

TD : Analyse des Données Textuelles en Python pour le NLP

Objectifs du TD

- Explorer des méthodes d'analyse des données textuelles.
- Visualiser des informations clés (fréquences, nuages de mots).
- Utiliser des outils pour effectuer une analyse de sentiment.

Durée : 1h30

Prérequis

- Données textuelles nettoyées à partir du TD précédent.
- Python 3.7+
- Bibliothèques nécessaires : Pandas, Matplotlib, WordCloud, TextBlob.

Plan du TD

1. Exploration statistique
2. Visualisation
3. Analyse de sentiments

1 Exploration statistique

Dans cette section, nous allons explorer les données textuelles nettoyées pour en extraire des informations clés comme les fréquences de mots.

Étapes à suivre :

- Charger les données nettoyées depuis un fichier CSV.
- Identifier les mots les plus fréquents en utilisant une bibliothèque d'analyse de texte.
- Calculer les statistiques de base sur les fréquences des mots.

2 Visualisation

Étapes à suivre :

- Générer un nuage de mots pour visualiser les mots les plus fréquents.
- Utiliser une bibliothèque de visualisation comme Matplotlib pour afficher les résultats.

3 Analyse de sentiments

Dans cette partie, nous utilisons un outil comme **TextBlob** pour effectuer une analyse de sentiments des textes nettoyés.

Étapes à suivre :

- Appliquer une fonction d'analyse de sentiment sur les textes nettoyés pour obtenir un score de polarité.
- Visualiser la distribution des scores de sentiment à l'aide d'un histogramme.
- Comparer les résultats obtenus avec des exemples concrets pour valider les interprétations.

Pour aller plus loin :

- Récapitulez les étapes principales de l'analyse textuelle.
- Appliquez les méthodes d'analyse et de visualisation à un corpus de votre choix.
- Comparez les résultats de l'analyse de sentiments avec votre propre interprétation.

Ressources complémentaires

- Documentation [Pandas](#)
- Documentation [Matplotlib](#)
- Guide [TextBlob](#)
- Documentation [WordCloud](#)