

SPRINT 7.1 - Aplicacions de ML al món real: el cas de la Botiga de roba a mida

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 oa 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?

Per al problema de predir la despesa anual dels clients, s'utilitzarà un algorisme supervisat de regressió, ja que l'objectiu és predir una variable contínua (la despesa anual estimada). Es consideren les següents opcions, amb justificació:

- 1. Random Forest**
 - **Justificació:** És un model robust que gestiona bé relacions no lineals i pot identificar quines variables (com l'estat de membresia, el temps d'interacció o el nombre de compres) tenen més impacte en la despesa. A més, evita problemes de sobreajustament gràcies al seu enfocament en ensembles.
 - **Cas d'ús:** Ideal si es requereix una predicció precisa i informació sobre la importància de les característiques.
- 2. Gradient Boosting (XGBoost o LightGBM)**
 - **Justificació:** És altament precís per a tasques de predicció i és especialment efectiu amb patrons complexos en dades estructurades. Aquest tipus d'algorisme equilibra bé el poder predictiu amb la capacitat de captar interaccions subtils entre variables.
 - **Cas d'ús:** Útil per millorar les prediccions en situacions on els patrons no siguin evidents.
- 3. Regressió Lineal**
 - **Justificació:** Un model senzill i interpretable, útil per entendre si hi ha una relació lineal entre atributs dels clients (temps d'interacció, estat de membresia, etc.) i la despesa.
 - **Cas d'ús:** Adequat com a punt de partida per obtenir una base d'interpretabilitat i comparació amb models més avançats.

Mètriques d'Avaluació

Per avaluar el rendiment del model, s'utilitzaran diverses mètriques per obtenir una perspectiva completa:

1. **Mean Absolute Error (MAE)**
 - **Per què:** Proporciona l'error mitjà en unitats interpretables (dòlars/euros), ajudant l'equip de negoci a comprendre l'impacte dels errors.
2. **Root Mean Squared Error (RMSE)**
 - **Per què:** Penalitza els errors grans, fet que és útil en un escenari on les prediccions excessivament altes podrien ser costoses o enganyoses.
3. **R-squared (R^2)**
 - **Per què:** Mesura la proporció de variància explicada pel model, indicant com de bé captura els patrons generals.
4. **Mean Absolute Percentage Error (MAPE) (opcional)**
 - **Per què:** Permet expressar l'error en termes percentuals, facilitant la comparació entre clients amb nivells de despesa diferents.

Aquest conjunt de mètriques assegura una visió equilibrada, prioritzant tant la precisió com la interpretabilitat. Així, es podrà seleccionar el millor model per optimitzar les decisions empresarials i augmentar la despesa anual mitjana per client.

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? L'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?

Privadesa de les dades

Atès que el projecte implica dades sensibles sobre els clients, és essencial protegir-ne la privadesa i complir amb les normatives de protecció de dades. Això inclou:

- **Detalls del perfil del client:** Informació com els patrons de despesa, preferències personals de roba, mides corporals o estils de vida inferits a partir dels seus historials de compra. Aquests detalls són altament confidencials, especialment si s'utilitzen per personalitzar experiències o campanyes de màrqueting.
- **Dades de consulta i preferències d'estil:** Les consultories personalitzades amb estilistes poden generar registres de preferències, opinions i motivacions personals dels clients, reflectint aspectes íntims com el seu estil personal, la seva autoimatge, o esdeveniments importants de la seva vida (per exemple, casaments o canvis de carrera).
- **Historial de compres i transaccions:** Les dades de despesa ofereixen una visió detallada dels hàbits financers del client i, possiblement, del seu nivell socioeconòmic. Els patrons de despesa poden reflectir capacitat financera o preferències per certs productes i rangs de preus.

2. Seguretat i equitat en l'ús de dades

- **Dades d'interacció:** La informació sobre el temps que els clients passen al lloc web o aplicació, les seccions que visiten, els productes que veuen amb freqüència i la freqüència de les visites poden revelar aspectes de les seves rutines, hàbits o possibles limitacions financeres. Aquestes dades s'han de gestionar amb precaució per no traspasar els límits de la privadesa.

- **Estat de membresia i fidelització:** La informació sobre l'estat de membresia pot reflectir el nivell de compromís i lleialtat dels clients amb la marca, però també podria indicar hàbits de despesa o valors personals.

A més, s'han d'abordar les **pràctiques ètiques** següents:

- **Transparència i Explicabilitat:** Els clients haurien de poder entendre com i per què reben certes recomanacions o contingut dirigit. Utilitzar models interpretables i comunicar clarament el propòsit dels algoritmes milloraria la confiança en la marca.
- **Màrqueting Responsable:** La predicció i possible **influència sobre la despesa dels clients** comporta un **risc ètic**. Les recomanacions han d'enfocar-se a millorar la satisfacció del client i no a fomentar despeses excessives