

Mémoire de fin d'études

Pour l'obtention du diplôme d'Ingénieur d'État en Informatique

Option : Systèmes Informatiques et Logiciels

Analyse des Sentiments sur les avis des clients de la marque Brandt avec prise en compte de la crédibilité

Réalisé par :

Mme. SLIMANI Wassila Maria

Encadré par :

Dr. SAID EL HADJ Lynda (ESI)

Mme. CHEMMOUL Ghania (Cevital)

Organisme d'accueil : Cevital

Promotion : 2022/2023

Dédicace

“

Je dédie ce mémoire,

À ma chère mère, pour son soutien, ses encouragements et son amour inconditionnel,

À mon père, pour sa confiance en moi et ses encouragements,

À ma chère mamie Halima, pour ses sacrifices, pour les longues heures qu'elle a priées pour mon bien-être et ma réussite et pour tout l'amour qu'elle m'a donnée,

À ma petite sœur Cerine et mon petit frère Yanis pour leur amour, leur soutien et les moments de joie que nous avons partagés,

À la mémoire de ma tante Kenza, la personne la plus gentille au monde, que ton âme repose en paix,

À mon cher oncle Abderezak pour avoir toujours été là quand j'avais besoin de lui sans jamais se plaindre,

À ma chère tante Wahiba et son mari Sidamer, pour leur soutien, leur aide, leur confiance et leurs prières.

À Sarah, Roumaissa et Ahlem, pour leur amitié, leur soutien et toutes nos belles aventures.

À Wisseme et Rania pour leurs conseils et leur aide,

À tous ceux qui ont cru en moi, À tous ceux qui me sont chers, À vous tous,

Merci.

”

- Maria

Remerciements

“

“No one who achieves success does so without the help of others. The wise and confident acknowledge this help with gratitude.”

-Alfred North Whitehead

”

En préambule à ce rapport, je souhaite exprimer ma reconnaissance envers toute personne ayant contribué à la réussite de ce projet.

Avant tout, je remercie Allah tout-puissant de m'avoir mis sur ce chemin et de m'avoir donné la patience et la volonté nécessaire afin de réaliser ce travail.

Je tiens à remercier toutes les personnes qui ont contribué à la réussite de ce projet de fin d'études. Particulièrement, j'adresse mes gratitude les plus sincères à mon encadrante, madame **Lynda Said Lhadj** (Maitre de conférence à l'ESI). Tout au long de la période de mon stage, j'ai pu compter sur son expertise, sa bienveillance et ses conseils. Je suis reconnaissante du temps qu'elle m'a consacré ainsi que des ressources qu'elle m'a fournies.

Je souhaite aussi exprimer ma reconnaissance envers toutes les personnes qui m'ont accueilli chaleureusement au sein de CEVITAL et de Brandt. Je tiens particulièrement à remercier Monsieur Nadjib Benmessaoud (Directeur Applications Métiers et Centre de Compétences SAP) pour son aide précieuse, et pour avoir été toujours disponible pour répondre à mes questions et m'accueillir avec bienveillance. Un grand merci également à Madame Amina Ait Djaoud (Assistante exécutive | Ressources Humaines) pour son soutien lors du processus d'entretien et pour l'aide apportée dans les formalités administratives. Je tiens aussi à exprimer ma gratitude envers Madame Ghania Chemmoul (Cevital) et Fatima Souaci (Brandt), mes promotrices, pour le temps qu'elles m'ont consacrée et pour leurs réponses enrichissantes.

Je remercie également toute l'équipe pédagogique de l'École nationale supérieure d'informatique (ESI), non seulement pour avoir assuré la méthodologie de recherche et de rédaction tout en long de la formation master, ce qui m'a été d'une grande aide durant mon épreuve de master et ce stage de fin d'études, mais aussi pour avoir assuré une formation de qualité durant mon cursus dans cette école.

Et enfin, je ne serais remercié assez ma famille et mes amies pour leur soutien et encouragements au fil de mes études.

Résumé

De nos jours, les entreprises sont souvent soucieuses de leur identité en ligne, car elles sont conscientes que leur réputation peut être affectée ou même détruite par un seul avis négatif, c'est pourquoi elles essaient d'analyser et de comprendre les avis en ligne selon deux axes : (i) Pour estimer leur réputation (ii) Pour les utiliser dans l'estimation de la satisfaction client et l'amélioration de leur expérience. Toutefois, les avis ne sont pas forcément crédibles. En effet, des spameurs, qui peuvent être des particuliers ou des entreprises, rédigent expressément des avis négatifs/positifs pour nuire ou promouvoir une marque. À la suite d'un projet de recherche précédent sur l'analyse de la crédibilité des opinions, nous poursuivons l'investigation à travers l'analyse de la réputation en prenant Brandt comme cas d'étude. À cet effet, nous avons élaboré une revue de littérature des approches de détection de la crédibilité des avis à partir de laquelle nous avons identifié les défis à résoudre pour analyser l'e-réputation chez Brandt. Nous avons investigué plusieurs modèles d'analyse de sentiment et de crédibilité en considérant les défis de la langue, ensuite, nous avons proposé un framework d'analyse d'e-réputation qui met en œuvre tout le processus d'analyse allant de la collecte au choix de modèles d'identification de langue, et de prédiction de polarité et de crédibilité jusqu'aux calculs des indicateurs d'e-réputation. Ce framework permet aussi l'évaluation des modèles de façon régulière afin d'actualiser les indicateurs d'e-réputation (Aperçu de l'authenticité des avis, corrélation entre le spam et la polarité,...). Nos modèles les plus performants atteignent une exactitude (accuracy) de 98% pour l'identification de la langue, de 78% pour l'analyse des sentiments multilingues et de 66% pour la détection du spam multilingue.

Mots clés : Fouille D'opinion, Détection De Spam d'opinion, Crédibilité, E-Réputation.

Abstract

In today's world, businesses are often concerned about their online identity because they are aware that their reputation can be affected or even destroyed by a single negative review. This is why they try to analyze and understand online reviews along two main dimensions : (i) to assess their reputation, and (ii) to use them in estimating customer satisfaction and improving their overall customer experience. However, not all reviews can be considered trustworthy. Indeed, spammers, whether they are individuals or businesses, deliberately write negative or positive reviews to harm or promote a brand. Building on a previous research project focused on analyzing the credibility of opinions, we continue our investigation by examining reputation through a case study of the Brandt company. To do this, we have conducted a literature review of various approaches for detecting the credibility of reviews. From this, we have identified the challenges we need to address in analyzing e-reputation for Brandt. We have explored several models for sentiment analysis and credibility assessment, taking into account language-related challenges. Subsequently, we have proposed a framework for e-reputation analysis that encompasses the entire analysis process, from data collection and language identification to polarity and credibility prediction, as well as the calculation of e-reputation indicators. This framework also facilitates regular model evaluation to update e-reputation indicators, including an overview of review authenticity, the correlation between spam and polarity, and more. Our most effective models achieve an accuracy of 98% for language identification, 78% for multilingual sentiment analysis, and 66% for multilingual spam detection.

Keywords : Opinion Mining, Opinion Spam Detection, Credibility, E-Reputation.

ملخص

في الوقت الحاضر، تُظهر الشركات بشكل متكرر اهتماماً كبيراً بهويتها على الإنترنت، حيث تدرك أن سمعتها يمكن أن تتأثر أو حتى تتضرر بواسطة تقييم سلبي واحد، ولهذا السبب تسعى جاهدة لتحليل وفهم التقييمات على الإنترنت من خلال محورين رئيسيين: (i) لتقدير سمعتها (ii) لتقدير رضا العملاء وتحسين تجربتهم. ومع ذلك، ليس جميع التقييمات موثوقة، حيث يقوم الرسائل العشوائية، سواء كانت من أفراد أو شركات، بكتابة تقييمات إيجابية أو سلبية بشكل متعمد للتأثير على سمعة العالمة التجارية بشكل سلبي أو للترويج لها. بعد مشروع بحث سابق حول مصداقية التقييمات، نواصل التحقيق من خلال تحليل سمعة براندت كدراسة حالة. لتلبية احتياجات براندت، أجرينا استعراضًا للأدلة حول أساليب اكتشاف مصداقية التقييمات، ومن ثم حددنا التحديات التي يجب حلها لتحليل سمعتها على الإنترنت. قمنا أيضًا بفحص العديد من نماذج تحليل المشاعر والمصداقية، مع مراعاة التحديات المتعلقة باللغة. بناءً على ذلك، قدمنا إطاراً لتحليل سمعة العالمة التجارية على الإنترنت يشمل عملية جمع البيانات، واختيار نماذج تحديد اللغة، والتبؤ بالمشاعر والمصداقية، بالإضافة إلى حساب مؤشرات سمعة العالمة التجارية. يمكن استخدام هذا الإطار أيضًا لتقدير النماذج بانتظام لتحديث مؤشرات سمعة العالمة التجارية، مثل تقدير أصالة التقييمات والارتباط بين الرسائل العشوائية والمشاعر، وغيرها. أفضل نماذجنا تحقق دقة بنسبة 98% في تحديد اللغة، 78% في تحليل المشاعر بعدة لغات، و66% في اكتشاف الرسائل العشوائية بعدة لغات.

كلمات مفتاحية : التنبؤ عن الآراء، الكشف عن الآراء الكاذبة، المصداقية، السمعة الإلكترونية

Table des matières

Dédicace	I
Remerciements	II
Résumé	III
Abstract	IV
V	ملخص
Table des matières	VI
Table des figures	IX
Liste des tableaux	XI
Liste des sigles et acronymes	XII
Introduction générale	1
I Étude bibliographique	4
1 L'analyse des sentiments	5
1.1 Introduction	5
1.2 Définitions	5
1.2.1 Opinion	5
1.2.2 L'analyse des sentiments	8
1.2.3 E-réputation	8
1.3 Processus de l'analyse des sentiments	10
1.4 Types d'analyse de sentiments	11
1.4.1 Les niveaux de granularité	11
1.4.2 La tâche effectuée	14
1.5 Approches d'analyse des sentiments	15
1.5.1 Approches basées sur l'apprentissage automatique	16
1.5.2 Approches basées sur le lexique	18
1.5.3 Approches hybrides	18
1.6 Jeux de données disponibles	19
1.7 Métriques et méthodes d'évaluation	19
1.7.1 Méthodes de tests statistiques	19
1.7.2 Méthodes de retour de pertinence	22
1.8 Défis et discussions	22

Table des matières

1.8.1	Dépendance du contexte	22
1.8.2	Dépendance du domaine	23
1.8.3	Fautes d'orthographe	23
1.8.4	Langage figuratif	23
1.8.5	Crédibilité de l'opinion	23
1.9	Conclusion	24
2	La détection de Spam d'opinion	25
2.1	Introduction	25
2.2	Motivation	25
2.3	Définitions	27
2.3.1	La crédibilité	27
2.3.2	Le spam	27
2.3.3	Le spameur	30
2.4	Processus de la détection de spam d'opinion	32
2.5	Jeux de données	33
2.5.1	Méthodes de construction des ensembles de données	33
2.5.2	Ensembles de données disponibles	33
2.6	Approches de prétraitement	34
2.7	Approches de représentation	35
2.7.1	Ingénierie des caractéristiques	35
2.7.2	Apprentissage des représentations	42
2.8	Approches de classification des spams d'opinion	49
2.8.1	Modèles ML traditionnels	49
2.8.2	Modèles de Deep Learning	54
2.9	Défis et discussions	60
2.9.1	L'acquisition des données :	60
2.9.2	Les caractéristiques :	61
2.9.3	Les modèles de classification :	61
2.10	Conclusion	61
II	Contribution	62
1	État des lieux	63
1.1	Introduction	63
1.2	Présentation de l'organisme d'accueil	63
1.2.1	Laboratoire LCSI	63
1.2.2	Groupe Cevital	64
1.2.3	Brandt	65
1.3	Conclusion	67
2	Solution proposée	68
2.1	Introduction	68
2.2	Présentation de la solution	68
2.3	Vue conceptuelle de la solution	69
2.3.1	Entrées du système	69
2.3.2	Le cœur du système	70
2.4	Conclusion	76

Table des matières

3 Expérimentations et résultats	77
3.1 Introduction	77
3.2 Identification de la langue	77
3.3 Analyse des sentiments	78
3.4 Classification du type	83
3.5 Detection de spam	84
3.6 Agent conversationnel	85
3.7 Conclusion	86
4 Prototype de la solution	87
4.1 Introduction	87
4.2 Architecture de la solution	87
4.2.1 Architecture au niveau général	87
4.2.2 Architecture au niveau logique	88
4.3 Tableau de bord	90
4.3.1 Acteurs	91
4.3.2 Collecteur d'avis	91
4.3.3 Indicateurs clés de performance	92
4.3.4 Plateforme collaborateur	94
4.3.5 Plateforme administrateur	100
4.4 Page d'accueil conversationnelle	102
4.4.1 Acteurs	102
4.4.2 Interface de la page d'accueil	103
4.4.3 Agent conversationnel	107
4.5 Conclusion	110
Conclusion et perspectives	111
Bibliographie	114
Annexes	123
A Charte graphique	124

