

# Introduction à l'apprentissage non supervisé

Examen du 23 janvier 2025

Durée : 2h.

Vous commencerez par répondre aux questions théoriques, sur papier, sans aucun support documentaire. Ensuite, vous enchainerez sur la partie pratique, où là les supports seront autorisés.

## 1 Questions théoriques (30')

- Qu'est-ce que l'apprentissage non supervisé et comment se distingue-t-il de l'apprentissage supervisé ?
- Qu'est-ce que l'algorithme K-means et quelles sont ses principaux avantages ? Lui voyez-vous des limitations ?
- Qu'est-ce qu'une composante principale en ACP ? A quoi peut-elle servir ?
- Décrivez les différences entre une méthode de classification ascendante hiérarchique (CAH) et la méthode des K-means.
- Quels sont les avantages de DBSCAN par rapport à K-means ?
- Quels sont les défis liés à l'évaluation des performances d'un modèle d'apprentissage non supervisé ?
- Comment peut-on déterminer un nombre optimal de clusters avec l'algorithme des K-means ?

## 2 Application (90')

Le fichier des données suivant contient, pour un grand nombre de pays et pour chaque jour du 3 janvier 2020 au 13 mai 2022, le nombre de nouveaux cas de COVID19, le nombre cumulé de nouveaux cas, le nombre de nouveaux décès, ainsi que le nombre cumulé de décès.

<https://eric.univ-lyon2.fr/jjacques/Download/DataSet/2022-05-13-WHO-COVID-19-global-data.csv>

Sur la base de ces données, réaliser un clustering de ces pays à l'aide de la méthode de votre choix, et donner une interprétation des clusters. Vous pourrez vous aider de représentations graphiques.

Votre rendu sera un fichier pdf contenant les résultats de votre analyse ainsi que les codes R utilisés, à renvoyer par mail à [julien.jacques@univ-lyon2.fr](mailto:julien.jacques@univ-lyon2.fr)