

La fouille de données ou "data mining" est un processus qui permet de découvrir des connaissances inconnues au préalable, enfouies dans une masse volumineuse de données.

Dans le cadre du tp N°2 il est question de :

- lire un benchmark dans le but d'effectuer le clustering de ses données : il faut évidemment procéder à la phase de preprocessing des données d'abord
- tracer la courbe d'elbow dans le but de trouver un nombre de clusters enfouis dans la population
- appliquer l'algorithme K-Means avec $k = \text{nbre de clusters trouvés par la courbe d'Elbow}$

Puis calculer les mesures de performances des clusters obtenus

-appliquer l'algorithme K_Medoids pour le meme parametre K

Puis calculer les mesures de performances des clusters obtenus

-etablir un histogramme des inerties des deux methodes et comparer ainsi les performances de chaque methode.

Rapport à remettre le 17 /03/2024