

# Predicting Effectiveness of Bank Marketing

Group 14: Mahima Mago, Ninad Khare, Shreyansi Jain, Subasish Behera

## 1 Introduction

A Portuguese banking institution launched directed marketing campaign to promote their products. These marketing campaigns were based on telephonic calls. It is important for the institution to know whether the campaign is effective in converting clients thus data was collected about the subscription of the product by the clients contacted. In this report, we aim to apply different classification techniques to the data gathered to predict the success of this bank marketing campaign.

## 2 Description of the Dataset

The dataset under study contains 10,000 observations with the following variables:

Table 1: Description of the Bank Marketing Dataset

Variable Name	Description
age	age at the contact date (Numeric)
job	type of job (11 Categories)
marital	marital status (3 categories)
education	education level (7 Categories)
default	has credit in default? ('no', 'yes')
housing	has housing loan? ('no', 'yes')
loan	has personal loan? ('no', 'yes')
contact	contact communication type ('cellular', 'telephone')
month	last contact month of year (10 Categories)
day_of_week	last contact day of the week (5 Categories)
duration	last contact duration, in seconds (Numeric)
campaign	number of contacts performed during this campaign and for this client
pdays	number of days that passed by after the client was last contacted from a previous campaign
previous	number of contacts performed before this campaign and for this client
poutcome	outcome of the previous marketing campaign (3 Categories)
emp.var.rate	employment variation rate – quarterly indicator (Numeric)
cons.price.idx	consumer price index – monthly indicator (Numeric)
cons.conf.idx	consumer confidence index – monthly indicator (Numeric)
euribor3m	euribor (Euro Interbank Offered Rate) 3 month rate – daily indicator (Numeric)
nr.employed	number of employees – quarterly indicator (Numeric)
y	has the client subscribed a term deposit? ('no', 'yes')

*It is important to note that the original dataset gathered by the Portuguese researchers contains over 40,000 observations, however this analysis is based on a subset of the complete dataset containing randomly selected 10,000 observations. Thus, the analysis in the report has less predictive power and accuracy.*

Table 1 gives the details about each variable in our dataset including the description, type of variable and the number of categories present. Note that there are missing values for some variables which will be dealt with in the further sections and have not been mentioned in the table above.

### **3 Exploratory Data Analysis**

Descriptive Analysis has been performed on the dataset.

### **4 Statistical Modelling**

**4.1 Model 1:**

**4.2 Model 2:**

**4.3 Model 3:**

**4.4 Model 4:**

**4.5 Model 5:**

**4.6 Model 6:**

### **5 Results**

### **6 Conclusion**