Analysis of Marketing Campaigns for Banking

DATA 5000 Project Proposal

Norbert Eke
School of Computer Science
Carleton University
Ottawa, ON
norbert.eke@carleton.ca

Olivia Faria
School of Journalism and Communications
Carleton University
Ottawa, ON
olivia.faria@carleton.ca

ABSTRACT

This project proposes to use a variety of statistical methods including clustering of bank telemarketing customer data to achieve the following two tasks: 1) Classification: to predict if a client will subscribe to a term deposit based on the bank's telemarketing customer data and 2) Prediction: to potentially identify a target demographic for future campaigns.

KEYWORDS

Data Mining, Data Analysis, Statistical Machine Learning, Telemarketing, Targeted Advertising, Classification, Clustering, Supervised and Unsupervised Learning, Banking

ACM Reference Format:
Norbert Eke and Olivia Faria. . Analysis of Marketing Campaigns for Banking: DATA 5000 Project Proposal. In *Proceedings of* . ACM, New York, NY, USA, 2 pages.

1 INTRODUCTION

A good marketing strategy is one of the essential tools required for any successful business. In order to develop said marketing strategy, it is critical to determine the audience or demographic for a specific product. Who will this product appeal to? Who is most likely to buy this product? Who is the least likely to buy this product? What will the success rate of a specific advertising medium (commercial, telemarketing campaign, web advertisement) be? All of these questions are important in informing a business' marketing strategy. One of the ways to answer these questions is to examine previous data to help inform the new project's research efforts. In the case of this project, data from a Portuguese national bank's 2008-2013 telemarketing campaign will be analyzed using a variety of methods to try and find the most suitable demographic for future campaigns.

2 MOTIVATION

The motivation for this project is to see if any hidden relationships between the attributes emerge when certain statistical machine learning models are employed and if any particular method may enhance marketing strategy insights in identifying key demographics for advertising.

3 PROBLEM STATEMENT

Previous approaches took a data mining approach with only a few statistical models being used for the data analysis process [2]. The models originally used for analysis were logistic regression, decision trees, neural network and support vector machines [1].

We believe that further analysis is needed in order to successfully identify the key factors of a successful telemarketing strategy and to gain insight into the campaign target demographic. This goal should be reachable by employing a larger variety of both supervised and unsupervised methods and statistical models to gain a more holistic understanding of the data.

4 OBJECTIVES

The main questions we seek to answer in this project are the following:

- Which attributes help to predict if the client will subscribe to a term deposit?

- Which attributes help to predict whether or not someone will pick up the phone?

Additional questions we hope to answer include:

- Which demographic groups respond well/badly to this method of marketing?
- Are there any hidden relationships within the attributes that have not been discovered through the methods previously used?

5 DATA SOURCE

The dataset was made available by researchers Sergio Moro, Raul Laureano and Paulo Cortez via the UCI Machine Learning Repository [1]. The data is based on a series of direct marketing campaigns by a Portuguese bank between 2008-2013. The telemarketing campaign attempted to get clients to sign up for a bank term deposit. The data set contains 20 input variables in the following categories: bank client data, last contact of the current campaign, other attributes and social and economic context attributes. It also includes one output variable (if the client subscribed to a term deposit). It is interesting to note that this data was collected during a recession.

6 METHODOLOGY

The data analysis conducted in this project will be focused on two main tasks:

(1) Clustering of bank telemarketing customer data to identify a target demographic for future campaigns
(2) Predict if a client will subscribe a term deposit (yes/no) based on bank telemarketing customer data

Firstly, proper data cleaning and validation techniques will be applied to the dataset. The dataset does not contain missing values, therefore there is no need to manage missing data.

Secondly, exploratory data visualization will be completed in order to explore and visualize dependencies between features, and even potentially find problems with the data. Visualization will also help tell the story behind the data.

Thirdly, unsupervised learning in the form of clustering techniques will applied to the data in order to find potential groups in the data. This form of exploratory analysis could be helpful for identifying key demographic groups within the customers participating or interacting with the telemarketing campaign.

Some of the unsupervised clustering techniques that will be applied to the data set include:

- Hierarchical clustering
- K-means clustering
- Mixture Model based clustering

Potentially, a dimensionality reduction technique like Principle Component Analysis could also be applied to the dataset in order to gain some insight about the attributes of the dataset by taking a look at the principle component loadings.

Lastly, for the supervised binary classification task of predicting whether or not a client will subscribe to a term deposit the training and testing of the following models is proposed:

- Logistic regression
- naive Bayes Classifier
- k-NN Classifier
- Linear Discriminant and Quadratic Discriminant Classifier
- Random Forrest Classifier
- Boosting Classifier
- Bagging Classifier
- Support Vector Classifier
- LASSO/Ridge Classifier
- Neural Network based Classifier
- Decision Tree based Classifier

The data set will be split into train and test sets according to common best practices in order to evaluate the models.

REFERENCES

[1] Sérgio Moro, Paulo Cortez, and Paulo Rita. 2014. A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems* 62 (2014), 22–31.
[2] Sergio Moro, Raul Laureano, and Paulo Cortez. 2011. Using data mining for bank direct marketing: An application of the crisp-dm methodology. In *Proceedings of European Simulation and Modelling Conference-ESM'2011*. EUROSIS-ETI, 117–121.