UNIVERSITEIT GENT

FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

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

WORKING PAPER

Predicting Customer Profitability During Acquisition: Finding the Optimal Combination of Data Source and Data Mining Technique

Jeroen D'Haen[1]

Dirk Van den Poel[2]

Dirk Thorleuchter[3]

October 2012

2012/818

[1] PhD Candidate at Ghent University, Faculty of Economics and Business Administration, Tweekerkenstraat 2, 9000 Gent, Belgium, http://www.crm.UGent.be
[2] Professor of Marketing Analytics at Ghent University, Faculty of Economics and Business Administration, Tweekerkenstraat 2, 9000 Gent, Belgium, http://www.crm.UGent.be
[3] Fraunhofer INT, Appelsgarten 2, 53879 Euskirchen, Germany

D/2012/7012/51

Predicting Customer Profitability During Acquisition: Finding the Optimal Combination of Data Source and Data Mining Technique

Jeroen D'Haen
PhD candidate, Ghent University, Faculty of Economics and Business Administration, Department of Marketing, Tweekerkenstraat 2, 9000 Gent, Belgium

Dirk Van den Poel (corresponding author)
Professor of Marketing Analytics, Ghent University, Faculty of Economics and Business Administration, Department of Marketing, Tweekerkenstraat 2, 9000 Gent, Belgium, URL: http://www.crm.UGent.be
E-mail: dirk.vandenpoel@ugent.be
Phone: +32 9 264 89 80
Fax: +32 9 264 42 79

Dirk Thorleuchter
Fraunhofer INT, D-53879 Euskirchen, Appelsgarten 2, Germany

Abstract
The customer acquisition process is generally a stressful undertaking for sales representatives. Luckily there are models that assist them in selecting the 'right' leads to pursue. Two factors play a role in this process: the probability of converting into a customer and the profitability once the lead is in fact a customer. This paper focuses on the latter. It makes two main contributions to the existing literature. Firstly, it investigates the predictive performance of two types of data: web data and commercially available data. The aim is to find out which of these two have the highest accuracy as input predictor for profitability and to research if they improve accuracy even more when combined. Secondly, the predictive performance of different data mining techniques is investigated. Results show that bagged decision trees are consistently higher in accuracy. Web data is better in predicting profitability than commercial data, but combining both is even better. The added value of commercial data is, although statistically significant, fairly limited.

Keywords: marketing analytics; predictive analytics, data source; b2b; web mining; web crawling; bagging; profitability; customer acquisition; external commercial data

1. Introduction

The acquisition of new customers is considered a multi-stage process in which only certain leads become real customers (Cooper & Budd, 2007; Patterson, 2007; Yu & Cai, 2007). This process is generally a stressful undertaking for sales representatives. Fortunately, these sales reps are assisted by models that assist them in selecting the 'right' leads to pursue. Two factors are important in selecting the 'right' lead: the probability the lead will convert into a customer and the profitability of that lead once he/she is a customer. This paper focuses on the latter. The goal is to design a model that is able to predict a dichotomous version of profitability (i.e. yes a customer is profitable or no a customer is not profitable). Profitability models exist, however, the main bottleneck they have is a lack of quality data. A new data source is introduced to solve this problem and it is compared in its performance to a more traditional data source. Furthermore, we investigate the impact of the data mining technique utilized on the estimated models of both data sources and examine which combination provides the highest accuracy.

This paper investigates the impact of three techniques: logistic regression, decision trees and bagged decision trees. While logistic regression is a more basic data mining technique that is often used in research, (bagged) decision trees are more advanced and less popular. The reason to consider different data mining techniques is twofold. First, according to Neslin et al. (2006), which data mining technique is used has an impact on the predictive performance of the created models. So, employing different techniques is a way to increase the overall predictive performance by finding the optimal technique. Second, the data mining techniques are used as a proxy of data complexity and noisiness. Basic techniques are only capable of estimating simple, linear relations, while more advanced techniques are able to fit more complex, noisy data. If (bagged) decision trees are not able to perform better than logistic regression for a specific data source, we can conclude that this data source is most likely linear and noise-free in nature.

A quality model to predict profitability can only be constructed if quality data is available. Most models rely on commercial data purchased from specialized vendors (Rygielski, Wang, & Yen, 2002; Wilson, 2006). A relatively new and underinvestigated source of input for customer profitability models is textual information extracted from websites. Web mining and text mining can be used to gather this information from existing and potential customers' websites (Thorleuchter, Van den Poel, & Prinzie, 2012). However, textual information is seldom used as input for analyses in companies (Coussement & Van den Poel, 2009). The reason for this is that web data contains unstructured data that is hard to analyze. Nevertheless, latent indexing techniques can be used to make the data more structured and available as input for acquisition models (Thorleuchter et al., 2012).

