

## SPRINT 3.2 - Aplicacions de ML al món real: el món bancari

### Predicció de subscripcions a dipòsits a termini en clients bancaris.

---

#### *Nom del projecte*

Predicció de subscripcions a dipòsits a termini en clients bancaris.

#### *Introducció*

Aquest projecte pretén millorar l'eficàcia de les campanyes bancàries preveient quins clients subscriuran dipòsits a termini, optimitzant així els recursos i augmentant la rendibilitat.

#### *Objectius del Projecte*

1. **Objectius del negoci:** Augmentar la taxa de conversió i reduir el cost per client potencial en les campanyes de màrqueting.
2. **Millorar processos amb ML:** Automatitzar la identificació de clients amb alta probabilitat de subscripció per personalitzar l'abast de les campanyes.
3. **Alternativa manual:** Donat el volum i complexitat de les dades, una anàlisi manual seria massa lenta i poc precisa.

#### *Metodologia Proposta*

Es proposa utilitzar els models **Random Forest** i **Regressió Logística** per la seva eficàcia en tasques de classificació binària:

- **Random Forest** és robust davant dades heterogènies, variables categòriques i numèriques, i pot capturar relacions complexes entre els atributs.
- **Regressió Logística** és adequada per a interpretació i permet entendre clarament la contribució de cada variable, mantenint una alta precisió en la predicció de resultats binaris.

#### *Dades Disponibles*

El conjunt conté més de 4.500 registres amb atributs com edat, ocupació, estat civil, nivell educatiu, saldo, i variables de contacte amb el banc.

**age** (edat): L'edat del client.

**job** (ocupació): La professió o tipus de feina del client.

**marital** (estat civil): Estat civil del client (casat, solter, divorciat).

**education** (educació): Nivell d'educació del client (primària, secundària, terciària).

**default** (morositat): Si el client té deutes impagats (sí/no).

**balance** (saldo): Saldo bancari del client.

**housing** (hipoteca): Si el client té un préstec hipotecari (sí/no).

**loan** (préstec): Si el client té un préstec personal (sí/no).

**contact** (contacte): Mètode de contacte utilitzat per arribar al client (telèfon, mòbil, etc.).

**day** (dia): Dia del mes quan el client va ser contactat.

**month** (mes): Mes quan el client va ser contactat.

**duration** (durada): Durada de la conversa amb el client en segons.

**campaign** (campanya): Nombre de contactes realitzats durant la campanya actual.

**pdays** (dies des de l'últim contacte): Nombre de dies des del darrer contacte amb el client.

**previous** (contactes previs): Nombre de contactes realitzats en campanyes anteriors.

**poutcome** (resultat de la campanya anterior): Resultat de la darrera campanya (èxit, fracàs, etc.).

**deposit** (dipòsit): Variable objectiu, si el client ha subscrit un dipòsit a termini (sí/no).

## **Mètrica d'èxit del projecte**

La **mètrica d'èxit** per al projecte pot centrar-se tant en les **mètriques de rendiment** del model d'aprenentatge automàtic (com l'accuràcia, precisió, recall, F1-score) com en els **resultats relacionats amb el negoci** (per exemple, l'augment de la subscripció de préstecs o dipòsits). Les mètriques del model (que inclouen falsos positius, falsos negatius, etc.) són importants per avaluar la precisió de les prediccions, però l'**impacte empresarial**, com l'increment de la subscripció de préstecs, seria la mesura d'èxit final

Les mètriques de rendiment seran **Accuracy, F1 Score i ROC-AUC, matriu de confusió**. L'**Accuracy** ens mostra la proporció de prediccions correctes, mentre que el **F1 Score** ajuda a equilibrar precisió i sensibilitat, crucial per detectar clients d'alt potencial. La **ROC-AUC** indica la capacitat del model per discriminar entre els clients interessats i no interessats.

**Confusion Matrix.** La **matriu de confusió** ajuda a avaluar el rendiment d'un model de classificació mostrant quants casos s'han classificat correctament i incorrectament:

- **Veritables Positius (VP):** Prediccions positives correctes.
- **Falsos Positius (FP):** Prediccions positives incorrectes.
- **Veritables Negatius (VN):** Prediccions negatives correctes.
- **Falsos Negatius (FN):** Prediccions negatives incorrectes.

Aquestes dades permeten calcular mètriques com la precisió, el *recall* i l'accuràcia del model.

## **Responsabilitats Ètiques i Socials**

És essencial respectar la privacitat, assegurant que les dades siguin anonimitzades. També cal evitar biaixos que discriminin clients per factors com edat, ocupació o saldo, per evitar exclusions injustes en les campanyes.