

Projecte de ML per augmentar la subscripció de clients a dipòsits bancaris.

Introducció

Es tracta d'un banc de Portugal que realitza una campanya de màrqueting telefònic per augmentar la subscripció a certs dipòsits bancaris entre els seus clients.

Predir de manera efectiva si un client acceptarà o no un dipòsit ens permetrà millorar l'eficiència de les campanyes i optimitzar els recursos del banc, aconseguint així una major conversió.

Objectius del Projecte

1. Quins són els objectius del negoci?

Augmentar el percentatge de conversió de les campanyes de màrqueting telefònic dirigides a la subscripció de dipòsits per part dels clients.

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

Automatitzar la predicció de la conversió de cada client segons les seves característiques per ajudar-nos a millorar les campanyes telefòniques: personalitzant més les campanyes, dirigint-nos als clients de forma diferent segons la seva probabilitat de conversió, etc.

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

Sí, es podria fer la predicció de la conversió de cada client utilitzant Excel o altres eines però seria més lent i es necessitaria més personal. Utilitzar machine learning ens permetrà automatitzar aquest procés reduint temps i costos.

Metodologia Proposta

En el dataset tenim un conjunt de variables independents i una variable dependent o etiqueta (deposit) que només té dos valors: "Yes" i "No". Per tant farem servir un algoritme **supervisat de classificació**. Podem treballar amb diferents algoritmes com **Regressió Logística, SVM's, Random Forest, XGBoost, etc.** A continuació, els analitzarem i compararem per decidir quins tenen el millor comportament.

Dades Disponibles

El dataset té 45.211 trucades a clients i 16 característiques o atributs, que contenen informació demogràfica, laboral i financera dels clients, i altres dades relacionades amb les trucades telefòniques (com durada, nombre de contactes previs, etc.). La variable objectiu és la columna “deposit”, que indica si un client s’ha subscrit o no. Les variables són:

- **age, job, marital, education:** Característiques demogràfiques dels clients.
- **balance, housing, loan:** Dades bancàries i productes contractats pels clients.
- **contact, day of week, month, duration, campaign:** Detalls de les trucades telefòniques amb els clients
- **campaign, pdays, previous, poutcome:** Dades relacionades amb campanyes anteriors i l’actual.

Mètrica d'èxit del projecte

Podem utilitzar aquestes mètriques:

- **Accuracy:** Per determinar la quantitat de clients classificats correctament..
- **Curves ROC:** Ens permet avaluar el model mitjançant la taxa de falsos positius i la taxa de veritables positius.
- Si el dataset està molt desbalancejat (proporció entre nombre de si i no molt diferent a 1) podem fer servir la mètrica **F1-Score**.

Responsabilitats Ètiques i Socials

- **Privacitat de les dades:** Hem de garantir que les dades personals dels clients s'utilitzaran de manera ètica, anonimitzades i amb el consentiment explícit dels clients..
- **Transparència:** Hem de comunicar als clients com s'utilitzaran les seves dades de forma clara i senzilla.
- **Justícia:** Hem d'evitar els biaixos, hem d'analitzar els models per garantir que no existeixin biaixos que podrien generar discriminacions entre els clients.
- **Impacte laboral:** L'automatització d'aquest procés podria afectar a la feina d'alguns treballadors. Hem de valorar com implementar el model per no afectar de forma negativa a aquests treballadors.
- **Responsabilitat,** hem de nomenar un responsable de les dades que doni resposta per si hi ha un error o una queixa dels clients. En aquest cas podria ser el cap de departament de BI.