manuscript No.
(will be inserted by the editor)

Using Customer Lifetime Value and Neural Networks to Improve the Prediction of Bank Deposit Subscription in Telemarketing Campaigns

Sérgio Moro · Paulo Cortez · Paulo Rita

Received: 2014 / Accepted: 2014

Abstract Customer lifetime value (LTV) enables using client characteristics, such as recency, frequency and monetary (RFM) value, to describe the value of a client through time in terms of profitability. We present the concept of LTV applied to telemarketing for improving the return-on-investment, using a recent (from 2008 to 2013) and real case study of bank campaigns to sell long-term deposits. The goal was to benefit from past contacts history to extract additional knowledge. A total of twelve LTV input variables were tested, under a forward selection method and using a realistic rolling windows scheme, highlighting the validity of five new LTV features. The results achieved by our LTV data-driven approach using neural networks allowed an improvement up to 4 pp in the Lift cumulative curve for targeting the deposit subscribers when compared with a baseline model (with no history data). Explanatory knowledge was also extracted from the proposed model, revealing two highly relevant LTV features, the last result of the previous campaign to sell the same product and the frequency of past client successes. The obtained results are particularly valuable for contact center companies, which can improve predictive performance without even having to ask for more information to the companies they serve.

S. Moro
Department of Information Science and Technology, ISCTE - University Institute of Lisbon, 1649-026 Lisboa, Portugal
E-mail: scmoro@gmail.com

P. Cortez
Dep. of Information Systems/ALGORITMI Research Centre, University of Minho, 4800-058 Guimarães, Portugal

P. Rita
Business Research Unit (BRU-UNIDE), ISCTE - University Institute of Lisbon, 1649-026 Lisboa, Portugal

Keywords Customer Lifetime Value (LTV) · Multilayer Perceptron · Recency, Frequency and Monetary value (RFM) · Telemarketing · Bank Deposits · Data Mining

1 Introduction

Customer lifetime value (LTV) stands for the value of a customer in terms of expected benefits considering likely future interactions with the customer [9]. LTV can be regarded as a relevant construct for every decision to improve customer relationship and profitability.

The identification of the most profitable customers, aiming to redirect a larger amount of the marketing effort towards those customers, has been a holy grail of marketing [21]. It is directly associated with predicting future behavior of customers, meaning that a good computation of LTV can help serving such purpose.

Typically, one way of characterizing a database of customers is by computing their recency, frequency and monetary (RFM) characteristics. These allow to capture customer behavior in a very small number of features [18]. In effect, RFM can be used as a base to compute LTV [10]. Still, the relative importance among recency, frequency and monetary varies with the characteristics of the product and industry. In fact, feature weight has been a subject of research in order to improve classification accuracy [1]. There are several approaches to determine the relevance of each of the three RFM features. Liu and Shih [17] used the analytic hierarchy process to determine the relative weights of RFM variables in the evaluation of LTV. Considering more recent studies, the work of Cheng and Chen [5] used the K-means algorithm to build clusters by RFM attributes, resulting in an enhancement of classification accuracy when applied to a Taiwanese company operating in the electronic industry. Kwon and Lee [16] measured loyalty by using a case-based reasoning method to compute the relationship of a user through its RFM and a smart object that attributes weights to RFM features based on its goal.

While most studies focus on optimizing the relation between each RFM parameter, Chen et al. [4] targeted the discovery of future RFM values by applying sequential pattern mining. The research used a Taiwanese supermarket chain and the results of the proposed algorithm were compared with the generalized sequential pattern apriori algorithm, showing special evidence of improvement in RFM values by cutting off more uninteresting patterns.

Another trend is to increase RFM modeling knowledge by adding other problem characterizing features. Yeh et al. [27] included two additional parameters, time since first purchase and churn probability, to model the likelihood that a customer will buy next time. Their research used a blood transfusion service for empirical analysis and the results showed greater predictive accuracy than using single RFM traditional approaches.

The evaluation of LTV is a subject widely studied. Keramati et al. [15] conducted an assessment of weaknesses and risks in Customer Relationship

Management implementations when considering the return in terms of LTV. Customer analysis and segmentation may help to enhance future targeting in terms of customer responses to marketing campaigns [26], increasing LTV. Still, it is remarkably difficult to predict future behavior of customers with effective accuracy [19]. Simply there are too many variables to account for. Each customer is a specific individual and will take any opportunity to get more value for his/her money, considering his/her own personal benefit, which varies according to each person. Furthermore, each market and company has specific contexts which are huge in its influencing characteristics making them hard to be computationally modeled. Therefore, every research that shows an improvement in predictions for targeting is relevant in its context environment.

It is worth of noticing that decision support systems and neural networks usage and its effect on real business by enabling better decision making aligned with business needs has been considered a differentiating factor to increase the return on investment, particularly in the banking industry [3, 2, 25]. This emphasizes the value of such systems and justifies the research in an attempt of improvement the accuracy of the baseline system.

