

# Apprentissage non supervisé

## Projet sous le logiciel R - 2024-2025

L'objectif du projet est d'implémenter un algorithme des k-médoïdes sous R.

**L'algorithme** Les données sont supposées vivre dans  $\mathbb{R}^p$ , et l'algorithme de clustering considéré est l'algorithme des k-médoïdes présenté en cours. Votre fonction de clustering devra prendre en entrée :

- les données,
- le nombre  $K$  de clusters,
- le type d'initialisation : `random` ou `kmeans++`,
- le nombre d'initialisations,
- une variable booléenne `plot`, qui si elle est égale à `TRUE` permettra de représenter graphiquement les données (dans le premier plan de l'ACP) avec une couleur différente pour chaque cluster.

En sortie, l'algorithme devra retourner :

- les coordonnées des k médoïdes,
- l'affectation des données aux différents clusters.

**Expérimentation** Vous devrez expérimenter votre algorithme sur les données iris. Vous comparerez les différentes stratégies d'initialisation en terme de résultat et de temps de calcul.

**Livrable (par binôme pour le 28/02/2025)**

- un script Rmarkdown (ainsi que sa compilation pdf) qui contiendra l'ensemble des éléments demandés : le code de votre fonction et son expérimentation sur les données iris.

**Consignes** Il est interdit d'utiliser ChatGPT ou de récupérer un code sur internet. Vous devez créer votre propre code. Dans le doute, une audition pourra être organisée pour échanger sur le contenu du code.