This paper makes two main contributions to the existing literature. Firstly, it investigates the predictive performance of two sources of data: web data and commercially available data. The aim is to find out which of these two has the highest accuracy as input predictor for profitability and to research if they improve accuracy even more when combined. Secondly, the predictive performance of different data mining techniques is investigated. So the overall research question can be formulated as follows: which technique is most accurate in combination with which data source? These two main contributions also show in what way this paper is different from the one presented by Thorleuchter et al. (2012). It investigates and compares different data sources and data mining techniques instead of simply focusing on only web data using a logistic regression. In this way there is a clear benchmark (i.e. commercial data) to which web data can be compared. As a result, this paper can be seen as the first real test of using textual data extracted using web mining as input for profitability models. Furthermore, the results obtained in this paper are discussed in more detail.

The remainder of the paper is structured as follows. First, the web versus the commercially available data are discussed. Next we go deeper into the different data mining techniques. Third, a short description of the used data is given. Then, the results are presented. Finally, we end with a conclusion and discussion and we discuss the limitations of this paper and give suggestions for further research.

2. Web data versus commercially available data

Today, most companies construct huge databases containing a wealth of information on their customers and their buying behaviours (Shaw, Subramaniam, Tan, & Welge, 2001). In order to extract the knowledge hidden in these databases, data mining can be applied to them (Mitra, Pal, & Mitra, 2002). Nevertheless, this source of data is not applicable to identify new profitable customers (Arndt & Gersten, 2001). The databases constructed by companies represent company-internal information, which means that they only contain information on their own customers.

Most companies purchase lists of information on potential (i.e. new) customers from specialized external vendors (Wilson, 2006). These lists tend to be of poor quality. Superior quality lists exist, though at a much higher expense (Buttle, 2009; Shankaranarayanan & Cai, 2005). Inferior data will render inferior results: this is the so-called garbage in, garbage out rule (Baesens, Mues, Martens, & Vanthienen, 2009). The main quality problem of purchased data is the high amount of missing values.

An alternative to the commercially available data is the use of web mining to extract customer information data (Shaw et al., 2001). The challenge of web data is twofold (Stumme, Hotho, & Berendt, 2006). On the one hand, the data is so unstructured that only humans are capable of understanding it. On the other hand, the amount of data is too huge for humans to handle and it can therefore only be processed by computers. This challenge can be solved by combining web- with text- and data-mining. Web mining can extract different types of data: content, structure, usage and user profile data (Srivastava, Cooley, Deshpande, & Tan P.N., 2000). Content data is utilized in this paper as input to the proposed models. This type of data refers to the textual content that is seen when visiting a site. The textual information of customers' websites is consequently converted into term vectors in a term-space model (Thorleuchter et al., 2012). Latent semantic indexing is used to group related terms. Subsequently, singular value decomposition is applied to generate semantic generalizations. These generalizations are linked to the appearance of terms in similar web pages. Each generalization is a concept that refers to the hidden (latent semantic) patterns in the textual information. Companies get a score on each concept and these scores reflect how well a website loads on a specific concept (see Thorleuchter et al. (2012) for a more in-depth overview of this approach).

3. Data mining techniques

Data mining techniques are a way of extracting hidden knowledge in large databases (Ngai, Xiu, & Chau, 2009). Their importance is increasing in CRM analyses as the size of databases keeps growing (Ngai et al., 2009; Rygielski et al., 2002). Moreover, data mining is being used in the decision making process of companies (Baesens et al., 2009). The next part elaborates on the data mining techniques employed in this paper.

