* **VADER**

**Nature : Modèle basé sur un lexique. (Lexicon model)**

Approche : Utilise un dictionnaire prédéfini qui associe des caractéristiques lexicales à des intensités de sentiment.

**Heuristique :**

Ponctuation : Intensifie le sentiment sans en modifier l'orientation.

Majuscules : Amplifie l'intensité d'un mot pertinent pour le sentiment.

Modificateurs de degré : Des mots comme "très" ou "légèrement" qui ajustent l'intensité.

**Conjonctions contrastives** : Des mots comme "mais" qui modifient la polarité du sentiment.

**Traitement de la négation** : Identifier et ajuster la négation qui modifie la polarité.

Dictionnaire de source collective : La moyenne des intensités de sentiment est calculée à partir de plusieurs évaluateurs humains afin d'éviter les biais.

Score : Fournit des scores pour les sentiments positifs, neutres et négatifs, ainsi qu'un score composé qui représente le sentiment global.

* **Roberta**

**Nature : Modèle de langage à grande échelle basé sur un transformateur. (LLM)**

Approche : Utilise l'apprentissage profond et un pré-entraînement approfondi sur divers ensembles de données pour comprendre le contexte et le sentiment.

Compréhension du contexte : Grâce à son architecture profonde, il excelle dans la capture des nuances et des dépendances complexes dans le langage.

Mise au point : Nécessite une mise au point sur des tâches spécifiques d'analyse de sentiments pour obtenir des performances optimales.

* **Analyse comparative**
* VADER : Il donne de bons résultats sur les textes courts et les données des médias sociaux, pour lesquels une heuristique et des lexiques simples suffisent. Il est rapide et peu coûteux en termes de calcul.
* Roberta : précision supérieure dans la compréhension du contexte et le traitement des phrases complexes grâce à ses capacités d'apprentissage en profondeur. Meilleures performances dans les textes longs et variés.

🡪**Facilité d'utilisation :**

VADER : simple à mettre en œuvre avec des ressources informatiques minimales. Convient aux applications en temps réel et aux situations où l'interprétabilité et la simplicité sont cruciales.

Roberta : Nécessite d'importantes ressources informatiques pour l'entraînement et l'inférence. Le réglage fin peut être complexe et prendre du temps.

🡪**Robustesse et flexibilité :**

VADER : limité par son lexique et ses règles prédéfinies, qui peuvent ne pas s'appliquer à tous les domaines ou à toutes les langues.

Roberta : Très flexible et adaptable à différents domaines avec un réglage fin approprié. Robuste aux différents styles d'écriture et plus résistant au bruit dans les données.

🡪**Interprétabilité :**

VADER : grande facilité d'interprétation grâce à son approche basée sur des règles. Les utilisateurs peuvent facilement comprendre et retracer la manière dont les scores de sentiment sont dérivés.

Roberta : Interprétabilité plus faible en raison de la nature "boîte noire" des modèles d'apprentissage profond. Cependant, il peut fournir des informations plus nuancées grâce à sa compréhension complexe du contexte.