Expected Maximum Profit for Customer Acquisition

S. Rahman, M. Bogaert, S. Muylle

March 2023

1 Disclaimer

This document contains ongoing research and has not been peer-reviewed yet. The contents of this research are subject-to-change and might not reflect the final outcome.

2 Introduction

Within Customer Relationship Management (CRM) literature, researchers have explored employing predictive modelling techniques to support the full customer life cycle (De Bock and De Caigny, 2021; Meire et al., 2017; Raskutti and Herschtal, 2005). These techniques range from acquisition, cross- and up-selling, churn to win-back strategies. This paper will focus on customer acquisition. Winning new customers is a challenging process which can take up a lot of sales representatives' time. Time which could be used for other value-adding tasks. By automating parts of the lead prospecting process, companies are able to more effectively and efficiently spend their scarce resources. Customer acquisition modelling aims to predict the conversion of leads to customers, which aids sales representatives in their decision process. These models outperform approaches which rely on expert knowledge to qualify the best leads (Meire et al., 2017).

3 Related work

Traditionally, classification models within CRM literature are classified using statistical performance measures (e.g., Area Under the Receiver Operator Curve (AUC), top-decile lift), which may or may not correlate with the business objectives. Verbraken et al. (2012) proposed a new performance measure to evaluate classification models by the profit they could potentially generate for the business. The authors define the average classification profit of a classifier with threshold t as:

$$P(t; b_0, c_0, b_1, c_1) = b_0\pi_0 F_0(t) + b_0\pi_1(1 - F_1(t)) - c_0\pi_0(1 - F_0(t)) - c_1\pi_1 F_1(t)$$

with parameters b_0, c_0, b_1, and c_1 representing the benefits and costs associated with correct and incorrect classifications specified in Table 1. $F_0(t)$ and $F_1(t)$ represent the cumulative score distributions of the classifiers for the positive and negative class, respectively. This function can be optimized by finding the optimal threshold T which maximizes the profit of the classifier. This optimum, referred to as the maximum profit MP, can be used to discriminate between classifiers based on the profit they can generate for the business. Analytically the MP can be expressed as:

$$\text{MP} = \max_{\forall t} P(t; b_0, c_0, b_1, c_1) = P(T; b_0, c_0, b_1, c_1)$$

At the optimal threshold the following identity hold true which can be derived from the first order condition of optimality:

$$\frac{f_0(T)}{f_1(T)} = \frac{\pi_1(b_1 + c_1)}{\pi_0(b_0 + c_0)} = \frac{\pi_1\theta}{\pi_0}$$

		Classified into	
		Class 0	Class 1
Belongs to	Class 0	$c(0\|0) = b_0$	$c(1\|0) = c_0$
	Class 1	$c(0\|1) = c_1$	$c(1\|1) = b_1$

Table 1: Confusion matrix containing the cost and benefits associated to different actions

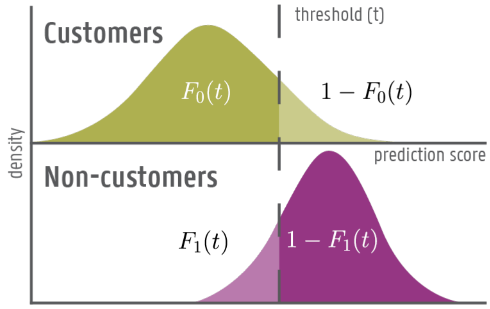

Figure 1: Visual representation of the score distributions of a classifier

with θ the cost-benefit ratio.

One limitation of this deterministic formulation is that in many cases the parameters b_0, c_0, b_1, and c_1 cannot be easily estimated. To capture this uncertainty, it is possible to define a probability distribution over these parameters and consider the *expected* maximum profit (EMP) of a classifier:

$$\int_{b_0}\int_{c_0}\int_{b_1}\int_{c_1} P(T; b_0, c_0, b_1, c_1) \cdot w(b_0, c_0, b_1, c_1)\, \mathrm{d}b_0 \mathrm{d}c_0 \mathrm{d}b_1 \mathrm{d}c_1$$

with $w(b_0, c_0, b_1, c_1)$ the joint probability distribution over the classification costs and benefits.