Logistic regression

Logistic regression is a regression analysis for categorical dependent variables and is based on the logit transformation of a proportion (Everitt & Skrondal, 2010; Field, 2009; Miguéis, Van den Poel, Camanho, & Falcao e Cunha, 2012). It is a standard parametric technique (Bellotti & Crook, 2008). The formula of a logistic regression is:

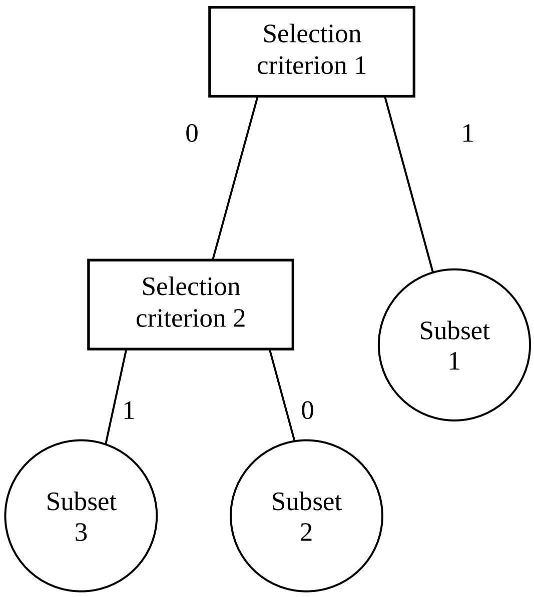
$$F(z) = \frac{1}{1+e^{-z}} \quad \text{where} \quad z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

(Blattberg, Kim, Kim, & Neslin, 2008a; Hansotia & Wang, 1997; Pampel, 2000; Thomas, 2010; Van den Poel & Buckinx, 2005) As logistic regression is an often used and well-known data mining technique we will not expatiate on this subject.

Decision trees

A decision tree divides a dataset in subsets, using the values of the independent variables as selection criteria, in order to predict the dependent variable (Blattberg, Kim, Kim, & Neslin, 2008b). The top of a decision tree is called the root node (Berk, 2008b). This root node contains the full dataset. The outcome of a decision at each node is called a split (Duda et al, 2001). Splits after the root node are termed branches and the final splits are the terminal nodes. All splits after the initial split imply interaction effects, unless they use the same predictor (Berk, 2008b). After the full tree is built, it needs to be pruned. Pruning is used to find the right size of the tree to avoid overfitting (Blattberg et al., 2008b; Duda, Hart, & Stork, 2001). The bigger a tree is, the less cases there are in the terminal nodes and the more chance there is of having an overfitted tree. Pruning a tree starts at the terminal nodes and works its way up to the top (Berk, 2008b). It eliminates nodes that do not reduce heterogeneity enough compared to the complexity they add to the tree. This is a version of Occam's razor that prescribes that researchers should prefer the simplest model that explains the data (Baesens et al., 2009; Duda et al., 2001). Decision trees have several specific advantages (Tirenni, Kaiser, & Herrmann, 2007). They are a non-parametric method, invariant to monotonic predictor transformations (i.e. no variable transformations are required). Parametric methods yield poor results when the dimensionality of data is high (as is in our case) (Petersen, Molinaro, Sinisi, & van der Laan, 2007). Furthermore, decision trees are robust to the effects of outliers. Figure 1 shows a graphical representation of a simple tree.

Figure 1 Decision tree



Bagging

A problem with a decision tree is that it has been shown to be unstable (Breiman, 1996b). This means that small changes in the training data (e.g. a different random selection) can cause large changes in the predictions. A method to overcome this instability is bagging, short for bootstrap aggregating, developed by Breiman (1996a). Bagging can be formalized as follows (Breiman, 1996a; Cunningham, Carney, & Jacob, 2000):

$$\hat{y}_{BAG} = \frac{1}{B}\sum_{b=1}^{B}\hat{\varphi}(x; T_b)$$

where *B* is the number of bootstrap samples of training set *T* and x is the input. $\hat{y}_{BAG}$ is the average of the different estimated trees (Fildes, Nikolopoulos, Crone, & Syntetos, 2008). A bootstrap sample is randomly drawn from the training set, but with replacement (Breiman, 1996a). Therefore, each observation can appear more than once in a single bootstrap sample or even not at all. The size of a bootstrap sample is usually chosen to be the same size as the training set (Martinez-Munoz & Suarez, 2010). It is important that when building bagged trees, the different trees are not pruned (Berk, 2008a). Unpruned trees have more variability which is needed to obtain a stable result when the trees are averaged. A bootstrap sample leaves out about 37% of the observations in the training data (Breiman, 1996a). There is no general rule as to how many bootstrap samples should be used. Breiman (1996a) found that in his case, 50 were enough, while 100 did not decrease the accuracy. That is why we decided to take 100 bootstrap samples. As each bootstrap sample is random, a bagged tree will be (slightly) different each time it is estimated. Twenty bagged trees are estimated and the minimum, maximum and average performances are reported.

