

Bank Marketing Campaign

Understanding the Problem:

- ABC Bank wants to sell its term deposit product to customers and before launching the product.
- They want to develop a model which helps them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution).
- Bank wants to use ML model to shortlist customer whose chances of buying the product is more so that their marketing channel (tele marketing, SMS/email marketing etc) can focus only to those customers whose chances of buying the product is more.

Understanding the DataSet:

- There are 2 dataset files: bank and bank-additional.
- Bank-additional with all examples (41188) and 20 inputs, ordered by date (from May 2008 to November 2010).

Try of Data:

1. **Input variables:**

- bank client data:

1 - age (*numeric*)

2- job (*categorical*: 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'services', 'student', 'technician', 'unemployed', 'unknown')

3 - marital (*categorical*: 'divorced', 'married', 'single', 'unknown')

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')

6 - housing: has housing loan? (*categorical*: 'no' , 'yes' , 'unknown')

7 - loan: has personal loan? (*categorical*: 'no' , 'yes' , 'unknown')

- related with the last contact of the current campaign:

8 - contact: (*categorical*: 'cellular' , 'telephone')

9 - month: (*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*).

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

- 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*: 'failure' , 'nonexistent' , 'success')

- 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*)

2. **Output variable (desired target):**

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

Problems in the Data:

Approaches: