

Bank marketing Project Introduction



1. Summary:

How the effect marketing on the clients and how the clients will be more committed with the loan I worked on this data to answer these questions.

2. Algorithms

- Data cleaning:

- i. Fixing the columns problem
- ii. Dealing with outliers with the help of quantiles

- Data validation:

- i. Checking no nulls
- ii. Checking no duplicates

3. Data:

3.1 Data Set Information:

The data is related to 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.

3.2 History: 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 inputs.
- 3) bank-full.csv with all examples and 17 inputs, ordered by date (older version of this dataset with fewer inputs).
- 4) bank.csv with 10% of the examples and 17 inputs, randomly selected from 3 (older version of this dataset with fewer 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).

3.3 Attribute Information: 3.3.1 Input variables:

bank client data:

1 - age (numeric)

2 - job : type of job (categorical: 'admin.', 'bluecollar', 'entrepreneur', 'housemaid', 'management', 'retired', 'selfemployed', '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

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)

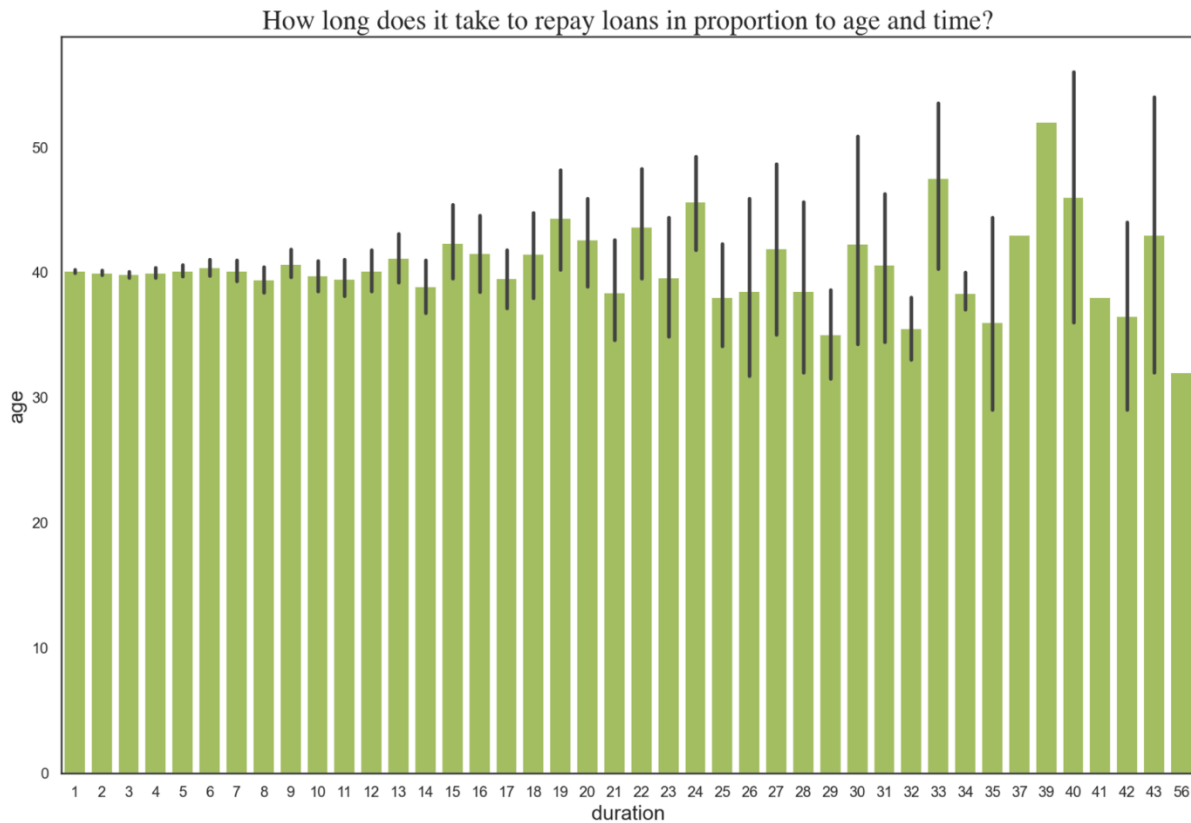
Output variable (desired target):

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

4. Tools

- NumPy, pandas
- Seaborn , sklearn , set_style
- set_context
- statsmodels.api , statsmodels.formula.api
- patsy , matplotlib.pyplot
- sklearn.linear_model > LinearRegression
- from sklearn.linear_model > RidgeCV
- Logistic Regression

5. Communication:



How the clients will be more committed to the loan based on material state, age and duration it takes to complete?

```
21]: <seaborn.axisgrid.FacetGrid at 0x7fb3775e66a0>
```