Evaluation criterion

The area under the receiver operating curve (also know as the 'AUC') is calculated to evaluate the quality of a model. AUC is a common metric to estimate the accuracy of a model (Chen, Hsu, & Hsu, 2011). It can vary from 0.5 to 1, with 0.5 being a random model and 1 being the perfect model (Baecke & Van den Poel, 2011; Blattberg, Kim, Kim, & Neslin, 2008c). The method presented by DeLong et al. (1988) is used to compute whether there is a significant difference between two AUCs.

4. Data

The test case of our approach was a German B2B mail order company. Profitability was defined as having a sales volume that was higher than a certain threshold that was specified by the mail order company itself. This reflects customers that were profitable to the mail order company at the time of data extraction. Those companies were selected that had a web site. The discussed web and text mining procedures (see Section 2) were applied to the websites of these companies to extract the web data. This resulted in 200 dimensions that reflect specific concepts present on the different customer websites. The commercially available data was extracted from a database containing comparable financial information for public and private companies in Europe. The selection criterion in this database was that the companies had to be located in Germany and had to have a website address or e-mail address in the database. A selection of variables had to be made, because most variables contained a high amount of missing values, showing the omnipresent quality problem in commercially available data. A list of variables such as total assets, sales and liquidity were retained. After matching both datasets and deleting some final missing observations, a final set of 2911 companies was retained of which 65% were profitable and 35% were not profitable. Two-thirds of the dataset was randomly divided in a training set and one-third in a test set as suggested by Blattberg et al. (2008c). The training dataset is chosen to estimate the models and the test set is used to calculate the predictive performance of the models.

5. Results

Table 1 shows the overall result of the different data mining techniques combined with the different sources of data. The overall impression in Table 1 is that bagging trees works best (it has the highest AUC). Also, web data renders better results than commercial data, but combining both data sources is even better. It might seem odd that combining commercial data and web data renders a lower AUC than solely using commercial data. This can be explained by the fact that a logistic regression analysis can not handle high dimensionality, even when selection methods are applied. As a result, the regression is fitting a suboptimal model. Furthermore, in the case of web data and combining the data, the more advanced technique bagging outperforms the regression analysis, while in the commercial data case there is no difference in predictive performance. So, it is likely that web data contains more noise than commercial data and that it is non-linear in nature. Further analyses will show if these results are statistically significant.

Table 1 AUC results

		Commercial data	Web data	Combined
Regression		0.6124	0.5568	0.5602
Decision Tree		0.5000	0.5000	0.5000
Bagged Tree	Min	0.6153	0.6827	0.7195
	Max	0.6312	0.7251	0.7564
	Avg	0.6236	0.7021	0.7367

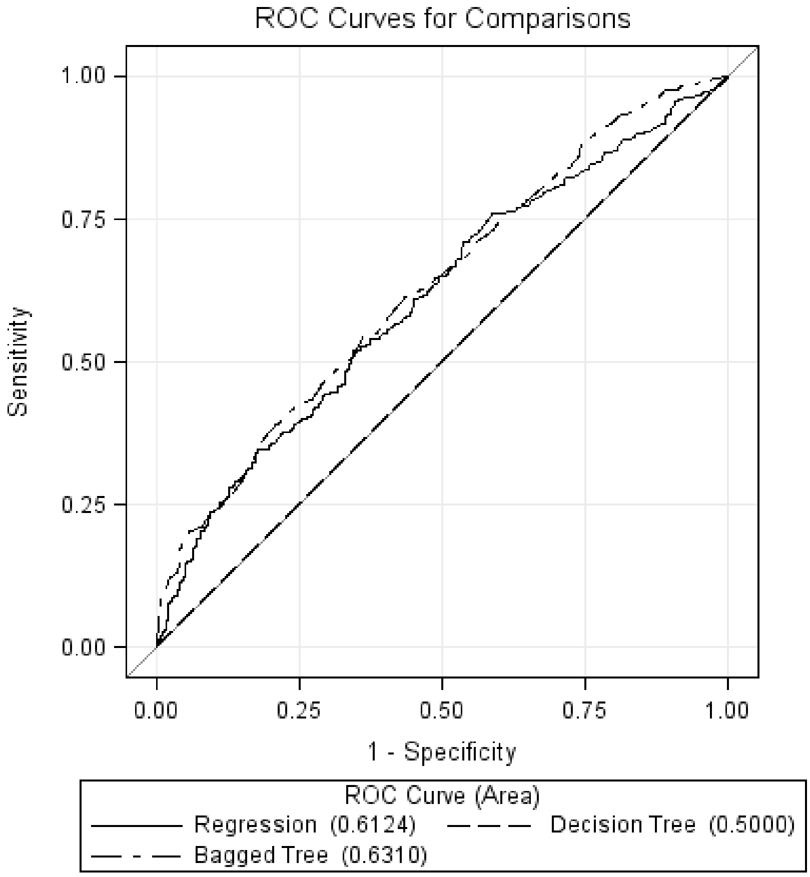
Decision trees rendered an AUC of 0.5, regardless of which data type was used (Table 1). The reason for this is that after pruning the tree, only the root nodes were retained. As a result, the decision trees gave just one constant value as prediction. In Table 2 we see that both regression and the bagged tree (examining the one with the highest AUC) have a significantly higher accuracy compared to a decision tree. The bagged tree and regression are not significantly different. This is also illustrated in Figure 2 where the lines of regression and the bagged tree intersect.

