**Bank Direct Marketing Campaign Prediction in Machine Learning & Causal Inferences**

**Qianhui Li**

**Introduction**

Nowadays, marketing has become an integral part of banks to promote their goods and services. However, banks always have limited visibility into their customers, which confines their ability to get the best out of the marketing business. The bank is an exemplary sector in which the promotional activity is highly competitive. The 2007-2009 global economic crisis increased competition among banks for deposit retention and marketing campaigns, due to credit restrictions pressure on banks (Moro, Cortez & Rita, 2012). Therefore, there is a need for efficient marketing campaigns with lesser contacts but keep the number of clients subscribing the deposit to some extent. Additionally, it is essential to identify the deterministic factor that has causal effect on customers’ responses.

There are two common approaches to promotions, which are mass marketing and direct marketing. In mass marketing, banks do not need to build direct relationships with customers. Instead, they broadcast their promotional message through television, radio, and newspapers. Due to the high ineffectiveness in mass marketing, banks are shifting to direct marketing. Direct marketing selects targeted customers and focuses more on customers’ specific needs for specific product and service offers (Elsalamony, 2014).

**Literature Review**

The machine learning method can be applied to direct marketing, which utilizes customers’ historical purchasing data and predictive models to measure whether a customer will respond to an offer or not (Sing’oei & Wang, 2013). It compensates the increasing costs in marketing promotion and decreasing customer response rates. Besides, although direct marketing, such as telemarketing is an interactive and powerful tool, it annoys customers sometimes (Vajiramedhin & Suebsing, 2014). The machine learning prediction can eliminate this problem by extracting knowledge from raw data and predicting customers’ responses more accurately.

The machine learning methods have rising popularity for prediction in bank direct marketing. Morgan & Sonquist (1963) first introduced classification trees in their work, and it gained popularities in marketing analytics. Grzonka, Suchacka, & Borowik (2016) reviewed tree-based classification methods, and the best predictive result was obtained from random forests. Also, Miguéis et al. (2017) used random forests to predict customers’ response to direct marketing campaigns. Besides, Artificial Neural Networks (ANNs) is wildly accepted in prediction and classifications. The previous study showed that neural networks has its advantage of allowing identifying links among factors, and not be based on “a priori” assumption (Bishop, 2005). In the research conducted by Ali & Özgür (2013), they compared the analytical results between ANN and logistic regression (LR), and it showed those two algorithms achieved the identical accuracy, but ANN ran faster than LR. Furthermore, Moro, Laureano & Cortez (2012) suggested in their study that the Support Vector Machine can achieve high predictive performances compared with Naïve Bayes and Decision Trees.

**Objective**

In my analysis, I plan to make use of Logistic Regression (LR), Lasso, Decision Tree (DT), and Neural Network (NN) to predict customers’ responses. I choose Lasso and NN, because they were seldom used in previous studies, and I would like to try something new. For LR and DT, they are popular in previous researches, and I would like to compare them with Lasso and NN. I proposed to compare the accuracy of those techniques and find the best model for predicting customers’ responses. Moreover, I would like to explore the causal inference and identify the deterministic covariant that has a causal effect on customers’ decision.

General comments:

The topic looks good! You may want to provide some details on why direct marketing becomes popular to banks now but not previously; why mass marketing becomes problematic now but not previously (or maybe it is never effective?). I can sense somewhat between the line how your study can be useful, but it is not explicitly stated. You may want to make a claim on who can use your study, and how they can benefit. I think if you provide enough details according to my two comments above, you may have an answer to the last question. Overall, much improved!

**References(Haven’t change style, incomplete…)**

Moro, Sérgio & Cortez, Paulo & Rita, Paulo. (2014). A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems. 62. 10.1016/j.dss.2014.03.001.

Morgan J. N., Sonquist J. A. (1963) Problems in the analysis of survey data, and a proposal, Journal of the American Statistical Association, Vol. 58, Issue 302, 415-434.

Elsalamony, H. (2013). Bank Direct Marketing Analysis of Data Mining Techniques. International Journal of Computer Applications (0975 – 8887), Volume 85, No 7

C. M. Bishop, Neural networks for pattern recognition, Oxford University press, New York, 2005.

S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, *Elsevier, 62*, 22-31, June 2014

Sing’oei L. & Wang J. (2013). Data Mining Framework for Direct Marketing: A Case Study of Bank Marketing. IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 2, March 2013

Vajiramedhin C. & Suebsing A. (2014). Feature Selection with Data Balancing for Prediction of Bank Telemarketing. Applied Mathematical Sciences, Vol. 8, 2014, no. 114, 5667 – 5672. Retrived from <http://dx.doi.org/10.12988/ams.2014.47222>

Grzonka, D., Suchacka, G. & Borowik, B. (2016). Application of Selected Supervised Classification Methods to Bank Marketing Campaign. Information Systems in Management. 5. 36-48.

Miguéis, V.L. & Camanho, A.s & Borges, José. (2017). Predicting direct marketing response in banking: comparison of class imbalance methods. Service Business, 11:831–849, DOI 10.1007/s11628-016-0332-3

Koç, A.A. & Yeniay, Özgür. (2013). A Comparative Study of Artificial Neural Networks and Logistic Regression for Classification of Marketing Campaign Results. Mathematical and Computational Applications, Vol. 18, No. 3, pp. 392-398

Moro, S., Laureano, R., & Cortez, P. (2012). Enhancing Bank Direct Marketing through Data Mining. 41th EMAC Conference Marketing to Citizens Going beyond Customers and Consumers. ISBN: 978-989-732-004-0