Previously we reported the research for a data mining approach to predict the success of telemarketing to sell bank deposits [20]. The goal for this problem was to reduce the number of contacts while at the same time minimizing the loss of successful contacts due to model error. Thus we worked towards a reduced number of false positives (FP) while maintaining a high number of true positives (TP), to the detriment of a higher number of false negatives (FN). The results achieved were considered of very good quality: by selecting the half better classified clients in terms of probability of subscribing the deposit, the best model would reach 79% of successful contacts from the total subscribers in the whole dataset. Still, the analysis did not include historical information from previous marketing contacts to the same clients. Rather than proposing new machine learning algorithms, this paper focuses on feature engineering, which is argued in Domingos [8] as a key issue for providing better predictive capabilities in real-world applications. In particular, in this paper we study the utility of customer telemarketing historical data, with a focus on LTV and related features, such as RFM. The main contributions of the research are:

- performing a feature forward selection procedure, in an attempt to enhance the original model (from the previous research) by adding client history features, such as RFM.
- comparing the proposed approach with previous research (used as a baseline) and analyze the improvement achieved in the light of LTV relevance.
- finally, showing how the model using customer history information might benefit the bank telemarketing business in terms of targeting.

This paper is organized as follows: Section 2 presents the case-study used for comparison and the techniques used; in Section 3, the experimental design is described and the obtained results are analyzed; finally, conclusions are drawn in Section 4.

2 Materials and Methods

2.1 Data Mining

For proper comparison purposes, we adopt the same methodology that was followed in our previous work [20] and that is described in detail in this section. For the experimental setup, we chose the rminer package of the R tool, which provides a simple set of coherent functions designed specifically for conducting data mining computation in a very intuitive way [6]. In Moro et al. [20], four data mining techniques were explored: logistic regression, decision trees, support vector machines and neural networks (NN). The best result was achieved by the NN, thus this is the only technique used for the experiments reported in this paper.

The NN is based on the popular multilayer perceptron with one hidden layer with H hidden nodes and one output node [14]. The input layer holds the input vector and then propagates the activations in a feedforward fashion, via weighted connections, through the entire network.

For a given input $\mathbf{x}_k$ the state of the i-th neuron (s_i) is computed by:

$$s_i = f(w_{i,0} + \sum_{j \in P_i} w_{i,j} \times s_j) \tag{1}$$

where P_i represents the set of nodes reaching node i; f is the logistic function; $w_{i,j}$ denotes the weight of the connection between nodes j and i; and $s_1 = x_{k,1}$, ..., $s_M = x_{k,M}$. The logistic function allows to model the output response as a probability, where the output response s_o should be 1 for a successful contact. The training of multilayer perceptron is not optimal, since the final solution is dependent of the choice of NN starting weights. To solve this issue, the **rminer** package uses an ensemble of N_r different trained networks and outputs the average of the individual predictions [13]. In [20], the NN ensemble is composed of $N_r = 7$ distinct networks, each trained with 100 epochs of the BFGS algorithm and the final ensemble response is given by $\sum_{j=1}^{7} s_{o,j}/7$, where $s_{o,j}$ denotes the output success probability for the j-th multilayer perceptron of the ensemble.

To evaluate the performance of predictions, two popular client targeting classification metrics are adopted: area under the curve (AUC) of the receiver operating characteristic (ROC) graph; and area of the lift cumulative curve (ALIFT). The receiver operating characteristic (ROC) curve shows the performance of a two class classifier across the range of possible threshold ($D \in [0, 1]$) values, where the NN predicted class is interpreted as positive if $s_o > D$, else it is interpreted as negative [11]. The ROC curve plots the false positive rate (FPR), in the x-axis, versus the true positive rate (TPR), in the y-axis As stated in Section 1, our main goal was to achieve a high TPR while reducing the total number of contacts, thus the D value point in the ROC graph should be in the lower values of FPR. The overall performance of the model can be measured by computing the area of the curve ($AUC = \int_0^1 ROCdD$), where a classifier is better if its AUC value is closer to 1.0.

Regarding the lift analysis, it sorts the records in a decreasing order of the predicted probability of success by dividing the population in deciles to facilitate selection of the most likely buyers, resulting in a popular measure in marketing [22]. The lift cumulative curve plots incrementally selected fractions of the population versus the real results, ordered from the most likely to the least likely buyers. Thus a lift cumulative curve area (ALIFT) closer to 1.0 presents a better model that is capable of selecting more buyers in a smaller fraction of the population.

Complex data-driven models, such as NN, tend to provide accurate predictions, but the obtained models are difficult to be understood by humans. To open the "black box", we adopt the Data-based Sensitivity Analysis (DSA) algorithm [7], which is a sensitivity analysis technique that works by analyzing the responses of a model when a given input is varied through its domain. The analysis of the sensitivity results allows to rank the input attribute influence and also show the average effect of the most relevant features in the model responses. Visually, the former can be analyzed using a input importance bar plot, while the later can be inspected using Variable Effect Characteristic (VEC) curve.

