We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Dual Approach to the Modelling Single Product Demand Curves in the Next Best Offer CRM Problem

Džulijana Popović
Zagrebačka Banka d.d., UniCredit Group
Croatia

1. Introduction

Customer Relationship Management (CRM) is a comprehensive approach for creating, maintaining and expanding customer relationships (Anderson & Kerr, 2002). The definition tells that CRM has to be a comprehensive way of doing business that touches all areas and does not belong to one department of an organization, such IT, Marketing or Sales. CRM is also a way of thinking and dealing with different aspects of relationship with customers and involves having a clear strategy and plan. In modern economy that relationship between a company and customer has various forms, from personal to completely virtual contacts with frequent customers who have never spoken to anyone from the company. Therefore, CRM also takes care about **creating the idea and strategy of high touch in a high tech environment. Technology must not stay on the way of applying CRM strategy and must not be confused with it.**

There are a number of important CRM questions and problems and the most common are:

- customer acquisition (problem of attracting new customers);
- customer retention (problem of preventing attrition of existing customers, especially the most profitable ones);
- cross-sell and up-sell (problem of making existing customers more profitable);
- risk approval and underwriting tools (problem of avoiding high-risk customers);
- lifetime value modelling (problem of recognizing new profitable customers);
- improved targeting and campaign response rates (problem of expense reduction and increasing sales and revenues);
- customer segmentation (problem of profiling customer database), etc.

Customers want to do business with organizations that know them, that understand what they want and need, and that continue to fill those needs (Anderson & Kerr, 2002) in changing environment. Modern companies collect various data about their customers, but these data rarely answer the direct questions such as "what will each of our customers need in the next future?" or "which are the product attributes that satisfy his/her needs in the way that results in product purchase?". The way to answer such questions is deep diving into available data, with strong understanding what the analyst is searching for, which

conclusions are valid syllogisms, and how to translate business goals mentioned above into analytical problems. The learning process about customers has to be an adaptive, flexible and feedback supplied process, in order to build dynamic knowledge about customers. In order to have the right proposal for each customer in each contact occasion, a company needs to solve the next best offer problem as well. The proposal has to be in form and content which meets both customer's and company's goals. While the company's business goals are well defined within the organization, the source of uncertainty in the next best offer problem are unknown product demand curves at the individual customer level, not only in their snapshot form, but rather in their dynamical form. Finally, the companies have to remember that data mining and statistical models only suggest, but business decides and supports the solution implementation. There is no much gain of having high tech tools and excellent predictive models, if the business environment and decision makers do not support and ensure their application.

2. Methodology

The chapter presents dual approach to the problem of offering the single product to the right customer at the right time and through the right channel, instead of trying to model the next best offer generally. The duality of the approach lies in the combination of product and customer perspective, as well as in the combination of classical statistical methods and modern methodologies. The model for the probabilities of buying the single product was built as the logistic regression model and interpreted as the single product demand curve. The model for customer clustering according to their channel usage and communication history was built with fuzzy c-means clustering algorithm and interpreted as the natural grouping of the customers that reveals their propensity to buy the product when reached through preferred channel. From the expert systems and knowledge discovery perspective the models are divided into two groups: **predictive models and classification models.** Regression model is a predictive model which gives the predicted values of the dependent (target) variable as output. Clustering is a classification model, since it gives the group flag or group membership function value to each observation. From the pattern recognition perspective **regression model is supervised learning technique**, since it uses the dependent variable for solution generation. **Clustering is unsupervised learning technique**, since it does not use the dependent variable for cluster detection and definition. But, discovered clusters can help in classifying new observations and in that way clustering becomes also predictive model (Cox, 2005). The flexibility of the final model is achieved in the second iteration, when the membership functions' values from the fuzzy c-means clustering algorithm enter the logistic regression model as additional input variables, improving the predictions of individual probabilities of product purchase. In that way all the information on the individual customer level that is comprised in the clusters' definition also enters the first model. The goal of the fuzzy clustering process is not to obtain another kind of general customer segmentation, but focuses on discovering the clusters according to their channel usage and communication history. In that way the model for the purchase probabilities is influenced by the membership functions that contain the information about communication channel that leads to the purchase. Since the clustering is not crisp, but fuzzy, it provides additional information on the extent to which the customer belongs to each of the channel fuzzy clusters. That information can be monitored over time, helping the company to recognize the changes in customer channel preferences before customer actually fall in

another cluster. The combination of the two underlying models can lead to more accurate predictions of sales (product perspective), to more responsive target of customers (marketing perspective), more predictive costs and revenues (budget perspective) and/or to the greater level in satisfying the customers' needs and loyalty (customer perspective). Generally, the dual approach can help the company to develop operationally possible and sustainable models to meet customer needs on time, to dynamically learn about them not only from their purchasing history, but from their feedback to every CRM intervention and still to reach financial and business goals.

2.1 Logistic regression analysis

The logistic regression model is a mathematical model used for the categorical data analysis and belongs to the class of generalized linear models (GLM). The class of generalized linear models extends the theory and methods of linear models to data with nonnormal responses. Differently from other extensions of linear modelling to nonnormal data, which all relied on transformations of the data, generalized linear models apply a transformation to the mean of the data. In that way the systematic component is a linear predictor $\eta = X'\beta$, where η, in contrast to linear models, does not represent the mean function of the data. The transformation called link function $g(\cdot)$ relates that linear predictor to the mean, $g(\mu) = \eta$. The link function is a monotonic and invertible function, so the mean can be expressed as the inversely linked linear predictor $\mu = g^{-1}(\eta)$. Generalized linear models allow the data to come from a distribution that is a member of the exponential family of distributions and link function provides a mapping between the linear predictor and the mean of the data.

Suppose x is a vector of independent or explanatory variables and $\pi = p(Y=1|x)$ is the response probability to be modelled. The linear model is not an appropriate model since the predicted values from a linear model can assume any value and probabilities are by definition bounded between 0 and 1. The other shortcoming of linear probability model is that the relationship between the probability of an outcome and independent variables is usually nonlinear. The one-unit change in an independent variable may have less impact when the probability is near 0 or 1 than when the probability is about 0,5. (Hosmer & Lemeshow, 2000). The appropriate model is a logistic model, which is based on the logistic function

$$f(z) = \frac{1}{1 + e^{-z}} \tag{1}$$

Equations (2) and (3) show that logistic function has the range of values between 0 and 1, regardless of the value of z (Kleinbaum & Klein, 2010):

$$f(-\infty) = \frac{1}{1 + e^{-(-\infty)}} = \frac{1}{1 + e^{\infty}} = 0 \tag{2}$$

$$f(+\infty) = \frac{1}{1 + e^{-(+\infty)}} = \frac{1}{1 + e^{-\infty}} = 1 \tag{3}$$

To obtain the logistic model from the logistic function, z is substituted by the expression

$$z = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} \tag{4}$$

and the equation that directly refers to the probability of the outcome is

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})}} \tag{5}$$

Equation (5) is **logistic regression model** and is equivalent to the **logit link transformation** expressed as natural log of the odds, which is the ratio of the probability of the outcome to the probability of no outcome:

$$\log it(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} \tag{6}$$

where β_0 is the intercept parameter and $\beta_1, ..., \beta_k$ are the slope parameters.

As mentioned earlier, logistic model is used for **categorical data analysis**, which is concerned with categorical responses, regardless of whether the predictor variables are categorical or continuous. If the set of response consists of only two values, then the response is dichotomous or binary and the appropriate model is **binary logistic model**. Therefore, the binary logistic regression model characterizes the relationship of several independent variables to a dichotomous dependent variable (Kleinbaum & Klein, 2010). It uses the categorical and/or continuous input variables to predict the probability of specific outcomes. Since the dependent variable in logistic regression is not continuous, but discrete or categorical, that makes logistic regression very useful in predicting discrete customer's actions such as a response to a buying offer (Parr Rud, 2001). The logistic regression model provides the estimates that lie in the range between zero and one and an appealing S-shape description of the combined effect of several independent variables (factors) on the probability of the dependent outcome (Kleinbaum & Klein, 2010). The methods employed in logistic regression analysis follow the same general principles used in linear regression analysis (Hosmer & Lemeshow, 2000) and are out of the scope of this chapter.

2.2 Fuzzy c-means clustering algorithm

Cluster analysis is a family of mathematical techniques that find groups of observations with similar characteristics (Parr Rud, 2001). The goal of clustering is to divide (in crisp clustering to partition) the input dataset into groups called clusters, according to observed characteristics, in the way that the observations within the same cluster are as much similar to each other as possible, while as much dissimilar to the observations in other clusters.

