

# Projet de l'Analyse de Données Textuelles : Journal

---

L'analyse de texte ou text mining est l'analyse des données « non structurées » contenues dans le texte de langage naturel en utilisant différentes méthodes, outils et techniques.

## Objectif :

L'objectif de ce projet est de vous initier aux étapes principales de la production d'un rapport d'analyse de données textuelles : du nettoyage et de l'importation des données brutes jusqu'à leur analyse descriptive et leur présentation graphique.

## Travail à effectuer

Lancer une analyse de texte du contenu des journaux en ligne marocains :

- Analyse Factorielle des correspondances
- Analyse des correspondances multiples
- Analyse discriminante
- Classification

Un tableau lexical sera rempli par la totalité des mots des pages d'un journal national et subi un prétraitement :

1. Les mots seront triés alphabétiquement
2. Les mots outils sont supprimés
3. Ces mots sont concaténés :
  - a. Singuliers et pluriels
  - b. Les conjugaisons du verbe
  - c. De même racine : produits, produire, ...
4. L'effectif maximal des mots est 1000 (à modifier selon les résultats).

Sur ce jeu de données vous allez :

- Traiter et présenter les données à analyser sous forme de tableaux appropriés ;
- Appliquer une analyse factorielle : Analyse en composantes principales ou Analyse des correspondances suivant la nature des données à analyser.
- Une typologie à l'aide d'une classification.
- Faire toutes les implications des notions vues dans le cours et interprétations utiles à la description des données.

## Critères d'évaluation

Le barème de notation englobera une note réservée à la présentation et la qualité des tableaux et graphiques contenus dans le rapport. Des points peuvent être enlevés en cas de copie évidente d'une page web ou d'une autre copie.

## REMISE DU PROJET

La date de remise du projet est fixée au **2 janvier 2023 au plus tard**.

**Aucun délai supplémentaire ne sera accordé.**