2.2 Bank Telemarketing Data

For our experiments, we use real data collected from a Portuguese bank, consisting in telemarketing campaigns to sell long-term deposits and encompassing a period of five years, from May 2008 to June 2013. All contacts are executed through phone calls with a human agent as the interlocutor. The vast majority of the contacts are outbound, while the few inbound contacts are used when the client calls the bank for any other reason and the agent takes advantage of the contact to try to sell the deposit.

The dataset consists of 52944 contacts executed through phone calls where only 6557 of them resulted in successful deposit subscriptions, thus it is an unbalanced dataset. The previous research [20] initially analyzed 150 characterizing features, but after careful selection, adopted only a subset of 22 features for prediction (listed on Table 1). This subset will serve as our baseline, for enrichment with additional LTV related features.

As stated previously, this study is an attempt of improving the previous work results [20], thus the data evaluation procedures are exactly the same, to allow a direct comparison of the results. The whole dataset is divided into two subsets: a training set (from May 2008 to June 2012), with 51651 of the oldest contacts; and a test set (from July 2012 to June 2013), which includes the most recent 1293 contacts, i.e., the ones that will effectively be used for prediction evaluation. To simulate a real runtime execution environment, a rolling window realistic scheme (of fixed-length of size W) is used, which performs several model updates and discards oldest data [24]. This method represents an on-going process that defines the model with the latest data, performs K predictions and feeds back the model with the results from the most recent K

Table 1 Initial set of features used for modeling a successful telemarketing contact

Attribute	Description
nat.avg.rate	national monthly average of deposits interest rate
suited.rate	most suited rate to the client according to its characteristics (e.g., if it holds a credit card, a mortgage account, assets above a certain threshold, etc.)
dif.best.rate.avg	difference between best rate offered for the deposit (independent of client characteristics) and the national average
ag.gender	gender of the agent (male/female) that made (outbound) or answered (inbound) the call
ag.generic	if generic agent, i.e., temporary hired, with less experience (yes/no)
ag.created	number of days since the agent was created
cli.house.loan	if the client has a house loan contract (yes/no)
cli.affluent	if the client is an affluent client (yes/no)
cli.indiv.credit	if the client has an individual credit contract (yes/no)
cli.salary.account	if the client has a salary account (yes/no)
call.dir	call direction (inbound/outbound)
call.nr.schedules	number of previously scheduled calls during the same campaign
call.prev.durations	duration of previously scheduled calls (in s)
call.month	month in which the call is made
cli.sec.group	security group bank classification
cli.agreggate	if the client has aggregated products and services
cli.profile	generic client profile, considering assets and risk
emp.var.rate	employment variation rate, with a quarterly frequency
cons.price.idx	monthly average consumer price index
cons.conf.idx	monthly average consumer confidence index
euribor3m	daily three month Euribor rate
nr.employed	quarterly average of the total number of employed citizens
outcome	unsuccessful or successful contact (output target)

predictions, discarding the oldest K, adapting the model to the most recent reality and keeping the process time to fit the model constant by maintaining the number of contacts used to build the model equal to W. The usage of only the most recent window of contacts in each prediction ensures that every modeling is computed with records that reflect a very recent past. In particular, we use a rolling window of size $W = 20000$, with the most recent W training examples at time t, in order to fit a NN and then predict the next $K = 10$ future contact outcomes. In the next iteration, the training examples are updated by discarding the oldest K contacts and incorporating K new real outcome values (available at time $t + K$). Then, a model update is fit and K new predictions are executed. This procedure is repeated, until there is a total of $U = L/K = 130$ model updates (i.e., NN trainings and predictions), where $L = 1293$ is the length of the test set. Using this procedure, the results achieved for the two metrics with the baseline method were 0.794 for AUC and 0.672 for ALIFT.

2.3 Customer Lifetime Value Features

As stated previously, the model built with the 22 input features listed in Table 1 constitutes the baseline model. To enrich it, we listed all possibly LTV characterizing features that can be made available based on the telemarketing history records from previous contacts. Table 2 defines the groups of features analyzed for RFM characteristics. Those are well documented and explored in the literature, thus are natural candidates for incorporating LTV in our decision support system. Only one attribute is used to represent the recency and frequency factors, while two input features are proposed to describe the monetary component. While the RFM concepts are widely known in the literature, authors have slightly different definitions for each of them. Some of the definitions are quoted from the references cited in the second column of Table 2. Besides the RFM related attributes, we also consider seven other LTV features that are described in Table 3. We should note that some of those features depend upon each other, representing variations which we wanted to test if they affected the model.