Classical clustering, also known as hard or crisp clustering assigns each observation to a single cluster, without information how far or near the observation is from all the other possible decisions. **Fuzzy clustering, based on the concept of fuzzy membership functions and the fuzzy set theory, allows entities to belong to many clusters with different degrees of membership** (Theodoridis & Koutroumbas, 2006). The basic steps in a cluster analysis are feature selection, proximity measure choice, clustering criterion, clustering algorithm, validation and interpretation of the clustering results. Different choices in these steps lead to completely different clustering results and the final answer to these questions is often

influenced by the expert knowledge. As stated in (Theodoridis & Koutroumbas, 2006), cluster analysis can be used for both **hypothesis testing and prediction based on groups**, which are the ideas used in this chapter. In context of the next best offer problem, one of the hypothesis is that the company can predict the individual customer probability of buying the product and the other one is that the company can influence the probability by choosing the most appropriate communication channel. Therefore, the cluster analysis should support the assumption that customers can be successfully clustered according to their purchase behaviour and channel usage, regardless other factors that influence their product demand curve. The clustering algorithm chosen for that purpose is **fuzzy c-means clustering (FCM)**, which is the most known method of fuzzy clustering. Generally, fuzzy clustering of X into p clusters is characterized by p membership functions μ_j, where

$$\mu_j : X \to [0,1], j = 1,\ldots,p \tag{7}$$

$$\sum_{j=1}^{p} \mu_j(x_i) = 1, \quad i = 1,2,\ldots,n \tag{8}$$

and

$$0 < \sum_{i=1}^{n} \mu_j(x_i) < n, \quad j = 1,2,\ldots,p \tag{9}$$

Membership functions are based on a distance function, such that membership degrees express proximities of entities to cluster centres, called **cluster prototypes**. Fuzzy c-means algorithm, initially proposed by Dunn and generalized by Bezdek (Cox, 2005), involves two iterative processes: the calculation of cluster centres and the assignment of the observations to these centres using some form of distance. Fuzzy c-means is attempting to minimize a standard loss function

$$l = \sum_{k=1}^{p} \sum_{i=1}^{n} \left[\mu_k(x_i) \right]^m \left\| x_i - c_k \right\|^2 \tag{10}$$

from which two fundamental equations necessary to implement fuzzy c-means are derived.

A new cluster centre value is calculated using the expression (11):

$$c_j = \frac{\sum_i \left(\left[\mu_j(x_i) \right]^m x_i \right)}{\sum_i \left[\mu_j(x_i) \right]^m} \tag{11}$$

and expression (12) is used to calculate the membership in the j-th cluster:

$$\mu_j(x_i) = \frac{\left(\frac{1}{d_{ji}} \right)^{\frac{1}{m-1}}}{\sum_{k=1}^{p} \left(\frac{1}{d_{ki}} \right)^{\frac{1}{m-1}}} \tag{12}$$

The symbols in the equations (10), (11) and (12) denote:

l   is the minimized loss value;
p   is the number of fuzzy clusters;
n   is the number of observations in the data set;
μ_k ( )   is a function that returns the membership of x_i in the k-th cluster;
m   is the fuzzification parameter;
c_k is the centre of the k-th cluster;
d_{ji} is the distance metric for x_i in cluster c_j;
d_{ki} is the distance metric for x_i in cluster c_k.

3. Next best offer problem in the CRM

In CRM literature and practice there are two well-known problems, oriented primarily to the sales and revenues increase: **up-sell and cross-sell.** Cross-sell is the event of a current customer buying a different product or service from the same company. Up-sell is the event of a current customer buying more of the same products and services of the company. Both events aim to reach product budget goals and increase the profitability of product lines. In the customer-centric strategy, the optimization of the value of each customer relationship and increase of the customer profitability comes to the first priority. In that situation, companies have to manage their offers more carefully to avoid over-soliciting their existing customers, while at the same time optimizing their product lines goals. Instead of answering the question: "to whom we can offer and sell product X in order to reach the budget?", the company tries to answer more complex and multi-criteria question: "which of my products will the customer need in the next future and how we have to offer it to him/her in order to reach the financial goals?". **Therefore, the next best offer problem in CRM is the decision problem of what to offer, when to offer and how to offer to each customer, in order to improve customer retention and loyalty, increase up-sell and cross-sell, reach customer profitability and product lines profitability goals.** To cross the way from the next best offer problem to the next best action, companies develop models and integrate expert knowledge with technology. Solving the general problem of the next best offer on the individual level is very challenging, because the complete CRM process does not finish with recognition of the product sequence for each customer. The complete solution should lead to the desired event of product purchase and revenue increase. It means that the company has also to decide which channel, what kind of message, which price and what timing to apply for each customer recognized by the model as the customer with high probability of buying the product. Failure to correctly define the goal can result in wasted money and lost opportunity (Parr Rud, 2001). These are the reasons why the company should know well the demand for each of its major products, as well as its dominant customer profile.

3.1 Purpose of the model

The main purpose of elaborated hybrid model is recognition of the customers who have high probability of buying the target product if contacted through the preferred channel. The first part is solved by modelling the single product purchase probabilities as the logistic regression curves, getting the independent customers' probabilities for each of the products. The second part is done by customers' clustering into the channels fuzzy clusters, according to their channels' usage data and communication history. The goals of these two models

Dual Approach to the Modelling Single
Product Demand Curves in the Next Best Offer CRM Problem
107

differ substantially, since the goal of the first model is predictive and the goal of the second is descriptive. The goal of the complete dual approach model is predictive.

It is important to note that the purpose of the statistical models in CRM is different from the purpose of CRM campaigns. The goal of a CRM campaign is to change behaviour and reaching an individual who is going to purchase the product anyway is only slightly more effective than reaching an individual who will not purchase despite having received the offer. Building the model which will recognize **likely responders** to the offer differs from the model which recognizes **likely buyers** of the product anyway. Therefore, the goal of CRM campaign is reaching individuals who are more likely to purchase **because of having been contacted through the right channel.** This kind of analysis is known as **differential response analysis** and compares the results from a treated group with results from a control group. The object of differential response analysis is to find segments with the greatest difference in response between the treated and untreated groups. Dual approach proposed here uses some ideas from differential response analysis, such as the idea of recognizing the customers with higher propensity to purchase if approached through their preferred channel.

3.2 General structure of the modelling database

After having defined the goal, the next step in model building is wide data selection and further data preparation. The questions in the process of data selection include, but are not limited to these: what is available in company's data warehouse and data marts, how much data will be enough and how much history is required, what these data must contain for the specific goal and/or problem, how many variables to include into initial database. Patterns in customer behaviour need time to become evident. Data warehouses need to support accurate historical data so that data mining process can pick up these critical trends (Berry & Linoff, 2004). After the initial data selection, the next step is getting to know the data, by examining distributions, validating initial assumptions, asking a lot of questions and ensuring very good preliminary data analysis before even starting to model. **The steps surrounding the model processing can be more critical to the overall success of the project than the technique used to build the model.** The organization of the input data here proposed is shown in the Figure 1.

The first modelling database is constructed by joining together the data from the four separate databases: customer data, product data, customer behaviour about the products data and channel usage data. Each of the input data tables contains the time variable and appropriate primary and foreign key used for joining. **Customer data** table contains the variables on the customer level and some of them are demographic, socio, economic and financial data. **Product data** table contains the variables on the each product level, for example the position of the product in the general product scheme of the company, product indicators for the time period such as product price, indicator whether the product was on promotion in that time period, total volume sold in the period, percentages of the volume sold in the regular sale and in the promotion campaigns, total revenues in the time period, etc. **Customer behaviour about the products** gives customer purchase history for each time period and each product. It contains the indicators whether the purchase of the concrete individual product happened, but also whether the purchase of any product from that product category happened, according to the company's product scheme. Finally, **channel**

