

Dataset 5. Sentiments des tweets

1. Description des données

On dispose de deux fichiers « train_tweets.txt » et « test_tweets.txt » contenant des données d'entraînement et de test respectivement. La description des fichiers :

- IDF : l'identifiant du tweet (entier)
- Polarité : le sentiment exprimé sur l'échelle [-1,0,+1] pour négatif, neutre et positif (entier signé)
- Texte : le texte brute du tweet (chaîne de caractère)

2. Tâches

Il s'agit de faire une analyse des données en combinant des **algorithmes d'analyse**.

Les tâches demandées sont les suivantes :

a. Tâche 1

1. Réaliser une analyse descriptive des tweets.

b. Tâche 2

1. Analyser les sentiments des tweets par des modèles d'apprentissage automatique.

c. Tâche 3

1. Analyser les métadonnées des tweets. Libre à vous de choisir une parmi les suivantes :
 1. Les URLs,
 2. Les émoticônes,
 3. La ponctuation,
 4. L'ironie,
 5. Les injures censurées,
 6. etc.