

# Mini-Projet 1

**Objectif :** Classification des sentiments en utilisant l'algorithme Naive Bayes.

**Source :** [Lien vers la source de données](#)

L'ensemble de données contient des phrases étiquetées avec un sentiment positif ou négatif.

|   |                |                      |      |                         |            |
|---|----------------|----------------------|------|-------------------------|------------|
| Caractéristiques de l'ensemble de données : | Text           | Nombre d'Instances : | 3000 | Domaine :               | N/A        |
| Caractéristiques des attributs :            | N/A            | Nombre d'attributs : | N/A  | Date du don             | 30-05-2015 |
| Tâches associées :                          | Classification | Valeurs manquantes ? | N/A  | Nombre de visites Web : | 217753     |

## Informations sur l'ensemble de données :

Cet ensemble de données a été créé pour l'article « *From Group to Individual Labels using Deep Features* », Kotzias et. Al., KDD 2015

Il contient des phrases étiquetées avec un sentiment positif ou négatif.

**Format:** score de la phrase

## Détails:

Le score est soit 1 (pour positif) ou 0 (pour négatif)

Les phrases proviennent de trois sites/champs différents :

- imdb.com
- amazon.com
- yelp.com

Pour chaque site web, il existe 500 phrases positives et 500 phrases négatives. Ceux-ci ont été sélectionnés au hasard pour des ensembles de données d'exams plus importants.

Nous avons essayé de sélectionner des phrases qui ont une connotation clairement positive ou négative, le but était qu'aucune phrase neutre ne soit sélectionnée.

## Informations sur les attributs :

Les attributs sont des phrases de texte, extraites de critiques de produits, de films et de restaurants.

**Outils :** Jupyter ou Spyder

**Bibliothèques :** NLTK ou Spacy

**Livrable :** Code source python (format pdf)

**Date Limite :** 29/10/2022 à 23h59.

**Email :** tt\_bouslimi.riadh@universitecentrale.tn