usage data is a database that contains channel usage flags for each customer and each time period, number of contacts during the period by the channel, ideally number of purchase events resulting from the contact, total amount of consumption in the time period, etc.
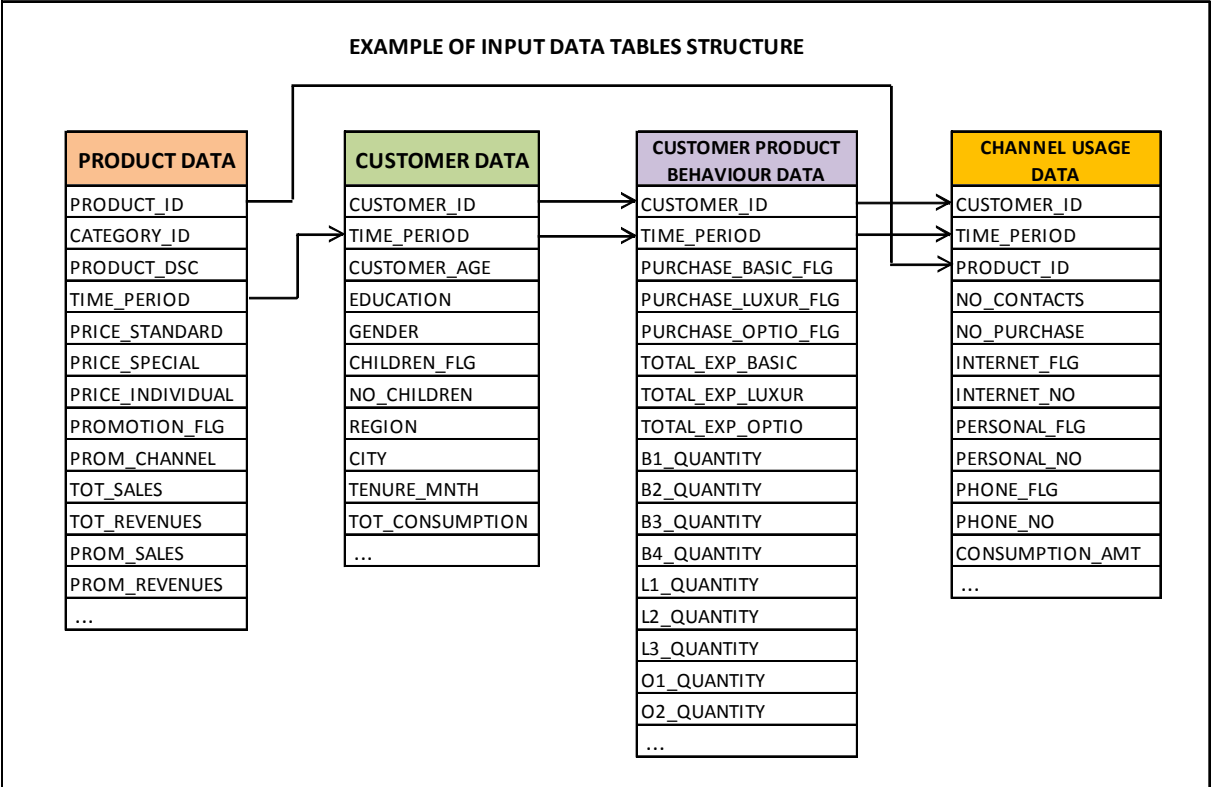

Fig. 1. Example of input data tables structure and data organization

Depending on the modelled event, these separate databases are joined together and outcome (target or dependent) variable is derived according to the event definition. In this study, the first modelled event is purchase of the chosen single product from the product scheme in the period T and comes in two versions. In the first version it is allowed to use all other data from the period T-1 as the independent input variables, apart from the data about the chosen product. This approach is testing the hypothesis that the probability for the purchase of the chosen product in period T can be derived from the customer's purchase history about other products. In the second version of the same event the history data about the chosen product are also allowed to enter the model as the predictors. This makes sense if the repetitive customer's need is being modelled and that is industry-dependent possibility. The second modelled event is purchase of the chosen product from the company's product scheme if customer was contacted in the same period through his/her preferred channel. The information about the preferred channel enters the model either in the form of original variables from channels' usage history or in the form of the fuzzy membership functions obtained from fuzzy clustering of the customer behaviour about channels and about products in the period T-1. One of our hypotheses is that the fuzzy membership functions can be better and more stable choice. The event is defined as the combination of the purchase and the decision of the company to contact the customer through the certain channel. The transformations that add target variable into the database are provided under the section 4. From the definition of the events it is obvious

that the modelling database must contain the data from at least three successive time periods. More available history leads to more accurate predictions. After joining input data tables into one modelling database, the process of initial data selection is finished and the process of data quality checks and data preparation process can start.

3.3 Data quality and data preparation

Data quality generally refers to the degree of excellence with which the data in the modelling database correctly represent the situation in the core company's information systems for each time period and satisfy relational integrities. The data are of good quality if they are complete, consistent, accurate, time-stamped and industry standard-based. If, in addition to that, the data satisfy special business need, the data are of high quality. For any model development the quality of the input data is crucial. The monitoring of data quality should become the integral part of the modelling processes and can include even development of data quality protocols. The data quality issues differ from the data adequacy for modelling defined event. Even when the initial data selection ensures appropriate variables for modelling target event and when the data are of high quality, further data preparation is needed. The process of data preparation for modelling consists of data cleaning, missing value treatment and imputation, outlier removal and preliminary variable selection. The realization of this process is problem-based and requires the decisions with regard to the modelling purpose for which the data are being prepared.

3.3.1 Data cleaning

Data cleaning, also known as data scrubbing, is a multiple steps process of maintaining the database clean, accurate, consistent and free of errors, mismatched or invalid information. There are various techniques of data cleaning, but generally all of them include following steps: error identification, error reporting, new data values check and data merging. The first step is identification of the errors and their classification into critical and non-critical errors. **The critical errors must be fixed**, not only because the dirty data can not be used for modelling purposes, but also because the critical errors can not remain in the database due to the regulatory reporting or other industry-dependent reasons. Non-critical errors do not cause operational problems and can be fixed periodically. After the errors were identified, whether using an automated system or a manual system, the errors must be verified. It means that humans need to look at the information and make a judgement whether it is truly an error or not, and whether the classification of the error as critical or non-critical is correct. This step is unavoidable before errors deleting, good data values identification and bad values replacement. After the cleaning process, error check must be repeated in order to verify that in the process of cleaning some new errors did not occur and that all bad data were correctly replaced by good data. The final step is data merging, since the data cleaning never takes place on actual working data in the production environment. What is being cleaned is the generated identical copy of the original dataset, which prevents unrepairable damage or loss of the original data. When audit verifies successful cleaning of the copy, the cleaned data can be used to replace original dataset.

3.3.2 Missing values treatment

Missing data occurs in a data set when at least one observation is missing a value on at least one variable. There are usually discovered during data cleaning process, when basic statistic for each variable is computed, showing the count and percentage of missing versus non-missing values and zero versus non-zero values. There can be a number of different reasons why the data are missing, from human's errors in entering the data correctly, to the equipment malfunctioning. **The important issue on missing values is whether they can be characterized as missing completely at random or missing not at random.** Data is missing completely at random when the probability that an observation X_i is missing is unrelated to the value of X_i or to the value of any other variable, otherwise data is not missing at random. The randomness of missing data is important for the later choice of missing values treatment. Generally, there are two approaches for dealing with missing values: **missing value elimination and missing value imputation.** The choice of the approach and the method within each approach depends not only on the randomness of the missing data, but also on the initial database size, the minimal number of observations needed for valid analysis of the concrete problem, occurrence of the missing values across all input variables in the dataset. In the elimination approach, if a missing value occurs on any of the variables the entire observation is eliminated. The consequence of this rigid approach is that even a dataset with modest total number of missing values suffers from substantial reduction in size, which can make the sample size insufficient for the later modelling purpose. Still, if data is missing completely at random this approach is the best solution, since the analysis and parameters estimates are unbiased by the absence of the data. When the data is not missing completely at random, exclusion of the incomplete observations leads to biased results. The only way to obtain unbiased parameters estimates is to model missingness and to incorporate that model into the model for estimation. This can be very complex task and belongs more to the research then to the business application field. The missing value imputation is the process of filling in the missing values of a variable, using one of the common imputation methods: substitution with a measure of central tendency, distribution based imputation, tree imputation, regression imputation, imputation using the expectation-maximization algorithm, multiple imputation. Substitution with a measure of central tendency is the simplest method and uses mean, median, midrange or mode value of the variable, estimated by the robust estimators, to substitute missing values. Better solution is distribution based imputation, which seeks to preserve the empirical distribution of the data and replacement values are calculated based on the random percentiles of the variable's distribution. Tree imputation and regression imputation replace missing values with the predicted values obtained from the decision tree and regression model, respectively. The expectation-maximization algorithm and multiple imputation are the finer methods, but based on the same ideas as the regression imputation which is appropriate for the most of the real datasets.