Over the past years, specific implementations of the EMP measure have been developed for the domains of customer churn (Verbeke et al., 2012; Verbraken et al., 2012; Janssens et al., 2022) and credit scoring (Verbraken et al., 2014). However, for customer acquisition an EMP framework has not been applied. This research aims to develop an appropriate profit function which can suitably capture the cost and benefits associated with customer acquisition.

4 EMP framework for customer acquisition

4.1 Marketing channels

Companies deploy a variety of marketing channels to reach their end customers. Any of which can be arbitrarily complex containing several actors and moving parts. Generally, marketing strategies are split up into three types: direct, indirect and mixed channels (Best, 2013). Direct channels allow companies to have vast control over their operations and the quality of the interface with the end-customer. Being the only actor, this strategy allows companies to capture most if not all value, but the companies bear the entirety of the costs, risks and responsibilities. Indirect channels employ one or multiple intermediaries to distribute the goods or services to the end customer. This results in lower out-of-pocket marketing and sales expenses, but the company partially relinquishes control and concedes a share of the value to intermediaries. Mixed channel approaches form a middle-ground, combining both strategies. Furthermore, direct and indirect channel strategies can be seen as special cases of the mixed channel strategy. Therefore, the next section will formulate a profit model for the mixed channel strategy, allowing flexibility to also model direct and indirect marketing channels.

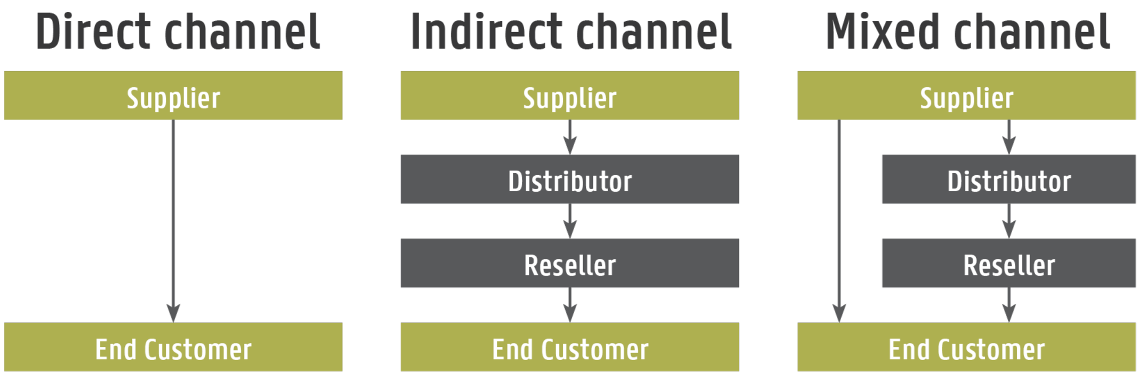

Figure 2: Visual representation marketing channels

4.2 The maximum profit measure for multi-channel customer acquisition

Figure 3 schematically displays the profit model for customer acquisition. Note that that the positive class is denoted with 0 (i.e., converting to a customer). A fraction of the leads classified as potential customers $(\pi_0 F_0(t) + \pi_1 F_1(t))$ are approached by the company (ρ) or an intermediary $(1-\rho)$. For acquired leads, the company receives the Customer Lifetime Value (CLV) minus the cost of converting that lead (Note that here the CLV does not contain acquisition costs, allowing a more nuanced distinction between direct and indirect channels, but deviating from the classic definition of CLV). The cost of converting a lead depends on whether they are targeted directly (s) or targeted indirectly (r). The fraction of leads falsely classified as customers $(\pi_1 F_1(t))$ eventually exit the sales funnel, accruing a fraction (γ) of the conversion cost s when targeted directly or r when targeted indirectly. For the latter, it is assumed that the administrative costs before referring to an intermediary do not change. Leads classified as non-customers $(\pi_0(1-F_0(t)+\pi_1(1-F_1(t)))$ have no impact on the profit of the acquisition campaign.

