

CIS 508 Online –Assignment #2

Direct Marketing

Application Area	Data Mining Task	Number of Instances	Number of Attributes
Banking, Business	Classification	45211	17

Abstract:

There are two main approaches to enterprise marketing: (1) mass campaigns, targeting several general customers, or (2) directed marketing, targeting a specific set of customers. In this competitive world, the mass campaign strategy is not very productive. Nevertheless, there are challenges to directed marketing: finding potential customers is not very easy, although data mining (DM) techniques are providing some assistance in that regard.

The given dataset is from a Portuguese banking institution and was used in their direct marketing campaign to sell term deposits to their customers. You can think of it as cross-selling. The campaign was mostly based on phone calls and the dataset stores general information about customers, details of contacts made with them and the output variable y which indicates whether a term deposit was subscribed to by the customer or not. There are two datasets –

1. Portuguese Bank Data – TRAIN, and
2. Portuguese Bank Data - TEST

Attribute Information:

Input variables:

bank client data:

- 1 - **age** (numeric)
- 2 - **job** : type of job (categorical: 'admin.', 'unknown', 'unemployed', 'management', 'housemaid', 'entrepreneur', 'student', 'blue-collar', 'self-employed', 'retired', 'technician', 'services')
- 3 - **marital** : marital status (categorical: 'married', 'divorced', 'single'; note: 'divorced' means divorced or widowed)
- 4 - **education** (categorical: 'unknown', 'secondary', 'primary', 'tertiary')
- 5 - **default**: has credit default? (binary/flag: 'yes', 'no')
- 6 - **balance**: average yearly balance, in euros (numeric)
- 7 - **housing**: has housing loan? (binary/flag: 'yes', 'no')
- 8 - **loan**: has personal loan? (binary/flag: 'yes', 'no')

related with the last contact of the current campaign:

- 9 - **contact**: contact communication type (categorical: 'unknown', 'telephone', 'cellular')
- 10 - **day**: last contact day of the month (numeric)
- 11 - **month**: last contact month of the year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
- 12 - **duration**: last contact duration, in seconds (numeric)

other attributes:

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

14 - **pdays**: number of days since the last contact with the client from a previous marketing campaign (numeric, -1 means client was not previously contacted)

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

16 - **poutcome**: outcome of the previous marketing campaign (categorical: 'unknown', 'other', 'failure', 'success')

Output variable (desired target):

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

Source:

[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.

Links:

Article - <http://archive.ics.uci.edu/ml/datasets/Bank+Marketing>