3.3.3 Removal of outliers

Many proposed definitions of outlier share the common idea that outlier is an observation which deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism (Ben-Gal, 2010). Although outliers are often considered as an error, here we distinguish the errors in the database records elaborated under the data

cleaning section from the observations that deviate markedly from other observations in the database. Since the outliers may lead to model misspecification, biased parameter estimates and incorrect results, it is important to identify them prior to analysis and modelling phase. Methods for detecting outliers can be divided into **univariate methods**, specific for the earlier researches and **multivariate methods** that belong to more recent works. Another taxonomy is between **parametric (statistical) and nonparametric methods** that are model-free. The difference between two is that statistical parametric methods assume a known distribution of the observations, which is often a violated assumption in the real-world data. According to parametric methods, those observations that deviate from the model assumption, such as normal distribution for example, are flagged as outliers. Within the class of non-parametric outlier detection methods there is a set of so called distance-based methods, which are based on local distance measures and can handle large databases. In that respect, outliers are defined as observations whose distance to the location estimate exceeds some multiple of the scale estimate. Location estimate can be, for example mean, median and trimmed mean, while scale estimate can be standard deviation, median absolute deviation, interquartile range or Gini's mean difference. These methods require no prior knowledge of the underlying data distribution and are used in this study. Outliers can be detected for each input variable and the intersection of these sets can be used to determine the final set of the observation considered and removed from the dataset as serious outliers. Although considered as errors, outliers can reveal some key information in the data generation mechanism or underlying behaviour that the analyst tries to model. These worthy case-studies can be afterwards incorporated into the research conclusions, even if the outliers were excluded from the modelling database since they lead the algorithms to incorrect or biased results.

3.4 Preliminary analysis and variables selection

Having a cleaned and prepared modelling customer database is still far away from having a problem solution. The preliminary analysis should take a place before modelling phase, since it helps in choosing appropriate methodology, speeds up the learning process and improves final model interpretability. When the modelling problem definition assumes the underlying differences between groups of observation, then **canonical discriminant analysis (CDA)** can be used as an auxiliary method for testing the separability of the groups. In the problem of predicting the probabilities for single product purchase, built as the binary logistic regression, there are at least two underlying groups of customers: the group of non-buyers and the group of product buyers. Therefore, canonical discriminant analysis can help in testing the separability between these two groups of customers, using different input variables combinations. **Canonical discriminant analysis is a multivariate, dimension-reduction statistical technique related to principal component analysis and canonical correlation.** It requires a nominal classification variable, for two or more groups of observations and several interval input variables. In our case the classification variable is derived in the way analogue to derivation of the target variable for the logistic regression. All other available interval variables can enter the canonical discriminant analysis, which then derives canonical variables as linear combinations of the interval variables with the scope to obtain the highest possible multiple correlation with the groups. The highest correlation is called the first canonical correlation, the coefficients of that linear combination are the canonical coefficients and the variable defined by that linear combination is the first canonical variable. The second canonical correlation is

obtained by finding the linear combination uncorrelated with the first canonical variable that has the highest possible multiple correlation with the groups. The process of extracting canonical variables is iterative and stops when the number of canonical variables reaches the number of original variables or the number of classes minus one, whichever is smaller. In the case of binary class variable only the first canonical variable would be of practical usage and in the case of defining several classes, according to the company's product scheme for example, k-1 canonical variables make sense, where k denotes the number of product categories. Generally, CDA helps not only in better understanding of the groups separability, but also in better recognition of input variables that mostly contribute in predicting the probabilities of buying the single product, which can be afterwards compared with the selection of the most powerful variables suggested by the logistic regression iterations. The preliminary analysis is not the substitution for the modelling, but is its complementary part in knowledge discovery process. Variable selection, also known as feature selection, can be done during the preliminary analysis or during the modelling phase, depending on the problem definition. **Variable selection is a technique of selecting a subset of relevant features for building robust learning models**, with many potential benefits such as: facilitating data visualization and data understanding, reducing storage requirements, reducing training time, avoiding the effects of the curse of dimensionality and improving prediction performance (Guyon & Elisseeff, 2003). Feature selection algorithms typically fall into two categories: **feature ranking and subset selection.** The first category algorithms rank the features by a metric and eliminate all features that do not achieve an adequate score. Algorithms for subset selection search the set of possible features for the optimal subset, according to the target problem. Both approaches can have their shortfalls. Selecting the most relevant variables is usually suboptimal for building a good predictor, particularly if the variables are redundant. At the other side, a subset selection may exclude many redundant, but still relevant variables and, since the approach searches for the optimal subset, the memory and processing time requirements may become unacceptable. The importance of the proper feature selection is especially visible in the problems where hundreds and even thousands of raw input variables are available as potential predictors. In the real business environment providing faster and more cost-effective predictors often becomes the objective itself and can determine the success of the modelling phase. Therefore, having a strategy for coming to the set of good predictors is highly recommended. In this study the most popular form of feature selection in statistics, called stepwise regression is used. **Stepwise regression is a natural choice for the problems solved by the logistic regression.** It is a greedy algorithm that adds the best feature and/or deletes the worst feature at each round, where the best and the worst are defined in terms of the variable contribution to the model significance. Therefore, the analyst chooses two p-values for the stepwise algorithm: p-value for the variable entrance into the model and p-value for the variable exclusion from the model. The process terminates when no significant improvement can be obtained by adding or by subtracting another variable. The subset of the variables already in the model at that moment is the best subset chosen by the stepwise procedure.

4. Dual approach model setup

The demand schedule represents the amount of some good that buyers are willing and able to purchase at various prices, assuming all other determinants of demand, such as income, tastes and preferences, the price of substitute goods, the price of complementary goods, the number of buyers etc., remain the same. It is graphically depicted as the two-dimensional

Dual Approach to the Modelling Single
Product Demand Curves in the Next Best Offer CRM Problem
113

downward sloping demand curve. The negative slope of the demand curve reflects the **law of demand**, which says that as price of the good decreases, consumers will buy more of the good, ceteris paribus (Samuelson & Nordhaus, 2000). The theory recognizes hypothetical types of the goods that have upward-sloping demand curves, such as Giffen goods and Veblen goods, where this general price rule is violated. The demanded quantity of good X for individual customer i is the function f_i:

$$Q_i(x) = f_i(p_x, p_s, p_c, I_i, T_i, P_i) \tag{13}$$

where p_x is the price of good X (it is not customer indexed, if the same price is applied to each buyer), p_s is the price of substitute goods, p_c is the price of complementary goods, I_i, T_i and P_i are the income, tastes and preferences of the buyer i respectively. Such a pure mathematical model, in which the relationships between inputs and outputs are captured entirely in deterministic fashion, is important theoretical tool, but is rather impractical for describing observational, experimental, or survey data. For such purpose, we need to allow the model to draw on stochastic as well as deterministic elements, changing it **from the deterministic into the stochastic model.** Since the final goal in this study is to estimate the form of the single product demand curves, that stochastic model will contain estimated parameters and becomes so statistical model. For these reasons, the equation (13) is transformed into:

$$Q_i(x) = f_i(p_x, wap_s, wap_c, I_i, PN_i, \sigma_i) \tag{14}$$

where wap_s represents weighted average of the most important substitutes in the market, rather than the price of one substitute or each of the substitutes exactly, wap_c represents weighted average of the most important complementary goods in the market, PN_i stands for the recognized and predicted needs of the customer i and replaces his/her practically unknown tastes and preferences, and σ_i represents stochastic effect on the demand of customer i. Note that the word "customer" replaced the word "buyer" as we move from theoretical framework to more practical demand function form.

If the equation (14) is accepted as the description of the purchased quantity of the product X by the customer i, then the following transformation introduces the first modelling event:

$$Purchase_{X(i)} = \begin{cases} 1, & for \quad Q_{X(i)} > 0, \\ 0, & otherwise. \end{cases} \tag{15}$$

The modelled event becomes a dichotomous event, represented by the binary variable $Purchase_{X(i)}$ and logistic regression model becomes the appropriate modelling method. In this phase the input dataset still does not include the information about customers' channel preferences, contained in the fuzzy membership functions. When this information enters the model in the second phase, the modelled event changes into the combined event, introduced by another transformation:

$$Purchase_{X(i)} = \begin{cases} 1, & for \quad Q_{X(i)} > 0 \quad and \quad \Pr efChannel = 1, \\ 0, & otherwise. \end{cases} \tag{16}$$

where PrefChannel is the variable having the value 1 if the customer was contacted through his preferred channel and the value 0 if not. The preferred channel is recognized as the fuzzy cluster for which the membership function reaches the maximum in the second sub-model. The final model consists of two sub-models: the first sub-model is logistic regression model for assessment of the single product demand curves and the second sub-model is fuzzy clustering model of the customers' behaviour on the channels. The next sections present the results obtained through the data analysis process, parameters' estimation and validation of the fuzzy membership functions contribution to the model. The results were obtained following the methodology described throughout the chapter, from the data cleaning phase to the logistic regression and fuzzy clustering part. The small portion of the real purchase data was collected in the inquiry and the rest of the data were simulated based on the collected sample, under the controlled conditions.

