

Projet : NLP pour l'analyse des avis clients en marketing

Contexte et Objectifs



L'objectif du traitement du langage naturel (**NLP**) est d'extraire des informations et de la valeur à partir d'un grand volume de données textuelles. **Pourquoi ?**

- Améliorer l'expérience client sur le site web, l'application mobile.
- Évaluer la satisfaction des clients différemment.
- Être plus disponible et accessible auprès des clients.
- Trouver de nouvelles solutions pour perfectionner les services et produits proposés.

Démarche et étapes de construction

Notre démarche applicative se décompose en 5 grands streams :

Étape 1 : Web Scraping

- Récouter et créer le schéma des données.
- Parser les avis clients pour enrichir la base de données : extraction du titre, de la description, la date, l'heure, le pseudo et la note.

Étape 2 : Analyse de sentiments et Scoring sentimental.

- Comprendre et sonder la satisfaction de chaque client.
- Scorer l'intensité et la polarité des sentiments à partir de la description de l'avis.

Étape 3 : Text mining et data cleaning

- Text cleaning adapté au domaine bancaire et au contenu général des avis.

Étape 4 : Topic Modeling (apprentissage non-supervisé)

- Pour améliorer la disponibilité et accélérer le temps de réponse, les avis peuvent être dissociés et triés par ordre de priorité en fonction du thème qu'ils abordent.

Étape 5 : Machine Learning (apprentissage supervisé)

- Sans lire les futurs avis, concevoir un modèle solide pour identifier le sentiment global exprimé par le client.

Analyse de sentiments

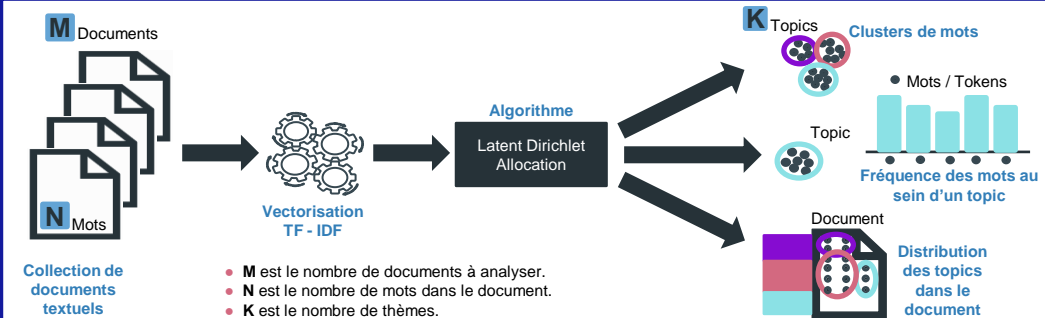
Valence Aware Dictionary and sEntiment Reasoner est un outil de NLP qui permet d'analyser des sentiments dans les textes en combinant :

- Des caractéristiques lexicales des sentiments (lexicons).
- Des règles grammaticales.
- Des règles syntaxiques pour exprimer la polarité (**positive** / **neutre** / **négative**) et l'intensité (**puissance**) des émotions.
- La prise en compte du contexte des mots, de la forme grammaticale, de la ponctuation et des majuscules.

$$S_i = \sum_{j=0}^N s_{i,j}, \text{ avec } S_i \in [-1; 1]$$

Où N est le nombre total d'avis, $s_{i,j}$ le score sentimental du $j^{\text{ème}}$ mot du $i^{\text{ème}}$ avis. Par exemple, la somme du score sentimental individuel de « j'aime », « j'apprécie », « je suis heureux » donnera $S_i > 1$.

Topic Modeling



Le modèle de classification final est une **méthode d'ensemble** qui combine performance, meilleures prévisions et performances que n'importe quel modèle individuel, et robustesse, réduit l'écart ou la dispersion des prévisions et des performances du modèle (compromis biais-variance).

Machine Learning