Table 2 RFM telemarketing features analyzed

Factor	**Reference**	**Citation**	**Application to telemarketing**
recency	[18]	"how recent is the last purchase?"	months since the last purchase up to date (for our case, 2008-2013, we choose months since they are enough for discriminating the records)
	[10]	"time of most recent purchase"	
	[17]	"period since the last purchase"	
	[23]	"the total days between the day of the latest purchase and analysis (days)"	
	[4]	"the period since a customers last purchase"	
frequency	[18]	"how often does a customer buy a product?"	number of times the client subscribed the deposit previously
	[10]	"number of prior purchases"	
	[17]	"number of purchases made within a certain period"	
	[23]	"consuming frequency (times)"	
	[4]	"the number of purchases made within a certain period"	
monetary	[17]	"the money spent during a certain period"	total amount of money the client subscribed in previous contacts
	[23]	"amount of money of total consuming"	
	[4]	"the amount of money that a customer spent during a certain period"	
monetary.successes.avg	[18]	"how much money does the customer spend per order?"	average value subscribed per success (up to date)
	[10]	"average purchase amount per transaction"	

Table 3 Other telemarketing client historical features analyzed

Feature	Description
last.result	last campaign result in which the client was contacted
last.result.prod	similar to last.result but considering only campaigns where the same product was being sold
prev.contacts.durations	total call durations for every previous contacts
prev.contacts.durations.avg	similar to prev.contacts.durations but considering average
total.contacts	total number of contacts
successes.per.contacts	total successes / total contacts
successes.minus.unsuc	total successes minus total unsuccessful contacts

To select which of these features do in fact provide added value to the predictive model, we first group the features into logical blocks. Then, we adopt a feature selection approach using the popular forward selection technique [12]. In each iteration of this forward selection, we add a few features (related with logical blocks, described in Section 3.1) to the original model. If its predictive performance improves (in terms of AUC and ALIFT metrics), the features are kept. Otherwise, these are discarded. Then, a next iteration is executed, in order to test a few more features. The procedure is repeated until all LTV features have been tested.

It is worth to be noticed that while some clients had previously been contacted which allows to provide history information, others did not. Thus, in order to fully evaluate the difference between using or not history information, we compute the metrics of AUC and ALIFT for all clients (overall, total of 1293 contacts in the test set), for clients with history (the ones with previous contacts, 353 of 1293 telemarketing calls) and with no history (950 of 1293 contacts).

3 Experiments and Results

3.1 Customer Lifetime Value Feature Analysis

All experiments described in this article were executed using the rminer package and R tool [6]. The NN ensemble is composed of $N_r = 7$ distinct networks, each trained with 100 epochs of the BFGS algorithm. For setting the number of hidden nodes (H), we performed a grid search where the number of hidden nodes was searched within the set $H \in \{0, 2, 6, 8, 10\}$. The rminer package applies this grid search by performing an internal holdout scheme over the training set (with 2/3 of the data), in order to select the best H value, that corresponds to the lowest AUC value measured on a subset of the training set, and then trains the best model with all training set data.

The rolling windows procedure was executed for the baseline model (with 22 features from Table 1) in order to extract the values of AUC and ALIFT metrics for both groups of contacts (with and without history). The results are shown on the baseline row of Table 4. We found only a very slight difference

(0.0116 in terms of AUC) between the prediction results for clients with and without telemarketing history. For comparison purposes, we also show the results for predictions without using the rolling windows procedure, to assess the benefits of adapting the model iteratively with the most recent data. The difference in the performance measurements of both AUC and ALIFT is quite significatively. Considering this remark, the rolling windows procedure was adopted for the remaining experiments.

Table 4 Results of the forward selection procedure used to select LTV features

Feature	**Overall**		**With history**		**No history**	
	AUC	ALIFT	AUC	ALIFT	AUC	ALIFT
baseline model [20]	0.7935	0.6718	0.8002	0.6701	0.7886	0.6746
baseline (no rolling windows)	0.7443	0.6718	0.7513	0.6444	0.7393	0.6472
recency, frequency, monetary	0.8142	0.6841	0.8382	0.6917	0.7868	0.6731
recency, frequency, **monetary.success.avg**	**0.8170**	**0.6857**	**0.8423**	**0.6940**	0.7886	0.6744
recency, frequency, monetary.success.avg, **last.result**	0.8166	0.6853	0.8398	0.6924	0.7904	0.6760
recency, frequency, monetary.success.avg monetary, **last.result.prod**	**0.8233**	**0.6893**	**0.8480**	**0.6972**	0.7943	0.6779
recency, frequency, monetary.success.avg, last.result, last.result.prod	0.8197	0.6873	0.8423	0.6940	0.7933	0.6773
recency, frequency, monetary.success.avg, last.result.prod, **prev.contacts.durations**	**0.8247**	**0.6901**	**0.8609**	**0.7044**	0.7825	0.6709
recency, frequency, monetary.success.avg, last.result.prod, **prev.contacts.durations.avg**	0.8227	0.6887	0.8553	0.7013	0.7853	0.6725
recency, frequency, monetary.success.avg, last.result.prod, prev.contacts.durations, **total.contacts**	0.8197	0.6871	0.8525	0.6998	0.7814	0.6701
recency, frequency, monetary.success.avg, last.result.prod, prev.contacts.durations, **successes.per.contacts**	0.8193	0.6868	0.8473	0.6966	0.7871	0.6738
recency, frequency, monetary.success.avg, last.result.prod, prev.contacts.durations, **successes.minus.unsuc**	0.8215	0.6882	0.8507	0.6986	0.7874	0.6740

The rows of Table 4 show the effect of adding each new group of features to the initial set of 22 features (from 1). The table is divided by thicker lines according to the logical blocks (after the baseline). The first block includes the standard RFM features, under two configurations (with monetary or monetary.success.avg); next, comes the last campaign result, followed by the duration of phone calls; finally, three isolated features appear, as each tries to add specific telemarketing associated knowledge.

