

Etude d'un modèle d'analyse des sentiments basé sur l'apprentissage automatique

Djuigouing Simo Daïna · Kappe Moko Priscille Laurette · Tanétertore Nzissié Géodan · Tchatchouang Meyindjou Farel Rosvelt

Résumé

L'analyse des sentiments est devenue une technique essentielle pour comprendre les perceptions et les opinions exprimées dans les données textuelles. Les opinions des personnes peuvent être bénéfiques aux entreprises, aux gouvernements et aux individus pour collecter des informations et prendre des décisions éclairées. Cette étude vise à explorer et à évaluer un modèle d'analyse des sentiments sur le dataset Stanford Sentiment Treebank. Deux modèles sont à cet effet utilisés et comparés : Le modèle TextBlob et le modèle VADER, qui tous deux permettent d'effectuer une classification des phrases en deux catégories de sentiments : Positif et négatif. Les résultats montrent que VADER présente de meilleures performances globales pour l'analyse des sentiments, pour ce qui est du dataset Stanford Sentiment Treebank. Cette étude souligne l'importance de l'évaluation comparative des outils d'analyse des sentiments pour garantir la fiabilité des analyses dans des contextes diversifiés.

Mots-clés : Analyse des sentiments, TextBlob, VADER, évaluation comparative, données textuelles

1 Introduction

L'analyse des sentiments, ou opinion mining, correspond au processus d'analyse de gros volumes de texte en vue de déterminer si ces derniers expriment un sentiment positif, négatif ou neutre. L'analyse des sentiments est un outil essentiel pour mieux comprendre les émotions et les perceptions des individus, et améliorer la prise de décision. Elle peut être utilisée dans de nombreux domaines, parmi lesquels : La compréhension des clients, l'analyse de l'opinion publique, la détection de la désinformation, la modération de contenu¹, l'amélioration des produits et des services, la prévision du comportement des consommateurs. Il existe de

nombreux modèles d'analyse des sentiments, dont les plus connus sont : Textblob et VADER qui sont des modèles basés sur le lexique, SVM, BERT, SVM, Random Forest. L'objectif de l'étude est d'entraîner un modèle d'analyse des sentiments performant sur le dataset Stanford Sentiment Treebank.

2 Méthodologie

❖ Description du dataset Stanford Sentiment Treebank

Le dataset Stanford Sentiment Treebank a été créé par l'Université de Stanford située à Silicon Valley. Il contient des phrases extraites de critiques de films publiées sur la chaîne Rotten Tomatoes. Chaque phrase a été annotée manuellement avec un sentiment sur une échelle allant de 1 (très négatif) à 5 (très positif).

Le dataset comprend des phrases complètes ainsi que leur décomposition en sous-phrases sous forme d'arbre syntaxique (treebank). Chaque sous-phrase hérite du sentiment attribué à la phrase complète. Cela permet d'évaluer la capacité des modèles à comprendre le sentiment à différents niveaux de granularité dans le texte.

Ce dataset compte : 11 855 phrases au total, 215 606 sous-phrases annotées, 6 920 phrases positives (4 ou 5), 2 210 phrases négatives (1 ou 2) et 2 725 phrases neutres (3).

Il est divisé en trois sous-ensembles : L'ensemble des données d'entraînement, celui des données de validation et celui des données test. Cette partition permet d'évaluer les performances des modèles sur de nouvelles données et de comparer différentes approches.

Le SST (Stanford Sentiment Treebank) est l'un des benchmarks (références) les plus utilisés pour l'évaluation des modèles d'analyse des sentiments au niveau de la phrase.

