Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

Using machine learning techniques to predict defection of top clients

W. Buckinx[1], B. Baesens[2], D. Van den Poel[1], P. Van Kenhove[1] & J. Vanthienen[2]
[1]*Ghent University, Dept. of Marketing, Belgium*
[2]*K.U.Leuven, Dept. of Applied Economic Sciences, Belgium*

Abstract

Fierce competition in many industries causes switching behavior of customers. Because foregone profits of defected customers are significant, an increase of the retention rate can be very profitable. In this paper, we focus on the treatment of companies' most promising current customers in a non-contractual setting. We build a model in order to predict churn behavior of top clients who will (partially) defect in the near future. We applied the following classification techniques: logistic regression, linear discriminant analysis, quadratic discriminant analysis, C4.5, neural networks and Naive Bayes. Their performance is quantified by the classification accuracy and the area under the receiver operating characteristic curve (AUROC). The experiments were carried out on a real life data set obtained by a Belgian retailer. The article contributes in many ways. The results show that past customer behavior has predictive power to indicate future partial defection. This finding is from a companies' point of view even more important than being able to define total defectors, which was until now the traditional goal in attrition research. It was found that neural networks performed better than the other classification techniques in terms of both classification accuracy and AUROC. Although the performance benefits are sometimes small in absolute terms, they are statistically significant and relevant from a marketing perspective. Finally it was found that the number of past shop visits and the time between past shop incidences are amongst the most predictive inputs for the problem at hand.

Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

1 Introduction

Since competition in many industries is increasing, life cycles of customers are becoming more and more transitory. Customers are offered a tremendous array of choices. Certainly in a non-contractual setting it is not straightforward that relationships are permanent. Customers split their purchases among several competitive companies [1]. Consequently, it becomes increasingly difficult to build long-term relationships with your customers.

Customers' switching behavior results into foregone profits. On top of the lost sales revenue, new customers need to be attracted which requires very costly actions. Advertising efforts as well as promotions and sales costs are significant but necessary expenses to fill up the customer base and establish new relationships [2]. Besides, new clients often are not profitable for some time. Moreover, defecting (dissatisfied) customers are convinced that your company offers inferior value and might persuade other customers by spreading a negative word-of-mouth [3].

The goal of this study is to identify and predict partial defection in a non-contractual setting. We focus on companies' most promising current customers by trying to identify them as early as possible in the process of defection.

This paper is organized as follows: Section 2 describes the relevance of churn analysis and our partial defection approach. The description of the data set as well as an overview of the attributes and the experimental setup used to predict customers' defective behavior is discussed in Section 3. Section 4 presents the results. Section 5 summarizes the conclusions and Section 6 ends the paper with suggestions for further research.

2 Defection of top clients: literature review

As pointed out in the introduction, a non–contractual setting suffers from the problem that customers have the opportunity to continuously switch their purchase behavior. Certainly in a FMCG (fast-moving consumer good) retailer environment (the setting of this study) competition is severe and customers have many alternatives.

Reducing customer defections can have an enormous impact on companies' results. An improvement in retention rate of only one percentage point results in a significant increase in profits. Suppose 25% of the top clients defect. Considering an average contribution of 250 Euro a year and a discount rate of 6%, an improvement of a companies retention rate with only one percentage will cause an increase of their profits by 12 510 Euro over 5 year per 1000 top clients (cf. Table 1).
First of all, by implementing retention programs, customers are confronted with increasing switching costs, giving them fewer incentives to change their current behavior [4]. Besides, competitors have a difficult time to gain market share and new entrants are deterred to enter the market because it looks less attractive [5]. Secondly, a firms' profitability is influenced by the length of customers'

Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

Table 1: Additional contribution calculation

Retention rate	**Number of customers left**				**Total contribution over 5 years (in Euro)**	**Additional contribution over 75% (in Euro)**
	Year 1	Year 2	...	Year 5		
75%	1 000	750	...	316	699 494	0
76%	1 000	760	...	333	712 004	12 510 Euro

relationship. The longer a customer stays, the more he/she spends at the company [6]. Buyers tend to purchase additional services (products) and are more likely to convince others about the positive value the company offers (word-of-mouth effect). They tend to be less price sensitive and exhibit a lower responsiveness to competitive pull. Retained customers produce higher revenues and margin than new customers. So, it is supported that companies first spend their marketing resources to keep existing customers ratter than to attract new ones [7]. In summary, customer retention is a valuable strategy to ensure long-term profitability and success of the company.

