|  |  |
| --- | --- |
| Université des Sciences et de la Technologie Houari Boumediene  Faculté d'Electronique et d'Informatique  Département d'Informatique | |
| Master 2 : SII | Année Universitaire : 2020/2021  Module Data Mining |

**Etude Exploratoire d'un Dataset**

Partie 1

Le but de ce premier projet est de mettre en application le prétraitement des données concernant l'extraction des caractéristiques descriptives des données. A cet effet considérer le Dataset "Thyroid\_Dataset.txt" et exécuter les actions suivantes :

1. Etudier le benchmark contenu dans le fichier "Thyroid\_Dataset.txt"
   1. Donner une description des entrées et de chacun de ses attributs
   2. Manipuler le data set (ouverture, lecture, affichage,...).
2. Effectuer une étude exploratoire des caractéristiques descriptives des données :
   1. Calculer la moyenne, la médiane et le mode pour les attributs 'T3-resin uptake test', 'Total Serum thyroxin' et 'Total serum triiodothyronine'.
   2. Les données de ces attributs sont-elles symétriques ?
   3. Pour chaque attribut, construire une boite à moustache. En déduire les données aberrantes (outliers).
   4. Fournir des histogrammes pour l'analyse des données.
   5. Construire des diagrammes de dispersion des données pour extraire des informations comme des regroupements, des corrélations et des données aberrantes.
   6. Existe-il des corrélations entre des attributs ?
3. Concevoir et programmer en Java une IHM permettant de visualiser le Dataset ainsi que les résultats des différentes statistiques effectuées.
4. Remettre un rapport du travail effectué le **20 Janvier 2021**.