Daniel Lesser

Marketing Analytics Final: Portuguese Bank Case Memo to Marketing Director

Many European banks, including our own, are under pressure to increase financial assets due to market competition and the recent financial crisis. At the same time, it is important to reduce costs and maintain a lean marketing strategy. Rather than use mass marketing campaigns which are costly and have low success rates, our bank has used direct telemarketing in recent years to increase customer deposits. This memo details why a data-driven approach can help better predict which customers are more likely to subscribe, increase the effectiveness of future campaigns, and ultimately increase profits.

Previous campaigns have seen 8% of contacted customers convert. In order to better understand our customer base, a few descriptive statistics are likely to be helpful. Previous marketing appears to have been more effective among younger customers, students, retirees, as well as customers who had longer calls with our marketing team. Based on this, it may be worthwhile targeting individuals who have more time on their hands and are available to listen to our offers. One caveat is that a long call duration is likely the result of campaign success rather than the other way around and would not be known before calling, so that feature will be removed for all further analyses.

In the next stage, we will apply more rigorous data mining methodologies to see whether these findings may help us predict which customers are more likely to respond positively to marketing campaigns. For each model, the data from previous marketing campaigns has been divided up into train, validation and test samples. The first sample will be used to train the model, the second sample will help us select the best model, and then the test data will be used at the end as an estimate of a model's effectiveness on a new, 'unseen' set of customers.

The first approach used is Logistic Regression. Logistic Regression is a linear classification method that maps customers to discrete outcomes, such as whether or not they agree to provide more deposits. In this case, it is useful to identify the average contribution that each feature provides to an

individual's likelihood to provide further deposits following a campaign. Several models were tried and while the best performing model was one with additional feature interactions, we nominate a simple 'full' logistic regression model for interpretability reasons. From this model, we learn that some features are strong indicators of campaign success and others are strong indicators of campaign failure, as can be seen in Figure 1. The most important positive drivers include success in previous marketing campaigns, certain months of the year, and if the individual is a student or has had post-secondary education. Certain features that indicate an individual is unlikely to respond to a marketing campaign include certain months of the year, if they have a loan of any type or if they are married. This confirms some of our earlier expectations based on exploratory data analysis, while opening new focal points for further consideration.

It is important to review the efficacy of the model. There are many measures of performance, though we feel the most relevant measures will be correctly identifying customers who will provide additional deposits (True Positive Rate/TPR), not mistaking customers for positive when in fact they are negative (False Positive Rate/FPR) and the ability to enhance responses for our target group versus the population average (Lift). We want a TPR as close to 1 as possible and an FPR as close to zero as possible. In our validation sample, our True Positive Rate is 37.1% and our False Positive Rate is 4.6%. This means we will capture about 37% of likely depositors and only capture about 5% erroneous customers in any future marketing campaign using this model. The lift was 4.17.

Another potential modeling approach is a decision tree. A decision tree seeks to successively split the data into smaller chunks that distill the different types of customers into individual leaves. It does this by identifying the feature at each split that provides the biggest increase in 'purity'. The results of the decision tree model were worse than that of logistic regression, with a TPR of 18.0%, a FPR of 1.3% and lift of 1. Due to the stronger performance and simplicity of logistic regression, we will be using that model for our marketing suggestions.

Previous marketing campaigns have centered around calling up individual customers and trying to convince them to sign up for attractive long-term deposit subscriptions at favorable rates. One alternative approach could be to do a mass-market advertising initiative on television, which would reach non-customers as well. However, success rates for these types of campaigns have diminished in recent times[1] and thus we do not suggest this approach. Instead, we propose a targeted marketing campaign during peak months of the year. Based on the results of the logistic regression model, the end of each quarter appears to be a good time to convince customers to provide additional deposits. In addition, we should focus on current and former students as well as retirees as they have the highest likelihood to convert. Those with financial obligations (loans) should be avoided. A traditional telemarketing campaign can be supplemented with mailings and emails at a lower cost to the bank.

Based on our test dataset of 4,489 customers, we would successfully target 201 (4%) of customers and incorrectly target 181 customers (4%). While this would mean we missed out on 8% of customers who would have responded positively to the promotion, we will save marketing costs on the other 84% of customers. Assuming each telemarketing agent is paid $15/hour, we know that each call should cost about $1.10 based on average call duration. We will also assume that each deposit received will earn the bank $10 in new revenue. In the old model, there was a profit of $532 for this sample (Figure 2). Using our new approach, the profit would be $1,590 (Figure 3). Should the actual costs and profits per customer be different, new figures can be plugged in and the outcome recalculated.

Based on our findings, a data-driven approach can be used to improve the results of future marketing initiatives. A logistic regression model helped identify likely depositors, reduce marketing costs, and increase overall profit.