The best results for each logical group of features are signaled through a gray background cell and white digits. As expected, the results of both AUC and ALIFT metrics do not change much for the contacts without history information, when compared with the baseline values (variation below 0.0129 for AUC and 0.0078 for ALIFT). However, for the group of 353 contacts with history information, the metrics are consistently increased as each group of features is added. In effect, AUC increases from 0.8002 to 0.8609, while ALIFT is enhanced from 0.6701 to 0.7044. The exceptions are the last three blocks, which included a single new feature each, and resulted in poorer performances for both metrics, thus being discarded. The selected model got enriched with new LTV features: recency, frequency, monetary value considering average of successful contacts, the result for the last campaign to sell the same product, and the total time spent on previous contacts for past campaigns.

To visually compare the quality of prediction results using client history information versus without history, we plot for both groups the ROC curves [11] and the cumulative Lift curves [22] respectively in Figures 1 and 2.

Considering the ROC curve, Figure 1 shows that the usage of the LTV history features benefits the results particulary for the lower values of FPR, meeting our goal, as stated in Section 2.1. The Lift figure shows that both curves are next to each other, although the largest curve area, which represents the prediction capability for contacts with history information, stands consistently above the baseline (without LTV features). For some client selection sample sizes' (x-axis), such as 60% and 70%, the difference is higher than 4 pp. This in an interesting discovery that directly benefits business, as we used solely telemarketing history information that is easily available at telemarketing service operators. Furthermore, contact center companies can use this type of information to enhance telemarketing campaigns without even having to ask for more information to their clients.

3.2 Explanatory Knowledge

In this section, we extract explanatory knowledge using a sensitivity analysis procedure. First, the best model (with 27 features, including the novel 5 LTV inputs) is fit with all data contacts. Then, the DSA algorithm [7], which is capable of measuring the global influence of an input, including its iterations with other attributes, is executed with its default parameters (e.g. use of 7 levels for numeric inputs) on the best selected NN. The respective DSA input importance bar graphic is plotted in Figure 3.

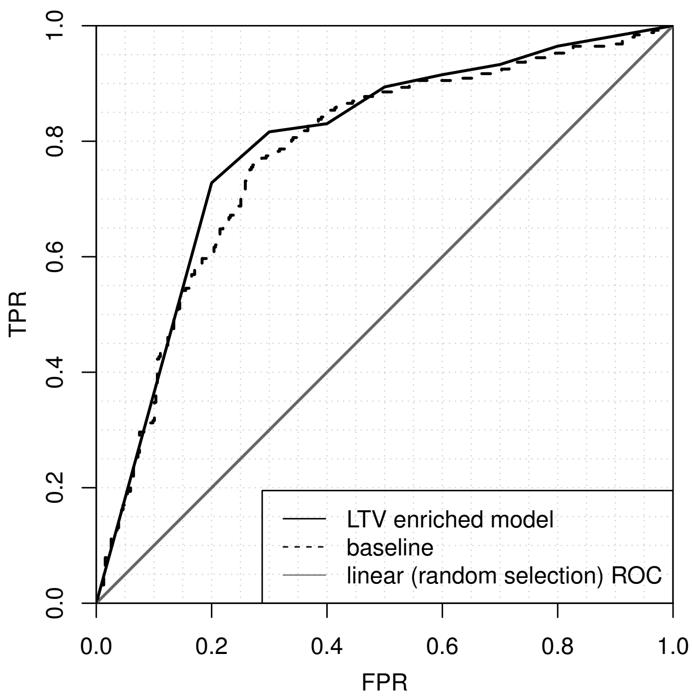

Fig. 1 Bank telemarketing ROC curves (with and without LTV history)

We detail our analysis to the top 5 most relevant features. Figure 4 compares the top 5 input features for the baseline and LTV enhanced model. The difference between the best rate offered for the deposit and the national average is now considered the most relevant feature, while previously it was the fourth in the rank [20]. This is a significant change for the model especially since this feature alone got a relevance higher than 15%, that is, 5 pp increase when compared to the initial baseline model. On the other hand, the euribor rate dropped from a relevance of 15% to slightly above 10%, dropping to second place in the rank. We note that the sensitivity method adopted (DSA) is capable of measuring the global influence of a feature in a predictive model, including its interactions with other features. Given that the main difference between the two predictive models is the inclusion of the new LTV inputs, obtained results suggest that the best predictive model performs a higher degree of interaction between the added LTV features with the difference rate attribute.

