

Projecte de ML per incrementar la contractació de dipòsits a termini fix

Introducció

Una entitat bancària portuguesa vol optimitzar les seves campanyes de màrqueting directe, que tenen com a objectiu aconseguir que els clients contractin dipòsits a termini fix. Tot i les diverses campanyes realitzades, el banc ha detectat que una part significativa dels esforços es dirigeixen a clients que no acaben contractant el producte.

Objectius del Projecte

1. Quins són els objectius del negoci?

Augmentar la taxa de contractació de dipòsits a termini fix identificant els clients amb més probabilitat d'acceptar la oferta i reduir els costos associats a contractar clients amb baixes probabilitats d'acceptació de dipòsits a termini fix.

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

Automatitzar el procés de determinar quins clients haurien de ser contactats durant una campanya, prioritzant aquells amb una alta probabilitat d'acceptar el dipòsit a termini fix. Això permetria maximitzar l'eficàcia de la campanya i reduir els costos associats a contactar clients amb baixes probabilitats de conversió.

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

Sí, el problema es podria resoldre manualment analitzant dades històriques, segmentant clients segons criteris fixos (com edat o diners al compte bancari) i definint regles simples per prioritzar contactes. No obstant això, fer-ho tot manualment és lent, poc precís i difícil d'escalar, especialment amb grans volums de dades o patrons complexos.

Metodologia Proposta

S'utilitzarà un model de classificació supervisada per predir la probabilitat que un client contracti un dipòsit a termini fix. En concret, s'utilitzarà el random forest regressor, ja que és un algorisme robust i interpretable que combina múltiples arbres de decisió per millorar la precisió i evitar el sobreajustament.

És molt efectiu per a problemes amb moltes variables i per manejar dades de tipus mixt.

A més, en el cas d'acceptació de dipòsits, pot haver-hi una distribució desequilibrada entre les classes (molts clients que no accepten l'oferta i pocs que sí). Random Forest és menys sensible a aquest desequilibri en comparació amb altres models.

Per últim, aquest algoritme permet identificar la importància de les característiques, que ajuda a entendre quins factors influeixen més les decisions del model (com l'edat, l'estat financer o les campanyes prèvies), que serà útil per decisions estratègiques.

Com a mètriques de rendiment tenim el àrea sota la corba ROC, que permet distingir entre les classes (accepta / no accepta), un valor proper a 1 indica un model perfecte i és útil amb dades desequilibrades. Altres mètriques de rendiment serien la precisió (classificats com sí accepta que realment ho fan), la sensibilitat i el F1-score. Aquest últim combina precisió i sensibilitat per fer un balanç i és útil quan les dades son desequilibrades.

Dades Disponibles

Les dades disponibles es divideixen entre demogràfiques (edat, ocupació, estat civil i educació), estat financer (si tenen deutes, saldo, hipoteca i préstec), detalls de la campanya (canal i data de contacte, durada de la trucada i número de contactes), històric de campanyes anteriors (pdays, número de contactes previs i resultats de campanya anterior), i variable objectiu (dipòsit).

Mètrica d'èxit del projecte

Augment de la contractació de dipòsits a termini fix. Aquesta mètrica reflectiria directament l'efectivitat del model en millorar les estratègies de màrketng.

Responsabilitats Ètiques i Socials

-Privacitat de les dades: garantir que es tractin de manera confidencial i acorde a les lleis.

-Impacte social: evitar biaixos que puguin discriminar per edat, gènere o ocupació, i garantir que tots els clients tinguin oportunitats de ser considerats.

-Transparència: informar de l'ús de les dades.

-Responsabilitat de les decisions automatitzades: s'ha d'assegurar que les decisions del model estiguin subjectes a revisió per part de persones, especialment en casos sensibles o d'impacte negatiu potencial. A més, s'han d'implementar mesures per identificar i corregir errors i establir qui assumeix la responsabilitat per les accions derivades de les prediccions del model, garantint que l'empresa no delegui les decisions crítiques a l'algoritme.