



Telemarketing bank campaign

Study case
Test for ML Analyst role
DAVIVIENDA

Carlos Garavito

Agenda

- Introduction
- Exploratory data analysis
- Model for clients classification
- Strategic recommendations



Introduction

Introduction

The data set contains information about marketing campaigns that were carried out through telephone calls from a banking institution to its customers. The objective of these campaigns is to encourage their clients to subscribe to a specific financial product of the bank (term deposit). After making each call, the client must inform the institution about his intention to subscribe to the product (indicating a successful campaign) or not (unsuccessful campaign).

The end result of this survey will be a binary result indicating whether or not the customer subscribed to the product.



Exploratory data analysis



Exploratory analysis

1. This data set contains 41.188 rows and 20 features: 10 numerical and 10 categorical.

2. Data set is imbalanced for target variable

- Positive response: 4.639 - (11.3%)
- Negative response: 36.546 - (88.7%)
- Total sample: 41.188

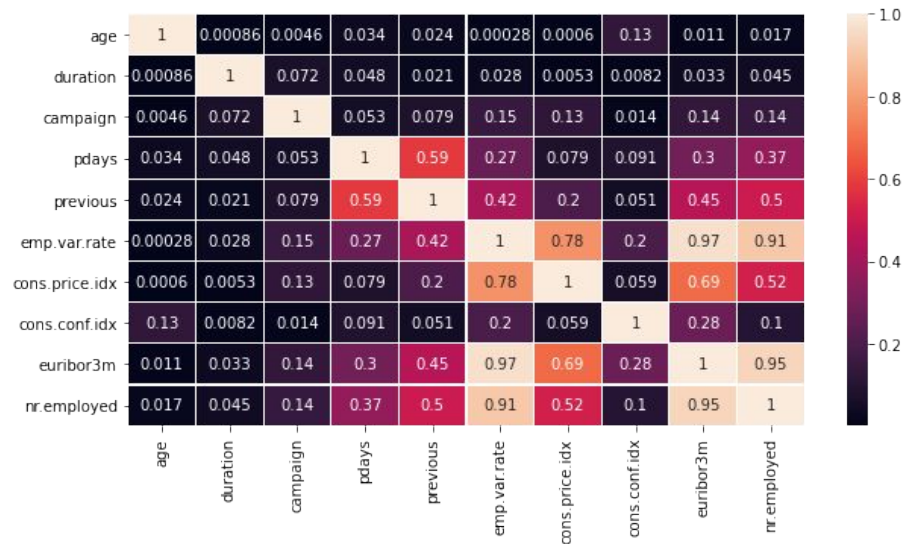
3. Features can be grouped as follows:

- Clients features
- Current campaign features
- Other current campaign features
- Social and economic features

Exploratory analysis

4. Correlated numerical features: Following variables must be not taking into account for modeling, due its high correlation and to avoid multicollinearity.

- 'emp.var.rate',
- 'nr.employed'
- 'cons.price.idx'

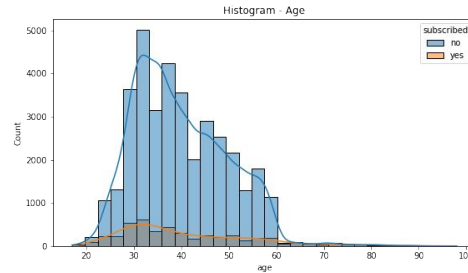


Exploratory analysis

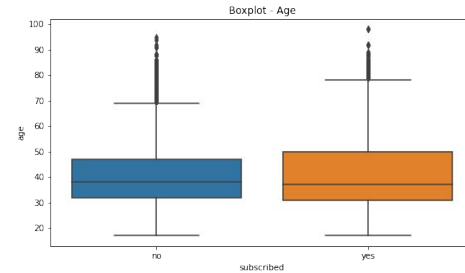
Clients features



- **age:** Clientes range age is from (30 - 50) years old
- **job:** Top tree subscribed clients are 'admin', 'blue-collar' and 'technician'.
- **marital:** most subscribed clients are married.
- **education:** most subscribed clients has an university degree, followed by high school degree.
- **default:** most subscribed clients are not in default.
- **housing:** housing loan is not quite different for subscribed or unsubscribed clients. However, there are more clients subscribed who has housing loans.



