

## Techniques de Datamining

### Partie 2

Le but de la deuxième partie du projet du cours de datamining est de mettre en application certaines techniques de datamining vues en cours. A cet effet, le dataset "Seeds.txt" étudié dans la première partie du projet, servira à l'implémentation de ces techniques. Le travail demandé est jalonné comme suit :

#### **A : Prétraitement des données**

- I. Programmer les deux méthodes suivantes de normalisation des données, puis les appliquer sur les attributs du dataset :
  1. Normalisation avec Min-Max.
  2. Normalisation avec Z-score.
- II. Programmer les deux méthodes suivantes de discrétisation des données, puis les appliquer sur les attributs du dataset.
  1. Discrétisation en classes d'effectifs égaux.
  2. Discrétisation en classes d'amplitudes égales.

#### **B : Extraction des motifs fréquents, des règles d'association et des règles de corrélation**

- I. Programmer l'algorithme Apriori, puis l'appliquer sur les instances du dataset.
- II. Programmer l'algorithme Eclat, puis l'appliquer sur les instances du dataset.
- III. Programmer l'algorithme d'extraction de règles d'association (support, confiance), puis l'appliquer sur les instances du dataset.
- IV. Programmer l'algorithme d'extraction de règles de corrélation (support, confiance, lift), puis l'appliquer sur les instances du dataset.

#### **C : Classification supervisée des instances du dataset**

- I. Programmer les deux algorithmes de classification Naïve Bayésienne et KNN, puis les appliquer sur les instances du dataset.
- II. Construire la matrice de confusion, puis comparer les deux algorithmes en termes d'Exactitude (accuracy), de sensibilité (sensitivity), de spécificité (specificity), de précision, de rappel et de F-score.

#### **D : Visualisation**

- I. Intégrer les programmes des sections A, B & C dans l'IHM développée dans la première partie du projet.
- II. Illustrer à travers l'IHM :
  1. Les motifs fréquents, les règles d'association et les règles de corrélation.
  2. La classification supervisée des instances.
  3. Le temps d'exécution des programmes.

Remettre un rapport du travail effectué le **02 Janvier 2021.**