**Predicting Customer Banking behavior from Campaigning**

**Raghu Raman Nanduri**

**DSC680 – Fall 2019**

<https://github.com/iamnrr/DSC680-Projects>

**Domain:**

I have picked banking domain data set for this project. I have worked for several top banks in US and it has always intrigued me, how they choose their prospective customers to cross – sell other banking products. So, I have picked a dataset from a banking campaign to predict whether the customer would subscribe to another product that is being marketed or not.

**Dataset:**

Source of this analysis can be found at the below link.

<https://www.openml.org/d/1461>

The data set consists of input variables – mainly demographic information like – age, education, martial status, some banking information like existing bank account balance, bank loans (if any) and some campaigning information like duration, date and month of the campaign call, along with some contact information. The target variable is client subscription, that indicates whether client subscribed to term deposit or not.

**Analysis & Research Proposal:**

The first rule of any business is to retain the existing customers rather than getting new customers. So, most companies target customer retention over customer acquisition. I want to analyze the success rate in customer retention through direct campaign marketing by a Portuguese Bank for which the data set pertains to. I mainly want to find out what parameters define the outcome of the direct campaign and what parameters are a must for a positive outcome leading to a subscription. The data set concentrates on only phone calls, so my research includes the effectiveness of the direct marketing campaign. The analysis includes feature engineering and feature selection as well to select features that define this success rate.

**Methods & Algorithms:**

The problem at hand is a binary classification, so I would like to consider classification algorithm. But, instead of selecting individual algorithm, I want to compare and contrast different algorithms like – logistic regression, Naïve Bayes, Support Vector Machines, Decision Tress. This gives me an ample opportunity to look at all classification algorithms again through this project If time permits, I would like to test the ensemble algorithms like Random Forest classifier as well.

**Challenges:**

The biggest challenge I worry about is the cleanliness of the data, which I did not get chance to analyze so far. Another challenge is in finding supplementary data sets to go with this data set. So far, I could not find any. But the data set as is, currently has 16 input variables and one target variable; so, I assume it is good enough for this analysis.

**Remarks:**

Through this analysis, I want to analyze the effectiveness of cross selling of the banking products through direct marketing campaign. In the age of digital marketing, it would give us some perspective on the telephonic marketing. Looking forward to identify the key factors for such a successful campaign.

**References:**

1) <https://www.openml.org/d/1461>

2) <https://thefinancialbrand.com/40264/direct-marketing-trends-in-banking/>

3) Chris Albon, April 2018, “Python Machine Learning Cookbook”, ISBN 9781491989388