4.1 Data analysis review

Before modelling, the input data quality, missing values existence and outliers' presence were checked. Variable CSI, which stands for Customer Satisfaction Index, had missing values on 37% of the customers. Therefore, the substitution with two different measures of central tendency, mean and median, were tried and both of them significantly changed the measures of the central tendency and the variable distribution after the imputation. The significantly high correlation between the variable TENURE_MNTH, which denotes the length of the relationship of the customer with the company, and CSI was discovered during the input variables analysis, showing that these missing CSI values were predominantly distributed among the customers with the shortest relationship with the company. That is a situation of the data not missing completely at random, therefore the better solution was to substitute missing values with the smallest value from the CSI variable domain, which was treated as the default value in research of customer satisfaction and correlated with the shorter tenure. That method affected much less the non-missing values distribution and was preserved as the best solution. The most serious outliers were detected for each input variable and each time period separately, using the standard deviation method. The intersection of these separate outliers' sets was found. The number of the occurrences of the CUSTOMER_ID as an outlier, the maximum multiple of the standard deviation obtained by CUSTOMER_ID and presence of the customer in the set of the most serious outliers for each time period was considered for the decision about their removal from the modelling dataset. The outliers occurred mostly on the variables describing their expenditure on luxury goods and seven customers showed up with maximum deviance in both time periods.

The outliers from the second time period are shown in the Table 1. The circled customers were found in the both outlier sets and were excluded from the modelling of single demand curves, as well as from the fuzzy clustering.

The complete correlation matrix of Pearson product-moment correlations was calculated for both time periods separately and for all the data together. The comparison of the correlation matrices showed time-consistent correlations between variables. The extraction from the correlation matrix for the first time period is given in the Table 2. The corresponding p-values of the correlations were used to test the null hypothesis that the correlation is 0, assuming independent and identically distributed observations from a bivariate distribution with at least one variable normally distributed. Table 2 shows that

Dual Approach to the Modelling Single
Product Demand Curves in the Next Best Offer CRM Problem
115

the correlation between Customer Satisfaction Index and Tenure in Months is positive and significant, which leads to conclusion that the customers with longer relationship with the company show higher satisfaction and are more likely to participate in the satisfaction inquiries, since the missing values of CSI were predominantly distributed among the customers with the shortest tenure. Correlations help not only in recognizing relationships among the variables, but also help in removing redundant variables from the later logistic model. For instance, discovered high and significant correlation between Customer Satisfaction Index and Tenure in Months means that we have to delete one of them from the final logistic regression model. Wald chi-square together with business experience helps in identifying which one to delete. Since these two variables are highly correlated, their correlations with other variables and with target variable bring similar information value to the model. However, it is more sensible to leave Tenure in Months and delete CSI, due to the fact that Tenure can be calculated for each customer and has no missing values, while CSI value depends on the customer's willingness to participate in the satisfaction measurements and is often missing.

TIME PERIOD	CUSTOMER ID	NUMBER OF OCCURRENCES AS AN OUTLIER	MAX MULTIPLE OF STD DISTANCE	VARIABLES WITH MAX DIFFERENCE
2Q	99	5	2,82	O4_EXP (O4_QUANTITY)
2Q	91	4	3,16	L1_EXP (L1_QUANTITY)
2Q	3	3	3,16	L1_EXP (L1_QUANTITY)
2Q	64	3	3,29	L3_EXP (L3_QUANTITY)
2Q	78	3	3,77	LUXURY_ACTIVITY
2Q	35	2	2,99	NO_CONTACTS
2Q	41	2	3,35	O5_EXP (O5_QUANTITY)
2Q	47	2	2,88	O5_EXP (O5_QUANTITY)
2Q	93	2	3,29	L3_EXP (L3_QUANTITY)
2Q	15	1	2,51	NO_CONTACTS
2Q	65	1	2,64	LUXURY_ACTIVITY

Table 1. Outliers in the second time period according to the standard deviation method

CSI vs.	NO_CHILDREN	TENURE_MNTH	EXP_LUXUR	NO_CONTACTS
PCC	0.13983	0.97974	0.45132	0.18264
p-value	0.18130	<0.0001	<0.0001	0.07970

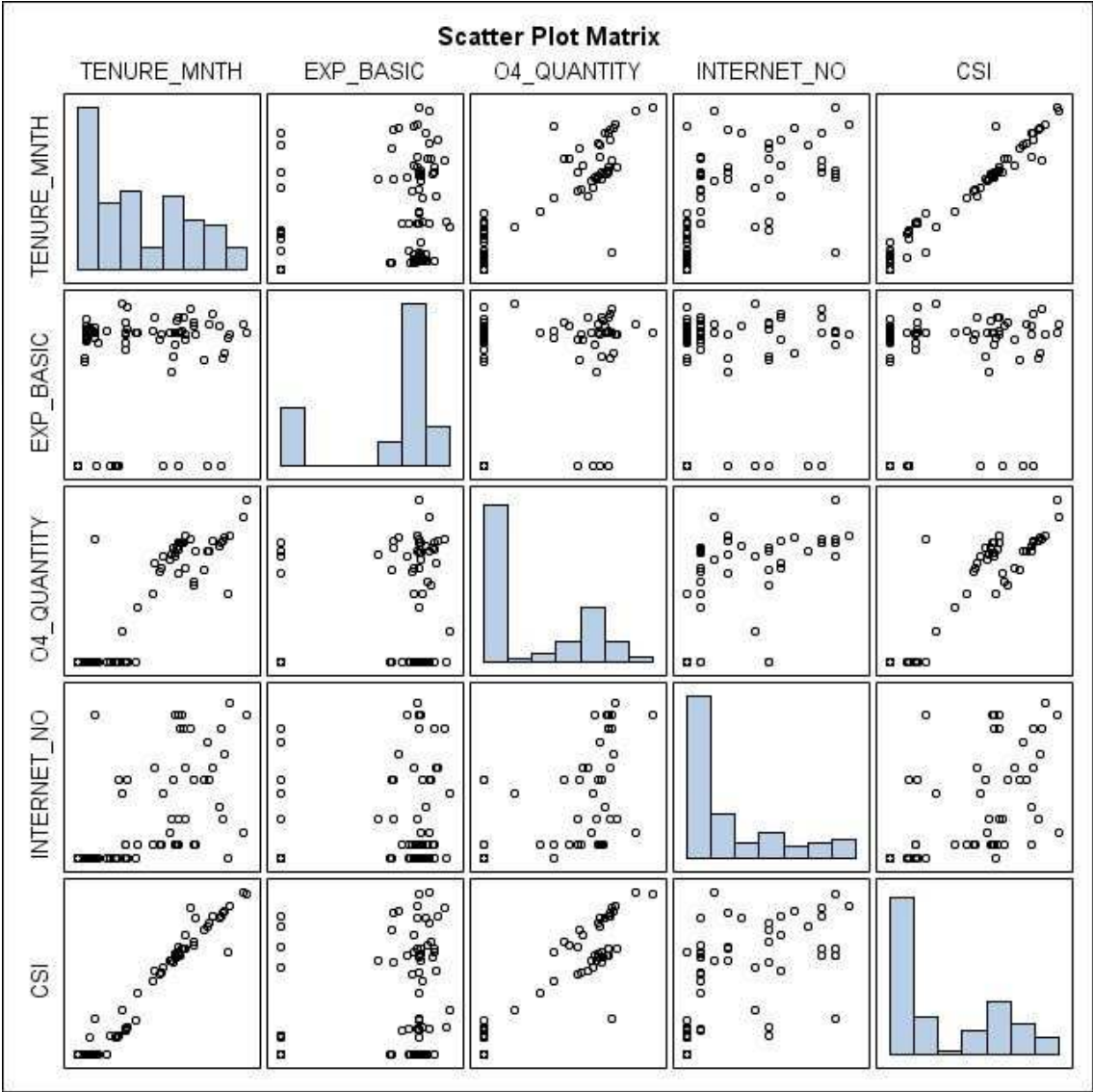Table 2. Extraction of Pearson Correlation Coefficients matrix for Customer Satisfaction Index

Fig. 2. Scatter plot matrix as an auxiliary tool in preliminary analysis

The correlation matrix is usually accompanied by the **scatter plots matrix**, which presents the relationship between each pair of the input variables graphically and is helpful tool in initial exploration of the input data. Figure 2 gives the scatter plots matrix of five variables of our interest for predicting the probabilities of product O4 purchase.

4.2 Single product demand curves

There were two products chosen for the demand modelling: product denoted as B1, belonging to the BASIC product category and product denoted as O4, belonging to the OPTIONAL product category according to the product scheme of the company. In order to prepare the data for the binary logistic regression, the target variables B1_FLG and O4_FLG were calculated using the transformation (15). In the first stage of the modelling only restricted number of variables were used and no data about the communication history or channels' usage were available. The stepwise regression method was used, with p-value for the variable entrance in the model set at the level of 0,3 and p-value for the variable to stay

Dual Approach to the Modelling Single
Product Demand Curves in the Next Best Offer CRM Problem
117

in the model set at the level of 0,35. The Figure 3 shows Receiver Operator Characteristics (ROC) curves of the model for B1 product, obtained without data about channels' usage. ROC curve is also known as Lift Curve and summarizes the performance of a diagnostic test with a positive/negative outcome. The final model was assessed in the nine steps.