Table des figures

.1	Résultats de l'enquête Ever-Growing Power of Reviews.	1
I.1	Le processus d'analyse des sentiments	10
I.2	Niveaux de granularité de l'analyse des sentiments	12
I.3	Taxonomie des approches adoptées dans l'analyse des sentiments	15
I.4	La courbe ROC (Receiver Operating Characteristic curve)	21
I.5	La métrique AUC (Area Under the receiver operating characteristics Curve)	22
I.6	Offres de rédaction d'avis trompeurs sur le site Fiverr	26
I.7	AISEO: Outil de génération d'avis	26
I.8	La taxonomie du spam dans la littérature	27
I.9	La relation entre la polarité des opinions spam et la qualité des produits selon (Jindal et Liu 2008)	30
I.10	Processus de détection de spam	32
I.11	Taxonomie des approches de représentation	35
I.12	Exemple du fonctionnement de Bag-of-Words (BOW)	37
I.13	Exemple du fonctionnement des n-grammes	38
I.14	Exemple du fonctionnement de la technique POS (Part-of-speech)	39
I.15	Architecture simplifiée d'un auto-encodeur	43
I.16	Architecture simplifiée des algorithmes Word2vec	45
I.17	Architecture simplifiée de ELMo	46
I.18	Exemple du fonctionnement de ELMo	47
I.19	Exemples d'affinage de BERT sur différentes tâches (Devlin et al. 2018b)	48
I.20	Architecture de l'approche adoptée dans (Rastogi et al. 2020)	50
I.21	Architecture de l'approche adoptée dans (Z. Wang et al. 2020)	53
I.22	Architecture générale de OpCNN (Zhao et al. 2018)	54
I.23	Architecture générale de l'approche RSBE de (Zeng et al. 2019)	56
I.24	Architecture générale de l'approche de (Harris 2022)	57
I.25	Architecture générale de l'approche de (Tang et al. 2020)	57
I.26	Architecture générale de l'approche de (Andresini et al. 2022)	59
II.1	Logo du groupe Cevital.	64
II.2	Logo du groupe Brandt.	65
II.3	Directions de la filiale Brandt Algérie.	66
II.4	Organisation de la Direction Marketing Opérationnel.	66
II.5	Vue conceptuelle de la solution	69
II.6	Pipeline de prediction.	71
II.7	Module d'identification de langue	71
II.8	Module d'analyse de sentiments	72
II.9	Module d'extraction de type	73
II.10	Module de détection de spam	73

Table des figures

II.11	Arborescence du répertoire d'un projet React	89
II.12	Architecture logique des serveurs de la solution	90
II.13	Aperçu d'un rapport périodique exemple	94
II.14	Interface de connexion de la plateforme collaborateur	95
II.15	Interface du tableau de bord de la plateforme collaborateur	95
II.16	Interface de la vue d'ensemble de la plateforme collaborateur	96
II.17	Interface de la vue d'analyse de sentiments de la plateforme collaborateur	96
II.18	Interface de la vue d'analyse de crédibilité de la plateforme collaborateur	97
II.19	Interface de la vue d'analyse de langues de la plateforme collaborateur	98
II.20	Interface de la vue de statistiques divers de la plateforme collaborateur	98
II.21	Interface de la vue du tableau d'opinions de la plateforme collaborateur	99
II.22	Interface de gestion de compte de la plateforme collaborateur	100
II.23	Interface de connexion de la plateforme administrateur	100
II.24	Interface de gestion de l'outil de la plateforme administrateur	101
II.25	Interface du tableau de bord de la plateforme administrateur	101
II.26	Interface du tableau de bord de la plateforme administrateur	102
II.27	Schéma de Zonage de la Page d'Accueil Conversationnelle	103
II.28	En-tête de la Page d'Accueil Conversationnelle	103
II.29	Menu de navigation de la Page d'Accueil Conversationnelle	104
II.30	Chatbot de la Page d'Accueil Conversationnelle	104
II.31	Section d'accueil de la Page d'Accueil Conversationnelle	105
II.32	Section de témoignages de la Page d'Accueil Conversationnelle	106
II.33	Section de contacts de la Page d'Accueil Conversationnelle	107
II.34	Pied de page de la Page d'accueil Conversationnelle	107
II.35	Exemples de Réponses du Chatbot	108
II.36	Exemples de réponses Chatbot - Scénarios Non Résolus	108
II.37	Processus de Réponse du Chatbot	109
II.38	Typographie de la solution	126
II.39	Palette de couleurs de la solution	127
II.40	Variations de logo de Brandt Insight	128
II.41	Bbot : Personnage de Brandt Bot	129
II.42	Variations de logo de Brandt Insight	129
II.43	Icônes utilisées pour la solution	130
II.44	Éléments d'Interface de la solution	130
II.45	Illustration pour la plateforme collaborateur de Brandt Insight	131
II.46	Illustration pour la plateforme administrateur de Brandt Insight	132
II.47	Illustrations pour la page d'accueil conversationnelle	133

Liste des tableaux

I.1	Exemples de types d'opinion	8
I.2	Avantages et inconvénients de différents outils de veille sur la e-réputation	9
I.2	Avantages et inconvénients de différents outils de veille sur la e-réputation	10
I.3	Résultats d'un exemple de fouille d'opinion sur divers niveaux d'analyse.	13
I.4	Quelques corpus utilisés dans les travaux d'analyse de sentiment	19
I.5	Table de confusion - Analyse des sentiments	20
I.6	Ensembles de données de référence (Gold-standard) utilisés dans les travaux de la littérature pour la détection de spam d'opinion.	34
II.1	Exemple de détection de polarité avec et sans prise en compte du contexte	72
II.2	Exemples d'entrées sorties du système	74
II.3	Répartition des langues dans le jeu de données d'entraînement pour le modèle d'identification de langues.	78
II.4	Résultats des expérimentations des modèles d'identification de langue.	78
II.5	Meilleurs résultats pour les expériences d'analyse de sentiment pour l'anglais. .	79
II.6	Descriptions des jeux de données pour le modèle d'analyse de sentiments.	80
II.7	Lemmatiseurs et Modèles BERT utilisés pour chaque langue.	80
II.8	Meilleurs résultats du modèle d'analyse de sentiments pour chaque langue.	81
II.9	Résultats des différentes expériences d'analyse de sentiments multilingues	82
II.10	Exemple de classification de dépendance/indépendance et d'accord/désaccord. .	83
II.11	Exemple de classification de type de texte.	83
II.12	Meilleurs résultats obtenus durant les Expériences de détection de spam pour l'anglais.	84
II.13	Meilleurs résultats obtenus avec le modèle de détection de spam pour chaque langue et pour la méthode d'ensemble.	85
II.14	Exemples de réponses du chatbot.	86

Liste des sigles et acronymes

AI	<i>Artificial Intelligence</i>
AMT	<i>Amazon Mechanical Turk</i>
ANN	<i>Artificial Neural Network</i>
ANOVA	<i>Analysis of variance</i>
AUC	<i>Area Under the receiver operating characteristics Curve</i>
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
BOW	<i>Bag Of Words</i>
CBOW	<i>Continuous Bag of Word</i>
CNN	<i>Convolutif Neural Network</i>
DL	<i>Deep Learning</i>
DR	<i>Dimensionality Reduction</i>
ELMo	<i>Embeddings from Language Models</i>
EUPHORIA	<i>nEural mUlti-view aPproacH fOr RevIew spAm</i>
FFNN	<i>Feedforward neural network</i>
FN	<i>False Negative (Faux négatif)</i>
FP	<i>False Positive (Faux positif)</i>
FSL	<i>Few-Shot Learning</i>
GAN	<i>Generative Adversial Network</i>
GloVe	<i>Global Vectors for Word Representation</i>
GMM	<i>Gaussian Mixture Modelling</i>
GPT	<i>Generative Pre-trained Transformers</i>
GRU	<i>Gated recurrent units</i>
ICA	<i>Independent Component Analysis</i>
KLDA	<i>Kernel Linear Discriminant Analysis</i>

Liste des tableaux

KPCA	<i>Kernel Principal Component Analysis</i>
LCSI	<i>Laboratoire de la communication dans les Systèmes Informatiques</i>
LDA	<i>Linear Discriminant Analysis</i>
LIWC	<i>Linguistic Inquiry and Word Count</i>
LR	<i>Logistic Regression</i>
LSI	<i>Latent Semantic Indexing</i>
LSTM	<i>Long-Short Term Memory</i>
MGSD	<i>Multi-iterative Graph-based opinion Spam Detection</i>
ML	<i>Mchine Learning</i>
MLM	<i>Masked Language Model</i>
MLP	<i>Multi-Layer Perceptron</i>
MMI	<i>Modified Mutual Information</i>
NB	<i>Naive Bayes</i>
NLP	<i>Natural Language Processing</i>
OOV	<i>Out Of Vocabulary</i>
P	<i>Précision</i>
PCA	<i>Principal Component Analysis</i>
PNN	<i>Perceptron Neural Network</i>
POS	<i>Part Of Speech</i>
PR	<i>Precision-Recall Curve</i>
	<i>Positive Unlabeled (Learning)</i>
R	<i>Rappel</i>
ReLU	<i>Rectified Linear Unit</i>
RI	<i>Recherche d'information</i>
RLOSD	<i>Representation Learning based Opinion Spam Detection</i>
RNN	<i>Recurrent Neural Network</i>
ROC	<i>Receiver Operating Characteristic curve</i>
SMOTE	<i>Synthetic Minority Over-sampling TEchnique</i>
SOM	<i>Self-Organizing Maps</i>
SVD	<i>Singular Value Decomposition</i>
SVM	<i>Support Vector Machine</i>

Liste des tableaux

TALN	<i>Traitemenat automatique du langage naturel</i>
TF	<i>Term Frequency)</i>
TF-IDF	<i>Term Frequency (TF)-inverse document frequency)</i>
TFP	<i>Taux de faux positifs</i>
TN	<i>True Negative (Vrai négatif)</i>
TP	<i>True Positive (Vrai positif)</i>
TVP	<i>Taux de vrais positifs</i>
Word2vec	<i>Word to Vector</i>

Introduction générale

Contexte

Grâce aux progrès de la technologie, en particulier du Web, on observe de nos jours une croissance sans cesse grandissante du commerce électronique, en particulier dans le marketing en ligne. Un des facteurs qui font émerger ce type de marketing et facilitent son déploiement est la disponibilité d'opinions publiques sur les entreprises-qui adoptent cette approche-et leurs produits/services, ce qui fournit aux clients une source d'information fiable pour la prise de décision.



FIG. .1 : Résultats de l'enquête Ever-Growing Power of Reviews.

Les opinions en ligne affectent les décisions d'achat des clients. La Figure .1 ci-dessus représente les résultats d'une enquête menée par The PowerReviews sous le nom de Ever-Growing Power of Reviews¹, en 2021, qui a interrogé 6 538 clients à travers les États-Unis. Selon cette enquête, les notes et les avis sont devenus le facteur le plus important qui influe sur les décisions d'achat, se classant au-dessus du prix, de la livraison gratuite, de la marque et des recommandations de la famille et des amis. Et ceci contrairement à des enquêtes similaires qu'ils ont menées en 2014 et 2018, où le prix était le facteur le plus important.

¹<https://www.powerreviews.com/insights/power-of-reviews-survey-2021/>

Introduction générale

En raison de l'impact significatif des avis clients sur les décisions d'achat, les entreprises accordent désormais une attention particulière à leur notoriété en ligne. Elles reconnaissent l'importance cruciale de leur e-réputation et cherchent activement à la comprendre et à la gérer. Cette préoccupation est particulièrement considérable pour les grandes marques, telles que notre organisation hôte, Brandt, qui doivent faire face à un volume important d'opinions en ligne.

Brandt, en tant que marque de renommée internationale, est confronté au défi de surveiller en permanence sa réputation numérique. La marque est consciente que la perception en ligne de ses produits et de ses services influe considérablement sur la fidélisation de sa clientèle existante et sur l'acquisition de nouveaux clients.

Problématique

Cette abondance d'information sur le web a fait émerger de nouveaux problèmes : il a été démontré que dans des conditions concurrentielles et classées, les opinions publiques peuvent être utilisées avec des intentions malveillantes qui pourraient être de promouvoir ou de rétrograder un produit ou un service cible dans le but d'induire en erreur les consommateurs pour un certain gain souvent financier. C'est ce qu'on appelle le spam d'opinion.

Cette augmentation des avis trompeurs sur le web a réduit la crédibilité des opinions publiques et a affecté la confiance des clients qui ne savent plus qui ou quoi croire. Ces avis dits spam peuvent être soit générés par des personnes payées pour cela ou par des algorithmes, ce qui rend la tâche de détection plus difficile puisque ces algorithmes peuvent être ajustés pour battre les techniques de détection de l'état de l'art.

Le spam d'opinion est un sujet d'intérêt qui a conduit à l'émergence de nombreuses approches qui visent à le détecter et le classifier pour faciliter la prise de décision. Cette problématique, qui revêt une importance capitale pour les entreprises et les marques, a été abordée pour la 1re fois par (JINDAL et LIU 2008). Cependant, pour des entreprises comme Brandt, qui accordent une grande importance à la veille sur la réputation en ligne, cette question s'étend bien au-delà de la simple détection du spam d'opinion. La surveillance et l'analyse approfondie de la réputation en ligne deviennent essentielles pour maintenir une image de marque positive et pour prendre des décisions stratégiques informées. C'est dans ce contexte que se situe la nécessité de développer des solutions plus complètes et personnalisées pour la gestion de la réputation en ligne, en prenant en compte les particularités propres à chaque entreprise, comme c'est le cas pour Brandt Algérie. Nous nous intéressons aux travaux de la littérature qui ont été proposés, depuis, avec un intérêt particulier aux questions suivantes :

- Comment peut-on définir la polarité d'une opinion ? Et est-elle un indicateur déterminant pour la crédibilité des opinions ?

- Compte tenu du contexte actuel, quelle serait la définition de la crédibilité ?

- Que peut-on définir en tant que spam d'opinion ?