---

[1] Ou, C., Liu, C., Huang, J. and Zhong, N. 2003. "On Data Mining for Direct Marketing". In *Proceedings of the 9th RSFDGrC conference*, 2639, 491–498.
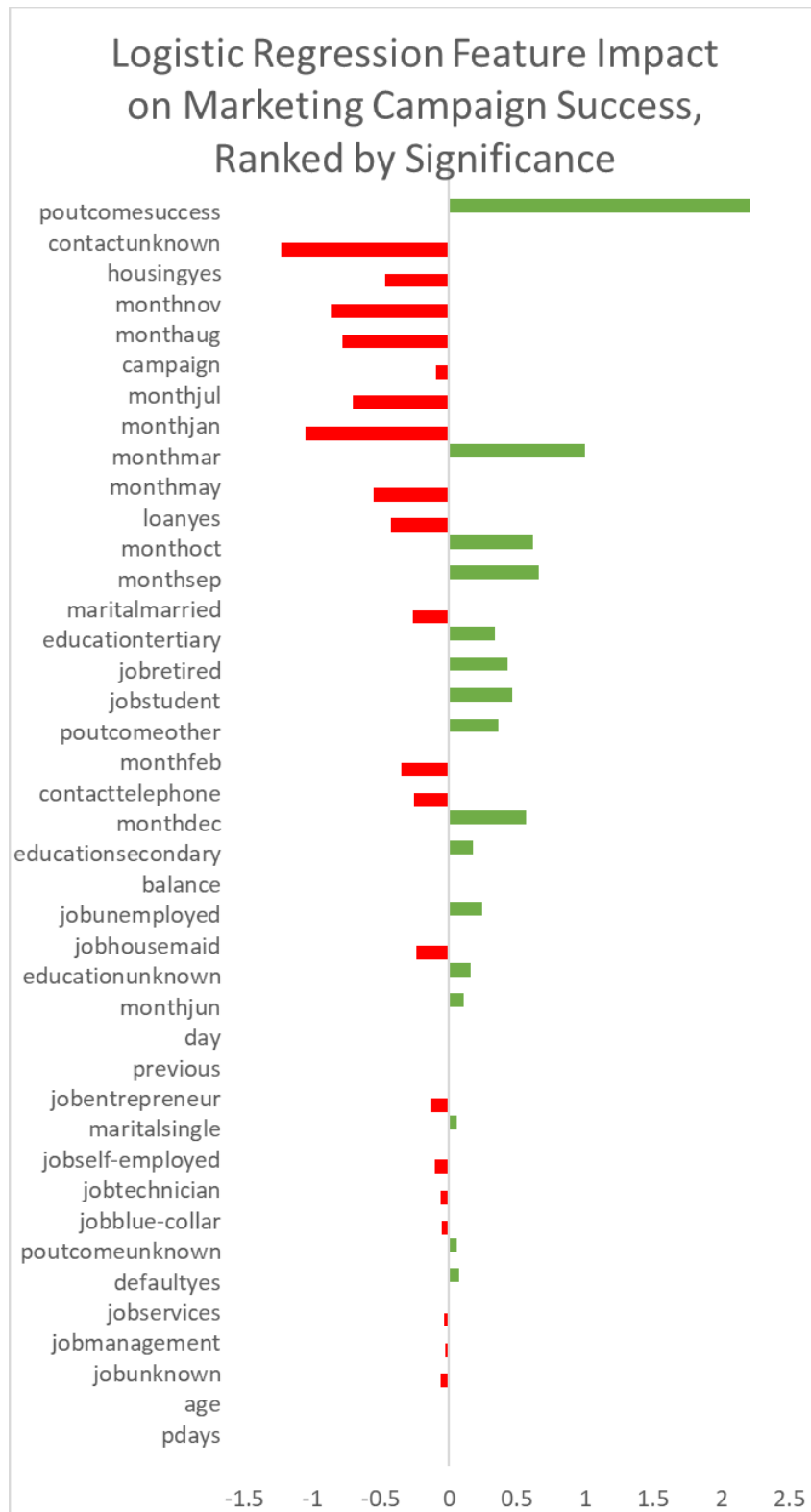
Appendix

Figure 1

Original Marketing Approach: Target Everyone		Actual Results	
		Not Adopter	Adopter
Predicted Results	Not Adopter	$ (4,137)	$ 3,079
	Adopter	$ (199)	$ 1,789

Figure 2

Original Marketing Approach: Target Everyone		Actual Results	
		Not Adopter	Adopter
Predicted Results	Not Adopter	$ -	$ -
	Adopter	$ (199)	$ 1,789

Figure 3