The topic of defective customer behavior has been discussed extensively in recent literature. Churn analysis typically tries to define predictors of customer retention. In all of the cases, however, switching behavior is synonymous with total defection. People totally interrupt their relationship with the company. In other sectors, it is more complex to determine when customers are leaving. They switch some of their purchases to another store, i.e. they exhibit partial defection. The danger exists that after a while they will switch completely to the competitor. So in the long run partial defection can lead to total defection. Literature shows that a retail environment is not the most popular industry to analyse churn. In a retail setting, total defection is relatively rare. Yet, not one academic article could be found in the retailer sector, investigating partial defection. All analyses consider total defection.

3 Empirical study

3.1 General

For our analysis, one of the largest retailers with worldwide operations offering fast moving consumer goods provided the necessary data. Different purchase occasions could be traced by means of a loyalty card. It concerns individual data of 158 884 customers from April 2000 until January 2001, which represented a random sample from the entire customer base.

This research does not focus on the entire customer base. We are interested in the core of a valuable customer base which consists of loyal customers [8]. In our case we focus our study on those who shop frequently and at the same time exhibit a regular buying pattern. To define that segment of clients we use two behavioral attributes: the frequency of purchases and the time between their

Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

purchases (interpurchase time or IPT). Both variables are commonly used to define good customers [8]. Consequently, the customers in our segment of attention satisfy the following conditions:

(1) The frequency of purchases is higher than average.
(2) The ratio of the standard deviation of the interpurchase time to the mean interpurchase time is less than average.

The first attribute provides an indication of a customer's loyalty [9] and potential profitability. The second attribute ensures that the time between the customer's visits is regular.

It is important to set a clear definition of (partial) defection. We again take into consideration both conditions of the previous paragraph that are used to define our segment of interest. We want to predict whether customers belonging to the promising segment, continue to be loyal towards the company. If one of the previous mentioned conditions is not fulfilled, we classify a customer as defective.

We use the first five months of the available data, from April until August, to define the customers we focus. Consequently, we select 32 371 customers, which we consider to be promising clients. Applying our partial defection definition 8 140 customers churn. This is 25.15% of the clients under investigation.

3.2 Predictors

The available data consists of behavioral information at the level of the individual customer. Prior research already supports the incorporation of past purchase behavior. A major part of the churn analysis focuses on demographics as antecedents of defection.

Table 2: Predictor categories

Variable category	Description	Number of inputs
Interpurchase-time related variables	Variables concerning the number of days between shop incidences	4
Frequency of purchases	Variables concerning the frequency of shopping	4
Monetary indicators	Variables concerning the spending of a customer	3
Categorical shopping behavior	Variables concerning the spending in 12 different categories	14
Brand purchase behavior	Variables concerning type of products (national vs retailer's brand)	3
Length of the Relationship	Number of days somebody is client	1
Timing of shopping	Variables concerning moment in time of shopping	3
Mode of payment	Variables concerning mode of payment	10
Promotional behavior	Variables concerning other promotional behavior	4

Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

We computed 46 variables to predict churn behavior and classified them in 9 variable categories. The calculation process is done in two steps. After having defined 32 variables in a first step, we used a second step to compile another 15 variants on the variables that showed significant predictive power. Table 2 summarizes the 9 categories that were used in the analysis.

3.3 Experimental setup

We will consider the following classification techniques: logistic regression, linear discriminant analysis, quadratic discriminant analysis, C4.5, neural networks and Naive Bayes. Logistic regression and discriminant analysis are popular statistical classification methods [9]. C4.5 is a well-known decision tree induction algorithm based on entropy measures to decide upon the splits [10]. The NN classifiers were trained using the Bayesian evidence framework of David MacKay with the ARD extension [11]. We used only 1 hidden layer influenced by theoretical works and varied the number of hidden neurons from 1 to 10. We then chose the network with the best training set accuracy for evaluation on the test set. For the Naive Bayes classifier, we use the kernel approximation for continuous attributes. We set the confidence level for the pruning strategy of C4.5 to 0.25 which is the default value that is commonly used in the literature.

