

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/249737446>

Marketing Segmentation Through Machine Learning Models: An Approach Based on Customer Relationship Management and Customer Profitability Accounting

Article in *Social Science Computer Review* · April 2008

DOI: 10.1177/0894439308321592

CITATIONS

16

READS

6,536

2 authors:



Raquel Florez

Universidad Pablo de Olavide

42 PUBLICATIONS 416 CITATIONS

[SEE PROFILE](#)



Juan M. Ramon

Universidad Pablo de Olavide

29 PUBLICATIONS 201 CITATIONS

[SEE PROFILE](#)

Social Science Computer Review

<http://ssc.sagepub.com/>

Marketing Segmentation Through Machine Learning Models : An Approach Based on Customer Relationship Management and Customer Profitability Accounting

Raquel Florez-Lopez and Juan Manuel Ramon-Jeronimo

Social Science Computer Review 2009 27: 96 originally published online 8 August 2008

DOI: 10.1177/0894439308321592

The online version of this article can be found at:

<http://ssc.sagepub.com/content/27/1/96>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Social Science Computer Review* can be found at:

Email Alerts: <http://ssc.sagepub.com/cgi/alerts>

Subscriptions: <http://ssc.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://ssc.sagepub.com/content/27/1/96.refs.html>

>> [Version of Record](#) - Dec 29, 2008

[OnlineFirst Version of Record](#) - Aug 8, 2008

[What is This?](#)

Marketing Segmentation Through Machine Learning Models

An Approach Based on Customer Relationship Management and Customer Profitability Accounting

Raquel Florez-Lopez

University of Leon, Spain

Juan Manuel Ramon-Jeronimo

University Pablo Olavide of Seville, Spain

Customer relationship management (CRM) aims to build relations with the most profitable clients by performing customer segmentation and designing appropriate marketing tools. In addition, customer profitability accounting (CPA) recommends evaluating the CRM program through the combination of partial measures in a global cost–benefit function. Several statistical techniques have been applied for market segmentations although the existence of large data sets reduces their effectiveness. As an alternative, decision trees are machine learning models that do not consider a priori hypotheses, achieve a high performance, and generate logical rules clearly understood by managers. In this article, a three-stage methodology is proposed that combines marketing feature selection, customer segmentation through univariate and oblique decision trees, and a new CPA function based on marketing, data warehousing, and opportunity costs linked to the analysis of different scenarios. This proposal is applied to a large insurance marketing data set for alternative cost and price conditions showing the superiority of univariate decision trees over statistical techniques.

Keywords: *artificial intelligence; customer relationship management; customer profitability accounting; cost–benefit analysis; marketing*

Customer relationship management (CRM) focuses on identifying the most profitable customers to build stable relationships and to optimize the financial results of the company (Malmi, Raulas, Gudergan, & Sehm, 2004; Van Raait, Vernooij, & Van Triest, 2003). The importance of these CRM programs have increased in past years due to a competitive environment that combined the sociodemographic characteristics of retail consumers and the specialization of sellers and buyers and forced the company to secure a dynamic management of clients to achieve higher profits and to gain a higher share of the market than its competitors (Kim, Street, & Menczer, 2001). To develop a complete CRM program, several stages need to be completed (Winer, 2001). From there, the initial segmentation of clients and the design of metrics for measuring the success of the CRM program are particularly critical, despite several limitations about their development.

Authors' Note: Please address correspondence to Raquel Florez-Lopez, Department of Economics and Business Administration, University of Leon, Leon, Spain; e-mail: raquel.florez@unileon.es.

Statistical techniques, such as cluster and principal component analysis (PCA) combined with discriminant analysis (DA) or logistic regression, have been traditionally used for building segmentation models, but the existence of large volumes of data linked to multiple-correlated features reduce the real fit, robustness, and interpretability of these models (Berger & Nasr, 1998; Rao & Steckel, 1995; Reinartz & Kumar, 2000; Schmittlein & Petersen, 1994). Contrary to statistical models, machine learning techniques do not consider restrictive a priori hypotheses about data and, combined with data warehousing from previous clients, might provide strategic information to optimize the efficiency of promotional politics, using previous behavioral trends to develop effective predictive models that forecast future customer decisions (Gath & Geva, 1988; Hruschka & Natter, 1999; Kim, Street, Russell, & Menczer, 2005). Decision trees are one of the most interesting machine learning alternatives for market segmentation: they are a good fit for different scenarios and generate logical rules that are clearly understood by managers, making them easier to reach effective management decisions (Breiman, Friedman, Olshen, & Stone, 1984).

With reference to metrics for measuring the efficiency of CRM programs, some measures have been developed (Winer, 2001), such as market share, profit margins, or customer acquisition costs. Nevertheless, these proposals are only partial metrics and must be combined with a final function to obtain a true measure of the success of CRM. Measurement of client overall profitability is the central aim of customer profitability accounting (CPA). However, there is a surprising lack of studies that combine both items, CPA and CRM (Malmi et al., 2004) and management literature has paid limited attention to customer profitability (Bouwens & Abernethy, 2000; Foster & Gupta, 1994; Guilding & McManus, 2002). Also, the limited literature combining CRM and CPA is very recent and mostly focuses on the comparison of the overall profitability of companies that develop (or not) CRM politics, but cost-benefit analyses have not been performed on the specific CRM tools to be implemented (Gerdin & Greve, 2004; Guilding & McManus, 2002; Malmi et al., 2004).

In this article, the development of a CPA model based on cost-benefit functions is analyzed and applied to measure the success of the CRM in view of specific promotional politics (mailing). This analysis includes mailing and data warehouse costs and relates them to the incremental profit derived from successful contacts and opportunity costs from unsuccessful contacts. Our approach differs from previous studies on direct marketing because of the consideration of a complex objective function and data reduction based on feature selection, together with searching for both accurate and understood segmentation models, which include statistical techniques, univariate, and oblique decision trees. Moreover, the impact of a CRM orientation on performance is analyzed and a new success measurement approach is provided for both CRM and CPA, which could be considered as a basis for further development.

This article is structured as follows. In the next section, CRM and CPA issues are analyzed and a theoretical approach to customer segmentation is provided. The section Methodological Analysis summarizes our methodological proposal to implement successful promotional politics, using univariate and oblique decision trees and cost-benefit functions for different scenarios. An empirical application is conducted in the subsequent section, using data from an insurance company. Conclusions are summarized in the final section.

Customer Relationship Management and Customer Profitability Accounting

In previous years, global markets, the increase in competitiveness, and the development of new methods to contact clients (based on marketing techniques such as the Internet, catalogs, and telephone sales, etc.), have caused a major change in buyers' behavior, which is more heterogeneous in nature and more difficult to predict than in earlier periods. In addition, the social and demographic patterns change more quickly each time and companies need to develop a dynamic customer management to identify clusters characterized by similar behaviors and responses to promotional politics (Kim et al., 2001). Considering the small attraction space provided by traditional physical-shop distribution channels and specialized supply and demand, retailing managers must focus on improving the efficiency of communication politics, by offering products and services that meet the clients' needs.

The theory about CRM suggests that buyers should identify their most profitable customers and should focus on building and nurturing these relationships well (Malmi et al., 2004). The allocation of the company's sales and marketing resources on some specific customers should be based on a cost-benefit analysis, linked to the careful identification of those segments of customers that have the best aptitudes for establishing future profitable relationships (Van Raaij et al., 2003). This article considers the development of a CPA model to achieve the best CRM, with the purpose of optimizing marketing resources in the launch of a new product.

Winer (2001) proposed a basic model, which takes into account what managers need to know about their customers and how that information is used to develop a complete CRM perspective, through seven different stages: (a) a database of customers' activities; (b) analyses of the database; (c) given the analyses, decisions about customers to be targeted; (d) tools for targeting the customers; (e) how to build relationships with the targeted customers; (f) privacy issues; and (g) metrics for measuring the success of the CRM program.