Table 2 AUC results commercial data

Contrast	Difference	χ^2	Pr > χ^2
	0.1124	35.4206	<.0001

Regression - Tree			
Tree - Bagged tree	-0.1310	49.0279	<.0001
Bagged tree – Regression	0.0186	0.9191	0.3377

Figure 2 ROC curves commercial data



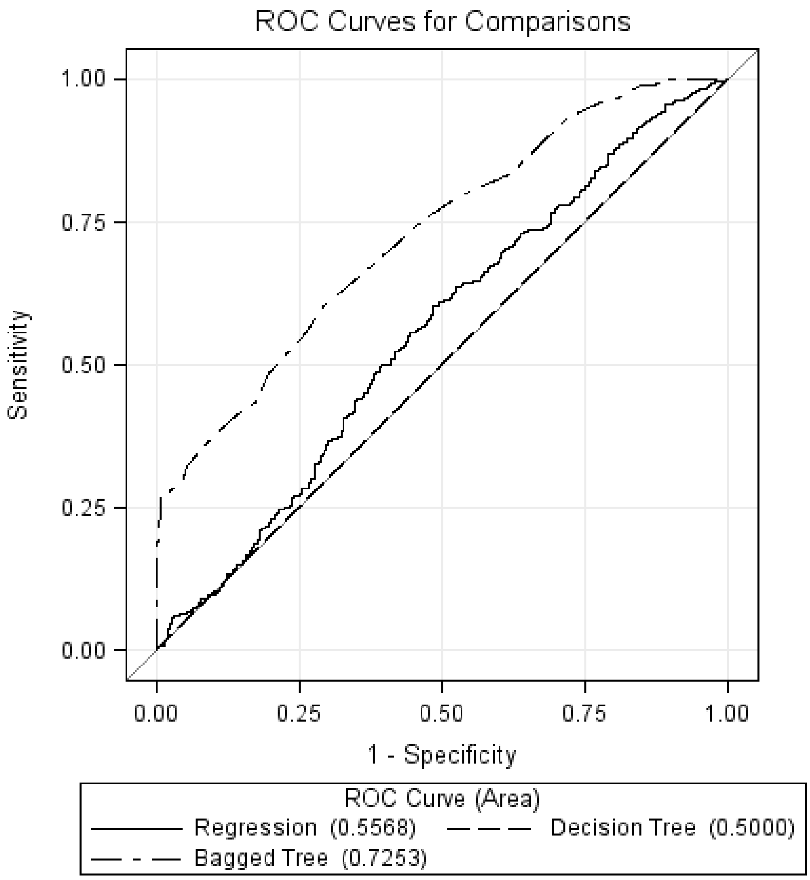
Regarding web data, it is clear that bagging has a significantly higher accuracy than regression and normal decision trees (Table 3). Figure **3** shows that there is no intersection between the bagged tree and any of the other data mining techniques. Regression is performing better than the decision tree, but it still has a relatively low accuracy (AUC of 0.56, Table 1).

Table 3 AUC results web data

Contrast	Difference	χ^2	Pr > χ^2
Regression - Tree	0.0568	7.9541	0.0048

Tree - Bagged tree	-0.2253	185.1293	<.0001
Bagged tree – Regression	0.1685	64.3068	<.0001