Histogram and boxplot for 'age'

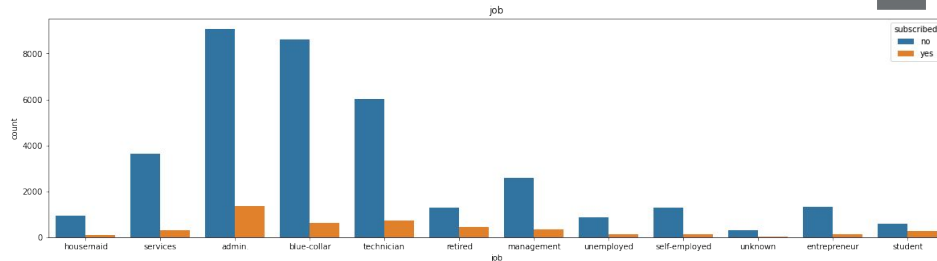


Exploratory analysis

Clients features



- age: Clientes range age is from (30 - 50) years old
- **job: Top tree subscribed clients are 'admin', 'blue-collar' and 'technician'.**
- marital: most subscribed clients are married.
- education: most subscribed clients has an university degree, followed by high school degree.
- default: most subscribed clients are not in default.
- housing: housing loan is not quite different for subscribed or unsubscribed clients. However, there are more clients subscribed who has housing loans.

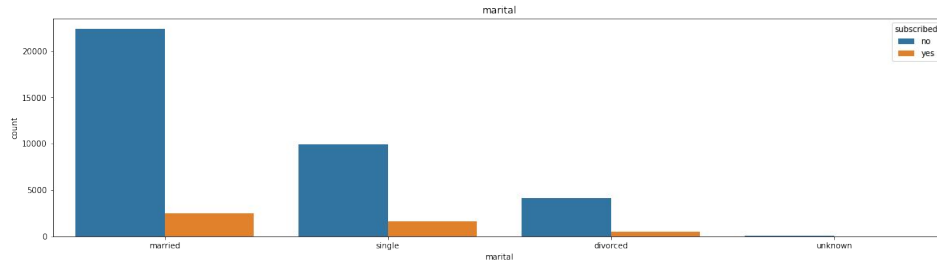


Exploratory analysis

Clients features



- age: Clientes range age is from (30 - 50) years old
- job: Top tree subscribed clients are 'admin', 'blue-collar' and 'technician'.
- **marital: most subscribed clients are married.**
- education: most subscribed clients has an university degree, followed by high school degree.
- default: most subscribed clients are not in default.
- housing: housing loan is not quite different for subscribed or unsubscribed clients. However, there are more clients subscribed who has housing loans.

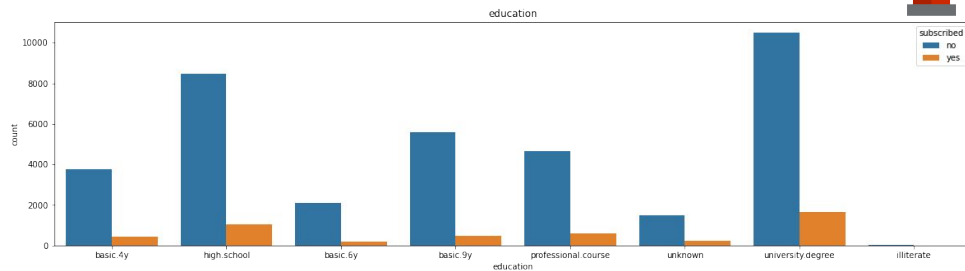


Exploratory analysis

Clients features



- age: Clientes range age is from (30 - 50) years old
- job: Top tree subscribed clients are 'admin', 'blue-collar' and 'technician'.
- marital: most subscribed clients are married.
- **education: most subscribed clients has an university degree, followed by high school degree.**
- default: most subscribed clients are not in default.
- housing: housing loan is not quite different for subscribed or unsubscribed clients. However, there are more clients subscribed who has housing loans.