C'est dans ce contexte que les collaborateurs de Brandt ont exprimé le besoin de suivre de près les avis en ligne concernant leur marque. Ils souhaitent être notifiés des nouvelles opinions, ainsi que de leur évolution. C'est précisément ce besoin qui nous a incité à relancer un projet précédent (MAAFI 2020) axé sur la détection de la crédibilité des opinions en ligne.

Objectifs

- Réaliser un état de l'art exhaustif présentant les différentes approches et techniques présentes dans la littérature afin de répondre aux problématiques posées.

Introduction générale

- Mener une étude exploratoire afin d'identifier les meilleures techniques permettant d'atteindre les meilleures performances lors de la résolution de ce problème.

- Proposer une solution répondant au problème posé.

- Développer un système, démontrant la faisabilité et l'intérêt de la solution proposée dans une situation réelle, tout en répondant aux besoins de l'entreprise.

Structure du mémoire

Ce mémoire est organisé comme suit :

Partie 01 : Étude bibliographique

La première partie “**Étude bibliographique**” synthétise les travaux de l'état de l'art pertinents aux problématiques posées.

Le premier chapitre couvre les concepts de base, les tâches, les sous-domaines et les défis rencontrés dans le domaine de l'analyse des sentiments.

Dans le deuxième chapitre, nous introduisons le domaine de la détection de spam d'opinions. Nous présentons les concepts de base de la détection de spam nécessaires pour la compréhension du domaine, son processus général, ainsi qu'une synthèse des travaux de la littérature.

Nous clôturons cette partie par une conclusion générale de nos recherches ainsi que les perspectives de développements mis en avant à travers les deux chapitres précédents.

Partie 02 : Étude de l'existant

La deuxième partie “**Étude de l'existant**” est divisée en deux chapitres. D'une part, le premier chapitre consiste à la présentation de l'organisme d'accueil. D'autre part, le deuxième représente une étude critique des solutions existantes sur le marché qui répondent à un ou plusieurs objectifs de ce mémoire.

Partie 03 : Contribution

La troisième partie “**Contribution**” quant à elle présente la contribution apportée par ce travail, elle est composée de trois chapitres. Nous introduisons, dans le premier chapitre, notre solution et les éléments qui la composent. Puis, dans le deuxième chapitre, nous présentons expérimentations que nous avons menées ainsi qu'une interprétation des résultats obtenus.

Dans le troisième chapitre de cette partie, nous détaillons également le système réalisé pour la marque Brandt. Nous analysons les besoins de l'entreprise ainsi que l'intérêt de notre solution, tout en présentant l'architecture, la conception et la réalisation des différents composants de cet outil.

Conclusion et perspectives

Enfin, dans cette partie, nous concluons notre mémoire par un bilan des résultats obtenus, des défis rencontrés ainsi qu'une présentation des perspectives de recherche ouvertes par notre travail.

Partie I

Étude bibliographique

Chapitre 1

L'analyse des sentiments

1.1 Introduction

Plusieurs enquêtes telles que celle de The PowerReviews¹ ont montré l'effet majeur des opinions en ligne sur les décisions d'achat des consommateurs et sur la réputation de l'entreprise. Ces opinions, autrefois des opportunités pour lesquelles les entreprises, commerciales en particulier, dépensaient des montants colossaux afin de les collecter, sont désormais disponibles instantanément pour peu qu'elles identifient les plus pertinents ou les plus crédibles.

L'objectif de ce chapitre est de donner un état des lieux des approches suivis dans l'analyse des opinions (sentiments) exprimées dans les avis des consommateurs vis-a-vis des entreprises, des produits et des marques. Dans ce qui suit, nous allons tous d'abord définir les concepts clés de l'analyse des sentiments ainsi que les étapes de ce processus (Section 1.2 et 1.3). Nous poursuivrons dans la section 1.4 par la classification de la fouille d'opinion selon plusieurs critères dont les niveaux de granularité et les objectifs de l'analyse. Nous nous intéresserons ensuite aux approches existantes (1.5), les jeux de données disponibles (1.6) ainsi que les métriques d'évaluations (1.7). Nous conclurons par une discussion des défis rencontrés ainsi que les axes de recherches futurs (Section 1.8).

1.2 Définitions

Dans ce qui suit, nous allons définir certains concepts de base nécessaires à la compréhension de notre domaine de recherche, principalement, l'opinion, l'E-réputation et la fouille d'opinion.

1.2.1 Opinion

Nous commençons par la définition de la notion qui sera le cœur de nos études, l'**opinion**. Selon le dictionnaire Larousse², Une opinion est un « *Jugement, avis, sentiment qu'un individu ou un groupe émet sur un sujet, des faits, ce qu'il en pense* »; ou l'« *Ensemble des idées d'un groupe social sur les problèmes politiques, économiques, moraux, etc.* ». Selon le dictionnaire Le Robert³ quant à lui, l'opinion, c'est la « *Manière de penser, de juger* », c'est l'avis, la conviction, la croyance, l'idée, le jugement, la pensée et le point de vue.

En ce qui est du domaine informatique, une opinion peut être caractérisée comme une croyance sur des sujets généralement considérés comme subjectifs ; c'est le résultat d'une émotion ou d'une interprétation des faits, elle peut être positive, négative ou neutre (GHELANI et BHALODIA 2017). Similairement, l'étude de (PADMAJA et FATIMA 2013) définit une opinion

¹<https://www.powerreviews.com/insights/power-of-reviews-survey-2021/>

²<https://www.larousse.fr/dictionnaires/francais/opinion/56197>

³<https://dictionnaire.lerobert.com/definition/opinion>

comme un jugement de personnes formé sur une chose particulière et n'est pas nécessairement basé sur des faits ou des connaissances.

1.2.1.1 Composantes d'une opinion

Une opinion fait généralement référence à la composition des éléments suivants :

- **Une entité cible**, ou *Opinion Target* en anglais, qui est la cible sur laquelle porte l'opinion, elle peut être un produit, un service, un sujet, un problème, une personne, une organisation ou un événement (LIU 2012). Dans l'exemple qui suit, nous prenons un "Laptop Asus" comme entité cible de notre opinion.

- **Un aspect**, vu que l'opinion peut soit porter sur l'entité cible en général et l'ensemble de ces caractéristiques (Ce qu'on appelle aspect « GENERAL »), soit elle se concentre sur une de ses propriétés, un de ses attributs ou même un de ses composants. La capacité de stockage, celle de la batterie, la résolution de l'écran, etc sont toutes des aspects de l'entité "Laptop Asus".

Posons ces deux exemples d'opinion portant sur l'entité "Laptop Asus" :

Exemple 1 : "*I love everything about this laptop, from its screen's resolution to its speed to its sleek design. Marvellous machine 10/10*"

Exemple 2 : "*I'm not too fond of this laptop's graphical card. When I'm gaming on ultra it makes my games lag which cost me few victories here and there...*"

Le premier exemple critique l'ensemble des caractéristiques de l'entité "Laptop Asus" en général, en revanche, le deuxième porte seulement sur l'aspect 'Carte graphique' de l'entité.

- **Un titulaire**, ou *Opinion Holder* en anglais, l'auteur, le porteur, le détenteur ou la source de l'opinion : c'est la personne/organisation ayant exprimé l'opinion.

- **Un instant**, c'est le moment où l'opinion a été exprimée par la source (l'auteur).

- **Une orientation**, également appelée orientation de sentiments, orientation sémantique, ou polarité, sur l'aspect de l'entité. La polarité peut être définie par des catégories telles que « positif », « neutre », et « négatif », ou encore par des nombres dénotant le degré de positivité ou de négativité.

L'exemple 1 présenté précédemment exprime une orientation positive envers l'aspect général de l'entité "Laptop Asus", par contre, l'exemple 2 est de polarité négative concernant l'aspect "Carte graphique" de l'entité "Laptop Asus".

1.2.1.2 Différents types d'opinions

D'après (LIU 2020), les opinions peuvent être classées selon de nombreuses dimensions. Dans cette partie, nous discuterons de certaines des classifications principales :

- **Individuelle ou Collective**

Une première dimension de classification des avis que nous pouvons conclure dès la définition de l'opinion dans les différentes encyclopédies (tel que celle de Larousse citée auparavant) est celle concernant le titulaire, ou l'auteur, de celle-ci.

Dans un premier temps, l'opinion est définie au niveau de l'individu comme une pensée d'une seule personne, c'est le cas des avis en ligne critiquant les différents produits ou services offerts sur des sites tel que Yelp, TripAdvisor, Amazon,... La seconde définition fait référence à une opinion partagée entre un groupe de personnes ayant les mêmes idéologies sociales, politiques, économiques, etc.

- **Régulière ou Comparative**

Dans la littérature, une opinion **régulière** est souvent désignée simplement comme une opinion. C'est une pensée exprimée envers une certaine entité, donc « I really enjoyed this experience, i'm definitely coming again! » est une opinion régulière. Comme nous le verrons dans le titre suivant, ce type d'opinion a deux sous-types principaux : Direct ou Indirect (LIU 2006, 2011, 2012, 2020).

L'opinion **comparative**, quant à elle, exprime une relation de similitudes ou de différences entre deux ou plusieurs entités et/ou une préférence du détenteur de l'opinion basée sur certains aspects communs des entités (JINDAL et LIU 2006a,b ; LIU 2012, 2020). Par exemple, les phrases « Product A is much better than product B » et « Product A is the best » expriment deux avis comparatifs. Une opinion comparative est généralement exprimée en utilisant la forme comparative ou superlatrice d'un adjectif ou d'un adverbe, mais pas toujours (par exemple, « I prefer using product A instead of product B. »).

- **Directe ou Indirecte**

Revenons aux opinions régulières, comme nous l'avons déjà mentionné, ces dernières peuvent être classifiées selon (LIU 2006, 2011, 2012, 2020) sous deux sous-types principaux : une opinion (régulière) directe ou indirecte. Une opinion **directe** est exprimée directement sur une entité ou un aspect de celle-ci contrairement à l'opinion **indirecte** qui se base plutôt sur l'effet positif ou négatif de l'entité (ou d'un de ses aspects) sur d'autres entités.

Par exemple, l'opinion « This laptop's screen resolution is great. » est une opinion directe portant sur l'aspect de la résolution de l'écran de l'entité laptop. Tandis que l'opinion « ...The product gave me extreme rashes and my face was itchy for the next couple of days... » (Exemple extrait du dataset (Jessica LI 2018)) exprime l'effet indésirable du produit (Adapalene / benzoyl peroxide) sur une autre entité (la peau du titulaire) et donc elle exprime indirectement un avis négatif sur l'entité (le produit en question).

- **Implicite ou Explicite**

Le dernier type d'opinion dont nous allons discuter classe les opinions (régulières ou comparatives) comme implicites ou explicites (LIU 2012).

Une opinion (régulière ou comparative) est dite **explicite** quand elle exprime une déclaration subjective en utilisant des mots d'opinion. Par exemple, « This product is **terrible** and it left a **horrendous** taste after I took it, it made me feel even more nauseous than before » exprime une opinion négative sur le produit à travers les mots d'opinion “Terrible”, “Horrendous”,...

Cependant, Une opinion **implicite** est une déclaration objective (régulière ou comparative) exprimant, généralement, des faits désirables ou indésirables, e.g. l'opinion “Bactrim cleared my skin up in less than a week.” (extraite du dataset (Jessica LI 2018)) ne contient aucun mot d'opinion explicite, mais il est facile de déduire qu'elle exprime un effet désirables (élimination de l'acné) et donc qu'elle est positive.

Comme il est plus facile de détecter les opinions explicites que les opinions implicites, la plupart des recherches présentes dans la littérature se sont concentrées sur les avis explicites (LIU 2012).

Notons que dans la suite de ce mémoire, le terme « Opinion » fera référence aux opinions régulières. En addition, le tableau I.1 est un tableau récapitulatif contenant des exemples d'opinions de différentes catégories afin de mieux comprendre les dimensions de classification citées dans cette partie.

TAB. I.1 : Exemples de types d'opinion

Document	Type d'opinion
I love this product ! it's quite cheap and its quality is great .	Positive, Directe, Explicite
This serum should be removed from the market before influencers start spreading it around !!!	Négative, Directe, Implicite
This product made my <i>skin</i> look Amazing , my <i>complexion</i> looks better than ever !	Positive, Indirecte, Explicite
After using this serum, my <i>skin</i> became oilier and I got my first breakout in years !	Négative, Indirecte, Implicite

1.2.2 L'analyse des sentiments

Il existe plusieurs termes qui renvoient à l'analyse des sentiments (Sentiment analysis), dont « Opinion mining », « Review mining », « Subjectivity analysis », « Sentiment AI », et d'autres termes qui sont maintenant tous connus sous l'appellation d'analyse des sentiments. Le terme « Sentiment Analysis » est apparu pour la première fois dans (NASUKAWA et YI 2003), alors que « Opinion mining » a été utilisée pour la première fois dans l'article de (DAVE et al. 2003). Il existe certaines différences entre toutes ces dénominations, mais elles restent insignifiantes vu qu'elles réfèrent toutes à l'analyse des sentiments subjectifs dans un texte. Toutefois, le terme « Sentiment analysis » (Analyse des sentiments) est plus couramment utilisé dans l'industrie tandis que le terme « Opinion mining » (Fouille d'opinion) est fréquemment utilisé dans le milieu académique (LIU 2012).

L'analyse des sentiments, ou la fouille d'opinion, fait référence au domaine qui croise plusieurs disciplines telles que le traitement automatique du langage naturel (TALN ou NLP), la recherche d'information (RI), la fouille de texte ainsi que le web mining. Le but principal de l'analyse des sentiments est de catégoriser les opinions exprimées au sein d'un texte selon le critère d'orientation (ou polarité) dans, généralement, trois classes distinctes Positive, Négative ou Neutre. (LIU 2012, 2020) définit l'analyse des sentiments, ou la fouille d'opinion, comme le domaine d'étude qui analyse les opinions, les sentiments, les appréciations, les attitudes et les émotions des personnes envers les entités et leurs attributs exprimés dans un texte écrit.

