

Analyse de la Satisfaction des Clients avec NLP

Realiser par :
Mohamed MALLOUKI

Jury par:
Pr.Charaf hamid
Pr. S.El gaou

Encadrer par:
prof.C.hamdi



plan

1-Introduction

2-Objectifs de l'Étude

3-Méthodologie

4-Résultats

5-Conclusion

1-Introduction

L'analyse de la satisfaction client est devenue une composante cruciale pour les entreprises cherchant à comprendre les besoins et les sentiments de leur clientèle. Dans un monde de plus en plus connecté, où les clients partagent leurs expériences en ligne, il est essentiel de tirer parti du Traitement Automatique du Langage Naturel (NLP) pour extraire des informations significatives à partir de vastes ensembles de données textuelles.

2-Objectifs de l'Étude

L'objectif de cette étude est d'appliquer des techniques de NLP à l'analyse des avis clients, en se concentrant particulièrement sur les commentaires liés aux produits alimentaires sur la plateforme Amazon. Les avis en ligne fournissent une mine d'informations, mais leur analyse manuelle peut être fastidieuse et chronophage. En utilisant des méthodes automatisées, nous cherchons à extraire des tendances, des sentiments et des informations utiles à partir de ces avis.



3-Méthodologie

1. Collecte de données.
2. Prétraitement des données.
3. Application de techniques de NLP.
4. Évaluation des résultats.

DATA

► Reviews : 568 454

ID	ProductID	UserID	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
1	B001E4NF08	A330XHTALHUR0W	demanian	1.0	1.0	5.0	1.303902e+09	Good Quality Dog Food	I have bought several of the Vitality canned food...
2	B00613GRG4	A10RTF6ZCVESAK	dil pa	0.0	0.0	1.0	1.3402975e+09	Not as Advertised	Product arrived labeled as Jumbo Sliced Porkch...
3	B000LQOCH0	ABXLMNXXXXAN	Natalia Correa "Natalia Correa"	1.0	1.0	4.0	1.219010e+09	"Delight"	This is a correction that has been around a lo...
4	B000UASQIQ	A398BOROFOV0XV	Karl	3.0	3.0	2.0	1.307902e+09	Cough Medicine	If you are looking for the secret ingredient i...

DATA

- ▶ Reviews : 568 454
- ▶ users : 256 059

id	ProductID	UserID	ProfileName	helpfulnessNumerator	helpfulnessDenominator	Score	Time	Summary	Text
1	B001E4NF08	A330XHTALHUR0W	demanian	1.0	1.0	5.0	1.303902e+09	Good Quality Dog Food	I have bought several of the Vitality canned &...
2	B00613GRG4	A10RTFGZCV5NAK	dil pa	0.0	0.0	1.0	1.3402975e+09	Not as Advertised	Product arrived labeled as Jumbo Sliced Protein...
3	B000LQOCH0	ABXLMWJXX0AN	Natalia Correa "Natalia Correa"	1.0	1.0	4.0	1.219010e+09	"Delight"	This is a correction that has been around a lo...
4	B000LW8QKQ	A398BORO6F0V0XV	Karl	3.0	3.0	2.0	1.307902e+09	Cough Medicine	If you are looking for the secret ingredient i...

DATA

- ▶ Reviews : 568 454
- ▶ users : 256 059
- ▶ Products: 74 258

id	ProductId	UserId	ProfileName	helpfulnessNumerator	helpfulnessDenominator	Score	Time	Summary	Text
1	B001E4NF08	A330XHTALHUR0W	demanian	1.0	1.0	5.0	1.303902e+09	Good Quality Dog Food	I have bought several of the Vitality canned &...
2	B00613GRG4	A10RTFGZCVESAK	dil pa	0.0	0.0	1.0	1.3402975e+09	Not as Advertised	Product arrived labeled as Jumbo Sliced Protein...
3	B000LQOCH0	ABSLMNVJXX0AN	Natalia Correa "Natalia Correa"	1.0	1.0	4.0	1.219010e+09	"Delight"	This is a correction that has been around a lo...
4	B000LUAQIQ	A398BOROF0V0XV	Karl	3.0	3.0	2.0	1.307902e+09	Cough Medicine	If you are looking for the secret ingredient i...

