

# BDMA Hackathon

2021 Edition

# Planning

Jour 1 - 19/04	
8H30	« Arrivée / connexion » des étudiants présentation des consignes et du sujet
9H00	Ouverture du serveur d'évaluation / leaderboard
12H00-1 3H15	Pause déjeuner
18H00	Fermeture du serveur d'évaluation
19H00	Le code doit être "zippé" et rendu sur celene, si trop lourd (> 20Mo) déposer un lien git
Jour 2 - 22/04	
13H00	Début des soutenances (15 minutes par groupe)

# The problem

## “Churn”

We are working for a telecom “Green”, their goal is to predict customers that will leave for their competitors soon and deploy marketing tools (special promotions, etc) to avoid this

## Task

Predict if the customer is leaving or not.

## Data

- 5786 instances :
  - 1534 Yes
  - 4252 No
- Additional data for some customer (who have an internet plan)

# The data - Main Table

id\_client: identifiant unique du client

mariee: le client est il marié

retraite: le client est-il à la retraite

a\_charge: le client a t il un ou plusieurs enfants à charge

genre: Homme ou Femme

facture\_mensuelle: total de la dernière facture mensuelle du client

telephone: le client a t il une ligne téléphonique

plusieurs\_numeros: le client a t il plusieurs lignes téléphoniques

internet: le client a t il une connexion internet si oui de quel type

total\_factures: la somme de toutes les factures du client

contrat: type d'engagement

facture\_par\_mail: le client souhaite recevoir ses factures par mail (au lieu de la poste)

type\_de\_paiement: moyen de paiement utilisé par le client

client\_depuis\_mois: durée depuis laquelle le client est chez "Green"

**Sortie\_client -> la valeur à prédire, le client est il partis ?**

# The data - services table

id\_client: identifiant unique du client

Sauvegarde: le client a souscrit à l'option sauvegarde en ligne

Antivirus: le client a souscrit à l'option antivirus

Assurance: le client a souscrit à l'assurance de son mobile

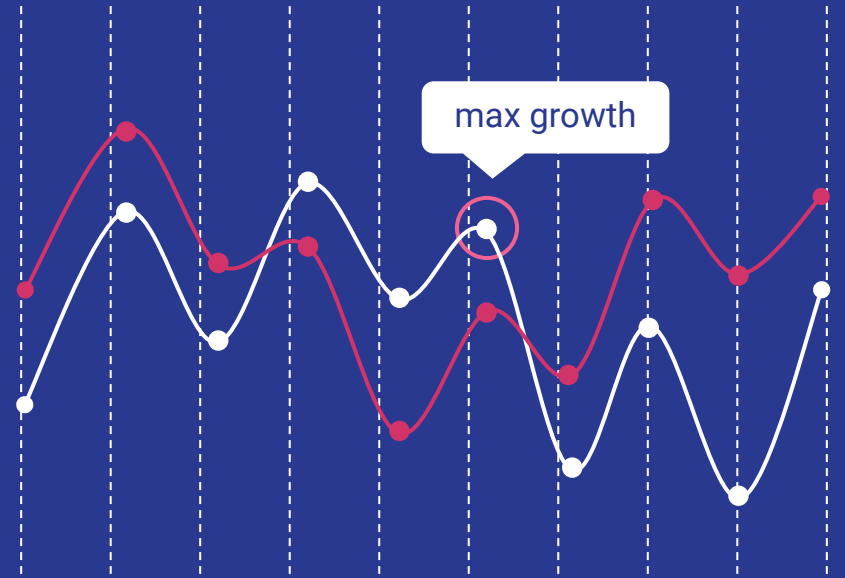
Streaming film: le client a souscrit à l'option streaming de films

Streaming series: le client a souscrit à l'option streaming de séries

Support technique: le client a souscrit à l'option assistance technique

# Evaluation

F1 measure



# F1 Score Explained

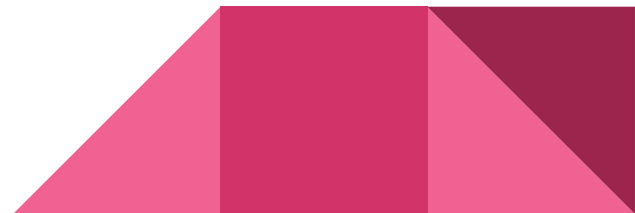
- Harmonic mean between precision and recall
- Allows to balance recall and precision
  - Indeed losing customer is expensive so a good recall is necessary
  - Bu marketing is expensive too and we don't want to send special offers to everyone if we can spare it

$$F_1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{\text{tp}}{\text{tp} + \frac{1}{2}(\text{fp} + \text{fn})}$$

# Evaluation

- Baseline /5 : 2.5 pts if  $F1 > 0.525$ , 5 pts if  $F1 > 0.575$
- Competition /5: 1st place 5 pts, 2nd 4 pts, 3rd 3 pts, ...
- Oral presentation /10: (individual) see grid below

Compréhension des modèles utilisées	3
Qualité du code (commentaires, orga du git, ...)	3
Capacité à travailler en équipe (répartir les tâches ...)	2
Compréhension et regard critique sur le sujet	2





# Submit your results

- There are 200 points with unknown class in 'telecom\_users\_eval.csv'
- Create a .txt file with 2 line
  - First line is your group token (provided by a teacher)
  - Second line are the positive and negative class predictions represented by 0 and 1 in the same order as the 200 points from the csv
  - 'example.txt ' contains an example
- Submit your file on <https://hackathon.lifat.fr/>
- Celene link <https://celene.univ-tours.fr/course/view.php?id=4007#section-1>