The performance of all trained classifiers will be quantified using both the classification accuracy and the area under the receiver operating characteristic curve (AUROC). The classification accuracy is undoubtedly the most commonly used measure of performance of a classifier. It simply measures the percentage of correctly classified (PCC) observations. However, it tacitly assumes equal misclassification costs and balanced class distributions. The receiver operating characteristic curve (AUROC) is a 2-dimensional graphical illustration of the sensitivity ('true alarms') on the Y-axis versus 1-specificity on the X-axis ('false alarms') for various values of the classification threshold [12]. It basically illustrates the behavior of a classifier without regard to class distribution or misclassification cost. The AUROC then provides a simple figure-of-merit for the performance of the constructed classifier. An intuitive interpretation of the AUROC is that it provides an estimate of the probability that a randomly chosen instance of class 1 is correctly rated (or ranked) higher than a randomly selected instance of class 0 [13].

We used McNemar's test to compare the PCCs of different classifiers. This chi-squared test is based upon contingency table analysis to detect statistically significant performance differences between classifiers. In [14], it was shown that this test has acceptable Type I error which is the probability of incorrectly detecting a difference when no difference exists

Multiple tests have been devised to compare AUROC values derived from the same test observations. Hanley and McNeil [15] developed a parametric approach which uses a z-statistic based on an approximation procedure that used the Pearson correlation or Kendall tau to estimate the correlation of the two AUROC's. DeLong, DeLong and Clarke-Pearson [16] suggested a non-

Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

parametric approach whereby the covariance matrix is estimated using the theory on generalized U-statistics. After a rather lengthy mathematical derivation, they arrived at a chi-square test statistic to test the difference of the correlated AUROC's (see [16] for more details). We will use this test in the remainder of this text.

4 Results

4.1 Classifiers' performance

Table 3 shows the predictive performance of the classification techniques. The neural network yielded the best absolute performance in terms of both PCC and AUROC. Although the absolute difference with the logistic regression classifier seems small at first sight, it is statistically significant according to McNemar's test and the test of DeLong, DeLong and Clarke-Pearson. Furthermore, in section 2 it was pointed out that in a customer retention context, small increases in the retention rate may yield substantial profit gains. Hence, the small difference can become very relevant from a marketing perspective.

Table 3: Performance of classification techniques

	PCC_{test}	$AUROC_{test}$
Logistic regression	75.57	79.02
Linear discriminant analysis	75.50	78.74
Quadratic discriminant analysis	71.13	76.47
C4.5	71.45	62.76
Naive Bayes	74.20	76.36
Neural network	**76.23**	**79.72**
KNN10	72.34	74.27
KNN100	72.45	77.22

4.2 Predictors

As mentioned before, we used the ARD extension of the evidence framework of MacKay to train the neural networks [11]. This extension introduces a separate weight regularisation term for each input. Using bayesian reasoning, it can then be proven that the weight regularisation parameter is inversely proportional to the variance of the corresponding weights of a Gaussian prior with mean 0. Hence, large weight regularisation parameters indicate less relevant inputs. We used the values of these parameters to rank order all inputs according to their relevance. The ten most important inputs are depicted in Table 4. The results show that the most powerful variables are classified in the categories of frequency of purchases (4), interpurchase time related variables (3), length of relationship (1), mode of payment (1) and promotional behavior (1).

Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

More specifically, the number of shop visits ever made by a customer is the most predictive variable.

Table 4: Importance of inputs

Rank	Variable	Description
1	Counter	Number of shop incidences
2	Frequency	Number of shop incidences / length of relationship
3	Counter2	Number of shop incidences during the last month
4	Meanrec	Average interpurchase time (IPT)
5	Lor	Length of relationship
6	Stdrec	Standard deviation of the IPT
7	Bench	Standard deviation of IPT / average IPT
8	Mop10	Amount of empty bottles returned
9	Counter3	Number of shop incidences during the last week
10	Freq	Number of shop incidences where coupon was used

5 Conclusions

Using our methodology, we are able to predict partial defection of top clients in a non-contractual setting. The analyses show that customers' past purchase behavior has significant predictive power to indicate future defectors.