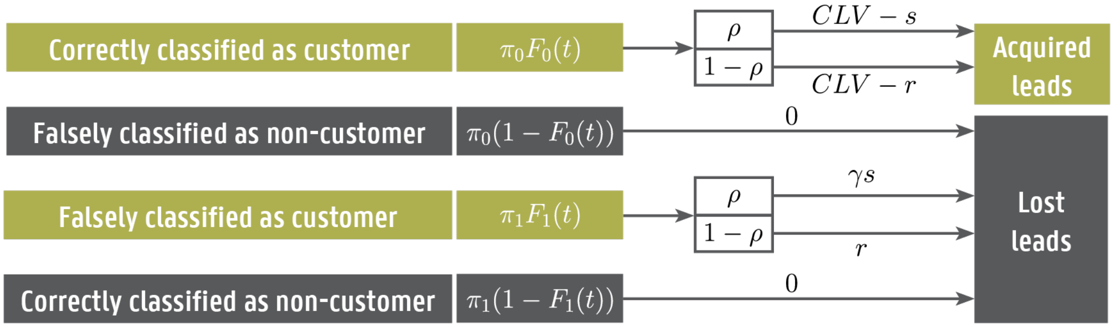

Figure 3: Schematic representation of customer acquisition profit model

Combining the costs and benefits defined above results in a profit function P_m for mixed-channel customer acquisition:

$$\begin{aligned} P_m(t;\rho,CLV,s,r,\gamma) &= [\rho(CLV-s)+(1-\rho)(CLV-r)]\pi_0 F_0(t) - [\rho\gamma s+(1-\rho)r]\pi_1 F_1(t) \\ &= [\rho(r-s)+CLV-r]\pi_0 F_0(t) - [\rho(\gamma s-r)+r]\pi_1 F_1(t) \end{aligned}$$

The profit model simplifies to a direct channel strategy P_d when all leads are handled by the company $(\rho = 1)$:

$$P_d(t;CLV,s,\gamma) = (CLV-s)\pi_0 F_0(t) - \gamma s \pi_1 F_1(t)$$

and to an indirect channel approach when all leads are handled by intermediaries $(\rho = 0)$:

$$P_i(t;CLV,r) = (CLV-r)\pi_0 F_0(t) - r\pi_1 F_1(t)$$

Hence, this specification of the profit function is flexible enough to model a variety of marketing channel strategies.

Analogously to the MP measure, the maximum profit for mixed-channel customer acquisition can be defined as:

$$\text{MPA} = \max_{\forall t} P_m(t;\rho,CLV,s,r,\gamma) = P_m(T;\rho,CLV,s,r,\gamma)$$

4.3 The expected maximum profit measure for mixed-channel customer acquisition

In the previous section, it was assumed that all parameters of the MP were deterministic. This only works if accurate estimates of the parameters can be computed. This is a reasonable assumption for fraction of the customers handled by the company ρ, internally handled lead conversion cost s, and externally handled lead conversion cost r. However, in acquisition the CLV of a lead is highly uncertain (Reinartz and Kumar, 2003) and so is the course that a bad lead takes throughout the sales funnel before dropping out (Duncan and Elkan, 2015), which manifests itself in γ. Hence, to support this uncertainty, the expected maximum profit measure for mixed-channel acquisition modelling EMPA is proposed:

$$\text{EMPA} = \int_{\gamma}\int_{CLV} P_m(T) \cdot h(\gamma) w(CLV)\, dCLV\, d\gamma$$

With $CLV \sim Gamma(\alpha_1, \beta_1)$ and $\gamma \sim Beta(\alpha_2, \beta_2)$. Both the beta and gamma distributions allow for enough flexibility to correctly model the uncertainty of CLV and γ. For simplicity, it is assumed that these stochastic variables are independent.

At the optimal threshold T the cost-benefit ratio can be defined as:

$$\theta_m = \frac{\rho(\gamma s - r) + r}{\rho(r - s) + CLV - r}$$

5 Methodology

The data is divided into a cross-validation and test set using a 60-40 split, stratified onto the distribution of converted leads. 2x5 stratified K-fold is used to tune the models' hyperparameters and to perform feature selection through Recursive Feature Elimination. The hyperparameter search is performed by the Optuna python package (Akiba et al., 2019). Continuous features are scaled by their mean and standard deviation, categorical features with more than two categories are encoded through the Weight of Evidence (WoE) (Smith et al., 2002).

The dependent variable is determined by whether identified leads in 2021 converted in 2022. The data contains socio-demographic data about the leads, and information discovered from early screening. Parts of the data are purchased from a third-party data provider.

The cost and benefit parameters of the profit function are estimated through available data (e.g., CLV distribution parameters estimated through Maximum Likelihood Estimation on the purchases of existing customers) and by discussing with domain experts.