❖ Choix du modèle d'analyse des sentiments

Il existe de nombreux modèles d'analyse des sentiments. Des exemples de modèles d'analyse des sentiments largement utilisés sont :

- **Les modèles basés sur le lexique:** Ils utilisent des dictionnaires de mots annotés avec des scores de polarité (positif, négatif, neutre). Nous avons par exemple les modèles **VADER**(Valence Aware Dictionary and sEntiment Reasoner) et **TextBlob**

- **Les modèles de langage pré-entraînés** et affinés sur des tâches d'analyse des sentiments, comme les modèles de type BERT, RoBERTa, XLNet, etc.
- **Les Modèles de transformer multi-tâches**, qui sont pré-entraînés sur plusieurs tâches linguistiques. Nous avons par exemple, le modèle T5 (Text-To-Text Transfer Transformer).
- D'autres modèles d'apprentissage automatique supervisé comme Naive Bayes, le modèle de régression logistique, SVM, et les réseaux de neurones

❖ Paramètres d'entraînement

L'échantillon d'entraînement représente environ 67% du dataset complet, soit environ 8 544 phrases, celui de validation, 8%, soit environ 1 101 phrases, et celui de test, environ 25% du dataset complet, soit environ 2 210 phrases. Cette répartition permet de disposer d'un ensemble de validation suffisamment important pour effectuer le réglage des hyperparamètres, tout en gardant un ensemble de test conséquent pour une évaluation finale fiable du modèle.

❖ Métriques d'évaluation :

Une métrique est une mesure quantitative utilisée pour évaluer les performances et la qualité du modèle. Les métriques utilisées pour évaluer notre modèle sont l'exactitude et le F1-score.

L'exactitude (accuracy) mesure la proportion de prédictions correctes parmi l'ensemble des prédictions. Le F1-score est la moyenne harmonique de la précision et du rappel, offrant un équilibre entre ces deux mesures.

3 Résultats

❖ Tableau comparatif des valeurs des métriques

Modèles	Métriques	Valeurs
Textblob	Exactitude	28,37%
	F1-score	24,68%

VADER	Exactitude	31,58%
	F1- score	31,36%

On constate que les valeurs de l'exactitude sont respectivement de 28,37% pour TextBlob et 31,58 % pour VADER, et que les valeurs du F1-score sont respectivement de 24,68% pour TextBlob et de 31,36% pour VADER.

❖ Graphiques

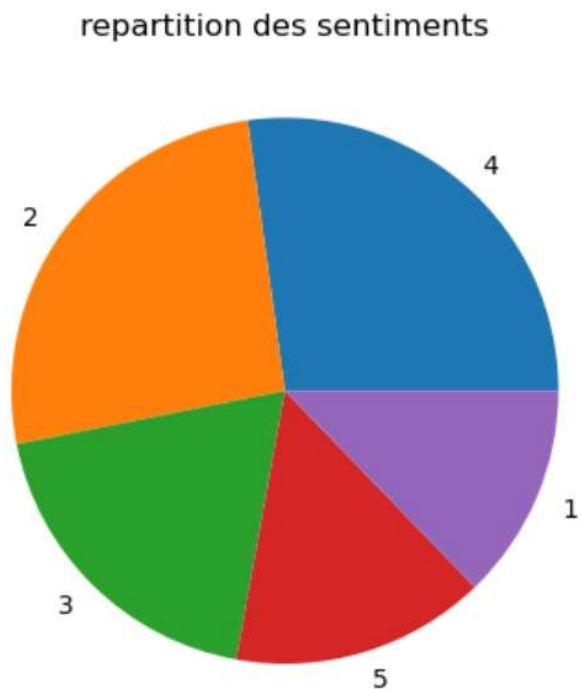


Figure 2 : Diagramme en secteurs présentant la répartition des différents sentiments dans le dataset

On constate qu'il y a une proportion plus grande de phrases positives.

