

Projet Data Mining

De l'Analyse des Données à l'Extraction de Connaissances

Le projet de Data Mining a pour but d'aider les étudiants à se familiariser avec quelques tâches typiques du métier de data analyst et data scientist. En plus de la phase d'analyse, nettoyage, Intégration, sélection et transformation des données, il existe trois principaux axes de data mining qui permettent l'extraction de connaissances, à savoir: l'extraction de motifs fréquents, la classification et le clustering.

Informations :

- Travail en binôme.
- La liste des binômes doit être envoyée par mail¹ au plus tôt.
- Rapport à remettre en version numérique + papier pour chaque partie.
- Une présentation de l'IHM (interface simple) et du code source sera organisée en fin de chaque partie du projet.
- Le travail doit être effectué en Java.
- Le projet est organisé en 5 parties conformément aux échéances suivantes:
 1. Analyse des données [15 jrs]
 2. Prétraitement des données [15 jrs]
 3. Extraction de motifs fréquents, règles d'associations et corrélations [15 jrs]
 4. Classification et Prédiction [21 jrs]
 5. Clustering [21 jrs]
- Il est important de fournir un rapport bien structuré et facile à lire. Les informations et conclusions importantes doivent être mises en valeur, fournir une analyse adéquate des graphes présentés, ...etc
- Deux datasets seront utilisés. Un premier pour les parties 1, 2, 4, et 5 et un deuxième pour la partie 3.
- Profitez des séances de TP pour poser vos questions plutôt que par email.
- Tout **plus** peut être récompensé !

Lien vers les datasets :

<https://drive.google.com/drive/folders/1fsOSlgANzv5ORuEf7Z2r3-UBVLtTJvF-?usp=sharing>

¹ HOUACINE Naila Aziza : n.a.houacine@gmail.com