L'analyse de sentiment peut aussi être définie comme un processus, qui prend en entrée une entité cible et donne en sortie un résumé des opinions vis-à-vis de l'entité ou de ses aspects (DAVE et al. 2003).

1.2.3 E-réputation

Un des buts principaux de l'analyse des sentiments pour les entreprises est de déterminer leur réputation en ligne. D'après le dictionnaire Larousse⁴, la réputation peut être définie comme une « *Manière dont quelqu'un, quelque chose est connu, considéré dans un public* » ou aussi comme une « *Opinion favorable ou défavorable du public pour quelqu'un, quelque chose* ».

En ce qui est de la réputation en ligne ou l'e-réputation, celle-ci se définit tout simplement comme la réputation, la notoriété ou l'image véhiculée d'une entité (une entreprise, une marque, un particulier, un produit...) sur les différents supports web (les moteurs de recherche, les réseaux sociaux, les plateformes d'avis...).

Il existe multiples solutions pour surveiller l'e-réputation sur le marché, allant de simples

⁴<https://www.larousse.fr/dictionnaires/francais/r%C3%A9putation/68543>

Chapitre 1. L'analyse des sentiments

solutions gratuites telles que Google Alerts⁵, aux tableaux de bord payants offrant une analyse en temps réel ainsi que des informations et des fonctionnalités intéressantes.

L'idée principale derrière ces solutions est simple : parcourir toutes les données disponibles publiquement pour certains mots-clés, collecter les données, et analyser leur sentiment pour fournir des informations pertinentes. Les tableaux de bord sont visuellement attrayants et présentent des informations intéressantes de manière organisée. Parmi les produits disponibles sur le marché, on peut citer SEMrush⁶, SentiOne⁷, Chatmeter⁸ et Meltwater⁹.

Cependant, ces solutions présentent également quelques inconvénients, parmi eux :

- Elles ne considèrent pas certaines sources uniquement, mais scannent l'ensemble du web. Cela signifie qu'en utilisant ces solutions pour Brandt Algérie, nous pourrions obtenir des informations provenant d'autres branches de l'entreprise ou même des noms de personnes qui partagent le nom "Brandt" sans rapport avec l'entreprise.
- Plusieurs de ces solutions ne sont pas multilingues, et même si elles le sont, elles ne prennent pas en compte le dialecte algérien, ce qui peut fausser les statistiques obtenues.
- Elles ne traitent pas le problème répandu des avis et des commentaires spam qui peuvent, eux aussi, fausser l'analyse et portent un grand risque sur la réputation en ligne de la marque.

Le tableau I.2 présente différents avantages et inconvénients des outils mentionnés.

TAB. I.2 : Avantages et inconvénients de différents outils de veille sur la e-réputation

outil	Avantages	Inconvénients
SEMrush	<ul style="list-style-type: none">• Boîte à outils marketing tout-en-un.• Large gamme de fonctionnalités SEO.• Bon outil de recherche de mots-clés.• Analyse détaillée des concurrents.• Support de qualité.	<ul style="list-style-type: none">• Prix élevé.• Courbe d'apprentissage modérée.• L'interface utilisateur n'est pas très conviviale pour les débutants.• Fournit des données uniquement pour Google.
SentiOne	<ul style="list-style-type: none">• Rapports et analyses.• Modèles personnalisables.• Analyse de campagnes.• Suivi de l'engagement.	<ul style="list-style-type: none">• Erreurs lors de la détection de la langue et du sentiment.• Mauvais support client.• Lenteur de certains graphiques.• Surcharge du tableau de bord par défaut.

⁵<https://www.google.com/alerts>

⁶<https://www.semrush.com/>

⁷<https://sentione.com/>

⁸<https://www.chatmeter.com/>

⁹<https://www.meltwater.com/en>

TAB. I.2 : Avantages et inconvénients de différents outils de veille sur la e-réputation

Outil	Avantages	Inconvénients
Chatmeter	<ul style="list-style-type: none"> Tableau de bord des avis unifié. Analyse des mentions sur les médias sociaux. Générateur automatisé de témoignages. Bon service client. 	<ul style="list-style-type: none"> Coûteux. Difficulté à répondre aux avis. Limitations dans la flexibilité des rapports. Conception de l'interface utilisateur à améliorer.
Meltwater	<ul style="list-style-type: none"> Bonne couverture. Interface utilisateur moderne. Facilité de configuration. 	<ul style="list-style-type: none"> Service client médiocre. Prix élevé. Courbe d'apprentissage complexe. Erreurs fréquentes.

1.3 Processus de l'analyse des sentiments

Le processus de fouille d'opinion le plus cité dans la littérature est celui du professeur Bing Liu, de l'université de l'Illinois à Chicago (UIC) ¹⁰, (LIU 2012, 2020), il est considéré comme le noyau principal du pipeline de l'opinion mining. Il est composé de trois étapes principales (I.1) :

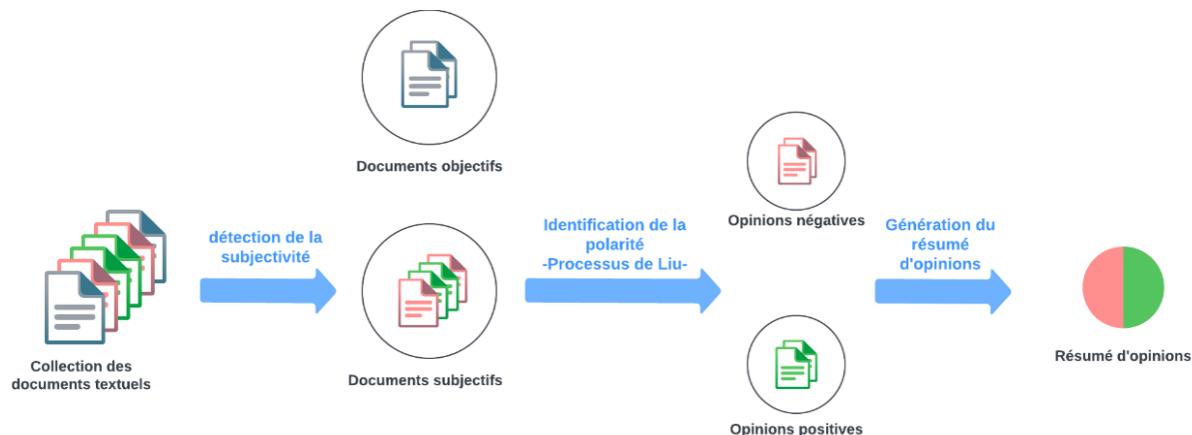


FIG. I.1 : Le processus d'analyse des sentiments

- Dans un premier temps, après la collecte des documents textuels pertinents vis-à-vis de l'entité cible, ces derniers sont analysés afin de détecter leur **subjectivité**.

- Ensuite, les documents identifiés comme subjectifs sont analysés à leur tour afin d'identifier leur polarité en se basant sur un processus proposé par (LIU 2012, 2020).

- Pour pouvoir enfin générer un résumé, visuel ou textuel, des opinions exprimées envers la cible.

Dans ce qui suit, nous allons décrire le processus d'identification de la polarité comme l'a

¹⁰ <http://www.uic.edu/>

formalisé le professeur Bing Liu (LIU 2012, 2020).

D'abord, le texte non structuré (l'opinion) est transformé en une forme plus structurée afin de pouvoir l'exploiter. Dans son livre « Sentiment Analysis and Opinion Mining » (LIU 2012), Liu définit une opinion par un quintuple $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$, où :

- **e_i** : Le nom de l'entité cible.
- **a_{ij}** : Un aspect de l'entité cible e_i . Si l'opinion porte sur l'entité e_i elle-même dans son ensemble, l'aspect a_{ij} est « GENERAL ».
- **s_{ijkl}** : Le sentiment, ou orientation, de l'opinion sur l'aspect a_{ij} de l'entité e_i .
- **h_k** : Le titulaire ou la source de l'opinion. Il peut faire référence à un individu ou un groupe d'individus.
- **t_l** : L'instant où l'opinion a été exprimée par la source d'opinion h_k .

Notons que dans cette définition, nous utilisons volontairement des indices pour souligner le fait que les cinq composants du quintuple doivent correspondre les unes aux autres. Donc un quintuple s_{ijkl} fait référence au sentiment de l'opinion exprimé par h_k à l'instant t_l concernant l'aspect a_{ij} de l'entité e_i .

Considérons maintenant un document subjectif D, l'objectif est d'identifier la polarité de toutes les opinions qui y sont exprimées. Chacune de celles-ci sera représentée par le quintuple de Bing Liu décrit ci-dessus. À partir d'un document subjectif **D** :

1. Extraire les entités et grouper les synonymes éventuels dans des clusters d'entités. Chaque cluster représente une entité e_i .
2. Pour chaque entité e_i , extraire tous ses aspects associés. Chaque cluster représente un aspect a_{ij} pour une entité e_i .
3. Extraire le titulaire h_k de l'opinion ainsi que la date t_l de son émission.
4. Pour chaque aspect a_{ij} d'une entité e_i , déterminer les orientations des opinions. La polarité s_{ijkl} peut être positive, négative ou neutre.
5. Générer l'ensemble des tuples $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ qui représentent toutes les opinions exprimées dans le document D, sur la base des résultats des tâches précédentes.

1.4 Types d'analyse de sentiments

La classification automatique de la polarité peut être catégorisée en fonction de diverses perspectives (YADOLLAHI et al. 2017), dans cette section (1.4), nous présentons une classification de l'analyse de sentiments selon deux critères principaux : d'abord la granularité, puis la tâche effectuée.

1.4.1 Les niveaux de granularité

La fouille d'opinion peut être effectuée à différents niveaux de granularité du texte de l'opinion (LIU 2012, 2020). Dans ce qui suit, nous présenterons les différents niveaux de l'analyse des sentiments comme illustré dans la figure I.2.

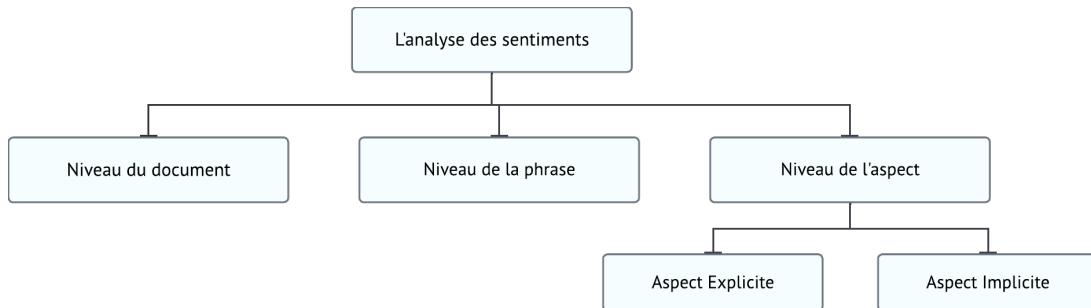


FIG. I.2 : Niveaux de granularité de l'analyse des sentiments

1.4.1.1 Niveau du document

L'objectif de la fouille d'opinion au niveau du document (macroscopique) est de déterminer la polarité d'un texte entier. L'hypothèse est que *le texte, qu'il soit court ou long, n'exprime qu'une opinion vis-à-vis d'une seule entité* (un seul produit, une seule marque...), par conséquent, ce type de fouille d'opinion ne s'applique pas aux documents qui évaluent ou comparent plusieurs entités, pour lesquelles une analyse plus fine est nécessaire.

L'analyse au niveau du document est importante, car elle donne une vue d'ensemble sur une entité cible, mais elle a ses lacunes (LIU 2012). Par exemple, même si ce type d'analyse est adapté aux avis qui ne ciblent qu'une entité spécifique, il est insuffisant lorsqu'il s'agit de non-avis tels que les discussions de forum, les blogs et les articles de presse qui pourraient comparer différentes entités. Aussi, ce type d'analyse n'extrait pas les détails qui pourraient intéresser certains utilisateurs, par exemple un utilisateur pourrait être intéressé de savoir quel aspect de son produit a été détesté ou aimé, etc. De plus, à cause de sa longueur, un document peut contenir des passages exprimant des informations factuelles objectives, ou encore des phrases non pertinentes au sujet.

1.4.1.2 Niveau de la phrase

Au niveau mésoscopique (niveau de la phrase), la polarité est calculée individuellement pour chaque phrase subjective contenue dans un document. Et comme l'affirme (LIU 2012, 2020), ce type d'analyse est étroitement lié à ce que l'on appelle la « classification de subjectivité » dont le but est de classer si une phrase est subjective ou objective.

L'analyse au niveau de la phrase se base sur l'hypothèse que *chaque phrase dans un texte ne contient qu'un seul avis concernant une entité unique*. L'un de ses inconvénients est le fait que le contexte entourant la phrase peut être assez définitif sur son sens et donc sa polarité : “This experience has left me speechless” pourrait être une déclaration positive ou négative selon la contexte dans lequel il est utilisé. De nos jours, La classification de polarité des tweets est l'application la plus intéressante de la fouille d'opinion au niveau de la phrase (YADOLLAHI et al. 2017).

1.4.1.3 Niveau de l'aspect

La fouille d'opinion au niveau de l'aspect est une analyse plus fine que les niveaux précédents, en se basant sur l'hypothèse qu'*une opinion possède une seule polarité et cible un aspect d'une entité (produit, service...)*.