It is related to the so-called "database marketing process" that considers five main stages in any marketing politics (Tooker, 2006): (a) readiness assessment for determining the company's database marketing needs, objectives, and capabilities, together with the review of products, offers, and communicational strategies; (b) database creation and management that considers different sources of customer information and selects the information to be stored and periodically updated; (c) campaign development and execution that develop customer segmentation through efficient models, decision and implementation of marketing strategies that the company needs to embrace to achieve its goals; (d) response management, related to the handling of responses and the building of measures on sales efficacy and sales losses; and (e) feedback management that uses modeling and response analytics to tune models over time to make them more efficient and profitable.

Previous proposals are focused on the identification of profitable clusters of clients or market segmentation and the building of metrics on clients' profitability, which could be implemented through CPA.

Characteristics of Market Segmentation

Despite the extensive literature on market segmentation, there is no overall consensus about the optimal segmentation methodology (Brijs, 2002). One reason for this diversity is the fact that segmentation can be observed from different perspectives.

Since Smith's (1956) pioneering article, different definitions of "market segmentation" have been proposed, which could be divided into two broad categories (Dibb & Stern, 1995): (a) segmentation as a strategy, related to targeting products to a selection of customers (Wind, 1978) and (b) segmentation as a methodology, related to techniques and methods employed for clients' clustering (Bass, Tigert, & Lonsdale, 1968; Green & Wind, 1975).

In a generic sense, market segmentation might be defined as the process whose aim is, "the partitioning of the market into homogeneous sub-markets in terms of customer demand, resulting in the identification of groups of customers that respond differently to the marketing mix" (Brijis, 2002, p. 94).

The principal benefits to be gained from the establishment of a market segmentation strategy are the efficient allocation of marketing resources and the design of specialized products and services adapted to buyers' needs. On this point, a recent international research focused on business to business direct mail estimated that the average true cost of poorly targeted marketing resources per company per annum might be in excess of 100,000 pounds sterling¹ (QAS, 2006) and data from consumer direct mail might be even higher. Other costs should also be considered, such as the damage to the company's reputation arising from poorly targeted mailings and the cost of warehousing and preparing the data for direct mail campaigns.

Focusing on marketing segmentation as a methodology, the intensive mining of customers' databases has become one of the strongest tools to be applied by companies that use previous experience to identify the more profitable buyers. It is particularly useful for the management of promotional politics such as mailing or direct marketing based on the World Wide Web.

Different attributes might be used for market segmentation, which include demographic, psychographics, and socioeconomic variables, collected through research, interviews, meetings, questionnaires, samples, and other techniques (Tseng & Huang, 2007). The choice for one or a combination of these segmentation bases largely depends on the business question under analysis (Wind, 1978). Such attributes can be classified according to two different criteria (Wedel & Kamakura, 2000): (a) general versus product-specific bases, which are independent or dependent (respectively) of products, services, and circumstances; and (b) observable versus nonobservable bases, which are directly measured or must be deduced respectively (Table 1).

With reference to methods used for market segmentation, the literature considers a wide variety of techniques. Statistical models have been traditionally employed for customer clustering, but in previous years new tools have been used, which consider both Bayesian models and machine learning paradigms (decision trees, artificial neural networks, genetic algorithms; Gath & Geva, 1988; Hruschka & Natter, 1999; Kim et al., 2005). In spite of this variety of methods, it is possible to classify these according to two dimensions (Table 1): (a) *a priori* segmentation versus *post hoc* segmentation and (b) predictive versus descriptive methods (Wedel & Kamakura, 2000).

Descriptive techniques analyze the associations across an overall set of variables (dependent and independent ones); cluster-based methods are classical descriptive techniques of market segmentation (Green, 1977; Wind, 1978). Opposite predictive techniques analyze the relationship between the set of dependent attributes and the set of independent variables.

Table 1
Classification of Segmentation Bases and Clustering Methods

Segmentation Bases	General	Product-Specific
Observable	Cultural, demographic, geographic, and socioeconomic variables	Usage frequency, brand loyalty, store loyalty and patronage, usage situations, purchase moment
Nonobservable	Psychographics, values, personality, and lifestyle	Psychographics, benefits, perceptions, elasticities, attributes, preferences, intentions
Clustering Methods	A Priori	Post Hoc
Descriptive	Contingency tables, log linear models	Hierarchical clustering, optimal clustering, latent class cluster models
Predictive	Cross-tabulation, regression, discriminant analysis, decision trees, heteroassociative neural networks	Autoassociative neural networks, latent class regression models

A priori segmentation is characterized by the definition of the type and number of segments by managers, using certain heuristics or business experience; on the contrary, post hoc segmentation uses statistical techniques to identify clusters of similar customers that have similar values on a number of segmentation variables (Brijs, 2002, p. 207).

If segmentation is related to marketing politics directed to households with no prior relationships with the company, decisions are based on the analysis of the relationship between independent attributes and the response to a mailing test taken from a representative household sample (Kim et al., 2005). It could be analyzed as an extreme form of a priori predictive segmentation, where the type (buyers vs. nonbuyers) and number of segments (two) are defined a priori and a set of independent variables is used to predict cluster membership. But it also includes a post hoc predictive segmentation, as once potential buyers are identified, several segments should be automatically obtained within them to develop specific marketing tools.

Statistical methods such as DA and logistic regression have traditionally dominated the group of techniques used by database managers in direct mail marketing (Berger & Nasr, 1998; Rao & Steckel, 1995; Reinartz & Kumar, 2000; Schmittlein & Petersen, 1994), even if a priori statistical assumptions were not fulfilled in presence on qualitative data (Dibb & Stern, 1995; Tseng & Huang, 2007).

Nevertheless, the current complex business environment, together with cheap data warehouse systems and increased computer power, have recently allowed the accumulation of large amounts of client data so that each customer is represented within the company through multiple attributes (even more than a hundred in many companies). This excess of data presents managers with three different problems:

- How to select relevant attributes for identifying potential buyers, from a wide group of potential independent variables.
- How to understand the logic of the model's predictions, if they include too many explicative attributes.

- How to verify initial hypotheses of statistical models in the existence of correlated attributes (independence, homoscedasticity, normality).

As a feature selection solution, PCA has been widely used to reduce data dimension in order to prepare new variables that are then used as predictors in a DA or a logistic regression (Dibb & Stern, 1995). Nevertheless, PCA has several drawbacks in terms of predictive modeling, interpretation of results and cost (Kim et al., 2005; Tabachnick & Fidell, 1983): (a) PCA does not consider the relationship between the independent and dependent variables in the process of data reduction, thus irrelevant variables could be included in the model; (b) PCA obtains components, which are difficult to interpret, in particular if a high number of variables exist; and (c) PCA requires to collect all the original predictive variables, which increases the data warehousing cost. The interpretability problem is for the most part critical, because the management utility of a particular segmentation is directly related to be simple to interpret (Anderberg, 1973; Doyle & Saunders, 1985).

In recent decades, feature selection has also received considerable attention from machine learning researchers and many models have been proposed. John, Kohavi, and Pfleger (1994) classify these proposals in filter and wrapper approaches; Blum and Langley (1997) also include a third category, the embedded approach.

Filter methods assess the relevance of features from data alone, independently of classifiers (Blum & Langley, 1997). Their advantages are lower computation time and their independence of classifiers, but they do not concern possible interactions between the training data and the induction algorithm.