The input variables chosen by the stepwise algorithm in the last step contained both Customer Satisfaction Index and Tenure in Months, recognized during the preliminary analysis as highly correlated variables. They satisfied the conditions for entrance and for stay in the model, improved fit statistics and obtained percent concordant pairs of 83,3%, percent discordant of 16,7%, percent tied of 0,00% and Somers' D of 0,66. The modeller, not the algorithm, brings the final decision whether to preserve both variables or to delete one of them and which one.
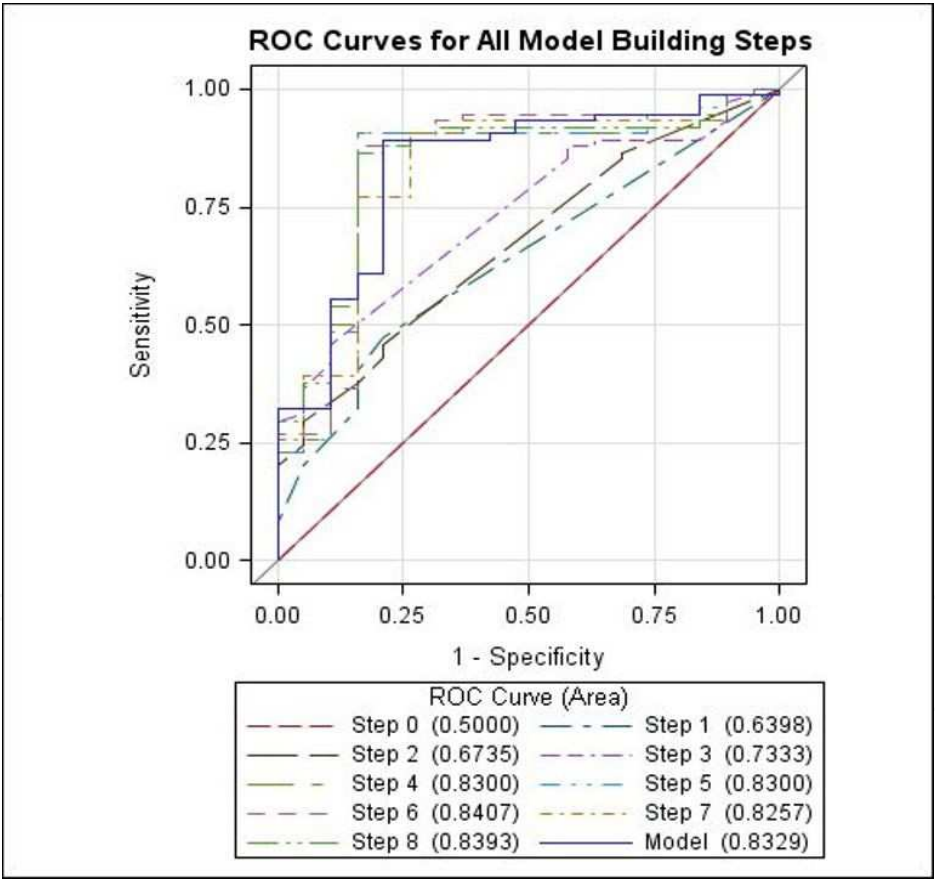

Fig. 3. Receiver Operator Characteristics (ROC) curves for the logistic regression on B1_FLG without channels data
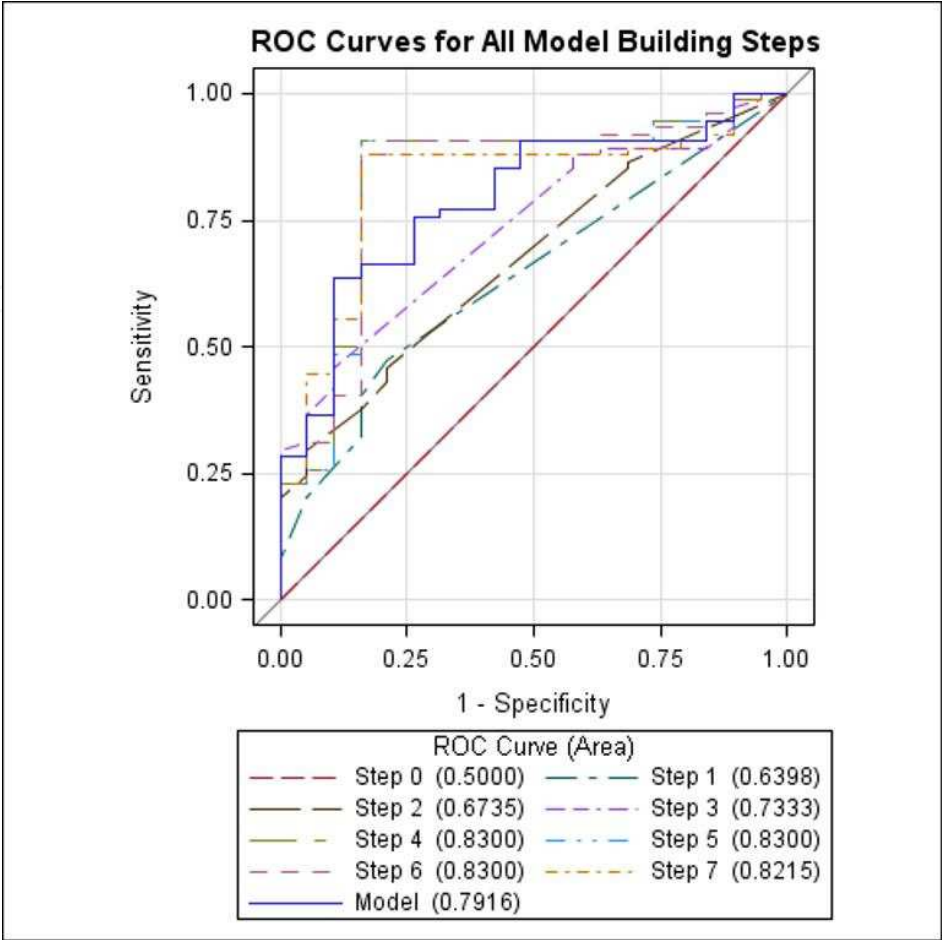
Fig. 4. Receiver Operator Characteristics (ROC) curves for the logistic regression on B1_FLG without channels data and CSI variable

Therefore, we run another logistic regression omitting Customer Satisfaction Index from the input dataset, which lowered Area Under the Curve (AUC) from 83,29% to 79,16% on the modelling data. Percent concordant dropped to 79,2%, percent discordant increased to 20,8%, percent tied remained 0,00% and Somers' D decreased to the value of 0,583, all showing lower fit on the modelling data. Figure 4 shows ROC curve for the logistic regression without channels data and Customer Satisfaction Index and Table 3 gives the variables, parameter estimates and p-values of the new model.

PARAMETER	VARIABLE DESCRIPTION	ESTIMATE	Pr > ChiSq
Intercept	-	1.3690	0.0238
L4_EXP	Expenditure on luxury product L4	-0.00916	0.0135
NO_CHILDREN	Number of children	-0.4044	0.2471
O1_EXP	Expenditure on optional product O1	-0.0137	0.0821
O5_EXP	Expenditure on optional product O5	0.0330	0.1863
Q_LUXUR	Total purchase of all luxury products	0.4287	0.0396
TENURE_MNTH	Relationship with the company in months	0.0167	0.0236

Table 3. Parameter Estimates for the logistic model on B1_FLG without channels data and CSI variable

Another trial to increase the model fit was to introduce the variables from the customers' communication history with the company, in their original form. When these variables entered the modelling database, they significantly influenced the stepwise process and lead to almost perfect predictions, with unreliable AUC of 0,9822. This happened mostly because the modelling sample was too small, the channels usage variables were mostly binary variables and some of them were highly correlated with already existing input variables describing customers' purchase history. Binary variables (flags) often bring a strong impact on the modelled event, but also lead to model over-fitting and questionable model validity. Three of five variables that remained in the model in the last iteration were number of internet contacts, number of total purchases and flag for phone contacts, which are correlated among each other rather significantly and therefore should not be in the model at the same time. In practice, although available, such discrete variables have to be avoided or substituted whenever possible, since the resulting model does not generalize well and can not be used for obtaining reasonable predictions. At the other side, the significant correlation between channel variables and purchase variables is desired for the final purpose of offering the right product through the right channel. The open question here is: "How to include valuable channels' usage information into the model, without corrupting its validity?" In the next step, the channel variables were used for the fuzzy clustering of the customers and the channels' usage information re-enter the logistic regression model in the form of the fuzzy membership functions.

