

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

Introducció:

Es tracta d'una botiga exclusiva especialitzada en confecció i venda de roba a mida. Els clients visiten la botiga per rebre assessorament directe d'estilistes experts que ajuden a crear peces úniques q s'ajustin perfectament a les seves preferències i mesures

Després de la sessió de consultoria, els clients poden fer comandes de roba per mitjà d'una app o a la web de la botiga

Objectius projecte:

Augmentar les vendes avaluant on concentrar els esforços: millorant l'experiència dels clients a l'app mòbil o a la web

Dades disponibles:

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 a l'app mòbil com a la web i l'estat dels membres

Que és vol millorar 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

La mètrica de l'èxit d'aquest projecte es l'augment en la despesa anual mitjana per client. Aquesta mètrica reflectiria directament l'efectivitat del model en millorar les decisions de l'empresa

Quin es l'algorisme de ML més adequat per resoldre aquest problema? Quines mètriques d'avaluació s'utilitzaran per mesurar el rendiment del model?

De la informació del projecte de la botiga, podem identificar quins són els paràmetres objectius plantejats:

- Predicció de despesa anual, a partir d'un model que estimi la despesa anual x client en funció del seu comportament
- Identificació de millores, determinant on s'han d'enfocar les actuacions prioritàries, en l'app mòbil o a la web
- Segmentació de clients en funció del seu estat (actius, inactius, categoria (VIP, normal, etc...))

A partir dels objectius, com el problema plantejat és una **predicció contínua** (estimació de la despesa anual del client), es tracta d'un problema de **regressió**.

El primer plantejament ens podria portar a plantejar un escenari amb un algorisme de **regressió lineal múltiple**, en funció de les relacions entre les variables independents (temps en l'app, temps a la web, estat de membres, etc.) i la variable dependent (despesa anual) ja que les dades no són massa complexes, el model és simple, interpretable i és el que consumeix menys recursos de càlcul.

Dividiríem les dades disponibles en un conjunt d'entrenament (80%) i un conjunt de prova (20%) i posteriorment, avaluaríem el model amb les dades de l'entrenament amb les mètriques apropiades per a un problema de regressió com serien el **MAE** (Mean Absolute Error), el **MSE** (Mean Squared Error) que penalitza errors més grans que MAE i el coeficient de determinació **R²**, que mesura quin percentatge de la variabilitat de la despesa anual explica el model (un valor proper a 1 indica que el model s'ajusta bé)

En aquest punt, faríem una interpretació del resultat de MAE, MSE i R² i les gràfiques de la variable dependent i independents per veure com s'ajusta l'algorisme seleccionat al problema a resoldre.

Si el resultat no fos l'adequat, utilitzaríem altres algorismes per al model de regressió com podrien ser **Random Forest**, **XGBoost**, etc..

Quines responsabilitats ètiques i socials és important tenir en compte?

És important tenir en compte aspectes ètics i socials durant el desenvolupament i la implementació del projecte, ja que s'utilitzen dades personals dels clients.

En concret, s'ha de assegurar la privacitat i protecció de les dades, de forma que el projecte compleixi amb el RGPD (Reglament General de Protecció de Dades) i s'anonimitzin les dades per evitar utilitzar dades identificatives que puguin comprometre la privacitat dels clients.

És bàsic informar als clients de forma clara com es faran servir les seves dades, demanar el seu consentiment explícit i indicar com exercir els drets recollits al RGPD com el dret a la rectificació, supressió, oposició, etc..

També assegurar que no es produeix cap discriminació cap a certs grups de clients, ni per qüestió de característiques físiques com edat, pes, raça ni en funció del seu status social o econòmic.

Per últim, s'ha de considerar que l'augment de la despesa anual per client no resulti d'estratègies per manipular al client per augmentar la despesa, sino que han de resultar de la millora de l'experiència global pels clients.