

## Bank Marketing Case Study

See the dataset at: <http://archive.ics.uci.edu/ml/datasets/Bank+Marketing> for context.

Use file **bank-additional.csv** from Blackboard.

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.

**Task: Classification - predict if the client will subscribe (yes/no) to a term deposit (y).**

### Input variables:

1. age (numeric)
2. job : type of job (categorical: 'admin.','blue-collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown')
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')
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: 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). 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 - You may want to research some of these variables for more context.*

- 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)

*Output variable (desired target):*

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

**Submission: Write a case-study report, including introduction, lit. review, methods, data exploration, data cleaning, data analysis, discussion of results, interpretation and conclusion. See Case Study Guidelines for details. Please put all codes in appendix or upload as a separate file.**