Wrapper methods employ classifier accuracy to judge if a feature subset is superior over another. They evaluate alternative feature subsets by running some induction algorithms on the training data and using the estimated accuracy of the classifier as its metrics. Usually, wrapper methods obtain better results than other methods, but they have a high computational cost and can produce overfitting because they employ the same algorithm for feature selection and classification tasks (John & Kohavi, 1995).

Embedded methods integrate the feature selection process within the basic induction algorithm. They use the specific structure of the model returned by the algorithm to get the set of relevant features. Embedded methods are easily computable, but the existence of irrelevant or correlated features can cause decreases in their accuracy (Blum & Langley, 1997; Kira & Rendell, 1992). Decision trees, analyzed later in the text, include embedded approaches.

At first glance, it will be necessary to employ a segmentation method that performs a robust feature selection process, avoids restrictive statistical hypotheses of data, and allows the efficient characterization of segments of customers. Additionally, this method has to develop decision rules that could be easily understood by managers and needs to use a reduced number of independent variables. The machine learning techniques known as decision trees are very attractive data mining methods for market segmentation, which generally achieve a higher performance than statistical models and have a lower degree of complexity.

Measuring the Success of CRM

Two different groups of metrics could be considered for measuring the success of a CRM program (Winer, 2001, p. 22):

1. Financial and market measures, such as market share, profit margins, etc.
2. Customer-centric measures, such as customer acquisition costs, conversion rates, retention/churn rates, existing customer sales rates, loyalty measures, etc.

Nevertheless, these metrics are partial solutions and should be combined in a final function to obtain a global index on CRM's success. Some proposals such as share of purchase, past customer value, or customer lifetime value, try to be consistent with this customer-centric paradigm of marketing (Kumar & Shah, 2004), but until the moment, there is not a clear formulation of the final success function to be considered.

The allocation of company sales and marketing resources from some clients rather than others should be based on careful analysis of real customer profitability. The company should analyze how current customer relationships differ in profitability and should use this information to identify which customer segments offer the highest potential for future relationships (Van Raaij et al., 2003). It is particularly important when launching a new product because the company needs to balance the cost of the marketing campaign and data warehouse, the profit of positive contacts, and the opportunity of noncontacted potential clients, which could define its competitive position in the medium term. Finally, several potential scenarios should be studied to obtain a global view of CRM's real success.

The design of measures and politics based on the profitability of clients is the main objective of CPA. Nevertheless, management accounting research literature has paid limited attention to customer profitability (Bouwens & Abernethy, 2000; Foster & Gupta, 1994; Guilding & McManus, 2002), and past research on management control systems has mainly considered the use of nonfinancial performance measurement such as quality measures (Davila, 2002; Ittner & Larcker, 1997; Mouritsen, 1997). Thus there is a surprising lack of studies on both CPA and CRM (Malmi et al., 2004). Some authors have analyzed the relationship between strategy and management control but they have focused on business level strategies (Langfield-Smith, 1997). Conversely, there is an absence of works on the operational perspective: how sales and marketing resources should be targeted to valuable clients and its effects on the company's operational results. More recently, some researchers have begun to integrate customer profitability into management control studies but the literature is still very limited (Gerdin & Greve, 2004; Guilding & McManus, 2002; Malmi et al., 2004). Research has focused on the comparison of the overall profitability of companies which develop (or not) CRM politics but no cost-benefit analyses have been performed on CRM tools to be implemented.

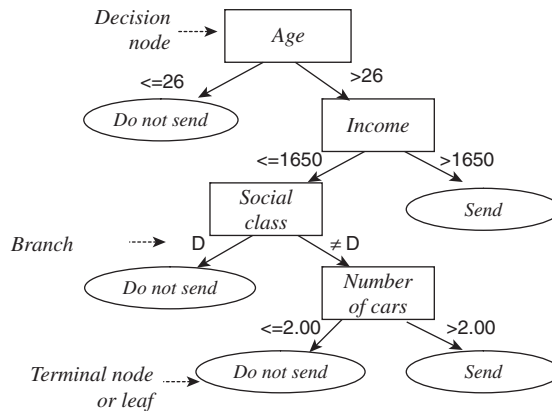
This article considers this latter topic that is, the establishment of a CPA model based on cost-benefit functions to select marketing features, design the best CRM, and optimize marketing resources by launching a new product. In particular, the analysis of the market segmentation problem within the CRM model is considered using both psychographic and socioeconomic attributes through statistical models and decision trees techniques.

Methodological Analysis

Decision Trees: An Overall View

Decision trees are hierarchical and sequential structures of classification that carry out recursive partitions from a group of individuals (data). They represent rules underlying data through three different components: decision nodes, branches, and terminal nodes or leaves.

Figure 1
A Decision Tree for the Mailing Problem



- *Decision nodes:* Each decision node (father) makes use of independent variables to implement a logical test that divide (split-up) data without ambiguity into two or more descendent nodes (children).
- *Branches:* Each branch connects a *father* node with a *children* node, and represent a logical conjunction operator (AND).
- *Terminal nodes or leaves:* Each terminal node or leaf is associated with a certain class which represents the maximum number of individuals in the node. To be viable, a decision tree should contain zero or more decision nodes and one or more terminal nodes.

As a result, a decision tree performs multistage hierarchical decision making (Figure 1).

The process known as *induction of decision trees* (IDT) refers to the construction of these models from given data. Most inductive methods employ the well-known *divide and conquer* proposal: starting from an initial group of training examples (Ω) successive partitions of cases are carried out until all subsets belong to the same class or some stopping rule is executed.

Decision trees' inductive algorithms have quickly evolved from the appearance of the concept learning system (Hunt, 1962; Hunt, Marin, & Stone, 1966) and can be grouped into two broad categories.²

Binary decision trees, which use a single feature (X_i) at each nonterminal decision node and divide data linearly into classes. They use a univariate splitting rule to assign each individual to the left or right branch by its value of that feature. Geometrically, it assigns the point to one side of a hyperplane that is parallel to one axis of the feature space:

$$X_i \leq C, \quad (1)$$

where C is a threshold.

Oblique decision trees, emergent models which use M features at each node to split data ($M > 1$). Each hyper plane is represented by a linear function of the feature components, so

that it is not necessarily parallel to any of the axes (Breiman et al., 1984; Heath, Kasif, & Salzberg, 1993):

$$\sum_{i=1}^M |\beta_i X_i| \leq C, \quad (2)$$

where β_i is the coefficient of the i th feature. Using oblique hyperplanes usually yields a smaller and much more efficient tree; nevertheless, it requires an efficient method to combine features and select hyperplanes,³ which is quite complicated and hinders their empirical application (Murthy, 1998). Also, decision rules from oblique trees tend to be less comprehensive than those from univariate trees.

Many algorithms can be applied for the development of decision trees (ID3, ChAID, ASSISTANT, C4.5, See5, etc.). Among them, the CART (classification and regression trees) model proposed by Breiman et al. (1984) is one of the most used algorithms for building decision trees in practical applications that include medical, economic, or physical problems, among others.

The CART algorithm is based on a two-stage procedure. In the first stage the building of the induction graph is done through a *top-down* process that performs successive partitions of data; in the second phase, a *down-top* pruning process is carried out to simplify the initial model at maximum, which improves both the capacity of generalization and the interpretability of the model.

Regarding the splitting-up criterion used by CART, two different approaches based on Bayesian risk could be considered: the Gini Index and the Twoing criterion. In reality, very small differences can be observed between them, but Breiman et al. (1984) pointed that the Gini Index usually generates slightly more effective partitions than the Twoing criterion. Both splitting criteria can be used to build univariate or oblique decision trees. With reference to the oblique variant, CART uses linear combinations of attributes and employs a heuristic hill climbing and backward feature elimination to find good linear combinations at each node.

