

Sujet du projet : Analyse des tweets

Ce projet est un travail pratique sur données réelles. Il s'agit de mettre en œuvre des méthodes vues en cours sur un jeu de données proposé pour chaque groupe (3étudiants maximum/groupe). Le projet doit suivre les étapes suivantes :

- analyser le jeu de données proposé afin de choisir l'approche la plus adaptée à votre problème,
- effectuer un prétraitement de votre base de données (si nécessaire),
- justifier le choix de votre approche,
- tester l'approche choisie sur l'ensemble de données en utilisant toutes les options vues en cours,
- analyser et commenter les résultats obtenus.

Présentation du rapport

Le rapport doit présenter de façon claire et concise:

- l'objet de l'analyse
- la description des données (individus/variables utilisées, variables supplémentaires etc.)
- l'analyse proprement dite
- les commentaires sur les résultats obtenus.

Ce rapport ne devrait pas dépasser 20 pages (les codes sources des programmes R ou Python utilisés peuvent être mis en annexe). Le projet sera jugé selon les critères suivants:

- Adéquation des méthodes utilisées aux données et problème étudiés.
- Richesse des analyses proposées (au delà du minimum requis).
- Justesse des commentaires sur les résultats.
- Qualité de la présentation du rapport.

Remise du projet

Vous devez envoyer votre rapport en format *.pdf* au plus tard le **15/06/2020 avant minuit** à l'adresse suivante l.labiod@gmail.com

Données

Tweets : La description des données est détaillée dans le tutoriel.

Démarche à suivre

- Proposer un sujet d'analyse (au choix)
- Extraction, prétraitement et manipulation des Tweets
- Exploration descriptive uni et bi variée + graphiques
- Exploration multidimensionnelle
 - AFC
 - K-means, CAH , NMF.....
 - Caractérisation des classes
- Analyse de sentiment

Analyse de Tweets avec python -- voir le lien : <https://github.com/bonzanini/Book-SocialMediaMiningPython>