Ce type d'analyse d'opinion vise à identifier les différents aspects ciblés d'une entité et de définir le sentiment que la source de l'opinion porte à son égard. Par exemple, une opinion qui déclare “I love this product ! ” est positive certes, mais elle n'apporte pas de détails e.g. des points que le client préfère vis-a-vis du produit. Aussi, même si une opinion est positive, elle

peut contenir des aspects négatifs sur l'entité et, par exemple, pour un consommateur intéressé par la qualité de la caméra d'un smartphone l'opinion “ I love this phone, it's performant and great, even if the quality of its camera sucks” l'influencera probablement à ne pas acheter ce smartphone, même si la polarité globale de l'opinion est positive. Voici deux exemples d'opinions C₁ et C₂ :

- C₁ : « I really like this laptop ! I got it a few weeks ago and had an enjoyable experience using it. Even though I think it's a bit overpriced... The screen's resolution is great, the design is quite stylish and I really enjoyed that I can run my favorite animation programs without lag ! However I'll have to mention that it weights a ton, and it's even smaller than what I'm used to. »

- C₂ : « I've had this phone for a while and it's the worst investment I've ever made. The pictures' quality is low, the battery doesn't hold for long and it lags a lot ! It's too expensive for such low performances... I won't lie though, the design was sleek and quite cute. But overall I don't think that this product is worth all the hype it got on social media. »

Considérons les comme entrée de trois modèles de fouille d'opinion M₁, M₂ et M₃. Chacun de ces modèles implémente un des niveaux de granularité décrits auparavant respectivement.

Le modèle M₁ traite le document C_i dans son entiereté. Il détermine, en sortie, la polarité de C₁ comme positive et celle de C₂ comme négative.

Le modèle M₂, qui implémente une fouille d'opinion au niveau de la phrase, extrait d'abord les phrases subjectives du texte du commentaire C_i avant de déterminer sa polarité. Par exemple, la polarité de la phrase « However I'll have to mention that it weights a ton » du commentaire C₁ est identifiée comme étant négative, tandis que la phrase « I won't lie though, the design was sleek and quite cute. » du commentaire C₂ est de polarité positive.

Pour finir, le modèle M₃, quant à lui, implémente une fouille d'opinion basée sur l'aspect de l'entité cible, i.e. il détecte d'abord les aspects présents dans le texte du commentaire C_i avant d'identifier la polarité des opinions les concernant. En ce qui est de l'aspect “Weight” (Poids) de l'entité Laptop dans l'exemple C₁, qui est de type **implicite**, le modèle déduit que les opinions envers elle sont de polarité négative. Contrairement à l'aspect “pictures' quality” (la qualité des images) de l'entité Phone du commentaire C₂, qui est de nature explicite et dont l'orientation est positive.

Le tableau I.3 résume les résultats de la fouille d'opinion effectuée sur les commentaires C₁ et C₂ avec les modèles M₁, M₂ et M₃.

TAB. I.3 : Résultats d'un exemple de fouille d'opinion sur divers niveaux d'analyse.

Fouille d'opinion au niveau du document		
Document en entrée	Polarité	
C ₁	Positive	
C ₂	Négative	
Fouille d'opinion au niveau de la phrase		
Document en entrée	Phrase (subjective)	Polarité
	I really like this laptop	Positive
	I got it a few weeks ago and had an enjoyable experience using it.	Positive
	Even though I think it's a bit overpriced	Négative

Chapitre 1. L'analyse des sentiments

C ₁	The screen's resolution is great, the design is quite stylish and I really enjoyed that I can run my favorite animation programs without lag!	Positive
	However I'll have to mention that it weights a ton, and it's even smaller than what I'm used to.	Négative
C ₂	I've had this phone for a while and it's the worst investment I've ever made.	Négative
	The pictures' quality is low, the battery doesn't hold for long and it lags a lot !	Négative
	it's too expensive for such low performances....	Négative
	I won't lie though, the design was sleek and quite cute.	Positive
But overall I don't think that this product is worth all the hype it got on social media		Négative

Fouille d'opinion au niveau de l'aspect

Document en entrée	Aspect	Type	Polarité
C ₁	GENERAL	Explicite	Positive
	Prix	Implicite	Négative
	Résolution de l'écran	Explicite	Positive
	Design	Explicite	Positive
	Performances	Implicite	Positive
	Poids	Implicite	Négative
	Taille	Implicite	Négative
C ₂	GENERAL	Explicite	Négative
	Qualité d'image	Explicite	Négative
	Batterie	Implicite	Négative
	Prix	Implicite	Négative
	Performances	Explicite	Négative
	Design	Explicite	Positive

1.4.2 La tâche effectuée

Généralement, l'analyse des sentiments se focalise sur l'identification de l'orientation d'un texte, cependant, elle peut aller au-delà de cette polarité pour détecter les sentiments et les émotions spécifiques (la colère, la joie, la tristesse, etc.) et même des intentions (intéressé ou pas intéressé) de l'auteur du texte. Selon la manière dont nous voulons exploiter les résultats de l'analyse, nous pouvons l'ajuster afin de répondre à ces besoins. L'analyse des sentiments peut donc être classifiée selon les tâches qu'elle vise à accomplir, parmi ces tâches, nous pouvons citer :

1.4.2.1 Analyse affinée

L'analyse fine des sentiments fournit un niveau de polarité plus précis en le décomposant en d'autres catégories, généralement de **très positives à très négatives**. Cela peut être considéré

comme l'équivalent d'une opinion sur une échelle de 5 étoiles (Très positive 5 étoiles – Positive 4 étoiles – Neutre 3 étoiles – Négative 2 étoiles - Très négative 1 étoile).

1.4.2.2 Analyse des sentiments multilingue

L'analyse des sentiments multilingue est souvent très difficile, car elle implique beaucoup de prétraitement et de ressources. C'est considéré comme un défi courant dans le domaine. Selon les langues, les ressources disponibles peuvent être limitées et nécessiter la création de corpus personnalisés. Par exemple, la langue **anglaise** est une langue dont les ressources sont largement disponibles, tandis que l'**arabe**, quant à elle, a des corpus limités et présente d'autres difficultés telles que les multiples dialectes disponibles et la romanisation de l'écriture.

1.4.2.3 Détection des émotions

La détection des émotions vise à identifier des émotions spécifiques plutôt que la positivité et la négativité. Les exemples pourraient inclure le bonheur, la frustration, la colère et la tristesse (NANDWANI et VERMA 2021).

Un des défis rencontrés dans ce type d'analyse est le fait que certains mots et expressions peuvent être utilisés afin d'exprimer différentes émotions. Par exemple, l'expression “ This experience had me shaking in my seat” renvoie, à première vue, à la **peur** cependant, dans un autre contexte, elle peut faire référence à la **surprise** et l'**enthousiasme**. Cette dépendance au contexte rend l'utilisation de simples lexiques (i.e. Liste de mots avec les émotions qu'ils véhiculent) insuffisante.

1.5 Approches d'analyse des sentiments

L'analyse des sentiments est un domaine multidisciplinaire qui comprend l'étude de divers domaines. Les différentes études de l'art (LIGHART et al. 2021b ; RAMESH et WEBER s. d. ; K. RAVI et V. RAVI 2015) classifient les approches adoptées dans l'analyse des sentiments en trois catégories : Basées sur l'apprentissage automatique, Basées sur le lexique et des méthodes Hybrides comme l'illustre la Figure I.3.

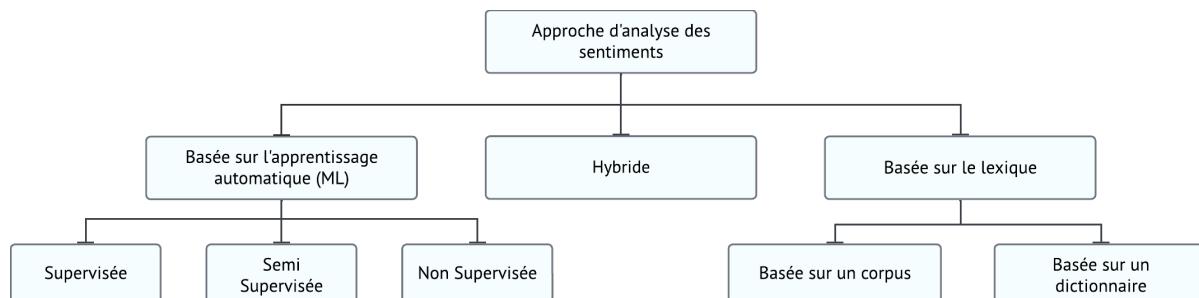


FIG. I.3 : Taxonomie des approches adoptées dans l'analyse des sentiments

Dans ce qui suit, nous présenterons ces différentes catégories d'approches en présentant diverses études de la littérature comme exemples. nous évaluerons les performances des modèles présentés en utilisant des métriques d'évaluation telles que F-score (qui est une valeur unique qui résume l'efficacité de la capacité d'un modèle à classer correctement les instances positives tout en maintenant un équilibre entre précision et rappel), l'exactitude (qui est une mesure de base de l'exactitude globale d'un modèle dans la classification des instances), etc Ces métriques seront présentées en plus de détail dans la section 1.7.

1.5.1 Approches basées sur l'apprentissage automatique

La classe des approches basées sur l'apprentissage automatique (ML) se divise à son tour en trois catégories : l'apprentissage supervisé, semi-supervisé et non supervisé.

1.5.1.1 Approches supervisées

La classification supervisée est l'approche la plus connue et la plus utilisée, elle consiste à entraîner un classificateur à partir d'un ensemble de données étiquetées permettant, par la suite, la classification de toute nouvelle entrée selon son orientation (Prédire la polarité d'une nouvelle donnée). Selon (MEHTA et PANDYA 2020), parmi les approches appartenant à cette catégorie on peut citer :

- a) **Les classificateurs probabilistes** : qui visent à prédire ou anticiper, étant donné une observation en entrée, une distribution de probabilité sur un ensemble de classes (plutôt que de ne fournir que la classe la plus probable à laquelle l'observation devrait appartenir). Parmi ces classificateurs, on trouve : Naïve Bayes, les réseaux bayésiens, l'entropie maximale, etc.

Par exemple, Naïve Bayes est considéré comme efficace et simple, et sensible à la sélection des caractéristiques (TIFFANI 2020). Dans (TIFFANI 2020), les auteurs ont appliqué le classificateur Naïve Bayes sur des données prétraitées. D'abord, les données ont été nettoyées (transformation des mots en minuscule, suppression des mots vides et radicalisation) puis parcourues par un N-gramme (ils ont testé l'Unigramme, le Bigramme et le Trigramme) avant de les classer à l'aide du classificateur. Les résultats étaient plutôt prometteurs, car l'exactitude (accuracy) du modèle où Naïve Bayes était appliqué à l'unigramme était de 81,30% et celles des deux autres n'étaient pas inférieures à 71%. Les auteurs de (SHAH 2021) ont appliqué ce classificateur avec un modèle bag-of-word sur un dataset publié par Amazon¹¹ sur kaggle¹² contenant des avis sur des produits smartphones (VIMAL et TARUN 2019) résultant en une exactitude de 91,76% et un F1-score de 94%.

- b) **Les classificateurs linéaires** : qui utilisent une combinaison linéaire des caractéristiques de l'opinion pour prendre une décision de classification et identifier la classe à laquelle elle appartient. On peut y trouver : les classificateurs SVM et ceux basés sur les réseaux de neurones.

En appliquant un classificateur SVM sur des tweets du dataset “Twitter US airline sentiment”, les auteurs de (RAVI KUMAR et al. 2021) ont abouti à une exactitude de 74,24% et un F1-score de 74%. D'autre part, les auteurs de (SHAH 2021) ont appliqué SVM avec un modèle bag-of-word (comme pour Naïve Bayes) et ont obtenu une exactitude de 93,58% et un F1-score de 96%, un meilleur résultat que celui obtenu avec Naïve Bayes.

Une autre méthode populaire est celle des **réseaux de neurones**, qu'ils soient artificiels (ANN), convolutifs (CNN), récurrents (RNN)... De nombreuses architectures basées sur cette méthode ont obtenu d'excellentes performances dans le domaine de l'analyse des sentiments (RAVI KUMAR et al. 2021 ; SHAH 2021).

Par exemple, dans (RAVI KUMAR et al. 2021), les résultats obtenus en implémentant les réseaux de neurones artificiels (ANNs) étaient plus performants que les autres modèles proposés (SVM, Arbres de décision), avec une exactitude de 75,99% et un F1-score de 74%. Pour (SHAH 2021), la meilleure performance a été obtenue en utilisant un réseau de neurones convolutif (CNN) avec FastText (BOJANOWSKI et al. 2017) (une extension du word-embedding Word2vec(MIKOLOV et al. 2013)), avec une exactitude de 94,62% et un F1-score de 96%.

¹¹<https://www.amazon.com/>

¹²<https://www.kaggle.com/>

- c) **Les classificateurs d'arbre de décision :** où les données sont continuellement divisées en fonction d'un certain paramètre, résultant, au final, en un arbre composé d'un nœud et de feuilles. Le principal avantage de cette classe est sa capacité à utiliser différents sous-ensembles de fonctionnalités et règles de décision à différentes étapes de la classification. Sa simplicité est aussi une des raisons de sa popularité.

Dans (RAVI KUMAR et al. 2021), l'utilisation des arbres de décision a donné une exactitude de 67.38% et un F1-score de 63%, c'était le modèle le moins performant comparé aux deux autres classificateurs utilisés (SVM et ANN comme cité précédemment).

