

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 mail^{1,2} 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 Python.
- Le projet est organisé en 2 parties conformément aux échéances suivantes :
 1. Partie 1 : [35 jrs]
 - 1.1. Analyse, prétraitement, visualisation des données (Dataset 1 et 2).
 - 1.2. Extraction de motifs fréquents, règles d'associations (Dataset 3).
 2. Partie 2 : Apprentissage supervisé et non supervisé [37 jrs]
 - 2.1. Classification, Prédiction (Dataset 1)
 - 2.2. Clustering (Dataset 1)
- 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
- Profitez des séances de TP pour poser vos questions plutôt que par email.
- Tout **plus** peut être récompensé !
- La note du projet est /15. Chaque partie est /7.5 :
 - Le rapport /4.5
 - La présentation du code source et l'interface /3

Lien vers les datasets :

https://drive.google.com/drive/folders/1dTS5gXz_-4EbOFwmsx5CVDsgQZtx3HBG