

Faculté des Nouvelles Technologies de l'Information et de la Communication

Département : Informatique Fondamentale et ses Applications

Année Universitaire : 2023/2024

Module : DAMI (Master 1 SDIA)

TD N° 2.2 : Méthodes de prétraitement et préparation de données

❖ **Exercice 1 : Caractéristiques des Données**

1. Donnez un exemple relatif pour chaque type de données étudiées en cours.
2. Peut-on convertir des données non structurées en données structurées ?
3. Dans le cas favorable, comment peut-on assurer cette conversion ?
4. Donner un exemple concret qui peut aborder cette conversion.

❖ **Exercice 2 : Caractérisation des Attributs**

1. Créez un tableau clair et bien structuré répertoriant les divers types de variables (attributs) en les définissant concrètement, et illustrez chaque type avec un exemple.
2. Soit une base de données concernant les employés. Donner le type de chacun des attributs suivants : ID, Nom, date de naissance, genre (Homme ou Femme), situation familiale, nombre d'enfants, poids, taille, degré de confort (moyen, bon, excellent ..)

❖ **Exercice 3 : Statistiques des Données**

1. Calculez la moyenne, la médiane, et la variance des âges suivants : 25, 30, 35, 40, 45.
2. Le tableau suivant représente le nombre d'appartement par occupation de personnes :

| | | | | | | | | |
|------------------------|-----|-----|-----|-----|----|----|---|---|
| Nombre de personnes | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Nombre d'appartement : | 201 | 238 | 137 | 142 | 55 | 15 | 4 | 2 |

Calculer la moyenne, la médiane, la variance et l'écart type.

❖ **Exercice 4 : Mesures de Distance et de Similarité**

1. Utilisez la distance euclidienne pour calculer la distance entre les points A (10, 30, 18, 30) et B (20, 15, 5, 50), puis avec les mêmes points utiliser la distance de Manhattan puis la distance de Minkowski avec $p=4$ puis $p=2$
2. Quelle est la différence entre la distance de Manhattan et la distance euclidienne ?
3. Calculez le coefficient de Jaccard entre les ensembles X (0, 1, 1, 0, 0) et Y (1, 0, 0, 1, 0), puis le coefficient de correspondance simple.

❖ **Exercice 5 : Réduction de Dimension**

1. Qu'est-ce que l'Analyse en Composantes Principales (PCA) et comment fonctionne-t-elle ?
2. Donnez un exemple de situation où la réduction de dimension à l'aide de PCA pourrait être bénéfique.

❖ **Exercice 6 : Transformation des Attributs**

1. Citez quelques techniques de transformation de Données, et donner un exemple pour chacune.