

# Campanya telefònica per contractar un dipòsit

## Introducció

L'èxit d'una campanya telefònica per vendre un producte financer depèn de les característiques de l'individu (any, ocupació, educació, estat civil, préstec personal, préstec hipotecari), i de si s'ha inclòs l'individu en campanyes anteriors, quan de temps fa en dies i si va contractar el producte financer objecte de la campanya anterior.

En les trucades es recollirà: tipus de telèfon de contacte, data de la darrera trucada, nombre de trucades a la campanya actual, durada i resultat de la campanya.

## Objectius del Projecte

### *1. Quins són els objectius del negoci?*

-L'efectivitat de la campanya per contractar el servei financer que s'oferta. Un primer objectiu d'assolir una taxa d'èxits alta en la contractació a la campanya actual, tot reduint el nombre de contactes telefònics totals (segon objectiu).

Consideracions inicials:

- La informació de l'èxit en alguna campanya anterior ha d'afectar en positiu o negatiu (no es pot saber sense analitzar les dades) el fet de contractar un dipòsit en una nova campanya.
- Hi ha característiques dels individus que els fan més susceptibles a contractar el dipòsit en una campanya.

### *2. Quines decisions o processos específics voleu millorar o automatitzar amb ML?*

- Determinar quines són les característiques dels clients potencials que afavoreien les possibilitats de contractació segons la informació de campanyes anteriors i reduir el nombre de trucades a realitzar en una campanya per assolir una contractació.
- Maximitzar la taxa de contractació del producte financer en una futura campanya.

### *3. Es podria resoldre el problema de manera no automatitzada?*

-Es podria fer una aproximació amb estadística descriptiva, però l'aprenentatge automàtic és d'esperar que millori els resultats en una campanya concreta considerant l'experiència de contractació en campanyes anteriors, el temps des de la darrera campanya, etc.

## Metodologia Proposta

Determinar les característiques individuals i l'efecte del temps des de la darrera campanya, la contractació positiva en una campanya anterior amb contacte, el nombre de trucades en una campanya anterior i l'actual, i la contractació positiva a la campanya actual. Després d'una exploració de les dades, es constata que el temps de contacte o el nombre de contactes en una anterior campanya no poden ser variables predictores en el model donat el nombre de mostres que no han participat en campanyes anteriors.

Una regressió logística per determinar la probabilitat de contractació segons les característiques de l'individu, l'èxit de l'anterior campanya o la no participació en una campanya anterior pot resultar ràpid en predicció. La durada de la darrera trucada no ha

d'incloure's com a variable explicativa, tampoc el nombre de contactes o dies des d'una campanya anterior.

## Dades Disponibles

-Les dades bank dataset contenen característiques i la variable target sembla ser *deposit*. Semblen dades d'una campanya telefònica, i resultats d'una campanya tancada. No hi ha registres repetits de la mateixa persona, però no es pot identificar l'individu per confirmar la hipòtesi. Descripció de les dades a:

<https://archive.ics.uci.edu/dataset/222/bank+marketing>

<b>age</b>	edat en anys
<b>job</b>	ocupació estudiant, aturat, treballador manual, retirat, administratiu, serveis, etc
<b>marital</b>	estat civil en solter, casat o divorciat
<b>education</b>	nivell educatiu en primària, secundària, superior o desconegut
<b>default</b>	has credit in default?
<b>balance</b>	average yearly balance
<b>housing</b>	has housing loan?
<b>loan</b>	has personal loan?
<b>contact</b>	Via de contacte telèfon mòbil, fix o desconegut
<b>day</b>	Dia del contacte
<b>month</b>	Mes del contacte
<b>duration</b>	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.
<b>campaign</b>	number of contacts performed during this campaign and for this client (numeric, includes last contact) (1 i 63)
<b>pdays</b>	number of days that passed by after the client was last contacted from a previous campaign (numeric; -1 means client was not previously contacted) (entre -1 i 854)
<b>previous</b>	number of contacts performed before this campaign and for this client (entre 0 i 58)
<b>poutcome</b>	outcome of the previous marketing campaign (categorical: 'failure', 'nonexistent', 'success')
<b>deposit</b>	has the client subscribed a term deposit? (si o no)

## Mètrica d'èxit del projecte

Les taules de confusió i les mesures de resultats F1 score i l'àrea sota la corba ROC són les mesures proposades per avaluar la qualitat del model de regressió logística.

## Responsabilitats Ètiques i Socials

Hi ha segments més probables de respondre trucades, no s'ha d'abusar i ser massa insistents si no és probable l'èxit en la contractació. El biaix a aturats i pensionistes cal evitar-lo, tampoc no és probable que contractin.

L'empresa que desenvolupi el projecte de Machines Learning ha de tenir la responsabilitat d'adoptar una política i unes pràctiques ètiques, que considerin la diversitat i la inclusió a l'equip de desenvolupament. Alhora que els algorismes desenvolupats evitin biaixos i discriminació, les dades d'entrenament han de pertànyer a diferents col·lectius segons gènere, grup d'edat, ètnia, religió, etc (característiques no recollides en les dades disponibles). De cap manera s'han de vendre les dades emprades en l'aprenentatge automàtic, ni donar recomanacions inadequades com per exemple, suggerir uns pantalons curts a una noia jove que va vestida amb un hijab o un vestit de nit a una menor.

**URL GitHub :** <https://github.com/lidiamontero/Project-Machine-Learning>