

## Challenge Task (Lesson 15)

**Scop:** Construirea unui model de clasificare utilizând un dataset de predicție a clienților băncii care vor intra în default și recapitularea cunoștințelor teoretice acumulate până în acest moment.

### Descriere:

Informații detaliate despre dataset pot fi găsite în notebook-ul lecției 15, accesibil prin acest [link](#). Datasetul este disponibil în aceeași mapă sub numele `data/CreditScoring_final.csv`.

### Etapele recomandate:

*Explorarea Datelor (EDA)* – Începeți cu analiza exploratorie pentru a înțelege structura și caracteristicile datelor.

*Feature Engineering* – Pe baza observațiilor din EDA, efectuați transformări și creați caracteristici noi care pot îmbunătăți performanța modelului.

*Modelare* – După ce datele sunt preprocesate, testați diverse modele de clasificare și evaluați performanța acestora folosind metrica *F1-score*. Experimentați cu diferite hiperparametri pentru optimizare. Nu există restricții cu privire la modelele pe care le puteți utiliza, însă trebuie să selectați un singur model pentru evaluarea finală.

### Atenție:

- Evitați *overfitting-ul*, asigurându-vă că modelul se generalizează bine atât pe seturile de training, cât și pe cele de validare.
- Consultați [acest video](#) despre procesul de cross-validare pentru a înțelege mai bine cum să alegeți parametri optimi ai modelului.
- Evitați *data leakage*, adică utilizarea informațiilor din setul de validare în timpul transformărilor efectuate pe setul de training.

**Evaluare:** Performanța finală va fi evaluată pe un set de date separat, care nu a fost folosit în antrenare sau validare. Va trebui să trimiteți notebook-ul complet utilizat în această sarcină, care să conțină toate transformările și procesul de antrenare al modelului, structurat astfel încât să îl pot rula pentru evaluare.

Updated 1: Pentru a asigura replicabilitatea transformărilor pe setul de date de test, **vă rugăm să utilizați pentru toate transformările pipeline din sklearn** sau să definiți o funcție separată care să efectueze toate transformările de preprocesare pe setul de date de test, evitând *data leakage*.