

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?

- **LinearRegression()**, seria adequat si el si s'eliminen del df variables irrelevantes, ja que no inclou mecanismes per evitar el sobre ajustament.

També provaria altres com:

Regressió Ridge, doncs pot ajudar a prevenir el sobre ajustament quan hi ha moltes característiques correlacionades. És robust i evita que els coeficients es disparin amb molta diversitat de dades.

Random Forest Regressor: útil si les relacions entre les variables i la despesa anual són complexes. Pot capturar interaccions entre variables que altres models més simples no podrien identificar. És robust amb dades mixtes (variables numèriques i categòriques), maneja bé dades amb valors perduts o poc equilibrades.

Les que no crec que necessitessin perquè el df no es veu massa complicat son: **Regressió Lasso** ni **Support Vector Regressor (SVR)**. Crec que són molt evidents les variables a tenir en compte i les que no ens aporten informació (color de l'avatar, correu electrònic, etc...).

Després d'entrenar el model amb **Linear Regression, Random Forest o Regressió ridge**, passariem a validar i avaluar el model utilitzant tècniques com la **cross-validation** (configurant el k-fold) i aplicant mètriques d'error com el **MSE** i el **R²** (creant scorers.).

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?

Completar pels estudiants