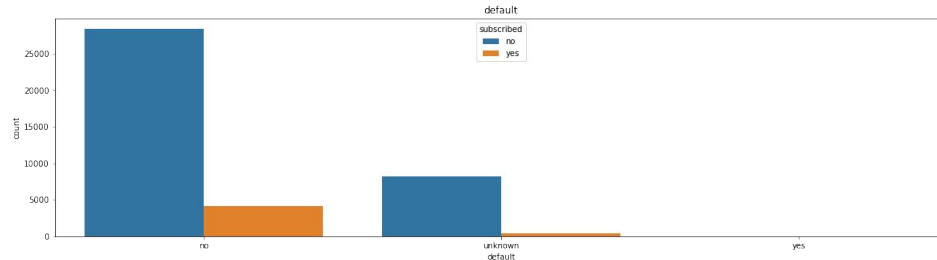


Exploratory analysis

Clients features



- age: Clientes range age is from (30 - 50) years old
- job: Top three subscribed clients are 'admin', 'blue-collar' and 'technician'.
- marital: most subscribed clients are married.
- education: most subscribed clients has an university degree, followed by high school degree.
- **default: most subscribed clients are not in default.**
- housing: housing loan is not quite different for subscribed or unsubscribed clients. However, there are more clients subscribed who has housing loans.

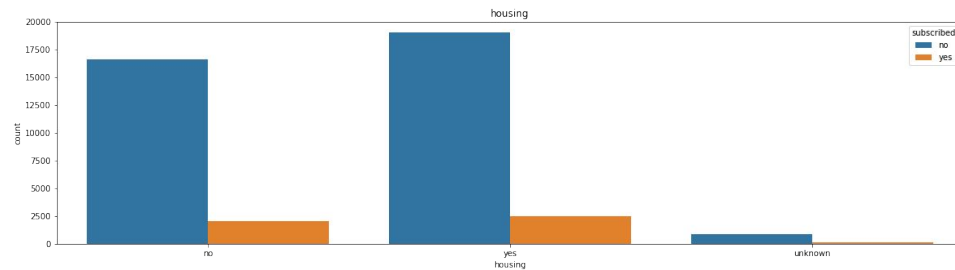


Exploratory analysis

Clients features



- age: Clientes range age is from (30 - 50) years old
- job: Top tree subscribed clients are 'admin', 'blue-collar' and 'technician'.
- marital: most subscribed clients are married.
- education: most subscribed clients has an university degree, followed by high school degree.
- default: most subscribed clients are not in default.
- housing: housing loan is not quite different for subscribed or unsubscribed clients. However, there are more clients subscribed who has housing loans.

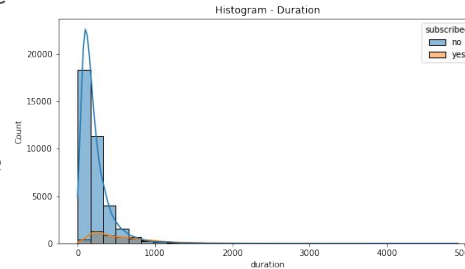


Exploratory analysis

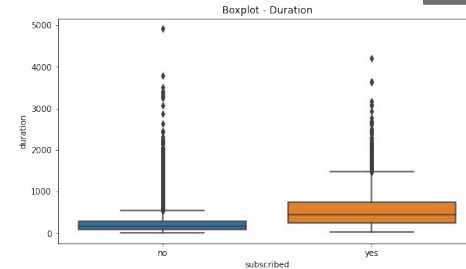
Current campaign features



- **duration:** clients subscribed tends to have more seconds duration during calls.
- **contact:** clients contacted through cellular are more subscribed
- **month:** the month where are more subscriptions is may, the month where are less is dec. Thus, there is a temporality impact to evaluate.
- **day of week:** subscriptions over day of the week are similar for all days. There is no temporality on days of the week.



Histogram and boxplot for 'duration'

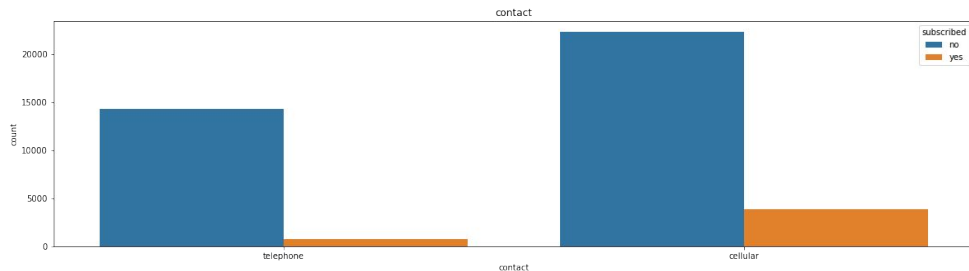


Exploratory analysis

Current campaign features



- duration: clients subscribed tends to have more seconds duration during calls.
- **contact: clients contacted through cellular are more subscribed**
- month: the month where are more subscriptions is may, the month where are less is dec. Thus, there is a temporality impact to evaluate.
- day of week: subscriptions over day of the week are similar for all days. There is no temporality on days of the week.

