

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 de ce tp il est question de :

- 0- Installer l'environnement anaconda qui contient le package SkLearn
- 1- Manipulation et exploration d'un fichier d'apprentissage
 - ouverture de benchmarks
 - lecture et affichage des données
 - nbre d'instances
 - nom des attributs et leur nbre et leur type
 - les 5 nbres pour un attribut :
 - Le Min, le max, le Median , Q1 et Q3
 - Tracer les boxplots de chaque attributs sur le meme graphe
 - Afficher le Scatter plot du benchmark
 - Afficher le mode et le mean et le median pour chaque attribut
 - Reperer les valeurs manquantes si elles existent et les remplacer
 - Normaliser les données par MIN/MAX et par Zscore
- 2- Préparer l'interface graphique du TP : IHM en python contenant les boutons suivants : - bouton « Ouvrir » (qui affiche un ensemble de benchmark) pour en choisir un et qui lancera l'application en meme temps, et executera cette partie du tp
 - Bouton « Clustering »
 - Bouton « Classification supervisée »
- 3- Ecrire un rapport détaillé des resultats de cette 1ere partie du TP, à rendre le DIMANCHE 10MARS 2024