Decision trees have been extensively analyzed in the past three decades; their effectiveness has been compared with other automated data exploration methods and to human experts and several advantages have been pointed out (Murthy, 1998): (a) knowledge acquisition from preclassified examples avoids the bottleneck of acquiring knowledge from a domain expert; (b) tree methods are exploratory and nonparametric due to only a few assumptions made about the model and sample, trees can model a wide range of data distributions; (c) hierarchical decomposition implies feature selection and computational efficiency in classification; (d) tree classifiers can treat unimodal as well as multimodal data (as opposed to some statistical models); (e) trees can be used in deterministic as well as incomplete problems.

Decision trees perform classification by a sequence of simple and easy-to-understand tests whose semantics are clear to domain experts. It simplifies their application on real domains where managers need to understand the model which they are using, such as market segmentation and CRM systems.

Cost–Benefit Analysis for CRM Evaluation

Cost–benefit analysis has been traditionally used by managers because, as management literature suggests, once you calculate your costs correctly you make the right decisions

(Cooper & Kaplan, 1988). As a consequence, the better the design of cost functions, the more likely the decisions will lead to a favorable income (Malmi et al., 2004).

Conventionally, the cost–benefit analysis of promotional politics has focused on two main variables which should be specifically adjusted according to promotional politics and to each product:

- The promotional cost per contacted customer, calculated by the unitary cost per contact and the estimated number of contacts per customer, as follows: $C_i = c_i m_i$.
- The final income per each positive contact, calculated by the unitary price per unit and the number of sold units: $R_i = p_i n_i$.

If the analysis focuses on direct marketing politics known as *mailing* and certain nonimmediate consumer goods are considered (only one unit of good is bought simultaneously), previous expressions could be simplified as follows:

- $C_i = c_i$, only one promotional contact is considered.
- $R_i = p_i$, only one unit is sold per positive contact.

Also, a simple cost–benefit function could be defined as follows:

$$P_i = A_{i1}p_i - N_{i1}c_i - A_{i2}(p_i - c_i), \quad (3)$$

$$N = N_{i1} + N_{i2}, \quad (4)$$

where P_i is the profit related to the i th scenario,⁴ N the number of potential clients included in the database, N_{i1} the selected customers who received the mailing, N_{i2} the customers that were not selected to receive the mailing, A_{i1} the number of positive contacts in N_{i1} , and A_{i2} the number of potential positive contacts the company would have obtained if these non-contacted clients would have received the mailing, $A_{i2} \in N_{i2}$.

The term $A_{i2}(p_i - c_i)$ represents a type of mailing opportunity cost and it is a critical issue to be measured by companies. The opportunity cost, economic cost, or alternative cost of using resources in a certain way is the value of what these resources could have produced if they had been used in the best alternative way (Mansfield, 1977, p. 11). Alternatively, it could be defined as “those benefits which could have been received had an alternative course of action been chosen” (Thompson, 1973, p. 263). The relevant opportunity cost can only be determined by considering the specific details of a specific problematical situation, which contributes to the complications associated with the measurement of relevant opportunity costs (Heymann & Bloom, 1990, p. 11).

The assessment of opportunity cost is fundamental to calculate the true cost of any action. Nevertheless, its measurement is quite difficult due to the existence of many alternative courses of actions and the uncertainty related to their future profit. As far as the marketing process is concerned, companies could consider setting up a control group of customers who are permanently excluded from any type of database marketing communications and compare their behavior to the rest of the customer universe; the control group need not be large (around 2% of the dataset to a maximum of 5%) and could provide a measure on the value of the database marketing (Tooker, 2006), and serve to estimate the opportunity cost of the marketing campaign.

Focusing on CRM, nearly 75% of companies admit that potential revenue is lost through missed business opportunities due to inadequate profiling of customers and prospective databases and, on average, the amount of potential revenue lost due to this could be close to 6% and could reach as high as 50% for financial and retail organizations⁵ (QAS, 2005).

Additionally, literature generally does not comment on the costs of developing and maintaining information systems, but they should be integrated in the CPA analysis. That is to say, the previous cost–benefit function should consider the expenses related to data warehouse which is directly related to feature selection: if a segmentation method needs only a few variables (n_i) for each client, then the cost of data warehousing would be cheaper than this for a model with a higher number of attributes. Also, feature selection should lead to the establishment of simple models more easily understood by managers (Tseng & Huang, 2007).

Being d_i the data warehouse cost for each attribute and client, the previous function should include a penalty term as follows:

$$P_i = A_{i1}p_i - A_{i2}(p_i - c_i) - d_i \times n_i \times N, \quad (5)$$

Independent from the selected scenario, the previous function moves between two well-defined limits:

If the mailing is absolutely right ($A_{i1} = N_{i1}$, $A_{i2} = 0$):

$$C_{\text{correct}} = N_{i1}(p_i - c_i) - d_i \times n_i \times N$$

If the mailing is absolutely wrong ($A_{i1} = 0$, $A_{i2} = N_{i2}$):

$$C_{\text{wrong}} = -N_{i1}c_i - N_{i2}(p_i - c_i) - d_i \times n_i \times N = -[N_{i1}c_i + N_{i2}(p_i - c_i) + d_i \times n_i \times N], \quad (6)$$

Also, it would be possible to relate the income obtained for each positive answer (p_i) and the unitary mailing cost (c_i), as follows:

$$t_i = \frac{p_i}{c_i}, \quad (7)$$

where t_i represents the times that price per unit covers mailing contacts; for example, $t_i = 20$ informs that the price per unit allows the finance of 20 mailing contacts, and so on.

Also, it is possible to relate the data warehousing cost (d_i) and the mailing cost (c_i) per client, as follows:

$$u_i = \frac{d_i}{c_i}, \quad (8)$$

Traditionally, w_i is much lower than 1 and it has decreased continuously in preceding decades.

From these relationships, previous limits of the cost–benefit function can be redefined:

If the mailing is absolutely right ($A_{i1} = N_{i1}$, $A_{i2} = 0$):

$$\frac{C_{\text{right}}}{c_i} = N_{i1}(t_i - 1) - w_i \times n_i \times N \quad (9)$$

If the mailing is absolutely wrong ($A_{i1} = 0, A_{i2} = N_{i2}$):

$$\frac{C_{\text{wrong}}}{C_i} = [N_{i1}1 + N_{i2}(t_i - 1) + w_i \times n_i \times N, \quad (10)$$

Finally, the relative profit of a specific scenario could be defined as

$$\frac{P_i}{C_i} = A_{i1}t_i - N_{i1} - A_{i2}(t_i - 1) - w_i \times n_i \times N \quad (11)$$

This last formula is very useful for making comparisons between different scenarios and models, to select the best clusters of potential clients.

Empirical Application

The mailing is a direct-marketing technique that has traditionally been one of the most effective methods to present a new product or service to the market. Nevertheless, many receptors will not be interested in the offer and will not pay attention to the mailing. This negative response increases the company's costs because promotional resources are directed to unprofitable consumers.

To address this situation, the company needs to acquire a better knowledge of their profitable clients, identifying both potential buyers and nonpotential ones, to improve the design of its promotional politics and to avoid wasting time and resources.

In this article, a three-stage methodology has been proposed to manage the mailing politics:

1. Identify the features that better characterize the customers' behavior (observable and nonobservable variables) using embedded algorithms included in decision tree based models (John et al., 1994).
2. Implement a segmentation method that provides both a high performance and a satisfactory degree of interpretability for managers, testing both statistical models (discriminant analysis) and decision trees (univariate and oblique ones).
3. Evaluate the profitability of alternative promotional politics, using a cost-benefit function based on the CPA approach.

This methodology is based on the well-documented and widespread accepted (Johnson, 1971) scheme,⁶ but we have tried to overcome several limitations derived from its assumptions (Dibb & Stern, 1995):

Market definition: The market boundaries are usually preestablished by marketing managers, but could not be meaningless to customers; in this article, we have considered a wide market without including a priori marketers' divisions.

