

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?

- Atacar els segments d'usuaris amb més possibilitats de contractació segons la informació de campanyes anteriors i reduir el nombre de trucades a realitzar en una campanya per assolir una contractació (hi ha segments que no val la pena insistir a partir d'un cert punt, cal saber el punt on parar les trucades quan no hi ha una resolució d'èxit o fracàs).

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

Clarament una primera exploració adreçada a la segmentació, és a dir classificació supervisada dels usuaris amb contractació positiva en alguna campanya. Ara bé, cal matitzar i veure quin és l'efecte del temps des de la darrera

campanya amb contacte i contractació positiva.

Una regressió logística per determinar la probabilitat de contractació segons les característiques de l'individu, el temps des de la darrera campanya i l'èxit de l'anterior campanya pot resultar ràpid en predicció. La durada de la darrera trucada no ha d'incloure's com a variable explicativa.

Dades Disponibles

-Les dades bank dataset contenen característiques i la variable target sembla ser *deposit*. Semblen dades d'una campanya telefònica, i resultats de la campanya anterior (i temps des de la darrera campanya amb contacte). No hi ha trucades repetides a 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>

age	edat en anys
job	ocupació estudiant, aturat, treballador manual, retirat, administratiu, serveis, etc
marital	estat civil en solter, casat o divorciat
education	nivell educatiu en primària, secundària, superior o desconegut
default	has credit in default?
balance	average yearly balance
housing	has housing loan?
loan	has personal loan?
contact	Via de contacte telèfon mòbil, fix o desconegut
day	Dia del contacte
month	Mes del contacte
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.
campaign	number of contacts performed during this campaign and for this client (numeric, includes last contact) (1 i 63)
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) (entre -1 i 854)
previous	number of contacts performed before this campaign and for this client (entre 0 i 58)
poutcome	outcome of the previous marketing campaign (categorical: 'failure', 'nonexistent', 'success')
deposit	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 poden donar una mesura de la qualitat del model de regressió logística de predicció.

Responsabilitats Ètiques i Socials

Hi ha segments més probables de recollir les 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.

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