4.3 Fuzzy clusters of the customers

In order to improve predictions for the purchase probabilities and to obtain model that generalizes well at the same time, we developed fuzzy clustering sub-models, one for each of chosen products B1 and O4. In development process, fuzzy c-means algorithm was applied only on the data about channels usage and communication history, testing different combinations of input parameters p, m and number of iterations. The final choice of FCM for product B1 was the combination of p = 3, m = 1,30 and number of iterations = 100 and the release is named FCM-p3. For product O4 the combination of p = 2, m = 1,25 and number of iterations = 20 was chosen and named FCM-p2. Informally, a combination of clustering parameters performs better than some other combination if the resulting clusters less overlap, if they are of similar size and obtain good separability of the clustered observations. If the resulting clusters are of very different size, and especially if one of them contains only a few observations, the number of clusters should be set to lower value (lower parameter p) and small clusters should be checked for outliers. All these steps were done in order to get the final releases FCM-p2 and FCM-p3. The standard performance metrics was calculated from confusion matrix of each release, according to the same criteria (variables). The following tables show why these releases were chosen and how their results enter the logistic models, leading to the final next best offer model. Table 4 gives the confusion matrices of the release FCM-p2 on the sample data, according to variables INTERNET_FLG and PHONE_FLG.

The fuzzy algorithm divided customers into two fuzzy clusters, as it was determined with parameter p = 2. In cluster profiling, which means discovering what makes the customers in one cluster similar to each other, on two variables performance metrics was good: INTERNET_FLG and PHONE_FLG. The algorithm correctly put the majority of internet users (73,9%) into the FUZCLUS_1 and the majority of non-users (89,4%) into the cluster FUZCLUS_2, according to calculated fuzzy membership functions $\mu_{FUZCLUS\_1}$ and $\mu_{FUZCLUS\_2}$.

CRITERION	FUZZY CLUSTER		
INTERNET_FLG	FUZCLUS_1	FUZCLUS_2	TOTAL
YES	**34**	12	46
NO	5	**42**	47
TOTAL	39	54	93
PHONE_FLG	FUZCLUS_1	FUZCLUS_2	TOTAL
YES	29	**42**	71
NO	**10**	12	22
TOTAL	39	54	93

Table 4. Confusion matrices of FCM-p2 according to INTERNET_FLG and PHONE_FLG

If we validate the same fuzzy clusters according to phone usage criterion, we see that FUZCLUS_1 which mostly contains internet users at the same time contains significant number of phone non-users (45,5%) and analogously FUZCLUS_2 which contains internet non-users at the same time contains the majority of phone users (59,2%). The performance metrics, given in the Table 5, shows that FCM-p2 performs very well in separating users and non-users on both channels, with more success for the internet channel. Therefore, the values of $\mu_{FUZCLUS\_1}$ and $\mu_{FUZCLUS\_2}$ calculated for each customer contain the information whether the customer prefers and uses more frequently internet or phone communication with the company. The advantage of fuzzy clustering over crisp clustering is that fuzzy membership functions enter the logistic regression as continuous input variables and not as class or binary variable and they change over time smoothly and not in discrete way. The next step is to recognize for which product this information is most useful, especially when focusing on the internet communication.

Metrics of FCM-p2 on	Hit Rate	False Alarm Rate	Accuracy	Specificity
INTERNET_FLG	73,91%	10,64%	81,72%	89,36%
PHONE_FLG	59,15%	54,55%	55,91%	45,45%

Table 5. Standard performance metrics of FCM-p2

FCM-p3 divided customers into three clusters of similar size, as it is shown in the Table 6. This release gives the advantage to discovering the phone users over the internet users, what can be seen from the performance metrics in the Table 7.

CRITERION	FUZZY CLUSTER			
INTERNET_FLG	FUZCLUS_1	FUZCLUS_2	FUZCLUS_3	TOTAL
YES	**27**	14	5	46
NO	2	**15**	**30**	47
TOTAL	29	29	35	93
PHONE_FLG	FUZCLUS_1	FUZCLUS_2	FUZCLUS_3	TOTAL
YES	23	**19**	**29**	71
NO	**6**	10	6	22
TOTAL	29	29	35	93

Table 6. Confusion matrices of FCM-p3 according to INTERNET_FLG and PHONE_FLG

Metrics of FCM-p3 on	Hit Rate	False Alarm Rate	Accuracy	Specificity
INTERNET_FLG	58,70%	4,26%	77,42%	95,74%
PHONE_FLG	67,61%	72,73%	58,06%	27,27%

Table 7. Standard performance metrics of FCM-p3

Since the product O4 has stronger correlation with INTERNET_NO (number of internet contacts in time period) then product B1, the resulting fuzzy membership functions from FCM-p2 enter the logistic model for O4, while the membership functions from FCM-p3 enter the logistic regression model for B1. In case of O4 variable $\mu_{FUZCLUS\_1}$ enters the logistic model as significant variable, while in the case of B1 variable $\mu_{FUZCLUS\_2}$ enters the model. Since the B1 model was elaborated previously in more detail, Table 8 gives parameter estimates for the logistic model on B1_FLG with fuzzy membership functions from FCM-p3. Figure 5 shows ROC curve for the new hybrid model, with improved AUC of 85,06% in comparison to the model without information on channels usage.

PARAMETER	VARIABLE DESCRIPTION	ESTIMATE	Pr > ChiSq
Intercept	-	0.3974	0.3716
L4_EXP	Expenditure on luxury product L4	-0.0129	0.0077
O2_EXP	Expenditure on optional product O2	-0.0392	0.1579
O3_EXP	Expenditure on optional product O1	-0.0877	0.2547
O5_EXP	Expenditure on optional product O5	0.0944	0.0215
Q_LUXUR	Total purchase of all luxury products	0.4602	0.0563
TENURE_MNTH	Relationship with the company in months	0.0298	0.0095
μFUZCLUS_2	Fuzzy membership function for fuzzy cluster 2 obtained by FCM-p3	3.2878	0.0233

Table 8. Parameter Estimates for the logistic model on B1_FLG with fuzzy membership functions from FCM-p3

The percent concordant is 85,1%, percent discordant 14,9%, percent tied 0,00% and Somers' D is 0,701. The Table 9 gives the final variable list and parameter estimates. According to p-values the most significant variables are now Expenditure on Product L4, Tenure in Months, Expenditure on Product O5, μFUZCLUS_2 fuzzy membership function and Total Purchased Quantity of Luxury Products, while Expenditure on Products O2 and O3 can be omitted. It is interesting that expenditure variables entered the model and not the purchased quantities of these products, although the both types of variables were in the modelling dataset. Since the expenditure on product in one time period equals purchased quantity multiplied by the product price at the moment of each purchase, this could potentially signalize that the prices of these products are in fact the factors that influence the probabilities of purchase of product B1, and not the general attributes of other products.
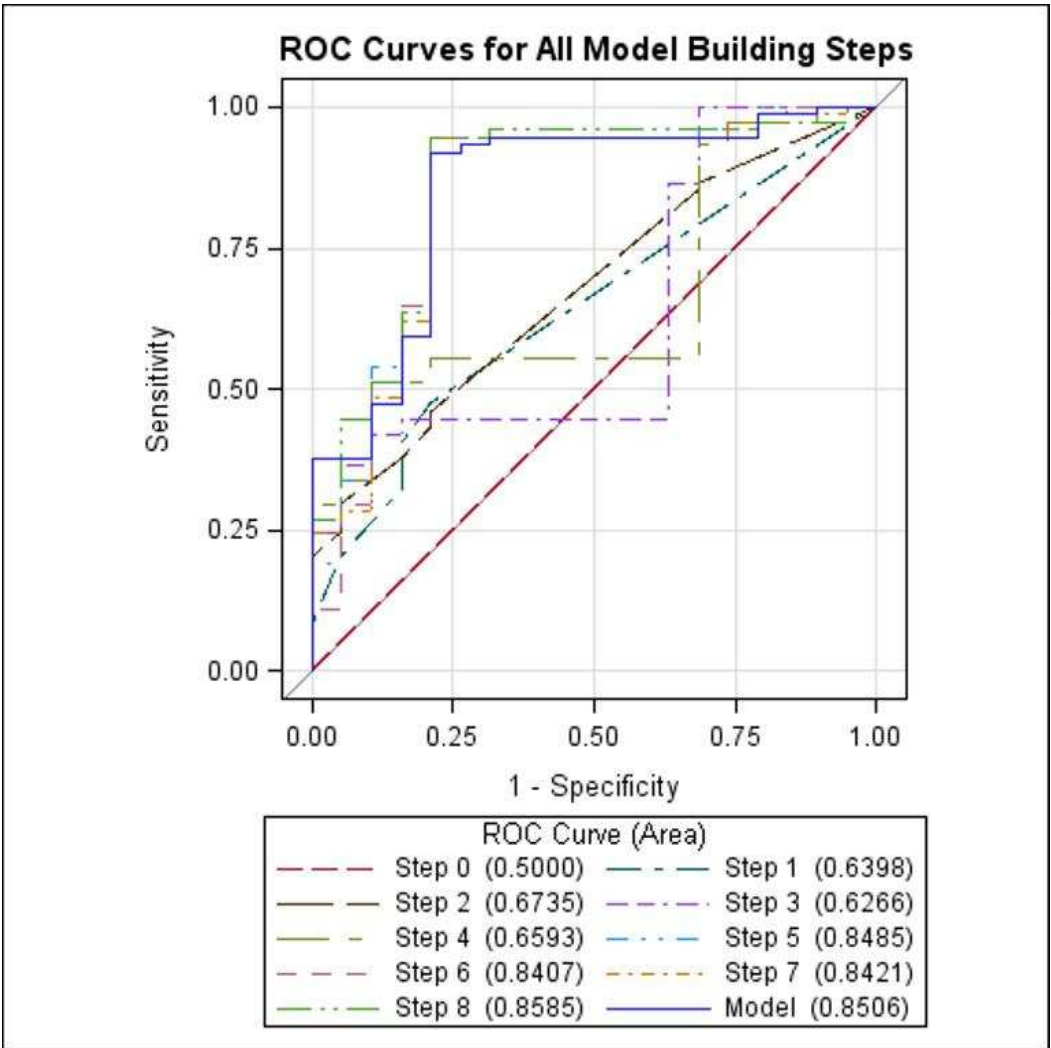
Fig. 5. Receiver Operator Characteristics (ROC) curves for the logistic regression on B1_FLG
with fuzzy membership functions from FCM-p3