More importantly, two of the newly proposed LTV inputs are highly ranked (denoted by a star symbol in Figure 4): the last result for previous campaign to sell the same product (last.result.prod) is ranked third, while frequency of successes is ranked as the fourth most relevant input. From the features added, it is worth to notice that while frequency is the fourth most relevant, recency only comes in eleventh place (Figure 3), with a relevance that is roughly half

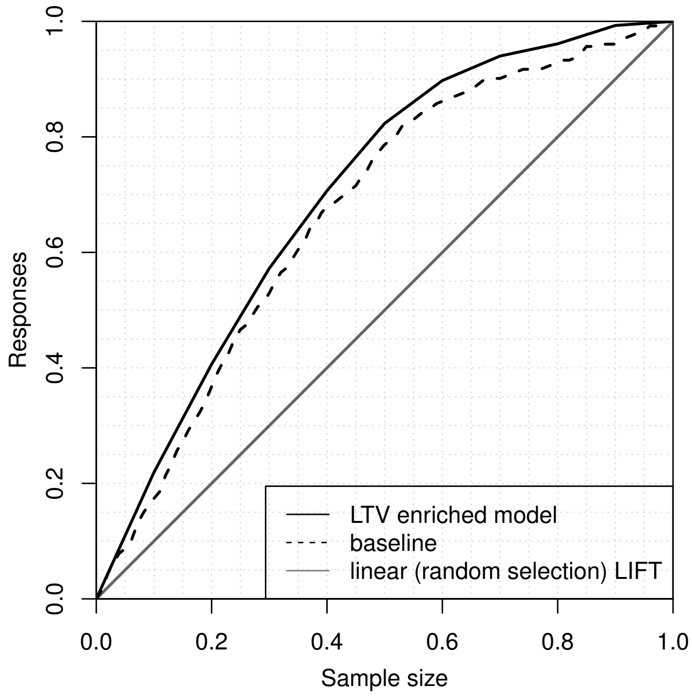

Fig. 2 Bank telemarketing lift cumulative curves (with and without LTV history)

of the one obtained by frequency. This is a different result when compared with the work of Liu and Shih [17], which points to recency as being more relevant than frequency.

Next, we analyze the average influence of the two most relevant LTV features. Figure 5 plots the respective VEC curves, where the x-axis denotes the range of values of the input and y-axis represents the expected average change in the output response. As suggested in Cortez and Embrechts [7], the x-axis is scaled, in order to compare the influence of two distinct inputs in the same graph. The obtained results are aligned with our expectations: if the client has subscribed the deposit in the last campaign through which he/she was contacted, then it is much more likely that he will subscribe it again. Indeed, there is an improvement of around 20 pp in the success probability when the last product result changes from unsuccessful to successful. Considering frequency, an increase in the number of successes also improves the probability for a next successful result. It should be noted that the improvement in the subscription probability is not linear, with the highest increase (around 10 pp) being set between the 0 to 1 frequency interval.

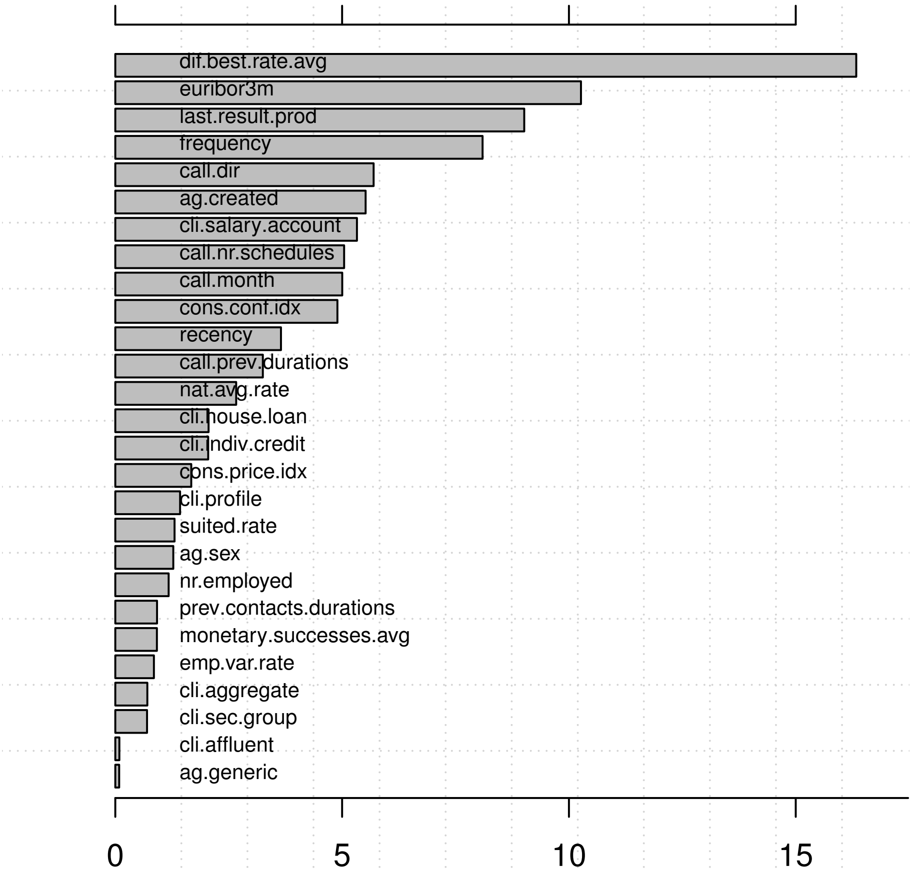

Fig. 3 Relative importance of input features to the data-driven model (in %)

4 Conclusions

In a mature on-going marketing business, usually it is available inside the company raw information that can potentially increase the lifetime value (LTV) of customers. The usage of LTV history information, such as recency, frequency and monetary (RFM) characteristics to enhance data-driven models is thus a key issue to improve prediction accuracy in marketing applications.