6 Preliminary results

Figure 4 displays the EMPA on the test set of models chosen by the highest AUC and EMPA on the cross-validation set. Preliminary results suggest that models chosen by EMPA generally outperform models chosen by AUC on a model-by-model basis. Additionally, the KNN model which results on the highest EMPA on the test set, is considered one of the worst choices when basing your decision on AUC. This indicates that model selection through AUC might lead to poor decisions for optimizing the business objectives. Interpretations should be cautious since these results are obtained on a single data set and might not generalize. This issue will be addressed in future research.

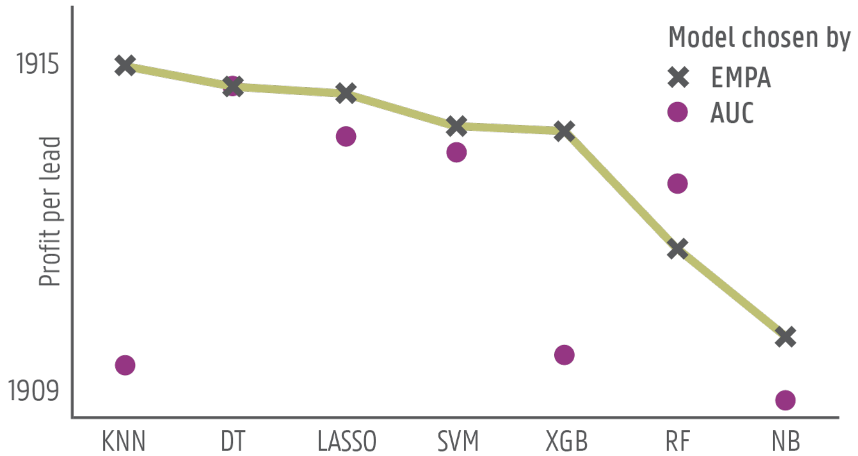

Figure 4: model benchmark comparing models selected by AUC and EMPA

7 Future research

Currently, the benchmarks are run based on one proprietary data set. However in the future the EMPA measure will be measure against additional 8 open-source data sets to increase the benchmark's robustness. These data sets cover a wider range of marketing channel strategies and parameter values. Open-source data sets do bring their own challenges, for instance, incomplete information about the cost and benefits parameters. In these cases careful assumptions need to be made about the value of these parameters which reflect reality.

Future research will include a wider range of benchmarked machine learning models and reported performance measures to give a better overview of the performance of the EMPA.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2623–2631.

Best, R. J. (2013). *Market-based management: Strategies for growing customer value and profitability.* Pearson.

De Bock, K. W., & De Caigny, A. (2021). Spline-rule ensemble classifiers with structured sparsity regularization for interpretable customer churn modeling. *Decision Support Systems*, *150*, 113523.

Duncan, B. A., & Elkan, C. P. (2015). Probabilistic modeling of a sales funnel to prioritize leads. *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 1751–1758.

Janssens, B., Bogaert, M., Bagué, A., & Van den Poel, D. (2022). B2boost: Instance-dependent profit-driven modelling of b2b churn. *Annals of Operations Research*, 1–27.

Meire, M., Ballings, M., & Van den Poel, D. (2017). The added value of social media data in b2b customer acquisition systems: A real-life experiment. *Decision Support Systems*, *104*, 26–37.

Raskutti, B., & Herschtal, A. (2005). Predicting the product purchase patterns of corporate customers. *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, 469–478.

Reinartz, W. J., & Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of marketing*, *67*(1), 77–99.

Smith, E. P., Lipkovich, I., & Ye, K. (2002). Weight-of-evidence (woe): Quantitative estimation of probability of impairment for individual and multiple lines of evidence. *Human and Ecological Risk Assessment*, *8*(7), 1585–1596.

Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European journal of operational research*, *218*(1), 211–229.

Verbraken, T., Bravo, C., Weber, R., & Baesens, B. (2014). Development and application of consumer credit scoring models using profit-based classification measures. *European Journal of Operational Research*, *238*(2), 505–513.

Verbraken, T., Verbeke, W., & Baesens, B. (2012). A novel profit maximizing metric for measuring classification performance of customer churn prediction models. *IEEE transactions on knowledge and data engineering*, *25*(5), 961–973.