4.4 Dual approach solution to the next best offer problem

The steps described in sections 4.2 and 4.3 are ideally repeated for all k products of
interest, resulting in k hybrid models. Each model gives not only the predicted purchase
probabilities, but also the information through which channel the customers should be
contacted in order to increase those purchase probabilities. The final solution which
product to offer to individual customer and through which channel to contact her/him is
obtained by descending sort of all single hybrid models' outcomes for each customer,
where the first remaining (or highest remaining) outcome wins, after eventual hard
filters and business rules were applied to the sorted queue. In our study only two
products were compared and resulting offers consisted of product O4 offered via
internet and product B1 offered by phone calls, depending on combination that resulted
in higher probability of purchase for concrete customer. The future changes in
customers' product preferences and purchase behaviour will affect the model through
the logistic regression sub-model, while the changes in communication preferences and
channels' usage will affect the fuzzy membership functions in fuzzy clustering sub-

model. Such a dual approach to modelling the problem of next best offer should raise the flexibility of the entire proposed model.

5. Conclusions and further work

The next best offer problem is well known problem in CRM community and there exist theoretical and practical solutions to this problem. Still, the literature describing in detail the methodology or new applicable scientific insights to the problem is scarce. The problem solutions are mostly spread as built-in solutions within software tools for statistical modelling and campaign management. The majority of them applies either the logistic regression or GLM to modelling the product demand curves and uses the sort algorithm as default method for choice of the product to be offered. Since the real business data does not behave nicely as artificial or simulated data, the improvement of the final solution is limited by the possibility of improvement of the underlying logistic regression model. The decision engine is too simplistic for growing business complexity, where the next best problem has already evolved into the next best action problem. The next best action is wider problem in the sense that it consist of several decisions: which product or service to offer to the customer, when ideally to offer, through which channel(s) to make the offer, at which price level and with which individualized message for each single customer. The growth of available software tools and IT applications for solving CRM problems is an advantage of modern world only if used in proper manner. Therefore, it becomes necessary to raise the level of deep understanding of such built-in solutions, to avoid misuse and to manage the expectations that companies have from the software itself. The level of specific statistical and business knowledge, required for modelling such complex problems, has to be recognized and constantly upgraded, if striving for better CRM solutions. The dual approach solution to the next best offer proposed in this chapter had and achieved the following goals: it explained in concise way the overall process of solving a general CRM problem, presented concrete statistical method (logistic regression) and knowledge discovery method (fuzzy clustering) with application to the next best offer problem, confirmed the hypothesis that fuzzy clustering on channels' data could result in fuzzy membership functions that are significant input variables for logistic regression and lead to better model performance. Finally, it gave an example of successful combining different methodologies into one hybrid solution. The further work on the topic might include extension of the solution to the cost-benefit analysis related to the elements of performance metrics of fuzzy clustering, applying different alpha-cuts on the membership functions in order to recognize the customers during their transition from one preferred channel to another preferred channel and testing the possibility of influencing such a transition by the company's actions. The greatest challenge in such CRM research remains the availability of real industry data. The pure researches rarely have access to real and fresh business data and analysts within the companies rarely have the possibility to devote their time to research, in order to shift the modelling horizon at the higher level. This chapter might give an idea how to sustainably introduce new methods into the already existing modelling framework, instead of forcing better results with exhausted possibilities or by buying new software. The CRM relies on knowledge discovery and knowledge discovery requires time to get know the data, to feel the data, to understand industry context and to acquire needed business wisdom.

6. References

Anderson, K. & Kerr, C. (2002). *Customer Relationship Management,* McGraw-Hill, ISBN 0-07-137954-1, New York, USA

Ben-Gal, I. (2010). Outlier Detection, In: *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers,* O. Maimon & L. Rokach, (Eds.), 117-128, Springer, ISBN 978-0-387-09822-7, New York, USA

Berry, M. J. A. & Linoff, G. (2004). *Data mining techniques: for marketing, sales, and customer relationship management,* Wiley Publishing, Inc., ISBN 0-471-47064-3, Indianapolis, USA

Cox, E. (2005). *Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration,* Morgan Kaufmann Publishers, ISBN 0-12-194275-9, San Francisco, USA

Guyon, I. & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research,* Vol.3, (March 2003), pp. 1157-1182, ISSN 1532-4435

Hosmer, D. W. & Lemeshow, S. (2000). *Applied logistic regression*, John Wiley & Sons, Inc., ISBN 0-471-35632-8, New Jersey, USA

Kleinbaum, D. G. & Klein, M. (2010). *Logistic Regression: A Self-Learning Text*, Springer, ISBN 978-1-4419-1741-6, New York, USA

Parr Rud, O. (2001). *Data Mining Cookbook Modeling Data for Marketing, Risk, and Customer Relationship Management,* John Wiley & Sons, Inc., ISBN 0-471-38564-6, New York, USA

Samuelson, P. A. & Nordhaus, W. D. (2000). *Ekonomija,* MATE d.o.o., ISBN 953-6070-10-3, Zagreb, Croatia

Theodoridis, S. & Koutroumbas, K. (2006). *Pattern Recognition,* Academic Press Elsevier, ISBN 0-12-369531-7, San Diego, USA
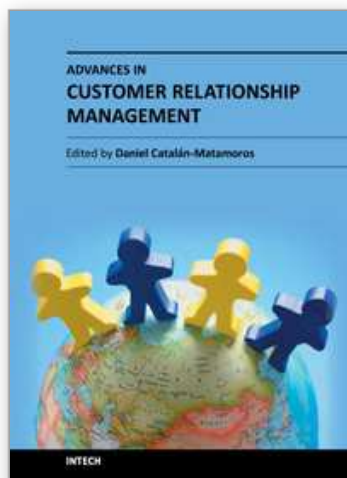
Advances in Customer Relationship Management

Edited by Dr. Daniel Catalan-Matamoros

Customer relationship management (CRM) strategies have become increasingly important worldwide due to changes in expectations from customers as well as changes in the nature of markets. This book puts forth a conceptualization that attempts to not only outline CRM's domain but also to reconcile the divergent perspectives found in the academic and popular literature. Readers can see through measurable data-containing examples how the theory is applied with great success by various real-life examples. This book presents innovative proven methods for determining whether a CRM strategy for changing the way a company provides service (by adding new technology, processes, and procedures) will realize the return on the investment projected. It could be a great help to CRM personnel, student, managers and any one that works directly or indirectly with customers.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Džulijana Popović (2012). Dual Approach to the Modelling Single Product Demand Curves in the Next Best Offer CRM Problem, Advances in Customer Relationship Management, Dr. Daniel Catalan-Matamoros (Ed.), ISBN: 978-953-51-0516-9, InTech, Available from: http://www.intechopen.com/books/advances-in-customer-relationship-management/dual-approach-to-the-modelling-single-product-demand-curves-in-the-next-best-offer-crm-problem

INTECH
open science | open minds