In this paper, we applied the concept of LTV by incorporating history information to enhance prediction capabilities of an already robust baseline decision support system using neural networks to sell bank deposits in a telemarketing campaign context. A forward selection technique was conducted, where twelve LTV candidate input features were tested. The evaluation procedure, using a robust and realistic rolling window scheme, and two metrics, favored a data-driven model that included five LTV features. When compared with the baseline model (with no LTV features), the enhanced LTV model produced an improvement of 6 pp in the area of receiver operating characteristic

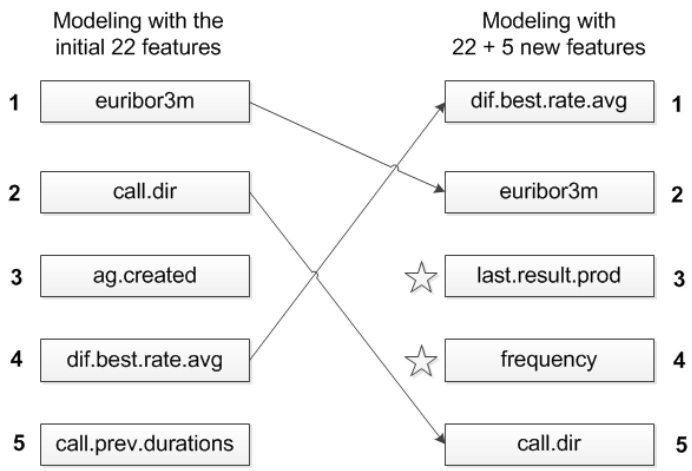

Fig. 4 Feature relevance evaluation (baseline and LTV enhanced models)

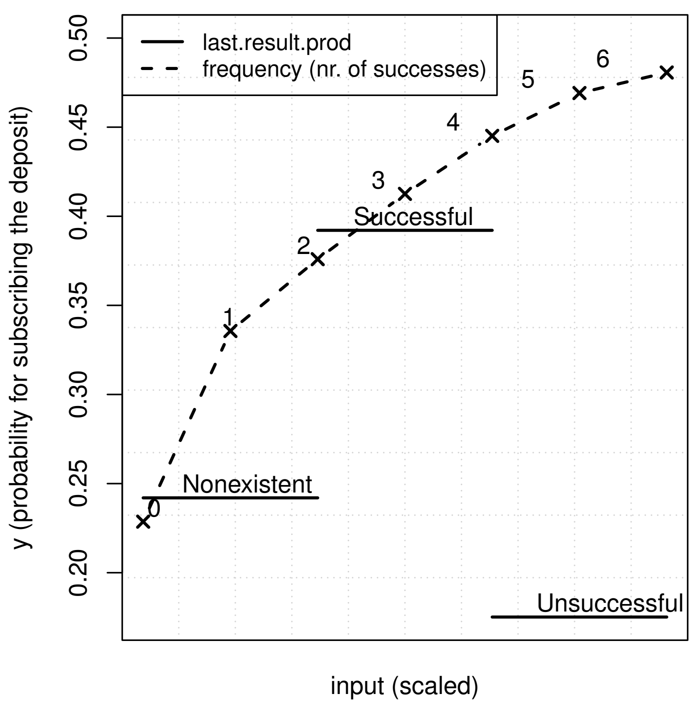

Fig. 5 Influence of the two LTV most relevant features (last.result.prod and frequency)

curve, with a total AUC=0.86, and 4 pp in the cumulative lift curve, with a total ALIFT=0.70, for clients with previous telemarketing history. Moreover, the improvement is perfectly consistent, meaning that whatever is the portion of clients selected for targeting, the estimated number of successes is always better than the baseline system. Furthermore, such LTV history information is easily available at marketing service operators and thus can be immediately used to benefit business (as opposed to external information, which might require requests and time for its acquisition and integration).

We also extracted explanatory knowledge from the LTV enhanced model, by using a sensitivity analysis procedure that allows to rank the inputs and show the global influence of each input in the data-driven model. Two of the newly LTV input variables were included in the top five most relevant features, confirming the utility of the proposed approach, namely the last result for previous campaign to sell the same product and the frequency of successes.

In future research, we intend to gather additional more recent data and perform new tests to address the limitation of using predictions of only 353 contacts. As time goes by and new telemarketing campaigns are executed, one can expect that more clients will be contacted again, allowing to benefit from past history information. It would be interesting to understand if the results achieved are time proof or additional history provides different results in terms of LTV features.

