

Projecte de ML per a incrementar la despesa anual de clients a la botiga

Introducció

Es tracta d'una botiga exclusiva especialitzada en la confecció i venda de roba a mida. Es destaca per oferir consultories altament personalitzades. Els clients visiten la botiga per rebre assessorament directe d'estilistes experts que ajuden a crear peces úniques que s'ajustin perfectament a les seves preferències i mesures. Després de la sessió de consultoria, els clients poden fer comandes de roba mitjançant una aplicació mòbil o del lloc web de l'empresa.

Objectius del Projecte

1. Quins són els objectius del negoci?

Augmentar les vendes avaluant on concentrar els esforços: millorant l'experiència dels clients al lloc web o l'aplicació mòbil

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

Es cerca optimitzar les decisions empresarials relacionades amb l'experiència del client i automatitzar la predicció de la despesa anual de cada client.

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

Si bé la predicció de la despesa anual de cada client es pot fer utilitzant fulls de càlcul o programari estadístics, l'ús de machine learning permetrà automatitzar aquest procés mitjançant entrenaments programats que incorporin ràpidament les transaccions diàries sense necessitat d'intervenció manual.

Metodologia Proposta

4. Quin és l'algorisme de Machine Learning més adequat per resoldre aquest problema? Com justifica l'elecció d'aquest algorisme? Quines mètriques d'avaluació s'utilitzaran per a mesurar el rendiment del model?

Per resoldre el problema de predir la despesa anual dels clients, recomanem l'ús d'algoritmes com Random Forest Regressor o XGBoost, ja que són capaços de capturar relacions no lineals i interaccions complexes entre variables, a més de proporcionar una bona robustesa i escalabilitat. Pel que fa a les mètriques d'avaluació, proposem principalment l'Error Absolut Mitjà (MAE) i l'Arrel de l'Error Quadràtic Mitjà (RMSE), ja que mesuren amb claredat la precisió del model en la mateixa unitat monetària i penalitzen de manera més forta els errors grans, respectivament.

Dades Disponibles

5. Quines dades estan disponibles per abordar aquest problema?

La botiga compta amb un conjunt de dades actualitzat que inclou informació identificativa de cada client, la suma anual que ha gastat a la botiga, el temps dedicat a interactuar tant al lloc web com a l'aplicació mòbil, i l'estat de membres.

Mètrica d'Èxit

6. Quina és la mètrica d'èxit per a aquest projecte?

Augment en la Despesa Anual Mitjana per Client. Aquesta mètrica reflectiria directament l'efectivitat del model en millorar les decisions de l'empresa.

Responsabilitats Ètiques i Socials

7. Quines responsabilitats ètiques i socials és important tenir en compte?

Per abordar les responsabilitats ètiques i socials del projecte, és important:

- Garantir la privacitat i protecció de dades complint amb lleis com el GDPR
- Obtenir el consentiment informat dels usuaris i ser transparents en l'ús de la informació dels clients
- Evitar biaixos i discriminació en el model de ML
- Implementar mesures robustes de seguretat de la informació
- Assignar una o més persones responsables del projecte per tal de donar resposta en cas d'ocasionar danys
- Assegurar la explicabilitat del model per entendre les decisions preses
- Establir un monitoratge continu per supervisar i ajustar el rendiment ètic del model

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 subconjunt addicional de validació o entrenar els models amb validació creuada
- Avaluació i comparació dels models segons les mètriques d'èxit definides. Les mètriques de rendiment que es contemplen són les següents:

- Recall (sensibilitat) per a la classe positiva ($y = \text{"sí"}$), ja que ens interessa no perdre oportunitats de subscripció de nous dipòsits
- F1-Score, sobretot si les classes estan descompensades, perquè combina precisió i recall
- AUC-ROC per avaluar la capacitat de separació de classes
- Precisió i altres mètriques complementàries si s'escau

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

L'èxit del projecte es mesurarà en termes de resultats de negoci, concretament l'augment de la taxa de contractació de dipòsits a termini en comparació amb la situació anterior a la introducció del model de ML. Aquesta mètrica reflecteix directament l'objectiu corporatiu de millorar l'eficàcia de les campanyes de màrqueting i augmentar la conversió de clients.

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 els resultats 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.