Notons que : 1= très négatif

2= négatif

3= neutre

4 = positif

5 = très positif

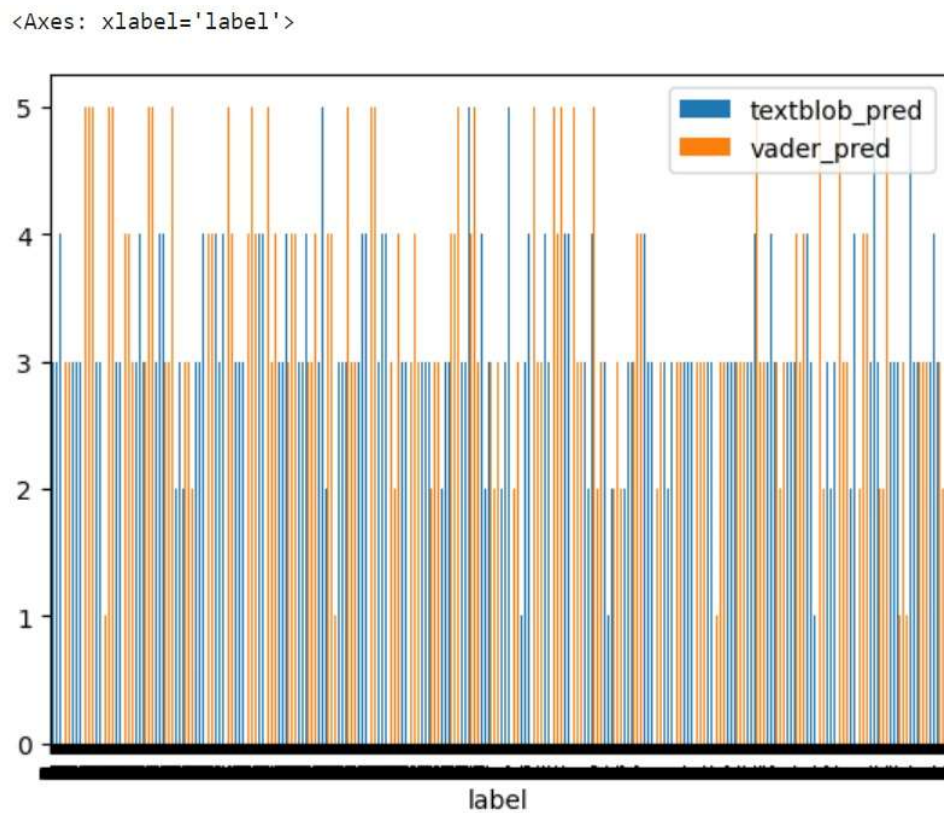


Figure 1 : Diagramme en bâtons de comparaison des performances de VADER et TextBlob

On remarque que comparé à TextBlob, VADER fait de meilleures prédictions.

❖ Analyses statistiques (résultats du modèle)

