

# Responsible and Explainable AI

## TP 1 –Données Tabulaires

Par groupes de 2 élèves, vous devez prendre en main le jeu de données que l'on vous met à disposition (sans le rendre public, sans le transmettre svp) et nous faire une restitution de vos travaux sous la forme d'un notebook.

Le jeu de données contient des données de ressources humaines anonymisées. Il est au format pickle et contient un DataFrame Pandas X et un vecteur Y avec les classes:

- la classe 0 indique que l'employé n'a pas démissionné au cours des 6 prochains mois
- La classe 1 indique que l'employé a démissionné au cours des 6 prochains mois.

Afin de vous guider dans la rédaction de votre notebook, nous vous demandons qu'il contienne **au moins** les parties suivantes:

### 1) Analyse du jeu de données

Utilisez ce que nous avons vu pendant les TPs. En utilisant les librairies panda et seaborn, indiquez par exemple si le jeu de données contient des données manquantes, explorez les différentes colonnes (type, modalités, plages de valeurs, cardinal, distributions, etc). Commentez.

Exhibez le parcours d'un employé qui a démissionné et celui d'un employé qui n'a pas démissionné. Commentez.

Affichez la matrice de corrélation. Commentez.

Toute autre analyse que vous jugez pertinente est la bienvenue (avec commentaires)

### 2) Apprentissage automatique

- Entraînez 3 modèles transparents au choix. (justifier votre choix)
- Décrivez et commentez les résultats.

### 3) Explication post-hoc

Appliquez une méthode d'explication post-hoc et réduisez les colonnes de votre jeu de données. Commentez.

### 4) Apprentissage automatique

- 5) Réentraînez les modèles avec le nouveau jeu de données réduit. Commentez les résultats.