Bank Marketing
Donated on 2/13/2012

The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variab...

Dataset Characteristics	**Subject Area**	**Associated Tasks**
Multivariate	Business	Classification

Feature Type	**# Instances**	**# Features**
Categorical, Integer	45211	16

Dataset Information ⌃

Additional Information

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.

There are four datasets:
1) bank-additional-full.csv with all examples (41188) and 20 inputs, ordered by date (from May 2008 to November 2010), very close to the data analyzed in [Moro et al., 2014]
2) bank-additional.csv with 10% of the examples (4119), randomly selected from 1), and 20

DOWNLOAD

IMPORT IN PYTHON

CITE

🗩 **9 citations**
👁 **328935 views**

Creators

👤 S. Moro

👤 P. Rita

👤 P. Cortez

DOI

10.24432/C5K306

License

This dataset is licensed under a **Creative Commons Attribution 4.0 International** (CC BY 4.0) license.

inputs.
3) bank-full.csv with all examples and 17 inputs, ordered by date (older version of this dataset with less inputs).
4) bank.csv with 10% of the examples and 17 inputs, randomly selected from 3 (older version of this dataset with less inputs).
The smallest datasets are provided to test more computationally demanding machine learning algorithms (e.g., SVM).

The classification goal is to predict if the client will subscribe (yes/no) a term deposit (variable y).
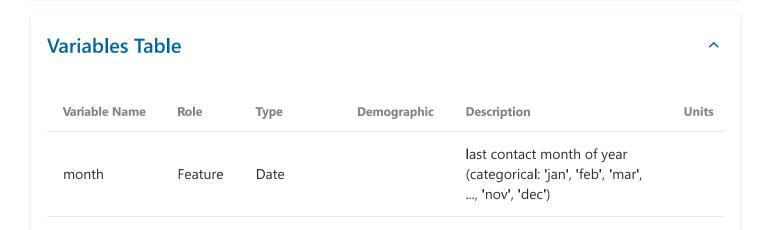SHOW LESS ∧

Has Missing Values?
No

Introductory Paper ∧

[A data-driven approach to predict the success of bank telemarketing](#)
By Sérgio Moro, P. Cortez, P. Rita. 2014
Published in Decision Support Systems

Variables Table ∧

Variable Name	Role	Type	Demographic	Description	Units
month	Feature	Date		last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')	

Variable Name	Role	Type	Demographic	Description	Units
duration	Feature	Integer		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.	
campaign	Feature	Integer		number of contacts performed during this campaign and for this client (numeric, includes last contact)	
pdays	Feature	Integer		number of days that passed by after the client was last contacted from a previous campaign (numeric; -1 means client was not previously contacted)	
previous	Feature	Integer		number of contacts performed before this campaign and for this client	

Variable Name	Role	Type	Demographic	Description	Units
poutcome	Feature	Categorical		outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')	
y	Target	Binary		has the client subscribed a term deposit?	

Rows per page (10)   10 to 17 of 17   [ ]   >

Additional Variable 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 in default? (binary: "yes","no")
   6 - balance: average yearly balance, in euros (numeric)
   7 - housing: has housing loan? (binary: "yes","no")
   8 - loan: has personal loan? (binary: "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 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 that passed by after the client was last contacted from a previous 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 a term deposit? (binary: "yes","no")

SHOW LESS ⌃

Papers Citing this Dataset ⌃

⚏ SORT BY YEAR, DESC

Fair Algorithms for Clustering

By Suman Bera, Deeparnab Chakrabarty, Nicolas Flores, Maryam Negahbani. 2019
Published in ArXiv.

Clustering with Fairness Constraints: A Flexible and Scalable Approach

By Imtiaz Ziko, Eric Granger, Jing Yuan, Ismail Ayed. 2019
Published in ArXiv.

Noise-tolerant fair classification

By Alexandre Lamy, Ziyuan Zhong, Aditya Menon, Nakul Verma. 2019
Published in ArXiv.

AdaFair: Cumulative Fairness Adaptive Boosting

By Vasileios Iosifidis, Eirini Ntoutsi. 2019
Published in

Quantification under prior probability shift: the ratio estimator and its extensions

By Afonso Vaz, Rafael Izbicki, Rafael Stern. 2018
Published in ArXiv.

Rows per page 5   0 to 5 of 9   < ▢

Reviews ⌃

4 ★★★★☆ (1 rating)

LOGIN TO WRITE A REVIEW

甯文駿 智慧商務系
★★★★☆   Nice
3/1/2024

THE PROJECT

About Us

NAVIGATION

Home

LOGISTICS

Contact

CML

National Science Foundation

View Datasets

Donate a Dataset

Privacy Notice

Feature Request or Bug Report