Figure 3 ROC curves web data



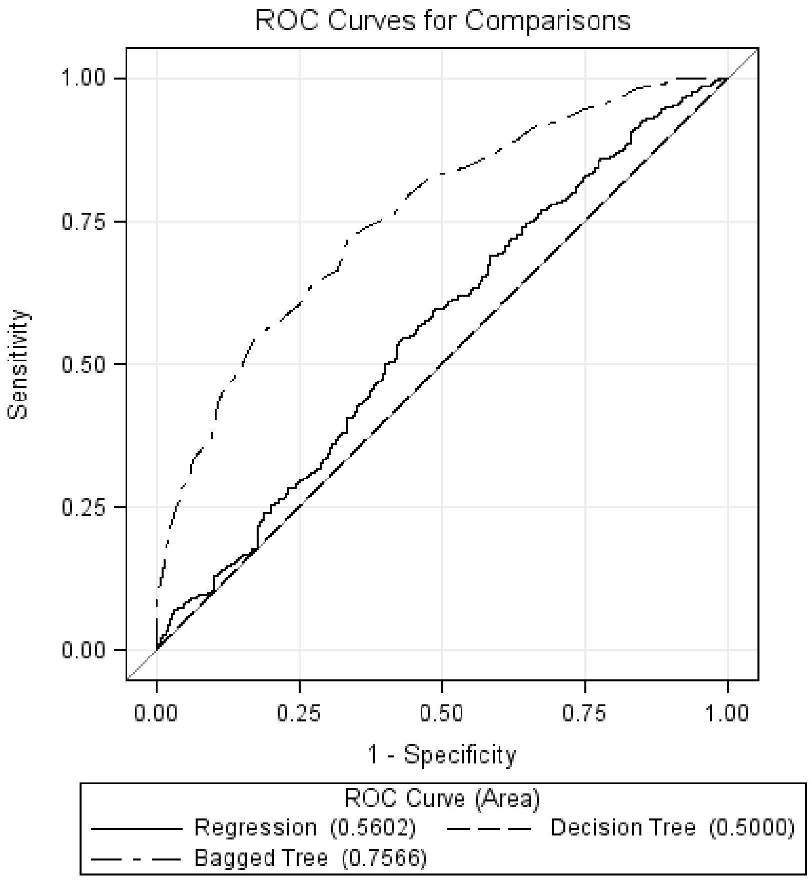
The next step is to combine both data sources (web + commercially available external data) and see what the predictive performance is of the different data mining techniques. Again, regression is doing significantly better than the decision tree (Table 4), although it still has a relatively low performance (AUC = 0.56, Table 1). Furthermore, when combining both data sources, regression is performing worse than when only the commercial data was used (Table 1). Bagging trees, that has the highest AUC, performs significantly better than both regression and normal decision trees (Table 4). This is also clearly shown in Figure 4.

Table 4 AUC results combined data

Contrast	Difference	χ^2	Pr > χ^2
Regression - Tree	0.0602	9.0263	0.0027
Tree - Bagged tree	-0.2566	256.1839	<.0001

Bagged tree – Regression	0.1965	87.2238	<.0001

Figure 4 ROC curves combined data



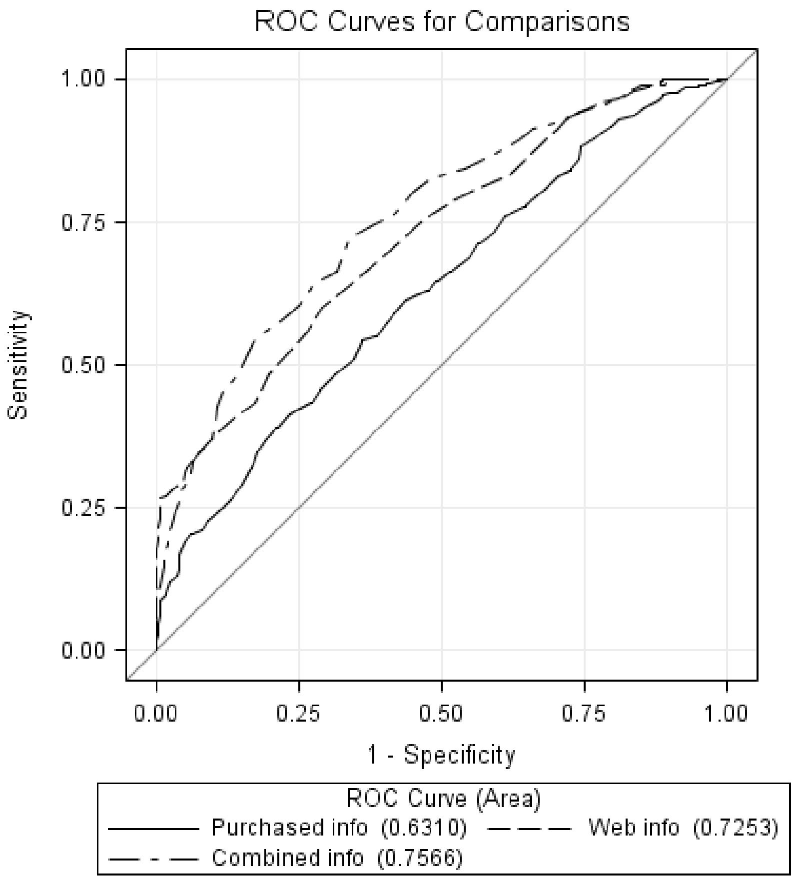
The final step is to compare the best data mining techniques for each data source (bagged trees in this case) and check which data type renders the best results. The web data has significantly better results than the commercial data, but combining both data types elevates the predictive performance even more (Table 5). Figure 5 shows this graphically. When bagging decision trees it is also possible to get a measure of variable importance. Most of the top 10 important variables were web data variables, but two of them were from the commercial data set. The loans and capital of a company were two important predictors in company profitability, being the fourth and ninth most important variables respectively.

