

Tasca 3.1: 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 a 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? Que mètriques d'avaluació s'utilitzaran per a mesurar el rendiment del model?

L'algorisme més adient es una **Regressió Lineal Múltiple**, per les següents raons:

- Donat que tenim un conjunt de dades etiquetades i volem maximitzar una d'aquestes dades, estem davant d'un aprenentatge supervisat.
- El conjunt de dades té 8 característiques, de les que 3 son de tipus text i la resta son de tipus numèric. Els tres de tipus text (Email, Address, Avatar) son només etiquetes que identifiquen al client i que no sembla a priori que tinguin influència en la despesa final. Addicionalment, com veurem mes endavant, són dades sensibles que caldrà anonimitzar.
- La resta de característiques son numèriques i descriuen el comportament del client. A priori, seran les que utilitzarem per entrenar el model i extreure els patrons de comportament, es a dir, seran les nostres 'X' i 'y' del model.
- L'objectiu fixat es "augmentar les vendes", el que equival a maximitzar la característica 'Yearly Amount Spent', es a dir, la 'y' del model.
- El model seleccionat, un cop entrenat, ens dona també informació de quines son les característiques que tenen mes importància en el model, mitjançant els coeficients de la regressió lineal. Això permetrà abordar la segona part del objectiu, que es avaluar on concentrar els esforços (web o aplicació mòbil), donat

que les 'X' tinguin un coeficient mes gran son les que tindran més impacte en el valor final de la predicció.

Per mesurar el rendiment del model utilitzarem les mètriques habituals de una regressió lineal: MSE, MAE, RMSE i R2. Cada mètrica ens dona informació diferent i complementaria però ens fixarem més en el **RMSE**, ja que té les mateixes unitats que el valor a predir 'y' i penalitza més els error grans. Aquestes mètriques es calcularan tant sobre les dades d'entrenament como sobre les dades de test, el que ens permetrà avaluar el rendiment del model i també si l'entrenament és insuficient (*underfitting*) o s'ha sobre-entrenat el model (*overfitting*).

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?

Tindrem en compte les següents responsabilitats ètiques i socials:

- Benefici humà, tant per la persona (millor experiència de compra) com per la botiga (increment de les vendes).
- Transparència i explicabilitat, gràcies a treballar amb models coneguts (regressió lineal) que es poden avaluar amb mètriques i aporten informació de com estan calculant les prediccions (coeficients de la regressió).
- Justícia i equitat, donat que no utilitzarem dades que puguin esbiaixar els resultats, como podrien ser el gènere, edat, religió o altres de similars. Així mateix, garantirem que tenim prou dades i amb elevada diversitat.
- Privacitat i seguretat, anonimitzant (o fins i tot eliminant) dades de caràcter personal (e-mail, adreça,...) o sensible (despesa total anual,...).
- Responsabilitat per part dels desenvolupadors del model de predicció i redempció de comptes en el cas de danys o conseqüències negatives derivades del model aplicat.