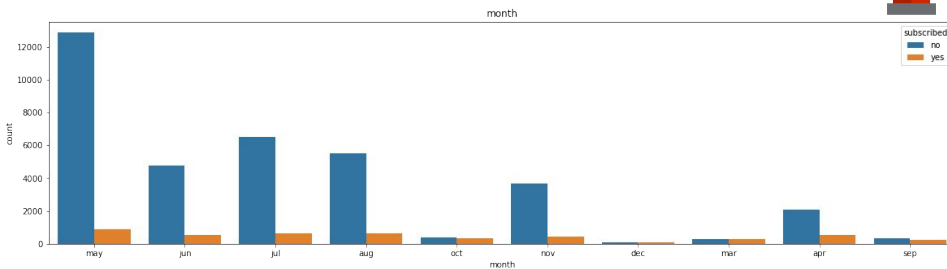


Exploratory analysis

Current campaign features



- duration: clients subscribed tends to have more seconds duration during calls.
- contact: clients contacted through cellular are more subscribed
- **month: the month where are more subscriptions is may, the month where are less is dec. Thus, there is a temporality impact to evaluate.**
- day of week: subscriptions over day of the week are similar for all days. There is no temporality on days of the week.

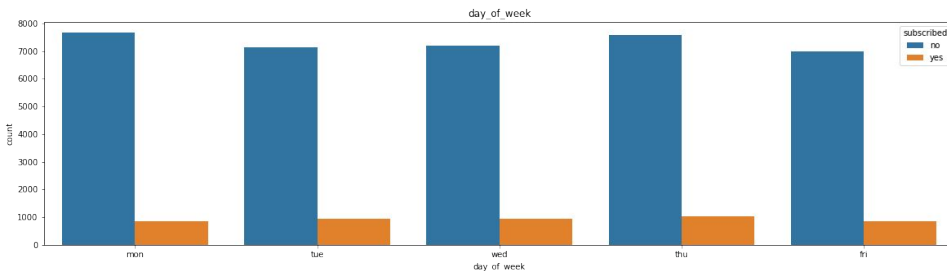


Exploratory analysis

Current campaign features



- duration: clients subscribed tends to have more seconds duration during calls.
- contact: clients contacted through cellular are more subscribed
- month: the month where are more subscriptions is may, the month where are less is dec. Thus, there is a temporality impact to evaluate.
- **day of week: subscriptions over day of the week are similar for all days. There is no temporality on days of the week.**

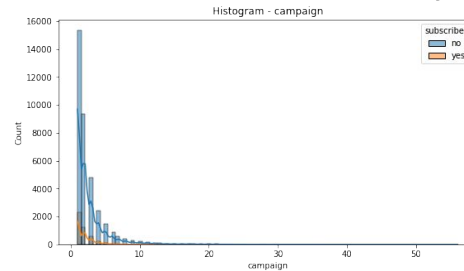


Exploratory analysis

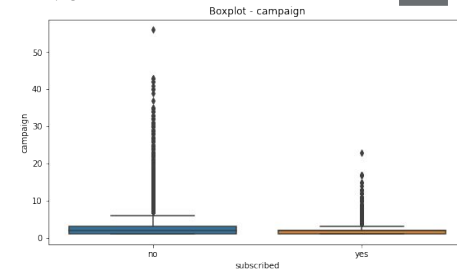
Current campaign features



- **campaign:** there are a high quantity of clients that were not contacted in this company. No contacted clients tend to not subscribe to the campaign. Thus, is important to take this feature into account for the model because is going to impact the prediction value.
- **pdays:** the major of clients has not be contacted by previous campaigns.
- **previous:** clients previously contacted due past campaigns, tends the more to subscribe.
- **poutcome:** there are a population of new clients that can be leverage. If the outcome is nonexistent, is due that client has not be contacted previously and the reason could be that is because is a new client.