Table 5 AUC results best data mining techniques

Contrast	Difference	χ^2	Pr > χ^2
Commercial – Web	-0.0943	14.1705	0.0002

Web – Combined	-0.0313	4.5761	0.0324
Combined - Commercial	0.1256	33.6046	<.0001

Figure 5 ROC curves best data mining techniques



6. Conclusion and discussion

The goal of this paper was to investigate which data mining techniques worked best in predicting customer profitability in combination with which data source. The techniques under investigation were logistic regression, decision trees and bagged decision trees. Two types of data were used: data originating from web mining and data purchased from a specialized vendor. The web data is free and available to anyone with internet access. Regardless of data source, it was the bagging of decision trees that provided the highest AUC (except for commercial data; in this case regression worked equally well). Web data had a higher predictive performance compared to commercial data, but the combination of both data types rendered the best results. This has the following managerial implications. Bagged decision trees should be preferred over logistic regression and normal decision trees to build a model. Moreover, web data is the ideal starting input for this model. If the budget is available to buy external data, this can be combined with web data to further increase the predictive performance of the model. However, a cost-benefit analysis should be done to find out whether the high cost of buying data is justified by the (relatively) small increase in predictive power.

7. Limitations and further research

The profitability definition that is utilized in this paper is a snapshot as variability can vary from one year to the other. Furthermore, the dataset will contain customers that have been with the company for a longer period and others who recently became customer. Further research should employ a definition of variability that covers this time aspect. A second limitation are the decision trees that always rendered an AUC of 0.5 because of the specific pruning method used. However, it is our conviction that, even with different pruning methods, a single decision tree will not produce surprising results. As explained in Section 1, two factors play a role in the customer acquisition process: the probability of converting into a customer and the profitability once the lead is in fact a customer. This paper focused on predicting customer profitability using a combination of data sources and data mining techniques. Further research should investigate whether web data can be implemented in customer acquisition models as this was outside the scope of this paper.

References

Arndt, P. & Gersten, W. (2001). External Data Selection for Data Mining in Direct Marketing. In *Proceedings of the International Conference on Information Quality* (pp. 44-61).

Baecke, P. & Van den Poel, D. (2011). Data augmentation by predicting spending pleasure using commercially available external data. *Journal of Intelligent Information Systems, 36,* 367-383.

Baesens, B., Mues, C., Martens, D., & Vanthienen, J. (2009). 50 years of data mining and OR: upcoming trends and challenges. *Journal of the Operational Research Society, 60,* S16-S23.

Bellotti, T. & Crook, J. (2008). Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society, 60,* 1699-1707.

Berk, R. A. (2008a). Bagging. In *Statistical Learning from a Regression Perspective* (pp. 169-192). Springer Verlag.

Berk, R. A. (2008b). Classification and Regression Trees (CART). In *Statistical Learning from a Regression Perspective* (pp. 103-166). Springer Verlag.

Blattberg, R. C., Kim, P., Kim, B. D., & Neslin, S. A. (2008a). Acquiring Customers. In *Database marketing: analyzing and managing customers* (pp. 495-514). Springer.

Blattberg, R. C., Kim, P., Kim, B. D., & Neslin, S. A. (2008b). Decision Trees. In *Database marketing: analyzing and managing customers* (pp. 423-441). Springer.

Blattberg, R. C., Kim, P. D., Kim, B. D., & Neslin, S. A. (2008c). Statistical Issues in predictive Modeling. In *Database marketing: analyzing and managing customers* (pp. 291-321). Springer.

