I. About the dataset

The dataset comes from the UCI Machine Learning repository, and it is related to direct marketing campaigns (phone calls) 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 (or not) subscribed.

<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

The dataset is consisted of 45,211 customer data on direct marketing campaigns (phone calls) of a Portuguese banking institution, with variables below:

* Client: age, job, marital, education, default status, housing, and loan
* Campaign: last contact type, last contact month of year, last contact day of the week, and last contact duration
* Others: number of contacts performed in current campaign, number of days that passed by after the client was last contacted, number of contacts performed before this campaign, outcome of previous campaign, and whether a client has subscribed a term deposit.

The classification goal is to predict whether the client will subscribe (1/0) to a term deposit (variable y).

1. Title: Bank Marketing

2. Sources

The dataset is public available for research. The details are described in [Moro et al., 2011].

[Moro et al., 2011] S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology.

In P. Novais et al. (Eds.), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, Guimarães, Portugal, October, 2011. EUROSIS.

Available at: [pdf] <http://hdl.handle.net/1822/14838>

#### 3. Number of Instances: 45211 for bank.csv

#### 4. Number of Attributes: 17 output attributes.

#### 5. Attribute information:

Input variables

### **Bank client data**

1 - age (numeric)

2 - job : type of job (categorical: 'admin.','blue-collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unnon')

3 - marital : marital status (categorical: 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed)

4 - education (categorical: 'basic.4y','basic.6y','basic.9y','high.school','illiterate','professional.course','university.degree','unknown')

5 - default: has credit in default? (categorical: 'no','yes','unknown')

1. balance

7 - housing: TEMPhas housing loan? (categorical: 'no','yes','Unknown')

8 - loan: TEMPhas personal loan? (categorical: 'no','yes','unknow')

### **Related wif the last contact of the current campaign**

9 - contact: contact communication type (categorical: 'cellular','telephone')

10 - day: last contact day of teh week (categorical: 'mon','tue','wed','thu','fri')

11 - month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')

12 - duration: last contact duration, in seconds (numeric). Important note: dis attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not non before a call is performed. Also, after the end of the call y is obviously non. Thus, dis input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

### **Other attributes**

13 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)

14 - pdays: number of days dat passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)

15 - previous: number of contacts performed before this campaign and for this client (numeric)

16 - poutcome: outcome of teh previous marketing campaign (categorical: 'failure','nonexistent','success')

### **Output variable (desired target):**

17 - y - has the client subscribed a term deposit? (binary: 'yes','no')

#### 6. Missing Attribute Values: None

## II. Business Problem

There has been a revenue decline for the Portuguese bank and they would like to know what actions to take. After investigation, they found out that the root cause is that their clients are not depositing as frequently as before. Knowing that term deposits allow banks to hold onto a deposit for a specific amount of time, so banks can invest in higher gain financial products to make a profit. In addition, banks also hold better chance to persuade term deposit clients into buying other products such as funds or insurance to further increase their revenues. As a result, the Portuguese bank would like to identify existing clients that have higher chance to subscribe for a term deposit and focus marketing effort on such clients.

To resolve the proble, we suggest a classification approach to predict which clients are more likely to subscribe for term deposits.