- d) **Les classificateurs basés sur des règles :** qui consistent en un ensemble de règles « SI-ALORS » (IF-THEN) obtenues en appréhendant statistiquement les données d'apprentissage. Nous pouvons produire les règles en fonction de nos besoins lors de la phase d'apprentissage (MEHTA et PANDYA 2020).

Les approches supervisées surpassent parfois les performances d'autres approches d'apprentissage automatique et sont plus faciles généralement à implémenter. Cependant, il faut noter que l'inconvénient majeur de ces approches est la dépendance aux ensembles de données annotées et la difficulté de se procurer ces ensembles. De plus, vu cette dépendance, la qualité des données annotées impacte considérablement la performance du modèle de classification.

1.5.1.2 Approches semi-supervisées

Pour la classification semi-supervisée, les modèles sont entraînés avec des ensembles de données non étiquetées augmentés par un ensemble, souvent limité, d'exemples de données étiquetées (VAN ENGELEN et Hoos 2020). Ce type d'approches délivre des résultats assez satisfaisants et nécessite moins d'effort humain par rapport aux méthodes d'apprentissage supervisé.

Les auteurs de (N. LI et al. 2020) proposent un cadre d'apprentissage multitâche semi-supervisé (appelé **SEML**) pour l'analyse des sentiments basée sur les aspects. Comme cette analyse implique à la fois l'extraction d'aspects et la classification des sentiments d'aspect, les modèles supervisés utilisés pour la résoudre nécessitent un grand nombre d'avis étiquetés qui sont très coûteux ou indisponibles. Pour cela, les auteurs ont opté pour une technique semi-supervisée (SEML) qui a 03 caractéristiques clés : une formation à vues croisées (Cross-View Training) pour permettre un apprentissage de séquence semi-supervisé sur un petit ensemble d'avis étiquetés et un grand ensemble d'avis non étiquetés du même domaine, une exécution en deux sous-tâches précédemment énoncées simultanément en utilisant 03 couches neuronales récurrentes bidirectionnelles empilées pour apprendre les représentations des opinions et enfin une unité récurrente attentive à fenêtre mobile pour ces couches afin d'améliorer l'apprentissage de la représentation et la précision de la prédiction.

Les résultats expérimentaux montrent que SEML surpasse de manière significative les modèles de l'état de l'art sur les mêmes jeux de données, pour la tâche d'extraction d'aspects, ils ont réussi à atteindre un F1-score de 83,37% sur D_L (des opinion sur des laptops), et 78,24% sur D_R (des opinion sur des restaurants), tandis que pour la tâche de classification des sentiments, ils ont atteint un F1-score de 59,85% sur D_L et de 69,17% sur D_R.

1.5.1.3 Approches non supervisées

La classification non supervisée vise à tirer des conclusions à partir d'un ensemble de données en entrée où ces données ne sont pas étiquetées, le résultat de cette classification serait le regroupement de ces données en des clusters selon leurs ressemblances. Ces méthodes aident à découvrir les modèles cachées dans les données.

L'analyse de sentiment basée sur des données est riche en ressources et a été bien développée,

cependant l'analyse non supervisée, particulièrement multi-source, est rarement étudiée. Cela est dû à de nombreux problèmes rencontrés par cette dernière, principalement dus au manque d'informations de supervision, aux écarts sémantiques entre les domaines et à la perte de connaissances. Pour atténuer à ces problèmes, (DAI et al. 2021) ont proposé un cadre d'adaptation de domaine en deux étapes. La première est une méthode d'adaptation de domaine sélective, qui transfère les connaissances du domaine source le plus proche. La deuxième, quant à elle, est une méthode d'ensemble axée sur la cible, dans laquelle les connaissances sont transférées via une méthode d'ensemble bien conçue. Les expériences approfondies montrent des résultats prometteurs surpassant les méthodes concurrentes non supervisées.

1.5.2 Approches basées sur le lexique

Ce sont les approches traditionnelles qui consistent à calculer le sentiment d'un document à partir de l'orientation sémantique des mots ou des phrases qui le composent, ces orientations sont documentées dans des lexiques de sentiments. Nous pouvons en distinguer deux catégories selon la méthode de création des lexiques de sentiments :

1.5.2.1 Approches basées sur un dictionnaire

Les approches basées sur un dictionnaire (par exemple WordNet¹³) commencent par la collecte manuelle d'un ensemble de mots de sentiments, appelés graines, puis par le développement itératif de cet ensemble à partir des synonymes et des antonymes énumérés dans le dictionnaire, et se termine quand plus aucun nouveau mot n'est trouvé par une inspection manuelle si voulue.

À titre d'exemple, nous pouvons citer (HOSSEN et DEV 2021), où ils ont utilisé une approche basée sur un dictionnaire sur un ensemble de données de critiques de films, ils ont utilisé les bibliothèques micro-WordNet-Opinion 3.0¹⁴. Ils ont eu de bons résultats : une exactitude de 72% dépassant les approches traditionnelles du lexique (qui atteignaient 51%). On peut également citer (ÖZÇELIK et al. 2021) où ils visaient à créer un nouveau lexique de polarité 'HisNet'. Cette recherche montre à quel point les ressources dans un tel domaine sont rares lorsqu'il s'agit de langues autres que l'anglais et, comme ils l'ont mentionné, il est vain d'essayer de traduire des lexiques anglais existants vers une langue cible.

1.5.2.2 Approches basées sur un corpus

Les approches basées sur un corpus, quant à elles, peuvent être adaptées à des domaines spécifiques. Ces méthodes commencent par une liste de mots de sentiments, souvent à usage général, puis en découvre d'autres en addition à leurs orientations à partir d'un corpus de domaine, en utilisant une approche **statistique** ou **sémantique** (AMINUDDIN et al. 2021 ; LIU 2020). Elles sont cependant plus difficiles à mettre en place et même dans un même domaine, le problème de dépendance au contexte persiste, i.e. un mot peut avoir une orientation différente selon le contexte d'où il est extrait. Les auteurs de (AMINUDDIN et al. 2021) ont utilisé une approche sémantique basée sur un corpus afin d'identifier des mots-clés dans les avis collectés, les catégoriser en fonction des sentiments et identifier la polarité des lexiques (positive, neutre, négative).

1.5.3 Approches hybrides

Les approches hybrides sont celles qui visent à combiner les techniques d'apprentissage automatique et celles basées sur le lexique afin de produire des résultats optimaux en utilisant un ensemble de caractéristiques tirées de ces deux méthodes. Ainsi, les lacunes et les limites des deux approches peuvent être surmontées. Par exemple, (PUTRA et al. 2021) ont utilisé une méthode

¹³<https://wordnet.princeton.edu>

¹⁴<https://github.com/aesuli/SentiWordNet/blob/master/papers/Micro-WNOp.pdf>

hybride combinant une méthode basée sur un dictionnaire (Sentiwordnet 3.0 (BACCIANELLA et al. 2010)) et une méthode basée sur l'apprentissage automatique (SVM), leurs résultats montrent que leur méthode hybride surpassé l'approche basée sur le lexique (dictionnaire) et l'approche SVM.

1.6 Jeux de données disponibles

Afin de pouvoir entraîner et tester les approches adoptées pour résoudre le problème de classification de la polarité d'opinion, nous avons besoin de jeux de données. Certains de ces ensembles de données ne sont pas accessibles, car ils sont constitués par les entreprises qui collectent les avis de leurs clients, d'autres (les corpus académiques) sont mis à la disposition de la communauté en vue d'une évaluation comparative des modèles.

Dans cette section, nous allons présenter des exemples de corpus académiques utilisés dans l'analyse des sentiments. Le tableau I.4 ci-dessus présente un ensemble de jeux de données disponibles publiquement et utilisés dans les différents travaux de l'analyse des sentiments.

TAB. I.4 : Quelques corpus utilisés dans les travaux d'analyse de sentiment

Dataset	Source	Volume	Domaine	Référence
Twitter US Airline Sentiment ¹⁵	Twitter	Environ 14,600 avis	Compagnies aériennes	(KAGGLE 2015)
Stanford Sentiment Treebank ¹⁶	Rotten Tomatoes	Plus de 10,000 avis	Films	(SOCHER et al. 2013)
Amazon Product Reviews ¹⁷	Amazon	233.1 million avis	Produits	(NI et al. 2019)
Internet Movie Database (IMDB) ¹⁸	IMDB	50.000 avis	Films	(MAAS et al. 2011)
LABR ¹⁹	GoodReads	63,257 avis	Livres	(ALY et ATIYA 2013)
Sentiment140 ²⁰	Twitter	1,600,000 avis	Produits	(GO et al. 2009)

1.7 Métriques et méthodes d'évaluation

Afin de déterminer l'efficacité d'une méthode et de la comparer avec d'autres méthodes, nous devons l'évaluer. Pour cela, plusieurs métriques d'évaluations ont été mises en place, et quel que soit le type du corpus, nous pouvons distinguer deux types de méthodes d'évaluation : les méthodes de tests statistiques ainsi que les méthodes de retour de pertinence.

1.7.1 Méthodes de tests statistiques

Les méthodes de tests statistiques supposent que les documents du corpus ont été classés manuellement et que l'annotation a été validée par des experts. L'annotation peut aboutir à un corpus équilibré, ou un corpus déséquilibré. À cet effet, nous pouvons diviser l'évaluation des performances des modèles en deux catégories selon la distribution des instances positives et

¹⁵<https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment>

¹⁶<https://nlp.stanford.edu/sentiment/treebank.html>

¹⁷<https://nijianmo.github.io/amazon/index.html>

¹⁸<https://ai.stanford.edu/~amaas/data/sentiment/>

¹⁹<https://github.com/mohamedadaly/LABR>

²⁰<https://www.kaggle.com/datasets/kazanova/sentiment140>

négatives :

- Pour les jeux de données équilibrés, la métrique du F-score est souvent utilisée.
- Pour les jeux de données déséquilibrés, la courbe ROC et AUC sont plus adaptées.

À noter que les métriques de test sont différentes des fonctions de perte (Loss functions). Tandis que, les fonctions de perte montrent une mesure des performances du modèle *pendant* l'entraînement de ce dernier, les métriques sont utilisées pour juger et mesurer les performances du modèle *après* l'entraînement.

1.7.1.1 Le F-score

Le problème d'analyse des sentiments revient à un problème de classification (LIU 2020). Pour mesurer l'efficacité d'un classificateur dans ce type de problème, trois mesures sont souvent utilisées : la précision **P**, le rappel **R**²¹ et le **F-score** (F-measure, F1 measure). Ces métriques doivent être contrôlées pendant toutes les phases de la chaîne de traitement et pas seulement dans la phase d'analyse de la polarité.

TAB. I.5 : Table de confusion - Analyse des sentiments

Classe Réelle Classe Prédite	Opinion positive	Opinion négative
Opinion positive	TP (Vrai Positif)	FN (Faux Négatif)
Opinion négative	FP (Faux Positif)	TN (Vrai Négatif)

Définissons d'abord les matrices de confusion. Une **matrice de confusion** est une matrice 2 x 2 montrant la distribution des performances d'un modèle de classification sur des données. Il nous montre à quel point le modèle fonctionne, ce qui doit être amélioré et quelle erreur il commet. Le tableau I.5 illustre la table de confusion, où :

- **TP** - Vrai positif (le résultat de classe positive (Opinion positive) correctement prédit du modèle)
- **TN** - Vrai négatif (le résultat de classe négative (Opinion négative) correctement prédit du modèle)
- **FP** - Faux positif (le résultat de classe positive (Opinion positive) mal prédit du modèle)
- **FN** - Faux négatif (le résultat de classe négative (Opinion négative) prédit de manière incorrecte du modèle)

Revenons au **F-score**. Ce terme a été introduit par Van Rijsbergen²¹, c'est une mesure statistique qui prend en compte à la fois le rappel **R**²¹! et la précision P. La **précision** (le degré de solidité) indique le degré de vérité des résultats obtenus, tandis que le rappel **R**²¹! (le degré d'exhaustivité) indique leur pertinence. La mesure de F-score est la moyenne harmonique entre la précision P et le rappel **R**²²!

Un modèle parfait a un F-score de 1. D'une manière générale, le calcul du F-score se fait selon la formule I.1 suivante :

$$F-score = 2 \times \frac{P \times R}{P + R} \quad (I.1)$$

La précision P et le rappel **R**²¹! sont calculés selon les formules suivantes :

²¹https://en.wikipedia.org/wiki/C._J._van_Rijsbergen

²²<https://www.expert.ai/glossary-of-ai-terms/f-score-f-measure-f1-measure/>

$$P_i = \frac{\text{nombre de documents correctement attribués à la classe } i}{\text{nombre de documents attribués à la classe } i} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (\text{I.2})$$

$$R_i = \frac{\text{nombre de documents correctement attribués à la classe } i}{\text{nombre de documents appartenant à la classe } i} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (\text{I.3})$$

1.7.1.2 Courbe ROC et AUC

La « **courbe ROC** » (de l'anglais Receiver Operating Characteristic curve) est une mesure de performance d'un classificateur binaire, souvent présentée sous la forme d'un graphique représentant les performances d'un modèle de classification pour tous les seuils de classification. La courbe ROC trace le taux de vrais positifs TVP (fraction des positifs qui sont correctement détectés) en fonction du taux de faux positifs TFP (fraction des négatifs qui sont incorrectement détectés) comme représentée dans la Figure I.4. Ces valeurs sont calculées selon les équations I.4.

