

Bank Marketing

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 (or not) subscribed.

Dataset information:

1) bank-full.csv with all examples, ordered by date (from May 2008 to November 2010).

The classification goal is to predict if the client will subscribe a term deposit (variable y).

- Number of Instances: 45211 for bank-full.csv

- Number of Attributes: 16 + output attribute.

- Attribute 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")
```

```
#####  
#####
```

```
# To DO:
```

- see if data needs to be pre-processed or not:
 - check for null values and handle it.
 - check if there a non-rational values or not.
 - label encode data.
 - scale data.
 - draw correlation map.
 - Visualize the data (optional).
 - split data into train and test sets.
 - use decision tree and random forest for data classification.
 - train models and evaluate them using Accuracy, confusion matrix, f1-score,etc.
- use GridSearchCV to get the best parameters for the tree. (bonus)