

Marketing Campaign Customer Selection

1. Problem Specification

Marketing campaign through phone calls is one of the most widely used strategies to reach out to potential customers and attract business across many industries. These phone-call based marketing campaigns are done by making phone calls to a pool of potential customers to promote certain products. Unfortunately, often it is necessary to contact the same client more than once. The biggest problem of this strategy is that while it takes a lot of time and energy to attract a new customer, more often people will feel uncomfortable about the repetitive phone calls and may even jeopardize the campaigning company's reputation.

The important question here is how to choose the pool of potential customer to increase success rate? If we can increase success rate of phone call marketing campaigns, less resource is required to attract each new customer and it is less likely to put the reputation of the company at risk.

Our approach to the problem is to analyze market campaign data with information about the background of customers and try to determine the specific profile of customers who will be more likely to subscribe. More specifically, our project is to propose a data mining approach to select the best set of clients that are more likely to subscribe a product through telemarketing calls. The classification goal is to predict if the client will subscribe a term deposit (variable y).

2. Description of Dataset

We selected the Bank Marketing dataset from UCI machine learning repository. The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

It is a multivariate dataset with 20 attributes and 45211 instances (dataset named as "bank-additional-full.xlsx" in repository). The introduction of the input variables is listed below.

Input variables (type):

- 1 - age (numeric)
- 2 - job : type of job (categorical)
- 3 - marital : marital status (categorical: 'divorced', 'married', 'single', 'unknown'; note: 'divorced' means divorced or widowed)
- 4 - education (categorical)
- 5 - default: has credit in default? (categorical)
- 6 - housing: has housing loan? (categorical)
- 7 - loan: has personal loan? (categorical)

related with the last contact of the current campaign:

- 8 - contact: contact communication type (categorical: 'cellular','telephone')
- 9 - month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
- 10 - day_of_week: last contact day of the week (categorical: 'mon','tue','wed','thu','fri')
- 11 - duration: last contact duration, in seconds (numeric).

other attributes:

- 12 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
- 13 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
- 14 - previous: number of contacts performed before this campaign and for this client (numeric)
- 15 - poutcome: outcome of the previous marketing campaign (categorical)

social and economic context attributes

- 16 - emp.var.rate: employment variation rate - quarterly indicator (numeric)
- 17 - cons.price.idx: consumer price index - monthly indicator (numeric)
- 18 - cons.conf.idx: consumer confidence index - monthly indicator (numeric)
- 19 - euribor3m: euribor 3 month rate - daily indicator (numeric)
- 20 - nr.employed: number of employees - quarterly indicator (numeric)