Data Visualization

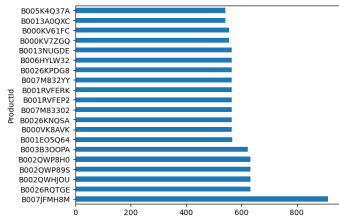


Figure: product

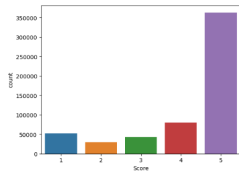


Figure: scor

Text preprocessing

- Suppression des balises HTML et des URL



Text preprocessing

- ▶ Suppression des balises HTML et des URL
- ▶ Suppression des mots vides (stop words)



Text preprocessing

- ▶ Suppression des balises HTML et des URL
- ▶ Suppression des mots vides (stop words)
- ▶ Stemming ou lemmatisation



Data

	Text	CleanedText
0	I have bought several of the Vitality canned d...	bought several vitality canned dog food produc...
1	Product arrived labeled as Jumbo Salted Peanut...	product arrived labeled jumbo salted peanuts p...
2	This is a confection that has been around a fe...	confection around centuries light pillowy citr...
3	If you are looking for the secret ingredient i...	looking secret ingredient robitussin believe f...
4	Great taffy at a great price. There was a wid...	great taffy great price wide assortment yummy ...
...
568449	Great for sesame chicken..this is a good if no...	great sesame chicken good not better resturant...
568450	I'm disappointed with the flavor. The chocola...	disappointed flavor chocolate notes especially...
568451	These stars are small, so you can give 10-15 o...	stars small give one training session tried tr...
568452	These are the BEST treats for training and rew...	best treats training rewarding dog good groomi...
568453	I am very satisfied ,product is as advertised,...	satisfied product advertised use cereal raw vi...

Figure: avant-après

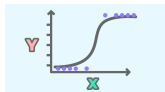
TF-IDF

- ▶ TF-IDF est une mesure statistique qui évalue l'importance d'un mot dans un document par rapport à une collection de documents (corpus).
- ▶ combine la fréquence du terme (TF), qui mesure à quelle fréquence un terme apparaît dans un document, avec l'inverse de la fréquence des documents (IDF), qui pénalise les termes courants dans de nombreux documents. .

la regression logistique

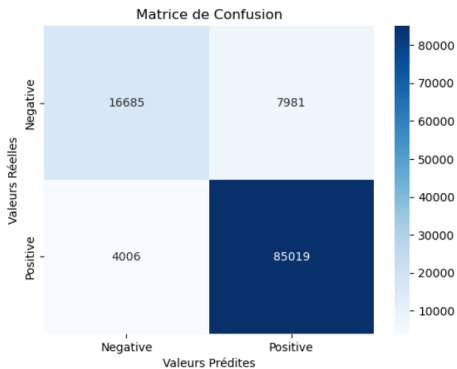
est spécifiquement conçue pour les problèmes de classification, où l'objectif est de prédire une variable binaire (deux classes).

La régression logistique utilise une fonction logistique pour modéliser la probabilité que la variable dépendante appartienne à une classe particulière.



4-Résultats

Classification Report:					
	precision	recall	f1-score	support	
0	0.81	0.68	0.74	24666	
1	0.91	0.96	0.93	89025	
accuracy			0.89	113691	
macro avg	0.86	0.82	0.83	113691	
weighted avg	0.89	0.89	0.89	113691	



Accuracy:0.89456

Conclusion

En résumé, la régression logistique est souvent choisie pour sa simplicité, son interprétabilité et sa capacité à gérer des problèmes de classification binaire, ce qui la rend adaptée à de nombreuses tâches d'analyse de texte, y compris l'analyse des critiques de produits. même tempe il y plusieurs model de classification par exemple Support Vector Machine (SVM), Naive Bayes, Decision Tree, Random Forest, Neural Network.