Breiman, L. (1996a). Bagging predictors. *Machine Learning, 24*, 123-140.

Breiman, L. (1996b). Heuristics of Instability and Stabilization in Model Selection. *The Annals of Statistics, 24,* 2350-2383.

Buttle, F. (2009). Managing the Customer Lifecycle: Customer Acquisition. In *Customer relationship management: concepts and technologies* (2 ed., pp. 225-254). Taylor & Francis.

Chen, W. C., Hsu, C. C., & Hsu, J. N. (2011). Optimal Selection of Potential Customer Range through the Union Sequential Pattern by Using a Response Model. *Expert systems with applications, 38,* 7451-7461.

Cooper, M. J. & Budd, C. S. (2007). Tying the Pieces Together: A Normative Framework for Integrating Sales and Project Operations. *Industrial Marketing Management, 36,* 173-182.

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

Cunningham, P., Carney, J., & Jacob, S. (2000). Stability problems with artificial neural networks and the ensemble solution. *Artificial Intelligence in Medicine, 20,* 217-225.

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,* 837-845.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Non-metric Methods. In *Pattern Classification* (2 ed., pp. 1-66). Wiley.

Everitt, B. S. & Skrondal, A. (2010). *The Cambridge Dictionary of Statistics*. (4 ed.) Cambridge University Press.

Field, A. (2009). Logistic Regression. In *Discovering Statistics using SPSS* (3 ed., pp. 264-315). Sage.

Fildes, R., Nikolopoulos, K., Crone, S. F., & Syntetos, A. A. (2008). Forecasting and operational research: a review. *Journal of the Operational Research Society, 59,* 1150-1172.

Hansotia, B. J. & Wang, P. (1997). Analytical Challenges in Customer Acquisition. *Journal of direct marketing, 11,* 7-19.

Martinez-Munoz, G. & Suarez, A. (2010). Out-of-bag estimation of the optimal sample size in bagging. *Pattern Recognition, 43,* 143-152.

Miguéis, V. L., Van den Poel, D., Camanho, A. S., & Falcao e Cunha, J. (2012). Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert systems with applications, 39,* 11250-11256.

Mitra, S., Pal, S., & Mitra, P. (2002). Data Mining in Soft Computing Framework: A Survey. *IEEE transactions on neural networks, 13,* 3-14.

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

Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification. *Expert systems with applications, 36,* 2592-2602.

Pampel, F. C. (2000). *Logistic Regression: a Primer*. Sage Publications.

Patterson, L. (2007). Marketing and Sales Alignment for Improved Effectiveness. *Journal of digital asset management, 3,* 185-189.

Petersen, M. L., Molinaro, A. M., Sinisi, S. E., & van der Laan, M. J. (2007). Cross-validated bagged learning. *Journal of Multivariate Analysis, 98,* 1693-1704.

Rygielski, C., Wang, J., & Yen, D. C. (2002). Data Mining Techniques for Customer Relationship Management. *Technology in society, 24,* 483-502.

Shankaranarayanan, G. & Cai, Y. (2005). A Web Services Application for the Data Quality Management in the B2B Networked Environment. In *Proceedings of the 38th Hawaii International Conference on System Sciences* (pp. 1-10).

Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems, 31,* 127-137.

Srivastava, J., Cooley, R., Deshpande, M., & Tan P.N. (2000). Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. *SIGKDD Explorations Newsletter, 1,* 12-23.

Stumme, G., Hotho, A., & Berendt, B. (2006). Semantic Web Mining: State of the art and future directions. *Web Semantics: Science, Services and Agents on the World Wide Web, 4,* 124-143.

Thomas, L. C. (2010). Consumer finance: challenges for operational research. *Journal of the Operational Research Society, 61,* 41-52.

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

Tirenni, G., Kaiser, C., & Herrmann, A. (2007). Applying Decision Trees for Value-based Customer Relations Management: Predicting Airline Customers' Future Values. *Database Marketing & Customer Strategy Management, 14,* 130-142.

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

Wilson, R. D. (2006). Developing New Business Strategies in B2B Markets by Combining CRM Concepts and Online Databases. *Competitiveness Review: An International Business Journal incorporating Journal of Global Competitiveness, 16,* 38-43.

Yu, Y. P. & Cai, S. Q. (2007). A New Approach to Customer Targeting under Conditions of Information Shortage. *Marketing intelligence & planning, 25,* 343-359.