Histogram and boxplot for 'campaign'

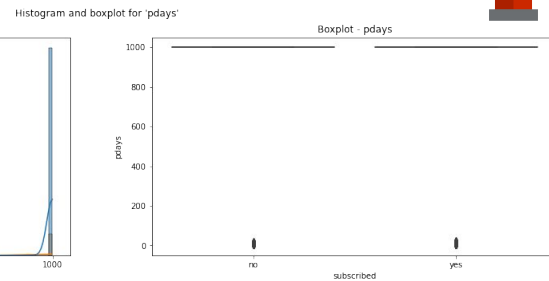
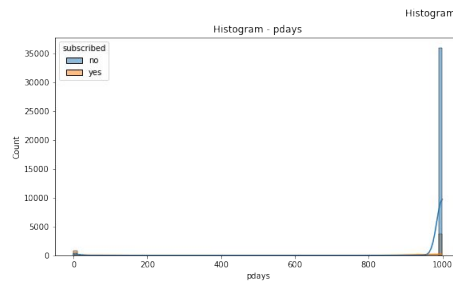


Exploratory analysis

Current campaign features



- campaign: there are a high quantity of clients that were not contacted in this company. No contacted clients tend to not subscribe to the campaign. Thus, is important to take this feature into account for the model because is going to impact the prediction value.
- **pdays:** the major of clients has not be contacted by previous campaigns.
- previous: clients previously contacted due past campaigns, tends the more to subscribe.
- poutcome: there are a population of new clients that can be leverage. If the outcome is nonexistent, is due that client has not be contacted previously and the reason could be that is because is a new client.

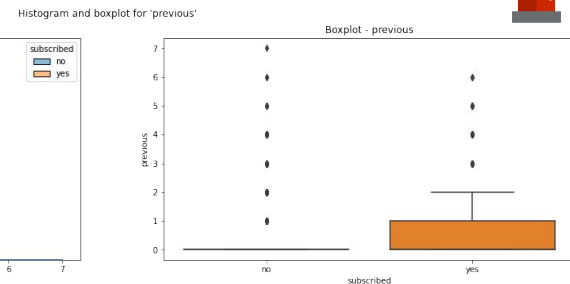
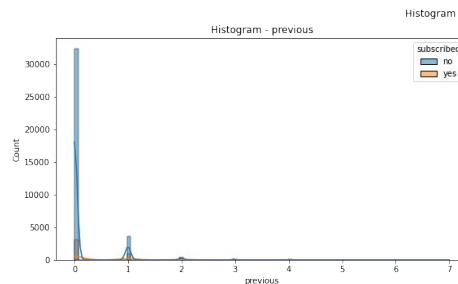


Exploratory analysis

Current campaign features



- campaign: there are a high quantity of clients that were not contacted in this company. No contacted clients tend to not subscribe to the campaign. Thus, is important to take this feature into account for the model because is going to impact the prediction value.
- pdays: the major of clients has not be contacted by previous campaigns.
- **previous: clients previously contacted due past campaigns, tends the more to subscribe.**
- poutcome: there are a population of new clients that can be leverage. If the outcome is nonexistent, is due that client has not be contacted previously and the reason could be that is because is a new client.