Market stability: The manager assumes that identified segments are stable with respect to both time and competitive actions, but in practice segment membership changes significantly with time; in this article, two different temporal samples are considered such that results assume this potential changes across time.

Attitudinal reliability: Many segmentation researches focus on consumer attitudinal responses rather than behavioral data; in this article, the final segmentation is based on real purchasing behaviors.

The validity of the previous proposal has been evaluated through the analysis of an empirical database which collects information on 5,822 customers from an insurance company with the aim of identifying the variables that better characterize clients who purchased a caravan insurance policy after a mailing process⁷ (Kim et al., 2005; He, Xu, Huang, & Deng, 2004; Huang, Yao, & Zhong, 2003).

This previous segmentation will be used to develop a new promotional mailing in a different geographical location, directed to 4,000 potential new clients, to improve the positive answers rate (only 5.97% for the first 5,822 individuals) and to optimize the final profit of the company.

As a consequence, the first dataset (5,822 customers) will be used as the training set to design the CRM-CPA model and the second group (4,000 customers) will be employed as the validation set to verify the true performance of the proposal.

Inside the database, each customer is characterized by 85 attributes containing sociodemographic data (derived from postal zip area codes, attributes 1-43) and product ownership (attributes 44-86). Data are based on an authentic business problem and includes highly correlated input variables, measured on a nominal (2), ordinal (59), interval (1), and ratio (23) scale.⁸

For the identification of the most informative variables and the establishment of a CRM-CPA model, we have checked two different segmentation methods:

A traditional statistical approach: DA. No PCA has been previously developed because it would require data warehouse of all potential variables and results could not be directly understood.

Univariate and oblique decision trees: The CART model (both univariate and oblique variants), using the *Gini* splitting rule and a pruning technique based on cross-validation. A minimum restriction of 5% examples per leaf has been imposed to reduce overfitting.

The comparative analysis of each method's performance on the validation set, considering different t_i and w_i quotients, provides an overall view of their particular suitability for the solution of this CRM problem. In this way, 15 different scenarios have been defined for each method, through the combination of two groups of parameters:

- Income (price) and mailing cost ratios: $t_i = 5$; $t_i = 10$; $t_i = 15$; $t_i = 20$; $t_i = 50$.
- Data warehouse and mailing cost ratios: $w_i = 0.0001$; $w_i = 0.001$; $w_i = 0.01$.

Table 2 and Figure 2 summarize the performance (profit) and selected features obtained for each method and scenario, together to a risk map, which includes the best method for each situation, considering the previous CPA function:

$$\frac{P_i}{c_i} = A_{i1}t_i - N_{i1} - A_{i2}(t_i - 1) - w_i \times n_i \times N, \quad (12)$$

The accounting performance, or $A_{i1}t_i - N_{i1} - w_i \times n_i \times N$ value, and the estimated opportunity cost, or $A_{i2}(p_i - c_i)$ value, are also included to measure the relative importance of the opportunity cost.⁹

As can be observed, both decision trees (univariate and oblique) achieve better results than DA for all scenarios and parameters (t_i and w_i). Moreover, the use of decision trees

Table 2
Profit for Each Method and Scenario (Validation Set)

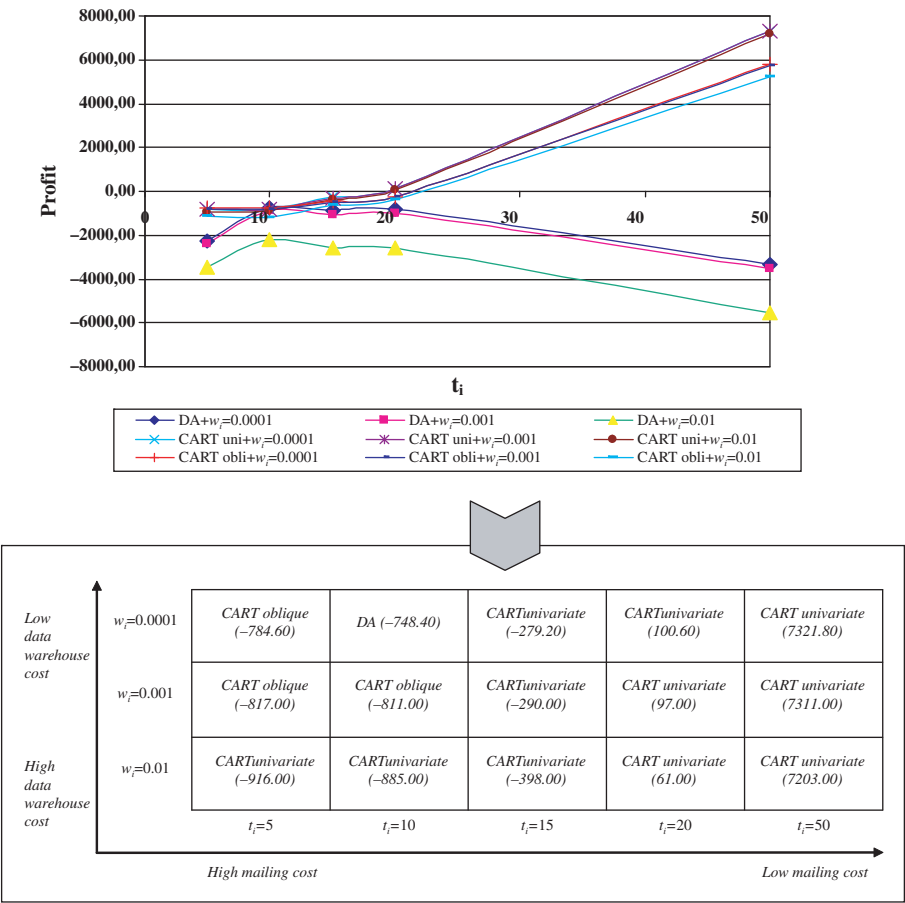
		$t_i = 5$	$t_i = 10$	$t_i = 15$	$t_i = 20$	$t_i = 50$
$w_i = 0.0001$	<i>Discriminant analysis</i>					
	Profit	-2292.00	-748.40	-890.80	-840.60	-3335.40
	Acc. perf.	-1480.00	340.60	901.20	1534.40	3916.60
	Opp. cost	-812.00	-1089.00	-1792.00	-2375.00	-7252.00
	N. features	30	35	42	44	56
	<i>CART univariate</i>					
	Profit	-836.80	-805.80	-279.20	100.60	7321.80
	Acc. perf.	-92.80	238.20	938.80	1582.60	8007.80
	Opp. cost	-744.00	-1044.00	-1218.00	-1482.00	-686.00
	N. features	2	3	3	1	2
	<i>CART oblique</i>					
	Profit	-784.60	-775.00	-456.60	-235.60	5768.40
$w_i = 0.001$	Acc. perf.	-120.60	350.00	887.40	1588.40	7532.40
	Opp. cost	-664.00	-1125.00	-1344.00	-1824.00	-1764.00
	N. features	9	10	4	4	14
	<i>Discriminant analysis</i>					
	Profit	-2400.00	-878.00	-1042.00	-999.00	-3537.00
	Acc. perf.	-1588.00	211.00	750.00	1376.00	3715.00
	Opp. cost	-812.00	-1089.00	-1792.00	-2375.00	-7252.00
	<i>CART univariate</i>					
	Profit	-844.00	-813.00	-290.00	97.00	7311.00
	Acc. perf.	-100.00	231.00	928.00	1579.00	7997.00
	Opp. cost	-744.00	-1044.00	-1218.00	-1482.00	-686.00
	<i>CART oblique</i>					
$w_i = 0.01$	Profit	-817.00	-811.00	-624.00	-394.00	5214.00
	Acc. perf.	-153.00	314.00	720.00	1430.00	6978.00
	Opp. cost	-664.00	-1125.00	-1344.00	-1824.00	-1764.00
	<i>Discriminant analysis</i>					
	Profit	-3480.00	-2174.00	-2554.00	-2583.00	-5553.00
	Acc. perf.	-2688.00	-1085.00	-762.00	-208.00	1699.00
	Opp. cost	-812.00	-1089.00	-1792.00	-2375.00	-7252.00
	<i>CART univariate</i>					
	Profit	-916.00	-885.00	-398.00	61.00	7203.00
	Acc. perf.	-172.00	159.00	820.00	1543.00	7889.00
	Opp. cost	-744.00	-1044.00	-1218.00	-1482.00	-686.00
	<i>CART oblique</i>					
	Profit	-1141.00	-1171.00	-624.00	-394.00	5214.00
	Acc. perf.	-477.00	-46.00	720.00	1430.00	6978.00
	Opp. cost	-664.00	-1125.00	-1344.00	-1824.00	-1764.00

