

Etude exploratoire d'un Dataset

Partie 1

Le but de cette première partie du projet est de mettre en application les concepts vus en cours quant à l'extraction des caractéristiques descriptives d'un jeu de données (Dataset). A cet effet, considérer le Dataset "Seeds.txt"¹ et exécuter les actions suivantes :

- I. Etudier le dataset "Seeds.txt" :
 1. Manipuler le dataset et visualiser son contenu.
 2. Fournir une description du dataset et de chacun de ses attributs.

- II. Effectuer une étude exploratoire des caractéristiques descriptives du dataset "Seeds.txt" :
 1. Pour chaque attribut :
 - Calculer les mesures de tendance centrale (moyenne, moyenne tronquée, médiane, mode et le milieu de l'étendue) et en déduire les symétries.
 - Calculer les mesures de dispersion (étendue, quartiles, écart interquartile, variance et l'écart type) et en déduire les données aberrantes (outliers).
 - Construire une boîte à moustache et afficher les données aberrantes.
 - Construire un histogramme et visualiser la distribution des données.
 2. Construire et afficher des diagrammes de dispersion des données et en déduire les corrélations entre les attributs.

- III. Concevoir et programmer en Java une application permettant de réaliser les actions I. et II.

- IV. Remettre un rapport du travail effectué le **25 Novembre 2021**.

¹ Lien de téléchargement : <https://archive.ics.uci.edu/ml/datasets/seeds>