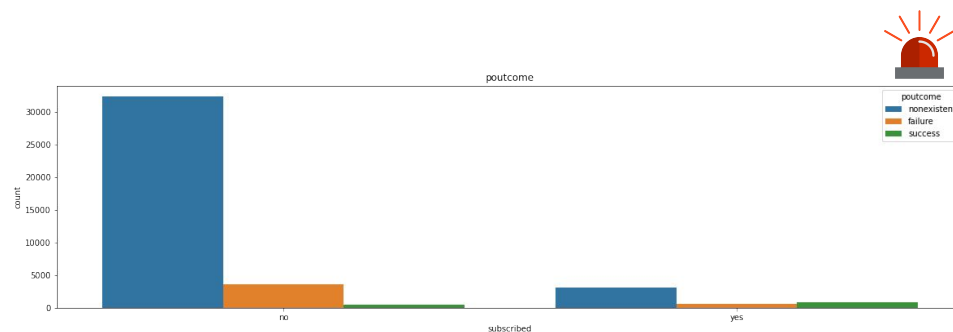


Exploratory analysis

Other current campaign features



- campaign: there are a high quantity of clients that were not contacted in this company. No contacted clients tend to not subscribe to the campaign. Thus, is important to take this feature into account for the model because is going to impact the prediction value.
- pdays: the major of clients has not be contacted by previous campaigns.
- previous: clients previously contacted due past campaigns, tends the more to subscribe.
- **poutcome:** there are a population of new clients that can be leverage. If the outcome is nonexistent, is due that client has not be contacted previously and the reason could be that is because is a new client.

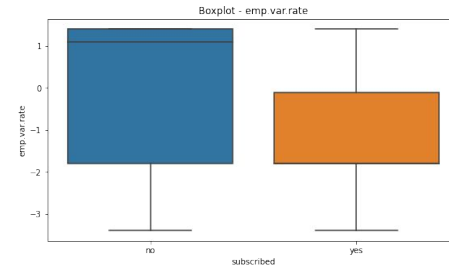
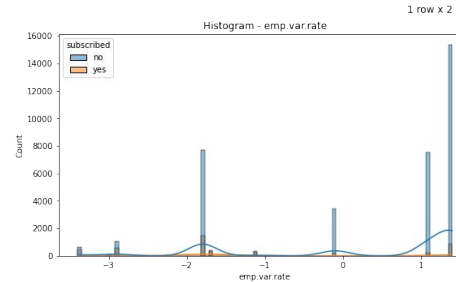


Exploratory analysis

Other current campaign features



- **emp.var.rate:** subscribed clients are in range $(-2) - 0$
- **cons.price.idx:** is similar for subscribed and non subscribed clients. Are in range (93-94)
- **nr.employed:** subscribed clients are in range (5050-5200), approx.

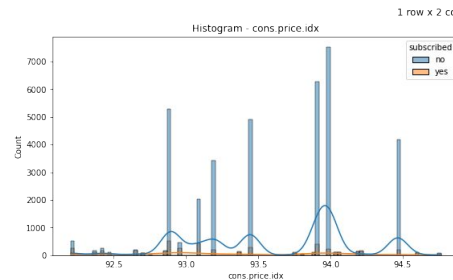


Exploratory analysis

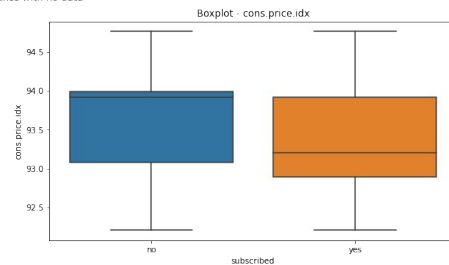
Other current campaign features



- emp.var.rate: subscribed clients are in range $(-2) - 0$
- cons.price.idx: is similar for subscribed and non subscribed clients. Are in range (93-94)
- nr.employed: subscribed clients are in range (5050-5200), approx.



1 row x 2 columns axes with no data

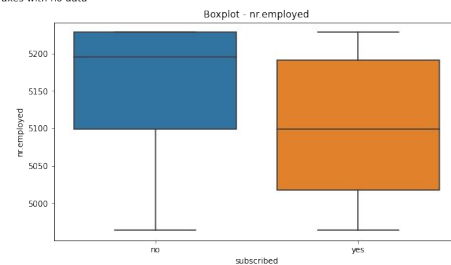
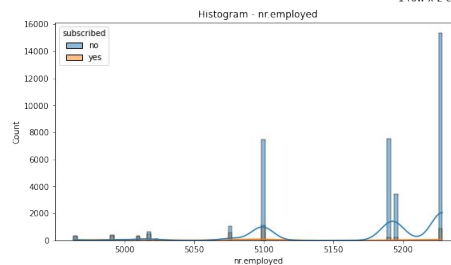


Exploratory analysis

Other current campaign features



- emp.var.rate: subscribed clients are in range $((-2) - 0)$
- cons.price.idx: is similar for subscribed and non subscribed clients. Are in range (93-94)
- **nr.employed:** subscribed clients are in range (5050-5200), approx.