Note: Acc. perf. = accounting performance; Opp. cost = opportunity cost; N. features = number of features; CART = classification and regression trees.

allows positive results to be achieved in some scenarios ($t_i \in \{20, 50\}$), but the DA method always gets negative results, that is to say, the company should not carry out this mailing under any circumstances.

If we were to focus on decision trees, it can be seen that the univariate model obtains better results than the oblique variant for all scenarios, considering both accounting performance

Figure 2
Relative Profit and Risk Map for Each Scenario



and opportunity costs, with the exception of the lowest data warehousing costs combined with the highest mailing costs ($t_i \in \{5,10\}$ and $w_i \in \{0.0001,0.001\}$); nevertheless, even in these cases, differences in profit are very reduced. Also, univariate decision trees are much simpler to understand than oblique trees due to fact that univariate rules only consider one variable per decision node.

Some conclusions could be reached if we focus on each particular method:

- DA improves its performance until some price-cost quotient is reached ($t_i = 10$), but any subsequent increment in this ratio gets worse results. Because of the method's difficulties in identifying positive clients in the database which forces it to maintain an increasing number of attributes for each potential customer, DA needs between 30 and 56 explicative attributes, an excessive number that increases the dimension of the problem, and the difficulties managers would have understanding the final model. Also, opportunity costs are the highest of

all methods. This behavior is not consistent with economic and cost theories which suggest that the company should manage any reduction in costs to achieve a better final result.

- CART univariate decision tree improves its performance continuously for higher price-cost quotients and low data warehousing costs which is consistent with economic and cost theories. Also, a much reduced number of explicative factors are selected, which makes results easy to understand by human managers and decreases data warehousing costs too. CART univariate has used between 1 and 3 variables for all scenarios, an optimal number if the large size of this data set is considered. Finally, opportunity costs increase more slowly if higher price-cost quotients are considered, a maximum for $t_i = 20$ is achieved, but it is significantly reduced for the highest value of the price-cost relationship. As a consequence, this univariate decision tree is particularly robust for the management of massive low-cost mailings.
- CART oblique decision tree also improves its profitability for higher price-cost ratios from a low number of selected attributes (between 9 and 14 features). As a result of the existence of the linear splitting rule, the number of selected attributes increases; thus, the oblique tree achieves the best performance of all methods for low data warehousing and low price-cost quotients; nevertheless, an overfitting process is observed for higher figures, as referenced. Opportunity costs behave in a similar way to those from the CART univariate decision tree but costs are much higher particularly for high price-cost relationships. As a consequence, this oblique decision tree is particularly robust for the management of specialized high-cost mailings.

Figure 3 summarizes decision trees related to $t_i = 10$, together to the discriminant function associated with this scenario (Class 1 = "Send the mailing to this client"; Class 2 = "Do Not send mailing to this client").

It seems clear that the previous CART univariate decision tree provides better comprehensive information than the discriminant function, which employs 35 variables and is nearly incomprehensible to the manager. Besides, the CART univariate decision tree can be converted in five simple decision rules as follows:

"If VAR47 ≤ 5.500 then DO NOT SEND MAILING TO THIS CLIENT"

"If VAR47 > 5.500 AND VAR1 ∈ {1,3,6,7,8,12,14,15,20,37} then SEND MAILING TO THIS CLIENT"

"If VAR47 > 5.500 AND VAR59 ≤ 2.500 AND VAR1 ∉ {1,3,6,7,8,12,14,15,20,37} then SEND MAILING TO THIS CLIENT"

"If VAR47 > 5.500 AND VAR59 ≤ 2.500 AND VAR1 ∈ {5,9,11,16,17,18,19,21,23,25,26,27,28,29,30,31,32,34,35,40,41} then DO NOT SEND MAILING TO THIS CLIENT"

"If VAR47 > 5.500 AND VAR59 > 2.500 then SEND MAILING TO THIS CLIENT"

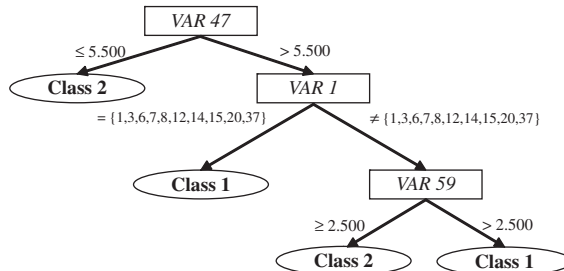
As a result of the fact that variable VAR47 represents the contribution to car policies, variable VAR59 represents the contribution to fire policies (which comes close to the risk appetite of the insurance cover) and variable VAR1 refers to customer subtypes, where the first categories (1, 2, 3, etc.) refer to highest income, urban and modern clients, and the final ones (27, 28, 29, etc.) refer to lowest income, rural and conservative people, it is possible to conclude that the mailing should be targeted to clients who have high premium car policies and a modern lifestyle or a high level of insurance cover.

On the other hand, the DA model does not allow any clear explicative results to be obtained about profitable consumers due to the existence of 35 different attributes; thus, managers would be forced to make ill-defined and uncontrolled decisions.

Figure 3
Examples of Discriminant Analysis and Decision Trees ($t_i = 10$)

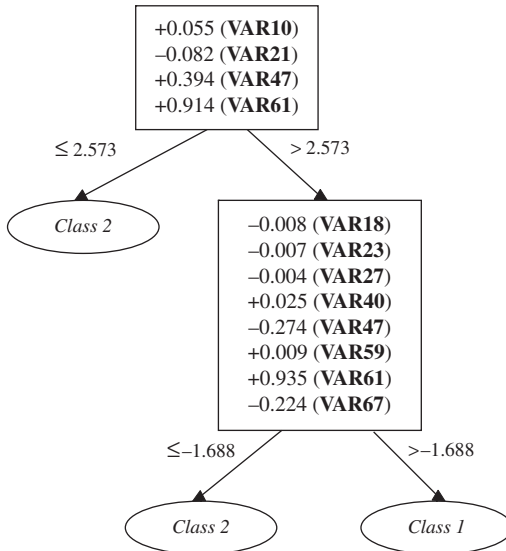
A

CART univariate decision tree



B

CART oblique decision tree



C

Discriminant function (canonical non-standardized coefficients):

$$Y = -0.212\text{VAR}2 + 0.224\text{VAR}4 - 0.073\text{VAR}6 - 0.036\text{VAR}9 + 0.082\text{VAR}10 - 0.039\text{VAR}14 \\ + 0.097\text{VAR}16 - 0.113\text{VAR}18 + 0.062\text{VAR}19 - 0.099\text{VAR}21 + 0.092\text{VAR}22 + 0.044\text{VAR}24 \\ + 0.126\text{VAR}28 - 0.038\text{VAR}30 + 0.060\text{VAR}32 - 0.046\text{VAR}36 - 0.027\text{VAR}37 + 0.070\text{VAR}40 - \\ 0.235\text{VAR}41 + 0.041\text{VAR}43 + 0.468\text{VAR}44 - 0.361\text{VAR}46 + 0.204\text{VAR}47 - 0.051\text{VAR}49 - \\ 0.302\text{VAR}50 - 0.360\text{VAR}53 + 0.957\text{VAR}58 + 0.270\text{VAR}59 - 0.647\text{VAR}65 - 0.402\text{VAR}75 - \\ 3.909\text{VAR}79 - 0.563\text{VAR}80 + 2.007\text{VAR}82 + 0.568\text{VAR}83 + 0.568\text{VAR}85 - 2.579$$