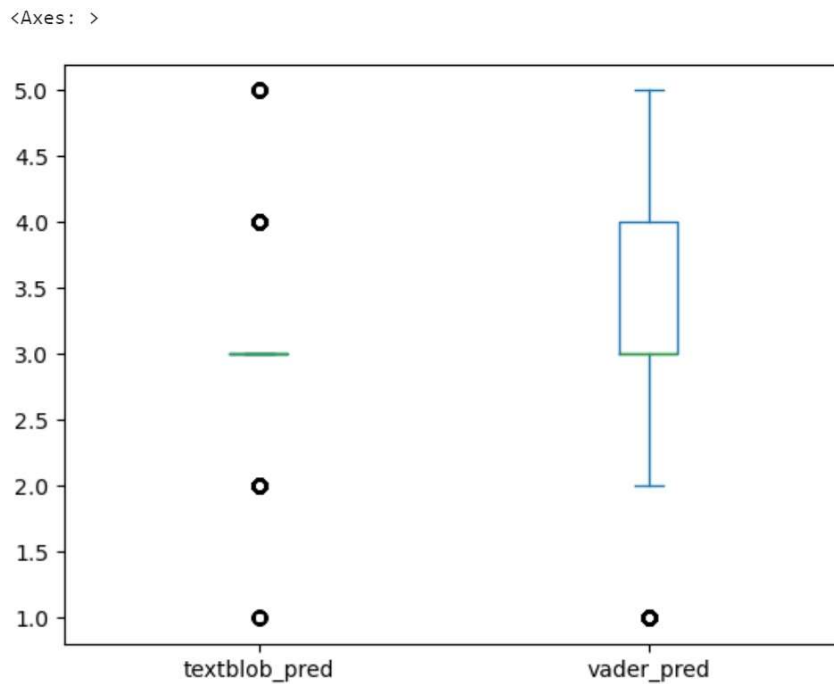


Figure 3 : Boîtes à moustaches

On remarque que 50% des prédictions de sentiments effectuées par le modèle VADER se situent entre les valeurs 2 et 3, et que pour ce modèle, il y a peu de valeurs aberrantes.

Pour ce qui est de TextBlob, on remarque que pratiquement toutes les valeurs prédites sont égales à 3. Le nombre de valeurs aberrantes au-dessus est égal au nombre de valeurs aberrantes en dessous de la médiane. Cette configuration indique une distribution parfaitement centrée et symétrique, avec une absence totale de variabilité dans la majeure partie des données, mais des valeurs aberrantes symétriques en haut et en bas.

4 Discussion

❖ Interprétation des résultats

Etant donné que VADER a une exactitude de 31,58%, tandis que TextBlob a une exactitude de 28,37% , le modèle VADER a de meilleures performances globales en termes de prédictions correctes par rapport à TextBlob. Le point fort de VADER est donc sa meilleure exactitude par rapport à TextBlob.

Également, étant donné que VADER a un F1-score de 31,36%, tandis que TextBlob a un F1-score de 24,68%, VADER a donc de meilleures performances en termes d'équilibre entre la précision (proportion de prédictions positives qui sont correctes) et le rappel ().

Le point fort de VADER est donc sa meilleure capacité à trouver un équilibre entre la précision et le rappel, ce qui se traduit par un F1-score plus élevé.

❖ Limites de l'étude

Le F1-score et l'exactitude pour ces deux modèles sont tous les deux très faibles; on peut conclure que les modèles Vader et TextBlob n'ont pas suffisamment appris.

❖ Pistes d'amélioration

Optimisation des hyperparamètres du modèle : On peut ajuster les hyperparamètres du modèle (taux d'apprentissage, nombre d'arbres, profondeur, etc.) à l'aide de techniques de validation croisée.

Collecte et ajout de nouvelles données d'entraînement pertinentes : Cela peut permettre aux modèles d'apprendre plus de motifs et de généraliser davantage.

❖ Perspectives de recherche future

Comme perspectives de recherches, il serait judicieux de pouvoir effectuer l'analyse des sentiments à l'échelle d'un long texte, sur des données visuelles également (images, vidéos), et de suivre l'évolution des sentiments sur de longues durées (commentaires sur les réseaux sociaux, etc). Il serait également très intéressant de pouvoir appliquer l'analyse des sentiments à des domaines tels que la psychologie, l'éducation, la finance et bien d'autres. Également, il serait bon de pouvoir obtenir des résultats beaucoup plus précis au terme de l'analyse des sentiments, ceci grâce à l'utilisation de modèles plus performants.

5 Conclusion

Cet article traitait de la réalisation d'un modèle d'analyse des sentiments, appliqué au jeu de données Stanford Sentiment Treebank. Pour mener à bien ce projet, nous avons jugé judicieux d'utiliser, puis de comparer les performances de deux outils de référence en terme d'analyse des sentiments : TextBlob et VADER. Il ressort de cette étude que le modèle VADER est plus performant, car donnant de meilleurs résultats. Il subsiste néanmoins des difficultés notables,

en ce qui concerne la précision des résultats fournis par chacun de ces outils. Les travaux futurs consisteront à élargir continuellement la zone de comparaison et les champs d'application des modèles d'analyse des sentiments. De ce fait, nous utiliserons d'autres modèles plus complexes, ceci afin d'affiner davantage les résultats.

Références

(IBM, 2024)[1]

(Kulkarni, 2022)

(Bernhard Kratzwalda, 2018)