Moreover, we are capable to track down partial defection, in contrast with past research that focused on total defection. This contribution is substantial because of several reasons. First of all, since we consider only promising clients the losses in terms of sales may be significant even if customers defect only partially. The average spending of a top client is 2 832 euro a year. Even if these clients switch only 10% of their expenditure to another store, the effect on companies' sales is remarkable. So avoiding this switching behaviour is valuable for the retailer (see Table 1: Additional contribution calculation). Secondly, partial defection can escalate and possibly lead to total defection in the long run. So being able to signal partial defection as early as possible will result in important returns and may even be of greater importance than predicting total defection. Consequently, marketing managers can define which of their customers do have a significant chance to decrease their loyal attitude towards the company. So they are able to execute specific marketing actions to these clients in order to prevent them from leaving.

The predictive performance of the classification technique is of crucial importance for the analysis and consequently for the manager to allocate his marketing activities. It was found that neural networks performed better than the other classification techniques in terms of both classification accuracy and AUROC. However, as we already indicated in part 2, though the performance

Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

benefits are sometimes small in absolute terms, they are statistically significant and relevant from a marketing perspective.

Finally, the number of past shop visits and the time between past shop incidences proved to be amongst the most predictive inputs for the problem at hand. Nevertheless, other predictor categories also seem to contain relevant variables.

6 Further research

Besides behavioral antecedents we can add inputs like demographics and perceptions in order to increase the predictive performance of the models. Unfortunately, the unavailability of this type of data for this anonymous retailer forces us to leave this as an issue for further research.

In order to make conclusions concerning the fundamental reasons of customer defection it will be interesting to open the neural network black box using e.g. rule extraction techniques. With these techniques it will be possible to highlight the reasons of defection and so it will be possible to customise marketing programs of the retailer more in detail.

References

[1] Reinartz, W.J. & Kumar, V., On the Profitability of Long-Life Customers in a noncontractual Setting: An Empirical Investigation and Implications for Marketing. *Journal of Marketing*, 64, 2000.

[2] Athanassopoulos, A.D., Customer Satisfaction Cues to Support Market Segmentation and Explain Switching Behavior. *Journal of Business Research*, 47, pp. 191-207, 2000.

[3] Mizerski, R.W., An Attribution Explanation of the Disproportionate Influence of Unfavorable Information. *Journal of Consumer Research*, 9, pp. 301-310, 1982.

[4] Jones, M.A., Mothersbaugh, D.L. & Beatty, S.E., Switching Barriers and Repurchase Intentions in Services. *Journal of Retailing*, 76(2), pp. 259-274, 2000.

[5] Rust, R.T. & Zahorik, A.J., Customer Satisfaction, Customer Retention, and Market Share. *Journal of Retailing*, 69(2), 1993.

[6] Zeithaml, V.A., Berry, L.L. & Parasuraman, A., The behavioural Consequences of Service Quality. *Journal of Marketing*, 60, pp. 31-46, 1996.

[7] Mozer, M.C., Wolniewicz, R., Grimes, D.B., Johnson, E., Kaushansky, H., Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry. *IEEE Transactions on Neural Networks*, 11(3), 2000.

[8] Ganesh, J., Arnold, M.J., Reynolds, K.E., Understanding the Customer Base of Service Providers: An Examination of the Differences Between Switchers and Stayers. *Journal of Marketing*, 64, pp. 65-87, 2000.

Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

[9] Couchen Wu & Hsiu-Li Chen, Counting your customers: Compounding customer's in-store decisions, interpurchase time and repurchasing behavior. *European Journal of Operational Research,* 127, pp. 109-119, 2000.

[10] Quinlan,J.R., C4.5 programs for machine learning, Morgan Kaufmann, 1993.

[11] MacKay, D.J.C., Bayesian methods for Adaptive Models, Phd. thesis, computation and Neural Systems, California Institute of Technology, Pasadena, CA, 1992.

[12] Egan, J.P., Signal Detection Theory and ROC analysis. Series in Cognition and Perception, Academic Press: New York,1975,

[13] Hanley, J.A. and McNeil, B.J., The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 148, pp. 839-843, 1983.

[14] Dietterich, T.G., Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7), pp. 1895-1924, 1998.

[15] Hanley, J.A. and McNeil, B.J., A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148, pp. 839-843, 1983.

[16] De Long, E.R. and De Long, D.M. and Clarke-Pearson, D.L., Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44, pp. 837-845, 1988.

[17] Swets, J.A., ROC analysis applied to the evaluation of medical imaging techniques. *Investigative Radiology*, 14, pp. 109-121, 1979.