

Projet : Partie 5 **Algorithmes de Clustering**

Le clustering est le processus de partitionnement d'un ensemble d'objets de données (ou d'instances) en sous-ensembles. Chaque sous-ensemble est un cluster, de sorte que les objets d'un cluster sont similaires les uns aux autres, mais différents des objets des autres clusters. Dans ce contexte, différentes méthodes de clustering peuvent générer différents clusterings sur le même ensemble de données. Le partitionnement n'est pas effectué par des humains, mais par l'algorithme de clustering. Par conséquent, le regroupement est utile dans la mesure où il peut conduire à la découverte de groupes jusque-là inconnus dans les données. L'analyse de cluster est largement utilisée dans de nombreuses applications telles que l'informatique décisionnelle, la reconnaissance de formes d'images, la recherche sur le Web, la biologie et la sécurité, etc.

De nombreux algorithmes de clustering ont été proposés afin de découvrir des regroupements utiles dans de larges banques de données. Ainsi, dans cette cinquième et dernière partie du projet, il vous est demandé d'implémenter, tester et comparer des algorithmes de clustering permettant de trouver des partitionnements intéressants à partir du dataset 1 après pré-traitement et en éliminant la classe, soit l'attribut "Attrition". Il vous est alors demandé de :

- A. Application d'algorithmes de clustering basé densité:
 - a. Programmer l'algorithme de clustering "DBSCAN"
 - b. Expérimentation des paramètres de DBSCAN sur les instances du dataset 1.
 - c. Évaluer, comparer et analyser les résultats de DBSCAN.
 - d. Illustrer par des exemples et graphes.
- B. Application d'algorithmes de clustering basé hiérarchie:
 - a. Programmer l'algorithme de clustering "AGNES"
 - b. Expérimentation des paramètres de AGNES sur les instances du dataset 1.
 - c. Évaluer, comparer et analyser les résultats de AGNES.
 - d. Illustrer par des dendrogrammes et graphes.
- C. Comparer les deux algorithmes de clustering DBSCAN et AGNES.
- D. Intégrer l'exécution de ces algorithmes à l'IHM.

Notes:

- Chaque binôme est tenu d'envoyer la version électronique du rapport au plus tard: **Samedi 07 Janvier 2023 à 23h59.**
- Chaque binôme devra présenter son interface et code source durant la semaine du **Dimanche 08 Janvier 2023 au Jeudi 12 Janvier 2023.**

Bon courage !