References

1. Hyunchul Ahn, Kyoung-jae Kim, and Ingoo Han. Global optimization of feature weights and the number of neighbors that combine in a case-based reasoning system. *Expert Syst*, 23(5):290–301, 2006.
2. Arash Bahrammirzaee. A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. *Neural Comput Appl*, 19(8):1165–1195, 2010.
3. Phillip R. Burrell and Bukola Otulayo Folarin. The impact of neural networks in finance. *Neural Comput Appl*, 6(4):193–200, 1997.
4. Yen-Liang Chen, Mi-Hao Kuo, Shin-Yi Wu, and Kwei Tang. Discovering recency, frequency, and monetary (RFM) sequential patterns from customers purchasing data. *Electron Commer R A*, 8(5):241–251, 2009.
5. Ching-Hsue Cheng and You-Shyang Chen. Classifying the segmentation of customer value via RFM model and RS theory. *Expert Syst Appl*, 36 (3):4176–4184, 2009.
6. P. Cortez. Data mining with neural networks and support vector machines using the r/rminer tool. In *Advances in Data Mining. Applications and Theoretical Aspects*, volume 6171, pages 572–583. Springer, 2010.
7. Paulo Cortez and Mark J Embrechts. Using sensitivity analysis and visualization techniques to open black box data mining models. *Inform Sciences*, 225:1–17, 2013. doi: 10.1016/j.ins.2012.10.039.
8. P. Domingos. A few useful things to know about machine learning. *Commun ACM*, 55(10):78–87, 2012.
9. F. Robert Dwyer. Customer lifetime valuation to support marketing decision making. *J Interact Mark*, 11(4):6–13, 1997.

10. Peter S Fader, Bruce GS Hardie, and Ka Lok Lee. RFM and CLV: Using iso-value curves for customer base analysis. *J Marketing Res*, pages 415–430, 2005.
11. T. Fawcett. An introduction to ROC analysis. *Pattern Recogn Lett*, 27 (8):861–874, 2006. doi: 10.1016/j.patrec.2005.10.010.
12. I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *J Mach Learn Res*, 3:1157–1182, 2003. doi: 10.1016/j.ejor.2010.10.019.
13. T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer-Verlag, NY, USA, 2nd edition, 2008.
14. Simon S Haykin, Simon S Haykin, Simon S Haykin, and Simon S Haykin. *Neural networks and learning machines*, volume 3. Prentice Hall New York, 2009.
15. A Keramati, S Nazari-Shirkouhi, H Moshki, M Afshari-Mofrad, and E Maleki-Berneti. A novel methodology for evaluating the risk of CRM projects in fuzzy environment. *Neural Comput Appl*, 23(1):29–53, 2013.
16. Ohbyung Kwon and Namyeon Lee. A relationship-aware methodology for context-aware service selection. *Expert Syst*, 28(4):375–390, 2011.
17. Duen-Ren Liu and Ya-Yueh Shih. Integrating AHP and data mining for product recommendation based on customer lifetime value. *Inform Manage*, 42(3):387–400, 2005.
18. Sara Madeira and João M Sousa. Comparison of target selection methods in direct marketing. In *European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems.* Citeseer, 2002.
19. Edward C Malthouse and Robert C Blattberg. Can we predict customer lifetime value? *J Interact Mark*, 19(1):2–16, 2005.
20. Sérgio Moro, Paulo Cortez, and Paulo Rita. A data-driven approach to predict the success of bank telemarketing. *Decis Support Syst*, 62:22–31, 2014.
21. Jeffrey S. Nevid. Introduction to the special issue: Implicit measures of consumer responsethe search for the holy grail of marketing research. *Psychol Market*, 27(10):913–920, 2010. ISSN 1520-6793. doi: 10.1002/mar.20365. URL `http://dx.doi.org/10.1002/mar.20365`.
22. D.L. Olson and B.K. Chae. Direct marketing decision support through predictive customer response modeling. *Decis Support Syst*, 54(1):443–451, 2012.
23. Chia-Cheng Shen and Huan-Ming Chuang. A study on the applications of data mining techniques to enhance customer lifetime value. *WSEAS Transactions on Information Science and Applications*, 6(2):319–328, 2009.
24. Leonard J Tashman. Out-of-sample tests of forecasting accuracy: an analysis and review. *Int J Forecasting*, 16(4):437–450, 2000.
25. Desheng Dash Wu and Jon G Hall. Special issue: business decision support systems. *Expert Syst*, 28(3):197–198, 2011.

26. Zhiyuan Yao, Peter Sarlin, Tomas Eklund, and Barbro Back. Combining visual customer segmentation and response modeling. *Neural Comput Appl*, pages 1–12, 2012.
27. I Yeh, King-Jang Yang, Tao-Ming Ting, et al. Knowledge discovery on RFM model using bernoulli sequence. *Expert Syst Appl*, 36(3):5866–5871, 2009.