

Predicció de subscripcions de dipòsits a termini

Introducció

Aquest projecte té l'objectiu de predir si un nou client contractarà un dipòsit a termini utilitzant dades d'altres clients. Aquesta predicció permet optimitzar esforços comercials, augmentar l'eficàcia de les campanyes i reduir costos, garantint que es destinen els recursos a clients amb més probabilitat de contractació.

Objectius del Projecte

1. Quins són els objectius del negoci?

- Millorar l'eficàcia de les campanyes de màrqueting directe
- Reduir els costos operatius evitant contactes amb clients amb baixa probabilitat de conversió
- En última instància, augmentar el percentatge de clients que contracten dipòsits a termini

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

- Automatitzar la selecció de clients potencials per a campanyes de màrqueting
- Prioritzar clients segons la probabilitat de contractació
- Identificar factors clau que contribueixen a l'èxit d'una subscripció

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

Tot i que seria possible analitzar patrons manualment, la complexitat i volum de les dades fan que un enfocament mitjançant ML sigui molt més eficient i robust.

Metodologia Proposada

S'utilitzarà un model supervisat senzill, com una regressió logística, i altres basats en arbres de decisió, com ara un Random Forest Classifier o un Gradient Boosting Machine. L'objectiu és predir la variable dependent (subscripció a dipòsit).

El flux de treball inclou:

- Preprocessament de dades (neteja de valors nuls o incomplets, codificació, estandardització, escalament...)
- Divisió del conjunt de dades en entrenament i test
- Estudiar la utilització d'un subset addicional de validació o entrenar els models amb validació creuada
- Avaluació i comparació dels models segons les mètriques d'èxit definides

Dades Disponibles

El conjunt de dades inclou les següents variables:

- Característiques demogràfiques (age, job, marital, education)
- Informació financera (default, balance, housing, loan)
- Històric de contactes (contact, day_of_week, month, duration, campaign, pdays, previous, poutcome)
- Variable objectiu (y)

Mètrica d'èxit del projecte

La mètrica clau serà la precisió de la predicció i la taxa d'encert (F1-Score) per a la classe positiva ($y = \text{sí}$). Aquesta mètrica és d'especial interès quan les classes de la variable objectiu estan descompensades i, per exemple, hi ha molts més “no”, que “sí”. També es mesurarà:

- AUC-ROC per quantificar la separació entre classes
- Cost-eficàcia, calculant el cost estalviat per contacte ineficient

Responsabilitats Ètiques i Socials

- Garantir la privacitat i protecció de dades: Assegurar la protecció de les dades dels clients complint amb les normatives vigents com el GDPR, mitjançant anonimització i control d'accés.
- Evitar biaixos i discriminació: Auditar el model per garantir que no discrimina per factors com edat, gènere o estatus social, promovent l'equitat en les prediccions. S'estudiarà la influència d'aquestes característiques en el resultat dels models entrenats.
- Consentiment informat i transparència: Recollir el consentiment explícit dels usuaris i comunicar clarament com es processen i utilitzen les seves dades.
- Explicabilitat del model: Garantir que les prediccions del model siguin comprensibles mitjançant tècniques que expliquin el paper de cada característica en les decisions.
- Monitoratge continu: Supervisar regularment el model per mantenir el seu rendiment, ajustar-lo si cal i assegurar-ne el compliment ètic.