Bank Marketing Data Set

Bank Marketing Data Set

This data set was obtained from the UC Irvine Machine Learning Repository and contains information related to a direct marketing campaign of a Portuguese banking institution and its attempts to get its clients to subscribe for a term deposit.

Source

This data set was obtained by downloading `bank-additional-full.csv` (contained in `bank-additional.zip`) from https://archive.ics.uci.edu/ml/datasets/Bank+Marketing.

The table contains 41,188 rows and 21 columns.

The path to this data set is `pub.demo.mleg.uci.bankmarketing`.

Input Variables

There are 20 columns in the table that provide information about each client, such as age, marital status, and education level. A subset of these are related to the last contact of the current campaign, such as the month and day of the week the last contact was made as well as the number of days since the client was last contacted in a previous campaign. There are 10 columns in the table that are categorial, meaning that they contain textual values that correspond to a particular category for a given variable.

Column Name	Description	Type
`age`	Age of the client	Numeric
`job`	Client's occupation	Categorial: • `admin` • `blue-collar` • `entrepreneur` • `housemaid` • `management` • `retired` • `self-employed` • `services` • `student` • `technician` • `unemployed` • `unknown`
`marital`	Marital status	Categorial: • `divorced` • `married` • `single` • `unknown` **Note:** `divorced` means divorced or widowed

Bank Marketing Data Set

education	Client's education level	Categorial: • basic.4y • basic.6y • basic.9y • high.school • illiterate • professional.course • university.degree • unknown
default	Indicates whether the client has credit in default	Categorial: • no • yes • unknown
housing	Indicates whether the client has a housing loan	Categorial: • no • yes • unknown
loan	Indicates whether the client as a personal loan	Categorial: • no • yes • unknown
contact	Type of contact communication	Categorial: • cellular • telephone
month	Month that last contact was made	Categorial: • jan • feb • ⋮ • dec
day_of_week	Day that last contact was made	Categorial: • mon • tue • wed • thu • fri

Bank Marketing Data Set

`duration`	Duration of last contact in seconds	Numeric **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`	Number of contacts performed during this campaign for this client (including last contact)	Numeric
`pdays`	Number of days since the client was last contacted in a previous campaign	Numeric **Note:** `999` means client was not previously contacted
`previous`	Number of contacts performed before this campaign for this client	Numeric
`poutcome`	Outcome of the previous marketing campaign	Categorial: • `failure` • `nonexistent` • `success`
`empvarrate`	Employment variation rate (quarterly indicator) **Note:** This column was named `emp.var.rate` in the original data set.	Numeric
`conspriceidx`	Consumer price index (monthly indicator) **Note:** This column was named `cons.price.idx` in the original data set.	Numeric
`consconfidx`	Consumer confidence index (monthly indicator) **Note:** This column was named `cons.conf.idx` in the original data set.	Numeric
`euribor3m`	Euribor 3-month rate (daily indicator)	Numeric

Bank Marketing Data Set

`nremployed`	Number of employees (quarterly indicator) **Note:** This column was named `nr.employed` in the original data set.	Numeric

Output Variable

There is one column in the table that corresponds to our target value.

Column Name	Description	Type
`y`	Indicates whether the client has subscribed for a term deposit	Binary (`yes` or `no`)

Dummy Variables

Since we cannot use textual data in our analysis, categorial variables are coded as dummy variables. Each dummy variable represents one of the categories in the categorial columns.

Column Name	Description	Type
`yy`	Client subscribes for a term deposit `y='yes'`	Boolean (`0` or `1`)
`hsng`	Client has a housing loan `housing='yes'`	Boolean (`0` or `1`)
`h_unk`	Unknown if the client has a housing loan `housing='unknown'`	Boolean (`0` or `1`)
`def`	Client has credit in default `default='yes'`	Boolean (`0` or `1`)
`d_unk`	Unknown if the client has credit in default `default='unknown'`	Boolean (`0` or `1`)
`loans`	Client has a personal loan `loan='yes'`	Boolean (`0` or `1`)

Bank Marketing Data Set

`l_unk`	Client has a personal loan `loan='unknown'`	Boolean (`0` or `1`)
`nonxst`	Previous outcome of marketing campaign is nonexistent `poutcome='nonexistent'`	Boolean (`0` or `1`)
`succ`	Previous outcome of marketing campaign was a success `poutcome='success'`	Boolean (`0` or `1`)
`blue`	Client occupation: blue-collar worker `job='blue-collar'`	Boolean (`0` or `1`)
`tech`	Client occupation: technician `job='technician'`	Boolean (`0` or `1`)
`j_unk`	Client occupation: unknown `job='unknown'`	Boolean (`0` or `1`)
`svcs`	Client occupation: services `job='services'`	Boolean (`0` or `1`)
`mgmt`	Client occupation: management `job='management'`	Boolean (`0` or `1`)
`ret`	Client occupation: retired `job='retired'`	Boolean (`0` or `1`)
`entr`	Client occupation: entrepreneur `job='entrepreneur'`	Boolean (`0` or `1`)
`self`	Client occupation: self-employed `job='self-employed'`	Boolean (`0` or `1`)
`maid`	Client occupation: housemaid `job='housemaid'`	Boolean (`0` or `1`)

Bank Marketing Data Set

`unemp`	Client occupation: unemployed `job='unemployed'`	Boolean (`0` or `1`)
`stud`	Client occupation: student `job='student'`	Boolean (`0` or `1`)
`marr`	Marital status: married `marital='married'`	Boolean (`0` or `1`)
`sgl`	Marital status: single `marital='single'`	Boolean (`0` or `1`)
`m_unk`	Marital status: unknown `marital='unknown'`	Boolean (`0` or `1`)