

PROJET DE TP

DATA MINING

1ere Partie : Phase Exploratoire des Données

On dispose d'un ensemble de données caractérisées par un ensemble d'attributs et on désire mener une fouille de données sur cet ensemble.

Avant d'entamer tout processus de fouille de données il est impératif d'explorer ces données et d'avoir une vue d'ensemble sur ces données. Pour cela, il faudra examiner les attributs qui caractérisent chaque instance de l'ensemble de données en question. Les attributs peuvent être nominaux, booléens, numériques,... et ils peuvent être erronés, manquer de valeurs, ...

Dans cette phase du TP il est question de :

- Recenser tous les attributs du DataSet, les normaliser et de remplacer les valeurs manquantes : celles ci sont à remplacer par la moyenne de l'attribut des instances appartenant à la même classe que l'instance qui possèdent la (ou les) valeur manquante.
- Lecture du dataset . Utiliser le package WEKA disponible sur le net gratuitement. (URL <https://www.cs.waikato.ac.nz/ml/weka/> )
- Affichage du Data set
- Affichage de quelques informations telles que le nombre d'instances, nombre d'attributs et leurs types
- Affichage des 5 chiffres pour chaque attribut : Min, Max, Median, Q1, Q3, ainsi que leurs MidRanges et Modes
- Utiliser MATLAB pour dessiner les BOXPlots pour chaque attribut et sur la meme courbe
- Fournir des histogrammes pour l'analyse des données
- Quels sont les attributs pour lesquels la fonction Mean-Mode = 3 (mode-median) ? conclure pour les données du data set : Symetric ? Positively skewed ? Negatively skewed ?

- Commencer par construire une IHM en JAVA permettant de visualiser toutes les opérations effectuées.

Cette 1ere phase est à finaliser et tester ENTRE LE 31 OCT et le 05 Nov 2018 avec remise d'un rapport détaillé.