As for the CART oblique decision tree, it generates complex rules because 10 different attributes are combined which limit its real utility for making decisions:

“If $0.055(\text{VAR}10) - 0.082(\text{VAR}21) + 0.394(\text{VAR}47) + 0.914(\text{VAR}61) \leq 2.573$ then DO NOT SEND MAILING TO THIS CLIENT”

“If $0.055(\text{VAR}10) - 0.082(\text{VAR}21) + 0.394(\text{VAR}47) + 0.914(\text{VAR}61) > 2.573$ AND $-0.008(\text{VAR}18) - 0.007(\text{VAR}23) - 0.004(\text{VAR}27) + 0.025(\text{VAR}40) - 0.274(\text{VAR}47) + 0.009(\text{VAR}59) + 0.935(\text{VAR}61) - 0.224(\text{VAR}67) \leq -1.688$ then DO NOT SEND MAILING TO THIS CLIENT”

“If $0.055(\text{VAR}10) - 0.082(\text{VAR}21) + 0.394(\text{VAR}47) + 0.914(\text{VAR}61) > 2.573$ AND $-0.008(\text{VAR}18) - 0.007(\text{VAR}23) - 0.004(\text{VAR}27) + 0.025(\text{VAR}40) - 0.274(\text{VAR}47) + 0.009(\text{VAR}59) + 0.935(\text{VAR}61) - 0.224(\text{VAR}67) > -1.688$ then SEND MAILING TO THIS CLIENT”

According to the previous rules some partial conclusions could be obtained: VAR10, VAR40, VAR59, and VAR61 are positively related to customers who buy insurance policies. These features refer to people who are married with high incomes and high premium fire policies and/or high premium boat policies respectively; in particular, VAR61 greatly affects the final buying decision.

On the contrary, VAR18, VAR21, VAR23, VAR27, and VAR67 are negatively related to customers who buy insurance policies and refer to those of a lower level education: farmer, skilled laborers, B2 social class, and a high number of third party insurance (agriculture) people.

Although both conclusions are coherent with previous CART univariate rules, the oblique tree also includes some contradictions because VAR 47 (high-premium car policies) affects both positively (first node) and negatively (second node) the buying decision. Nevertheless, it was previously identified as one of the most relevant and positively-related variables for caravan insurance buyers.

As a consequence, the CART univariate decision tree seems to be the most robust and understood model of all the alternatives. It also achieves the best results for the majority of analyzed scenarios (12 of 15), generates low data warehousing costs because very few attributes are needed to characterize clients and has low opportunity costs.

Conclusions

The importance of CRM has increased in previous years due to a competitive environment forcing the company to implement a dynamic management of clients to obtain higher profits and to acquire a higher market share than its competitors. One of the most important CRM decisions for retailers is the design of efficient direct marketing politics, especially, if a new product or service is launched into the market.

The development of a CRM program needs to fulfill several successive stages, customer segmentation and the measurement of the success of the CRM program being two of the most critical steps. Market segmentation has traditionally developed through statistical techniques such as PCA, DA, or logistical regression. These models consider restrictive a priori hypotheses that are not fulfilled in presence of high volumes of data and multiple-correlated

features, thus feature selection methods and alternative clusterization techniques must be considered to obtain robust segmentation. In this article, we proposed to combine statistical and machine learning techniques for market segmentation; in particular, the performance of decision trees (both univariate and oblique variants) is analyzed.

Although several metrics have been developed to measure the success of the CRM program, they are partial solutions that must be integrated in an overall measure. The overall measurement of client profitability is CPA's main aim. Considering CPA concepts, we have proposed a 3-stage methodology for CRM management that includes customer-preferred feature selection, several segmentation techniques (DA, univariate decision trees, and emergent oblique decision trees) and a global cost-benefit function for measuring the success of the program. This new function integrates marketing costs (derived from promotional contacts), data warehousing costs (from the management of databases within the company), potential profit (derived from positive responses from contacted customers), and opportunity costs (caused by potential positive responses from noncontacted customers).

Our proposal is applied to the analysis of a well-known dataset related to the mailing of a new caravan insurance policy. Several scenarios are considered related to different data warehousing costs, marketing costs, and prices. Results show that machine learning segmentation techniques achieve a higher performance than statistical methods for the majority of scenarios taking into account final profit, accounting performance, and opportunity costs. In particular, the CART univariate decision trees allow an efficient segmentation that directs low-cost mailings to a high percentage of positive-response buyers, maximizes profit, and minimizes opportunity costs. It also generates logical decision rules easily understood by managers and facilitates the tracking and nurturing of profitable clients. Finally, the CART oblique decision trees outperform other alternatives for specialized high-cost mailings.

Notes

1. A total of 800 respondents for the research were drawn equally from eight regions (Asia Pacific, Benelux, France, Germany, Scandinavia, Spain, the United Kingdom, and the United States); industry sectors represented by the samples include manufacturing, public sector organizations, retail, wholesale and distribution, financial services, business services, hotels and catering, professional services, utilities, and telecoms.

2. A graphical representation of decision regions from univariate and oblique decision trees is available at <http://www.upo.es/dde/personal/jmramjer>

3. Methods used in the literature for finding good linear tests include linear discriminant analysis, hill climbing search, linear programming, perceptron training, and others.

4. Each scenario makes reference to a specific segmentation method and market.

5. Research was carried out on 550 respondents in public and private organizations with more than 200 employees across 10 countries around the world (Australia, Belgium, France, Germany, Luxembourg, the Netherlands, Singapore, Spain, the United Kingdom, and the United States). Sectors represented by the sample were banking, insurance and finance; retail; telecoms; utilities; leisure, tourism and travel; and the public sector.

6. It includes four steps: (a) choose a basis for segmentation and select appropriate variables; (b) Use multivariate analysis to split up the market; (c) evaluate and validate the results; (d) analyze the results in economic terms.

