7), 86-94.

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

Risselada, H., Verhoef P.C., Bijmolt, T.H.A. (2010). Staying Power of Churn Prediction Models. *Journal of Interactive Marketing, 24*(3), 198-208.

Shin, H., & Cho, S. (2006). Response modeling with support vector machines. *Expert Systems with Applications, 30*(4), 746–760.

Simpson, J. A. (1987). The Dissolution of Romantic Relationships: Factors Involved in Relationship Stability and Emotional Distress. *Journal of Personality and Social Psychology, 53*(4), 683-692.

Steenburgh, T. J., Ainsle, A., & Engbretson, P. H. (2003). Massively categorical variables, revealing the information in ZIP codes. *Marketing Science, 22*(1), 40–57.

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

Van den Poel, D., & Buckinx, W. (2005). Predicting online-purchasing behaviour. *European Journal of Operational Research, 166*(2), 557-575.

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

Zahavi, J., & Levin, N. (1997). Applying Neural Computing to Target Marketing. *Journal of Direct Marketing, 11*(1), 5 –22.