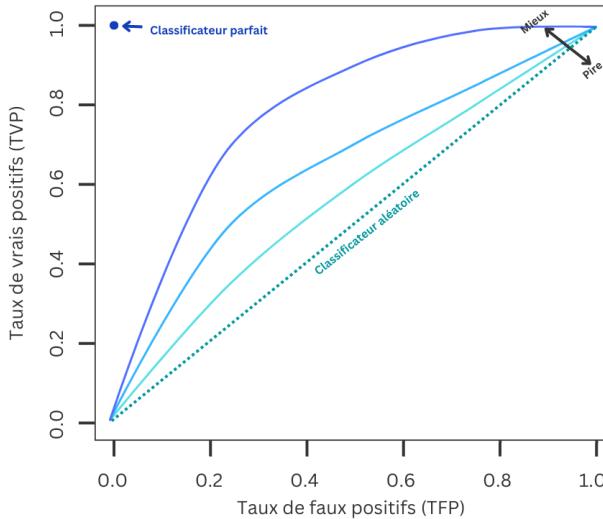


FIG. I.4 : La courbe ROC (Receiver Operating Characteristic curve)

$$\text{TVP} = \frac{\text{Vrai Positif}}{\text{Vrai Positif} + \text{Faux Negatif}} \quad \text{TFP} = \frac{\text{Faux Positif}}{\text{Faux Positif} + \text{Vrai Negatif}} \quad (\text{I.4})$$

En ce qui est de l'**AUC** (de l'anglais Area Under the receiver operating characteristics Curve), ce terme fait référence à l'aire sous la courbe ROC. La mesure AUC fournit une mesure agrégée des performances pour tous les seuils de classification possibles. Les valeurs d'AUC sont comprises dans une plage de 0 (toutes les prédictions sont erronées) à 1 (toutes les prédictions sont correctes).

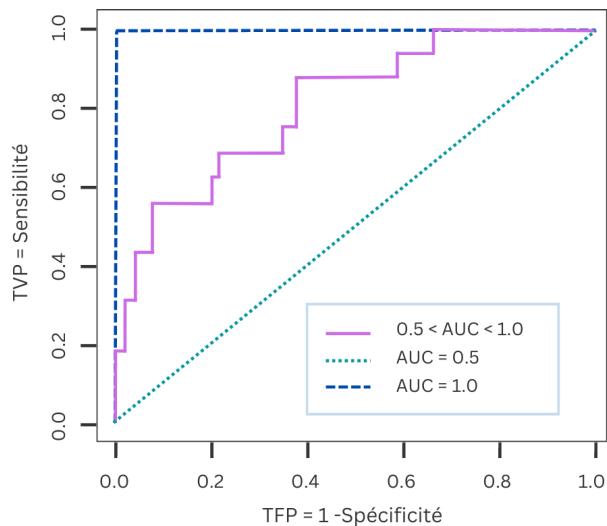


FIG. I.5 : La métrique AUC (Area Under the receiver operating characteristics Curve)

L'aire sous la courbe ROC (AUC) peut être interprétée comme la probabilité que, parmi deux opinions choisies aléatoirement, une positive et une non-positive (négative), la valeur du marqueur soit plus élevée pour l'opinion positive que pour l'opinion négative. Nous pouvons donc conclure que, comme illustré sur la figure I.5²³, une AUC de 0,5 (50%) indique que le marqueur est non informatif, et son augmentation indique une amélioration des capacités discriminatoires, (avec un maximum de 1,0 i.e. 100%).

1.7.2 Méthodes de retour de pertinence

Le retour de pertinence (Relevance feedback) est un domaine particulier de la recherche d'information, proposé en 1971 dans les travaux de Rocchio relevance-Feedback, c'est un moyen de prendre en compte le ressenti des utilisateurs vis-à-vis des informations renvoyées après une première recherche. Ceci est basé sur l'idée que l'utilisateur est la seule personne qui sait exactement ce qu'il recherche, et qu'il est donc le mieux placé pour juger de la pertinence des informations renvoyées par un système de recherche.

Ce principe de validation a été adopté dans le domaine de la fouille d'opinions. Il est utilisé en validant les premiers résultats par les experts, puis les réinjecter dans les systèmes eux-mêmes pour qu'ils soient réévalués par les utilisateurs.

1.8 Défis et discussions

L'analyse des sentiments est un domaine émergeant et dynamique, un tel dynamisme entraîne l'apparition de défis. Parmi les défis majeurs dans le domaine de l'analyse des sentiments, on peut citer la dépendance de la langue, la dépendance du contexte, la dépendance du domaine, les fautes d'orthographe ou de frappe, le langage figuratif, le dynamisme temporel des opinions et les fausses opinions (W. ZHANG et al. 2018). Dans ce qui suit, nous en détaillerons quelques-uns.

1.8.1 Dépendance du contexte

En général, tout propos subjectif est tenu dans un certain contexte, par exemple un mot donné peut être un indicateur de positivité dans un contexte et de négativité dans un autre (CHATURVEDI et al. 2018 ; DING et LIU s. d. ; HUSSEIN 2018 ; W. ZHANG et al. 2018). Ainsi, la prise en considération du contexte lors du pré-traitement ou le post-traitement des données

²³<https://www.idbc.fr/tutoriel-comment-lire-une-courbe-roc-et-interpreter-son-auc/>

est parfois cruciale pour l'obtention de résultats plus précis, surtout avec l'existence de mots polysémiques et même énantirosémique. Prenons, par exemple, ces deux expressions : « The instructor was so enthusiastic you can tell that he is an **amateur** about this! » et « This handmade candle is so poorly made, you can tell it was the work of an **amateur**... ». Dans la première expression, le mot ‘amateur’ fait référence à une personne qui aime un sujet et y est compétente, le deuxième par contre fait référence à une personne incompétente et exprime donc que l'entité cible n'est pas de bonne qualité.

1.8.2 Dépendance du domaine

La dépendance du domaine est partiellement une conséquence des changements de vocabulaire (HUSSEIN 2018). Ainsi, la même expression peut indiquer un différent sentiment en fonction du domaine. Certaines approches, et c'est le cas de la plupart, donnent de bonnes performances dans un domaine, mais fonctionnent atrocement dans un autre. Cela reste un problème difficile à contourner pour plusieurs.

1.8.3 Fautes d'orthographe

Un autre défi qui peut apparaître durant l'analyse des sentiments est celui des fautes d'orthographe. Il en existe deux types, celles qui sont volontaires et celles qui ne le sont pas. Les erreurs délibérément commises peuvent avoir un impact sur la polarité d'un commentaire, en fait une augmentation des voyelles et de ponctuations peut indiquer un sentiment fortement positif (par exemple « I'm soooooooooo HAPPPPY with this new product !!! ») ou fortement négatif (par exemple « This has been the WOOOORST EXPERIENCE EVEERRR »). En revanche, les erreurs causées par les méconnaissances ou les fautes de frappes n'ont aucun effet sur la polarité si on arrive à distinguer le mot d'origine (Exemple : « Skol » pour « School »). (OUESLATI et al. 2018) ont considéré les erreurs d'orthographe comme une caractéristique durant leur analyse des sentiments.

1.8.4 Langage figuratif

Le langage figuratif peut aussi poser problème lors de l'analyse des sentiments (HERCIG et LENCI 2017), or, il est difficile pour une machine de comprendre des choses que nous, les humains, peuvent distinguer facilement. Nous pouvons citer : les métaphores (« He is a shining star »), la métonymie (« La salle a applaudi l'orchestre ») et même l'ironie et le sarcasme (MAYNARD et GREENWOOD 2014) où les gens expriment leurs sentiments en utilisant le vocabulaire de polarité opposée, par exemple la phrase « L'expérience sur votre site web est MAGNIFIQUE ! Aucun bug en vue... » serait détectée comme une phrase positive par une machine, cependant un humain pourrait bien y soupçonner de l'ironie.

1.8.5 Crédibilité de l'opinion

Un autre défi majeur rencontré dans le domaine de la fouille d'opinions aujourd'hui est la détection de la crédibilité des opinions publiées et l'authenticité de leurs propriétaires (JINDAL et LIU 2008 ; OUESLATI et al. 2018). Cette problématique peut avoir des conséquences significatives sur l'e-réputation des entreprises. Pour cela, l'analyse de la crédibilité de l'opinion est cruciale lors de l'analyse sémantique d'un ensemble de documents, cependant, elle peut se montrer compliquée avec les différences de comportement entre les spameurs et l'apparition des algorithmes de génération de spam qui peuvent être facilement paramétrables pour battre les méthodes de l'état de l'art. Ce problème est devenu, ces dernières années, un sujet d'intérêt intéressant les chercheurs académiques comme ceux de l'industrie, c'est ce qui nous a poussés à en faire l'objet du prochain chapitre de cette partie du mémoire.

1.9 Conclusion

Dans ce chapitre, nous nous sommes intéressées au domaine de l'analyse des sentiments. Nous avons d'abord défini quelques concepts de base, puis nous avons modélisé la fouille d'opinion sous forme de processus, pour ensuite la classifier selon deux critères. Par la suite, nous avons présenté les jeux de données disponibles, quelques approches adoptées dans ce domaine et les métriques et méthodes d'évaluation. Pour finir, nous avons conclu ce chapitre en discutant des différents défis rencontrés dans ce domaine, parmi eux, on trouve la détection de spam d'opinion qui sera le sujet du deuxième chapitre de ce rapport vu qu'elle reste un défi ouvert dans le domaine, et comprendre la crédibilité d'une opinion est essentiel pour utiliser efficacement l'analyse de sentiment dans des applications réelles vu qu'elle détermine la fiabilité des résultats.

Chapitre 2

La détection de Spam d'opinion

2.1 Introduction

Il est vrai que la tache d'analyse de sentiment est une tache prépondérante pour la prise de décision que ce soit pour l'amélioration de la relation client ou de l'e-réputation. Cependant, avec l'ouverture du web, disponibilité de différentes plateformes de rédaction d'avis (les réseaux sociaux, les blogs...), les internautes peuvent partager leurs expériences et leurs opinions autour des divers produits, services, marques ou personnes.

Les avis en ligne disponibles en masse, peuvent influencer les décisions d'achat des clients et même totalement changer leurs points de vue envers une cible (une marque, un produit, un service...) ; les entreprises sont, de nos jours, plus conscientes que jamais de cette influence et de son impact sur leur réputation, voire leurs profits. D'une part, une réputation en ligne généralement positive (i.e. les avis disponibles sont majoritairement positifs) est considérée comme un point de force par rapport à la concurrence pouvant impacter positivement les ventes. D'autre part, une quantité de critiques négatives (même inférieure au taux d'opinions positives) peut nuire, voire détruire, la réputation d'une cible menaçant son existence en diminuant ses profits.

Dans ce chapitre, nous définirons les concepts principaux du problème de détection de spam d'opinion dans la section 2.3. Nous présenterons ensuite le processus de détection de spam d'opinion dans la section 2.4, les sections qui suivent détaillent les étapes de ce processus, à savoir la collecte et les jeux de données 2.5, le prétraitement 2.6, la représentation 2.7, la classification 2.8 et enfin l'évaluation ???. Pour conclure, nous discuterons des défis rencontrés dans ce domaine et des orientations pour les recherches futures dans la section 2.9.

2.2 Motivation

Conscients de l'effet des avis en ligne, certains marchands créent des fausses critiques dans le but de faire du profit, soit en créant des avis positifs afin de promouvoir leurs produits et services ou des avis négatifs pour nuire aux images de leurs compétiteurs. Certains de ces marchands optent pour le recrutement de rédacteurs professionnels et les chargent de rédiger des critiques sur leurs produits et services, plusieurs plateformes sont devenues source d'offres de rédaction d'avis pareilles tel que les groupes Facebook, les sites de travail libre comme Fiverr¹,... la figure I.6 illustre 05 offres de services de rédaction de critiques par des travailleurs indépendants sur la plateforme Fiverr, il est évident que ces rédacteurs n'ont aucune expérience reflétant la réalité des produits/services qu'ils critiquent et que leur seule motivation est le profit financier et non le partage d'informations utiles pour les clients.

¹<https://www.fiverr.com/>

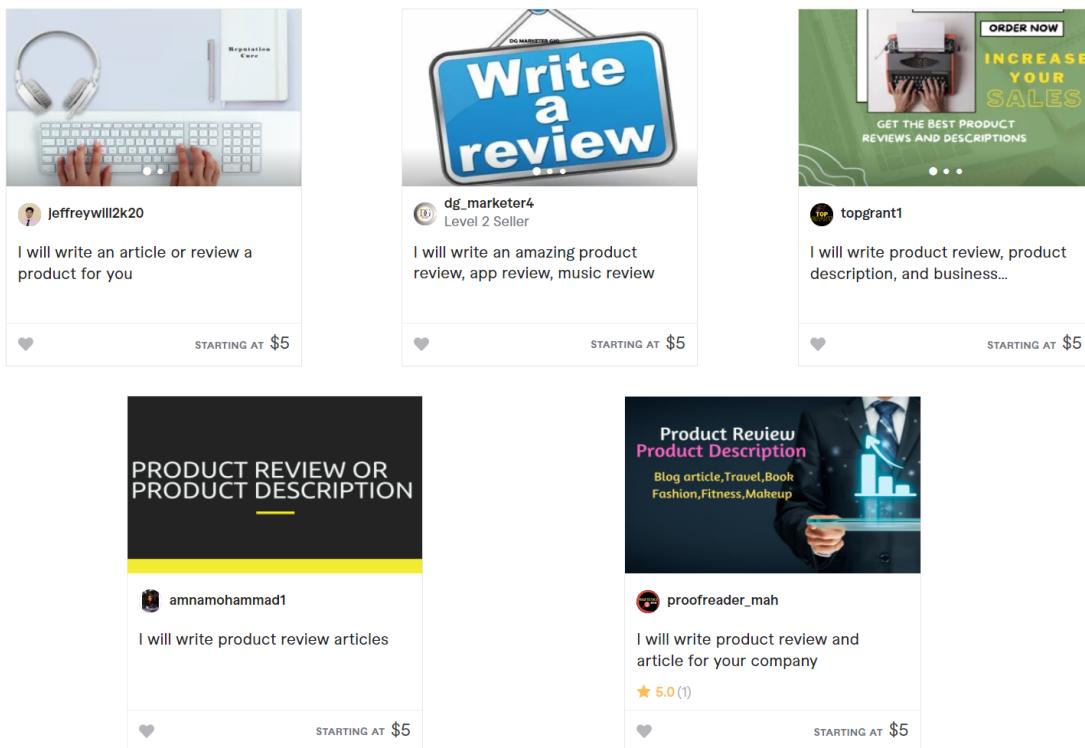


FIG. I.6 : Offres de rédaction d'avis trompeurs sur le site Fiverr

D'autres vendeurs utilisent des algorithmes qui génèrent ces avis automatiquement, la figure I.7 illustre un exemple : l'outil AISEO² qui génère des fausses critiques sur un produit, des algorithmes de la sorte sont facilement ajustables pour battre les techniques de détection de l'état de l'art. Ces opinions qui véhiculent de fausses informations sont appelées « spam d'opinion ».

