

Analyse de sentiments sur des critiques de cinéma

Objectif : effectuer une analyse de sentiments avec Python.

Travail à faire :

1. Lire le fichier **avis.csv** dans une variable Dataframe de pandas
2. Afficher un échantillon

- La description représente le commentaire de l'utilisateur
- La note est comprise entre 0 et 5

3. Ajouter une colonne « Avis » selon les notes : positif si la note ≥ 3 , négatif sinon
4. Préparer les données en utilisant les techniques suivantes :

- ✓ Des expressions régulières pour retirer les bruits (ponctuation, etc.) des commentaires.
- ✓ Du NLP pour tokeniser et réduire le corpus de chaque commentaire (afin par exemple de ne garder que les mots importants via les stopwords)
- ✓ Des sacs de mots TF_IDF afin de « transformer » nos mots en nombres qui pourront alors être exploités dans un algorithme de Machine learning tel qu'une régression logistique par exemple.

5. Diviser le jeu de données en un ensemble de training et un ensemble de test en utilisant sklearn train_test_split(). L'ensemble de training sera utilisé pour former le modèle et l'ensemble de test pour tester le modèle. Utiliser 30% pour le test et 70% le training.
6. Entraîner plusieurs modèles de classification en utilisant les algorithmes SVM, KNN, décision tree et la régression logistique
7. Créer un tableau comparatif en utilisant les mesures : accuracy ; precision, recall et F1
8. Tester le modèle performant pour un nouveau commentaire