

Projet : Partie 1

Analyse des Données

Les données du monde réel sont généralement bruitées, d'un volume énorme et peuvent provenir de sources hétérogènes. C'est pourquoi la première étape du Data Mining consiste en l'analyse (ou étude exploratoire) des données. Cela implique d'examiner de plus près les attributs et les valeurs des données.

L'analyse des données fait référence à: l'analyse des types d'attributs qui composent vos données, des types de valeurs de chaque attribut, de la nature des attributs (discrets ou continue), de la distribution des valeurs, de fournir des résumés visuels des données, de l'existence de valeurs aberrantes, de redondance, ou de valeurs manquantes, de la mesure de similarité de certains objets par rapport à d'autres, ...etc

Ainsi, dans cette première partie du projet, il vous est demandé de vous familiariser avec vos données et d'en extraire un maximum d'informations. Le travail requiert la conception et l'implémentation d'une application JAVA (IHM) permettant au moins de:

- A. Manipulation de datasets:
 - a. Importer et visualiser le contenu d'un dataset.
 - b. Fournir une description globale du dataset.
 - c. Fournir une description de chaque attribut.
 - d. Mise à jour d'une instance ou d'une valeur du dataset.
 - e. Enregistrement du dataset après mise à jour.
- B. Analyse des caractéristiques des attributs du dataset sélectionné:
 - a. Pour chaque attribut :
 - i. Calculer les mesures de tendance centrale et en déduire les symétries.
 - ii. Calculer les mesures de dispersion et en déduire les données aberrantes (outliers).
 - iii. Construire une boîte à moustache et afficher les données aberrantes.
 - iv. Construire un histogramme et visualiser la distribution des données.
 - b. Construire et afficher des diagrammes de dispersion des données et en déduire les corrélations entre les attributs.

Notes:

- Le rapport devra contenir l'analyse du **Dataset 1** transmis par mail.
- Chaque binôme est tenu d'envoyer la version électronique du rapport au plus tard: **Samedi 15 Octobre 2022 à 23h59.**
- Chaque binôme devra présenter son interface et code source le **Dimanche 16 Octobre 2022** durant la séance de TP.

Bon courage !