

## Metodologia proposta

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?

### Algoritme escollit

L'algoritme triat és el Random Forest Regressor, que es basa en una combinació d'arbres de decisió per millorar la precisió i la robustesa de les prediccions. Aquesta tècnica pertany a la categoria dels mètodes d'ensamble, ja que utilitza múltiples arbres per obtenir resultats més fiables.

Aquest mètode construeix diversos arbres de decisió independents, en el que:

- Se selecciona un subconjunt aleatori de dades: per la construcció de cada arbre es fa servir una mostra aleatòria del conjunt de dades.
- Hi ha un subconjunt de característiques: en cada node només es considera un subconjunt aleatori per determinar el millor punt de divisió.
- El resultat de la predicció és la mitjana de totes les prediccions individuals dels arbres per millorar la precisió.

### Per què escollim aquest algoritme

El Random Forest crea múltiples arbres per trobar patrons que associïn les variables d'entrada a la despesa anual, ponderant-les segons la seva influència. Així, es podrà veure quines d'aquestes variables són més importants per a la despesa.

Aquest algoritme té determinades característiques que fan idònia la seva aplicació a aquest problema concret:

- Dades heterogènies: les dades inclouen variables continues i categòriques (i.e. temps dedicat i estat del membre).
- No linearitat: la relació entre variables d'entrada pot no ser lineal, com el temps que s'empra a l'aplicació i l'estat del membre.
- Sobreajustament: per reduir aquest risc, utilitzen característiques aleatòries.
- Escalabilitat: pot utilitzar grans volums de dades
- Rellevància: si un conjunt de dades no és rellevant, l'arbre menys efectiu tindrà menys importància perquè el model final es basa en la mitjana.

### Mètriques d'avaluació per mesurar el rendiment del model.

- Coefficient de determinació. Indica la proporció de la variabilitat. Un valor proper a 1 indica que el model explica tota la variabilitat i un valor proper a 0 indica que el model no té cap poder predictiu.
- Root mean square error. Mesura l'error en les prediccions penalitzant més els errors més grans. Indica la precisió global del model.
- Error absolut mitjà. És menys sensible per outliers però més robusta que l'error mitjà.
- Error percentual mitjà absolut. Es calcula l'error absolut com un percentatge respecte als valors reals. S'utilitza per indicar el rendiment del model en termes relatius.
- Validació creuada. Divideix el conjunt de dades en diversos subconjunts i entrena el model en un conjunt a l'hora que el valida en un altre.

## **Responsabilitats ètiques i socials**

Quines responsabilitats ètiques i socials és important tenir en compte?

### **Transparència**

Privacitat i ús de les dades personals: L'ús d'IA per optimitzar les decisions empresarials exigeix que els models siguin transparents. Per tant, s'ha de garantir que la informació es recopila amb el consentiment explícit dels clients i que aquests saben com s'utilitzarà.

### **Privacitat i seguretat de les dades personals**

Confidencialitat: les dades personals dels clients han de ser protegides amb protocols de seguretat robustos per evitar accessos no autoritzats.

Compliment legal: l'ús de les dades ha de respectar normatives i les lleis, per assegurar que es gestionen de manera ètica i legal.

A més, l'empresa ha de garantir que les dades es facin servir exclusivament per millorar l'experiència dels clients i no per altres finalitats no autoritzades.

### **Impacte social**

S'han de tenir en compte les discriminacions, ja que el model resultant podria perjudicar alguns clients. Per tant, la botiga hauria d'implementar procediments per determinar els biaixos, garantint una experiència equitativa.

### **Responsabilitat de les decisions automatitzades**

Manipulació del comportament del client: un model de IA podria suggerir estratègies per augmentar la despesa mitjana per client, com ara promoure la venda d'articles més cars. No obstant això, l'empresa ha d'assumir la responsabilitat d'assegurar-se que aquestes estratègies no resultin manipuladores o deslleials. Això implica equilibrar els objectius comercials amb la satisfacció i els interessos reals dels clients.

Per exemple, l'algorisme pot identificar patrons de despesa i oferir recomanacions personalitzades, però s'ha d'evitar utilitzar aquesta informació per influir en el comportament del client. Per exemple, fomentar compres innecessàries.

### **Impacte social i econòmic**

Reducció de l'ocupació: tot i que es poden automatitzar tasques d'anàlisi de dades, la botiga pot utilitzar aquest model de IA per complementar i no substituir els estilistes experts, per reforçar el valor humà de la consultoria personalitzada.