

# Fouille de données

## Master 2 Informatique décisionnelle

### Projet – Analyse de données

#### Indications générales

Soutenance : dernière séance de cours  
Seul(e) ou en binôme

#### Travail demandé

L'objectif de ce projet est de permettre aux étudiants de mettre en pratique les concepts de fouille de données dans un contexte au choix en construisant une base de données réelle et en effectuant une analyse approfondie à l'aide de techniques de fouilles de données.

Les étapes du projet sont les suivantes :

1. Choix du domaine : sélectionner un domaine spécifique (santé, réseaux sociaux numériques, journalisme, ...).
2. Collecte de données : identifier et collecter des ensembles de données réels pertinents à partir de sources fiables dans le domaine choisi (pas depuis une plateforme comme Kaggle).
3. Prétraitement des données : effectuer des opérations de prétraitement sur les données, telles que le nettoyage, la transformation et la normalisation, afin de les rendre appropriées pour l'analyse.
4. Choix des méthodes de fouille de données : sélectionner des techniques de fouille de données (au moins l'une d'entre elle doit avoir été étudiée en cours cette année).
5. Application des méthodes : appliquer les méthodes de fouille de données choisies sur la base de données construite et évaluer les performances.
6. Analyse des résultats : interpréter les résultats obtenus à partir des méthodes de fouille de données et tirer des conclusions significatives.
7. Rapport et présentation : rédiger un rapport de projet d'environ 10 pages, décrivant en détail les étapes du projet, les choix effectués et les résultats obtenus. Préparer une présentation orale de 15 minutes qui met en évidence les points clés du projet.

L'évaluation se fera selon le barème suivant :

- Pertinence de la base de données : 5 points ;
- Pertinence du traitement des données : 5 points ;
- Qualité du rapport (organisation, lisibilité, contenu) : 5 points ;
- Qualité de la présentation (clarté, éloquence, respect du temps imparti) : 5 points ;

L'ordre de passage des groupe sera décidé aléatoirement au début de la séance. Un retard sera pénalisé de 5 points pour le ou les retardataires et de 2 points pour le ou les autres membres du groupe.