## Documentation des Bibliothèques Utilisées

### os

Le module 'os' fournit une interface pour interagir avec le système d'exploitation. Il permet d'effectuer des opérations comme la gestion des fichiers et des dossiers, vérifier l'existence de fichiers et créer des répertoires.

### pandas

Pandas est une bibliothèque puissante pour la manipulation et l'analyse de données. Elle offre des structures de données comme DataFrame et Series, adaptées au traitement de données structurées. Pandas permet de lire/écrire des fichiers CSV, de nettoyer les données et d'effectuer des analyses statistiques.

### joblib

Joblib est une bibliothèque légère pour la sérialisation d'objets Python. Elle est souvent utilisée pour sauvegarder et charger des modèles d'apprentissage automatique, particulièrement pour les fichiers volumineux.

### sklearn.feature_extraction.text.TfidfVectorizer

TF-IDF Vectorizer convertit les données textuelles en une matrice de caractéristiques basées sur la fréquence des termes (TF) et leur importance relative (IDF). Cela est utile pour les applications de traitement du langage naturel.

### sklearn.ensemble.RandomForestClassifier

RandomForestClassifier est un algorithme d'apprentissage automatique basé sur un ensemble d'arbres de décision. Il combine les prédictions de plusieurs arbres pour améliorer la précision et réduire le surapprentissage.

### sklearn.metrics

Ce module fournit des outils pour évaluer les performances des modèles d'apprentissage. Les métriques incluent les matrices de confusion, les rapports de classification et les courbes ROC-AUC.

### sklearn.model_selection

Ce module propose des outils comme train_test_split pour diviser les ensembles de données et GridSearchCV pour optimiser les hyperparamètres grâce à la validation croisée.

### sklearn.utils.class_weight

Utile pour calculer les poids des classes dans des jeux de données déséquilibrés. Cela aide à équilibrer l'impact des classes minoritaires lors de l'entraînement des modèles.

### matplotlib.pyplot

Pyplot, de Matplotlib, est utilisé pour créer des visualisations statiques, interactives et animées. Il permet de tracer des graphiques, personnaliser des éléments visuels et sauvegarder des figures.

### seaborn

Seaborn est basé sur Matplotlib et est spécialisé dans la création de graphiques statistiques attrayants et informatifs. Il s'intègre bien avec Pandas pour visualiser des données structurées.

### nltk

Natural Language Toolkit (NLTK) est une bibliothèque pour le traitement et l'analyse de données textuelles. Elle offre des outils pour la tokenisation, le lemmatisation, le stemming, et bien plus.

### nltk.corpus.stopwords

Ce module fournit une liste de mots vides (stop words) pour plusieurs langues. Ces mots communs (comme 'le', 'est') sont généralement supprimés lors du prétraitement des textes.

### nltk.stem.WordNetLemmatizer

WordNetLemmatizer réduit les mots à leur forme de base ou racine à l'aide de la base lexicale WordNet. Contrairement au stemming, il produit des formes de base significatives.

### re

Le module 're' permet d'utiliser des expressions régulières pour rechercher, faire correspondre et manipuler des motifs dans des chaînes de caractères.

### streamlit

Streamlit est un framework Python pour créer des applications web interactives et visuelles. Il est conçu pour les développeurs qui souhaitent intégrer rapidement des visualisations dans des applications.