7. This database was proposed in the CoIL Challenge (2000), available at <http://www.liacs.nl/~putten/library/cc2000/>

8. A detailed description of attributes is available at <http://www.upo.es/dde/personal/jmramjer>

9. The contingency table for each method and scenario, which includes positive and negative answers with independence of the w_i parameter, is available at <http://www.upo.es/dde/personal/jmramjer>

References

- Anderberg, M. R. (1973). *Cluster analysis for applications*. New York: Academic Press.
- Bass, F. M., Tigert, D. G., & Lonsdale, R. T. (1968). Market segmentation: Group versus individual behaviour. *Journal of Marketing Research*, 5, 264-270.
- Berger, P. D., & Nasr, N. (1998). Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing*, 12, 17-30.
- Blum, A., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97, 245-271.
- Bouwens, J., & Abernethy, M.A. (2000). The consequences of customisation on management accounting system design. *Accounting, Organizations and Society*, 25, 221-241.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. New York: Chapman & Hall.
- Brijs, T. (2002). *Retail market basket analysis: A quantitative modelling approach*. Unpublished doctoral dissertation, Department of Applied Economics, Limburg University, the Netherlands.
- Cooper, R., & Kaplan, R. S. (1988). Measure cost right: Make the right decision. *Harvard Business Review*, September-October, 96-103.
- Davila, T. (2002). An empirical study on drivers of management control systems' design in new product development. *Accounting, Organization and Society*, 25, 383-409.
- Dibb, S., & Stern, P. (1995). Questioning the reliability of market segmentation techniques. *Omega*, 23, 625-636.
- Doyle, P., & Saunders, J. (1985). Market segmentation and positioning in specialized industrial markets. *Journal of Marketing*, 49, 24-32.
- Foster, G., & Gupta, M. (1994). Marketing, cost management and management accounting. *Journal of Management Accounting Research*, 6, 43-77.
- Gath, I., & Geva, A. B. (1988). Unsupervised optimal fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 773-781.
- Gerdin, J., & Greve, J. (2004). Forms of contingency fit in management accounting research: A critical review. *Accounting, Organization and Society*, 29, 303-326.
- Green, P. E., & Wind, Y. (1975). New ways to measure consumers' judgements. *Harvard Business Review*, 53, 107-117.
- Green, P. E. (1977). A new approach to market segmentation. *Business Horizons*, 20, 61-73.
- Guilting, C., & McManus, L. (2002). The incidence, perceived merit and antecedents of customer accounting: An exploratory note. *Accounting, Organization and Society*, 27, 45-59.
- He, Z., Xu, X., Huang, J. Z., & Deng, S. (2004). Mining class outliers: Concepts, algorithms and applications in CRM. *Expert Systems with Applications*, 27, 681-697.
- Heath, D., Kasif, S., & Salzberg, S. (1993, August-September). Learning oblique decision trees. In R. Bajcsy (Ed.), *Proceedings of the 13th International. Joint Conference on Artificial Intelligence* (pp. 1002-1007). San Francisco: Morgan Kaufmann.
- Heymann, H. G., & Bloom, R. (1990). *Opportunity costs in finance and accounting*. Westport, CT: Quorum Books.
- Hruschka, H., & Natter, M. (1999). Comparing performance of feedforward neural nets and K-means for market segmentation. *European Journal of Operational Research*, 114, 346-353.
- Huang, J., Yao Y. Y., & Zhong, N. (2003, October). Attribute reduction of rough sets in mining market value functions. *Proceedings of the IEEE/WIC International Conference on Web Intelligence* (pp. 470-473). Washington, DC: IEEE Computer Society.
- Hunt, E. B. (1962). *Concept learning: An information processing problem*. New York: Wiley.
- Hunt, E. B., Marin, J., & Stone, P. J. (1966). *Experiments in induction*. New York: Academic Press.

- Ittner, C., & Larcker, D. (1997). Quality strategy, strategic control systems, and organizational performance. *Accounting, Organization and Society*, 22, 293-314.
- John, G. H., & Kohavi, R. (1995, August). Feature subset selection using the wrapper model: Overfitting and dynamic search space topology. *Proceedings of the First International Conference on Knowledge Discovery and Data Mining* (pp. 192-197). Menlo Park, CA: AAAI Press.
- John, G. H., Kohavi, R., & Pfleger, K. (1994, July). Irrelevant features and the subset selection problem. In W. W. Cohen & H. Hirsh (Eds.), *Proceedings of the Eleventh International Conference on Machine Learning* (pp. 121-129). San Francisco: Morgan Kaufmann.
- Johnson, R. M. (1971). Market segmentation: A strategic management tool. *Journal of Marketing Research*, 8, 13-18.
- Kim, Y., Street, W. N., & Menczer, F. (2001). An evolutionary multi-objective local selection algorithm for customer targeting. *Proceeding Congress on Evolutionary Computation*, 2, 759-766.
- Kim, Y., Street, W. N., Russell, G. J., & Menczer, F. (2005). Customer targeting: A neural network approach guided by genetic algorithm. *Management Science*, 51, 264-276.
- Kira, K., & Rendell, K. A. (1992). The feature selection problem: Traditional methods and a new algorithm. *Proceedings of the Ninth International Conference on Artificial Intelligence* (pp. 129-134), Cambridge: MIT Press..
- Kumar, V., & Shah, D. (2004). Building and sustaining profitable customer loyalty for the 21st century. *Journal of Retailing*, 80, 317-330.
- Langfield-Smith, K. (1997). Management control systems and strategy: A critical review. *Accounting, Organization and Society*, 22, 207-232.
- Malmi, T., Raulas, M., Gudergan, S., & Sehm, J. (2004). *An empirical study on customer profitability accounting, customer orientation and business unit performance* (Draft paper). Helsinki School of Economics & Business Administration.
- Mansfield, E. (1977). *Economics, principles, problems, decisions*. New York: Norton.
- Mouritsen, J. (1997). Marginalizing the customer: Customer orientation, quality and accounting performance. *Scandinavian Journal of Management*, 13, 20-35.
- Murthy, S. K. (1998). Automatic construction of decision trees from data: A multidisciplinary survey. *Data Mining and Knowledge Discovery*, 2, 345-389.
- QAS. (2005, September). *The impact poor data management has on your organisation*. Dynamics Markets Commissioned by QAS.
- QAS. (2006, September). *The hidden costs of poor data management* (International Research White Paper). Dynamic Markets Commissioned by QAS.
- Rao, V. R., & Steckel, J. H. (1995). Selecting, evaluating and updating prospects in direct mail marketing. *Journal of Direct Marketing*, 9, 20-31.
- Reinartz, W., & Kumar, V. (2000). On the profitability of long-time customers in a non-contractual setting: An empirical application and implications for marketing. *Journal of Marketing*, 64, 17-35.
- Schmittlein, D. C., & Petersen, R. A. (1994). Customer base analysis: An industrial purchase process application. *Marketing Science*, 13, 41-67.
- Smith, W. (1956). Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing*, 21, 3-8.
- Tabachnick, B. G., & Fidell, L. S. (1983). *Using multivariate statistics*. New York: Harper & Row.
- Thompson, A. A. Jr. (1973). *Economics of the firm: Theory and practice*. Englewood Cliffs: Prentice Hall.
- Tooker, R. N. (2006). *Bringing it all together. Understanding the database marketing process* (White Paper). KnowledgeBase Marketing Inc.
- Tseng, T. L., & Huang, C. C. (2007). Rough set-based approach to feature selection in customer relationship management. *Omega*, 35, 365-383.
- Van Raaij, V., Vernooij, M. J. A., & Van Triest, S. (2003). The implementation of customer profitability analysis: A case study. *Industrial Marketing Management*, 32, 573-583.
- Wedel, M., & Kamakura, W. A. (2000). *Market segmentation: Conceptual and methodological foundations*. Boston: Kluwer Academic.
- Wind, Y. (1978). Issues and advances in segmentation research. *Journal of Marketing Research*, 15, 317-337.
- Winer, R. S. (2001). Customer relationship management: A framework, research directions, and the future (Draft Paper). University of California, Berkeley.

Raquel Florez-Lopez (PhD Hon, MSc Hon) is an assistant professor of accounting and business at the University of Leon, Spain. She is the author of articles on international journals focused on applying information systems and machine learning models to business management and risk. She has been a participant in European research projects and has contributed to more than 30 international conferences.

Juan Manuel Ramon-Jeronimo (PhD Hon, MSc) is a lecturer in accounting and business at the University of Pablo Olavide, Seville, Spain. His research area focuses on management accounting, interfirm firm collaboration, and buyer–supplier relationships. He has participated on European research projects, contributions to several international journals and conferences.