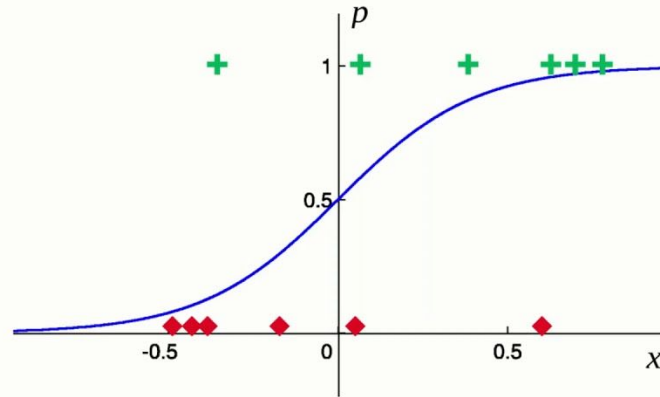
Clients classification model

Model

REGRESIÓN LOGÍSTICA INTERPRETACIÓN

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 x$$

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$



Model

Features that impacts the client decision

1. duration
2. euribor3m
3. month
4. cons.conf.idx
5. poutcome
6. pdays
7. campaign
8. job

```
For values of best alpha = 0.01 The train AUC is: 0.6814842408954781
For values of best alpha = 0.01 The cross validation AUC is: 0.6918247553116867
For values of best alpha = 0.01 The test AUC is: 0.6979729028239202
```

feature	weight	abs
duration	-1.138615	1.138615
euribor3m	-0.394409	0.394409
month_may	0.185938	0.185938
month_mar	0.182653	0.182653
cons.conf.idx	0.176865	0.176865
poutcome_success	0.149676	0.149676
month_jun	0.144241	0.144241
month_oct	-0.129165	0.129165
pdays	0.127246	0.127246
month_jul	-0.117679	0.117679
campaign	-0.116249	0.116249
poutcome_failure	0.108272	0.108272
job_student	0.069842	0.069842
education_university_degree	0.069842	0.069842
default_no	0.060313	0.060313
default_unknown	-0.060153	0.060153
job_retired	0.088466	0.088466
job_blue-collar	-0.085517	0.085517
education_basic5y	-0.064529	0.064529
education_professional_course	0.056120	0.056120
month_aug	0.054519	0.054519
month_apr	0.051609	0.051609
month_dec	0.050296	0.050296
education_basic4y	-0.050230	0.050230
month_sep	0.049222	0.049222
education_highschool	-0.048485	0.048485
day_of_week_wed	0.041523	0.041523
marital_divorced	0.039655	0.039655
job_housemaid	0.033359	0.033359
day_of_week_tue	0.032710	0.032710
education_basic3y	-0.030078	0.030078
marital_married	-0.028118	0.028118
contact_telephone	-0.026830	0.026830
contact_cellular	0.026830	0.026830
job_self-employed	-0.026287	0.026287
day_of_week_mon	-0.025146	0.025146
day_of_week_thu	-0.024920	0.024920
day_of_week_fri	-0.023479	0.023479
month_nov	0.022293	0.022293
job_entrepreneur	-0.021915	0.021915
job_admin	0.020874	0.020874
marital_unknown	0.020866	0.020866
education_unknown	0.017649	0.017649
age	0.015824	0.015824
job_unemployed	0.015639	0.015639
loan_yes	-0.011769	0.011769
job_management	-0.010717	0.010717
education_illiterate	0.009940	0.009940
default_yes	-0.009694	0.009694
poutcome_nonexistent	0.008706	0.008706
loan_no	0.008242	0.008242
housing_unknown	0.007084	0.007084
loan_unknown	0.007084	0.007084
job_technician	-0.006461	0.006461
previous	0.003752	0.003752
housing_no	-0.003053	0.003053
job_services	0.001658	0.001658
housing_yes	0.000857	0.000857
marital_single	0.000599	0.000599
job_unknown	-0.000027	0.000027

Strategic recommendations



Recommendations

1. month:
Develop campaigns that take into account the month.
2. duration:
Train call center staff to maintain a lasting and pleasant conversations.
3. campaign:
Focus on those customers who have already been contacted
4. pdays:
Develop a telemarketing plan to call customers who have not contacted.

Thanks!

Questions, recommendations or suggestions...

Carlos Garavito
carloss.garavito@gmail.com