The screenshot shows the AISEO Product Review Generator interface. It includes fields for Product Name (product_X), Notes (very easy to use, good price quality ratio, good pictures), and Keywords (fascinating, practical, unique, quality, camera). A sidebar provides suggested SEO keywords and a preview of generated reviews. One review template is shown:

Keywords to include: best camera
Product Review: **product_X** is a very simple and easy to use camera. It is perfect for the beginner user, but it can also double as a very simple surveillance camera.
From one beginner to the next, buy **product_X** !!!
##

I am new to the Movie making world and have been trying to learn this craft. I chose to use a camcorder as I am only a beginner. I chose this camcorder because of its easy to use. I watched a few video tutorials on YouTube and was able to figure out how to use it. I also bought an inexpensive tripod so that I would not have to worry about the camera wobbling while shooting. I then bought a few items of video editing software.

FIG. I.7 : AISEO : Outil de génération d'avis

Le problème de spam d'opinion a été décrit et formalisé pour la première fois en 2008 par (JINDAL et LIU 2008), il est généralement abordé comme un problème de classification. Mais avec les comportements et les styles différents, même imprévisibles, des rédacteurs, ainsi que la puissance des algorithmes de génération de spam d'opinion et la quantité limitée de données annotées, le problème de spam d'opinion s'est avéré difficile et complexe, ce qui a suscité

²<https://aiseo.ai/templates/review-generator.html>

beaucoup d'intérêt et est devenu un sérieux problème étudié tant au niveau des entreprises qu'au niveau académique. C'est ainsi qu'est né le besoin de disposer d'un outil offrant la possibilité de distinguer les opinions qui sont crédibles de ceux qui ne le sont pas.

Divers travaux sont proposés pour résoudre cette problématique ; devant ce nombre d'avis disponibles aujourd'hui, quelle est la meilleure approche afin de distinguer les avis crédibles des avis spam ?

2.3 Définitions

Dans ce qui suit, nous allons définir deux concepts essentiels à notre compréhension de ce domaine de recherche.

2.3.1 La crédibilité

Le terme « Crédibilité » remonte au milieu du 16^e siècle, selon le dictionnaire Larousse³, il est issu du mot latin médiéval « credibilitas » et signifie « Caractère de quelque chose qui peut être cru » ou « Caractère de quelqu'un qui est digne de confiance ». Une information, un propos ou un document peut être considérée crédible selon le niveau de fiabilité et de confiance que le récepteur accorde à sa source et à son canal de transmission.

Dans (CASTILLO et al. 2011), les auteurs ont employé la crédibilité dans le sens de la fiabilité, comme une qualité perçue composée de multiples dimensions offrant des motifs raisonnables d'être cru tout en n'étant évaluée qu'avec les informations disponibles. D'autre part, la crédibilité de l'information a été définie dans (R. LI et SUH 2015) comme étant la mesure dans laquelle on perçoit l'information comme étant crédible, et est un prédicteur puissant de l'action ultérieure d'un consommateur d'information, comme la recommandation ou la volonté d'adopter le point de vue de l'information reçue.

Dans le domaine de fouille d'opinion, la crédibilité est considérée comme une qualité importante qu'une opinion doit satisfaire. Une opinion crédible est une opinion qui reflète une expérience réellement vécue par le client qui l'exprime, elle ne doit pas contenir de propos mensongers concernant l'entité qu'elle cible, que ce soit un produit, une marque, un service ...

2.3.2 Le spam

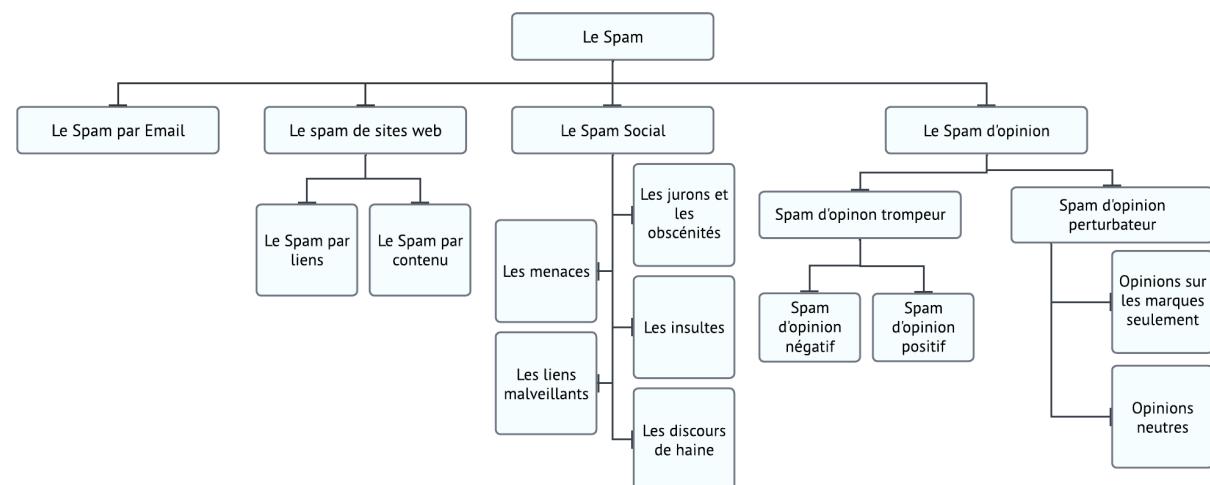


FIG. I.8 : La taxonomie du spam dans la littérature

³<https://www.larousse.fr/dictionnaires/francais/cr%C3%A9dibilit%C3%A9/20311>

En consultant les différents dictionnaires et encyclopédies, tels que Larousse⁴, nous trouvons que la majorité de ces derniers définissent le terme 'spam' comme du courrier indésirable (BECCHETTI et al. 2008) : « Courrier électronique non sollicité envoyé en grand nombre à des boîtes aux lettres électroniques ou à des forums, dans un but publicitaire ou commercial ». Or ceci n'est qu'un seul type de spam et il en existe plusieurs autres.

Parmi les autres types de spam, nous pouvons trouver : Le spam des sites web, le spam social ainsi que le spam d'opinion. La figure I.8 résume la taxonomie du spam dans la littérature comme nous allons l'aborder dans cette section (2.3.2).

2.3.2.1 Le spam de sites web

Le spam de sites web est le deuxième type de spam le plus répandu après le spam par e-mail, il représente une tentative de manipulation du classement des moteurs de recherche. Il en existe deux types principaux, à savoir le spam de lien et le spam de contenu (BECCHETTI et al. 2008).

2.3.2.2 Le spam social

Le spam social qui est un contenu indésirable apparaissant sur les services de réseaux sociaux, et tout site web incluant du contenu généré par l'utilisateur. Selon (WASHHA et al. 2017) il peut se manifester sous une énorme variété de formes, parmi elles nous pouvons citer : les jurons et les obscénités, les insultes, les discours de haine Les menaces et les liens malveillants.

2.3.2.3 Le spam d'opinion

Le spam d'opinion est un type de spam subtil, bien conçu et délibérément rédigé pour paraître authentique (OTT et al. 2011). Il vise à influencer les décisions d'achat des consommateurs et à changer la réputation d'une entité cible pour le meilleur ou pour le pire. C'est à ce type de spam que nous nous intéressons dans ce projet.

Le spam d'opinion se distingue du spam web du fait qu'il cible les clients et non pas les moteurs de recherche, et donc l'utilisation de liens hyperlinks peut être facilement détectée, et la répétition de mots clés non pertinents n'a aucun sens si on veut que le contenu du spam soit convaincant. D'autre part, il se démarque du spam social du fait qu'il vise à influencer l'opinion individuelle ou publique en faveur ou contre une entité.

En formalisant le problème de détection de spam, (JINDAL et LIU 2008) ont identifié trois types de spam d'opinion :

- **Les opinions mensongères ou trompeuses** : ceux qui trompent délibérément les lecteurs et ne sont pas fondées sur de vraies expériences et visent soit à nuire à l'image d'une entité (spam négatif) ou à l'améliorer en donnant des critiques positives non méritées (spam positif).

- **Les opinions qui portent sur les marques seulement** : ceux qui commentent sur la marque et non pas sur l'entité à évaluer. Ils peuvent être considérés comme spam, car ils n'apportent pas d'informations de valeur aux clients.

- **Les opinions neutres (non-avis)** : ce ne sont pas des avis, et il en existe 02 types : les annonces et d'autres textes aléatoires non pertinents. Ils n'ont aucun impact sur la réputation du produit/service.

Le deuxième et le troisième type de spam d'opinion sont appelés **spam d'opinion perturbateur**, car ils ne représentent pas de menace pour les clients, mais en masse, ils perturbent leur expérience.

En revanche, le premier type de spam d'opinion cité, à savoir celui des opinions mensongères,

⁴<https://www.larousse.fr/dictionnaires/francais/spam/10910104>

est-ce qu'on appelle le **spam d'opinion trompeur**, car il est difficile à identifier manuellement, et de nos jours, il est devenu de plus en plus répandu sur les réseaux sociaux, de plus en plus sophistiqué et subtil, sa détection est devenue un défi majeur. En effet, (OTT et al. 2011) ont confirmé que ce type de spam est si subtile que le consommateur a tendance à confondre une opinion spam pour une opinion authentique. Montrons cette subtilité à travers un exemple, voici deux opinions extraites du premier jeu de données public de référence (gold standard dataset) construit dans le domaine de la détection de spam d'opinion⁵ (OTT et al. 2011) :

Opinion 01 : « *Last week I stayed at the Hilton Chicago for 4 days and 3 nights and I was very pleased with the experience. As soon as I approached the front desk, I knew right away the staff was friendly and courteous. They had given me a list of local attractions such as the Museum of Science and Industry, the Broadcast Museum, the Ford Center For The Performing Arts and Willis Tower. I stayed in a junior suite which included a king size bed, a 27 inch television, non-allergenic feather pillows and a variety of other amenities. The indoor pool was rather large as well as the gym which includes countless treadmills and even a jogging track. Eating in the area is not a problem as the Hilton Chicago is home to Kitty O'Sheas which is a restaurant offering authentic Irish fare. In conclusion, I had a wonderful time staying here and I can not wait to plan my return trip to the Hilton Chicago. ».*

Opinion 02 : « *My family and I have just had a two week holiday in Chicago, and we stayed for a week at the Hilton Towers. We had a fantastic time and enjoyed every minute of our stay at this incredible hotel. The sheer size of the place is breathtaking and the atmosphere is very friendly and hospitable. If you prefer discreet modern boutiquey-type hotels, maybe this wouldn't be your thing. However if you enjoy old style glamour and glitz, you will be bowled over by this hotel. The place literally sparkles thanks to the enormous chandeliers that are everywhere in the lobby. Tons of celebs have stayed here over the decades and the history of the hotel and past guests etc is fascinating. We got a very reasonable rate for our stay and could not fault anything about this very special place ! ».*

Maintenant, si nous devions essayer de déterminer avec juste le contenu lequel des avis présentés est le faux et lequel est le vrai, quelle serait notre réponse ? Très probablement, il serait difficile, voire impossible, d'identifier lequel des deux est un spam. Dans l'ensemble de données construit dans (OTT et al. 2011), 'Opinion 1' a été annotée comme une opinion trompeuse, tandis que 'Opinion 2' a été jugée authentique.

Les deux exemples présentés étaient des opinions positives (qu'elles soient trompeuses ou non), mais il est important de noter que des opinions trompeuses de différentes polarités ont un impact différent, en effet, un nombre plus élevé d'avis positifs incite un client à acheter un produit et renforce les gains financiers des fabricants, tandis que les avis négatifs incitent les consommateurs à rechercher des alternatives, entraînant des pertes financières. De plus, la relation entre la polarité de l'opinion trompeuse et la qualité réelle du produit ciblé a un impact sur la gravité de la situation. la figure I.9 illustre une explication de cela telle que présentée dans l'étude (JINDAL et LIU 2008).

Les opinions appartenant à la première colonne (i.e. Les opinions trompeuses positives) ont pour objectif de promouvoir l'entité cible (le produit), et sont généralement écrites par les fabricants du produit ou par des personnes ayant un intérêt quelconque dans le produit (économiques ou autre). Notons que même si les avis appartenant à la première case (les avis spam positifs pour un produit de bonne qualité) puissent être corrects, cependant ils ne sont pas basés sur une expérience réelle et donc ils sont considérés trompeurs malgré tout.

⁵<https://myleott.com/op-spam.html>

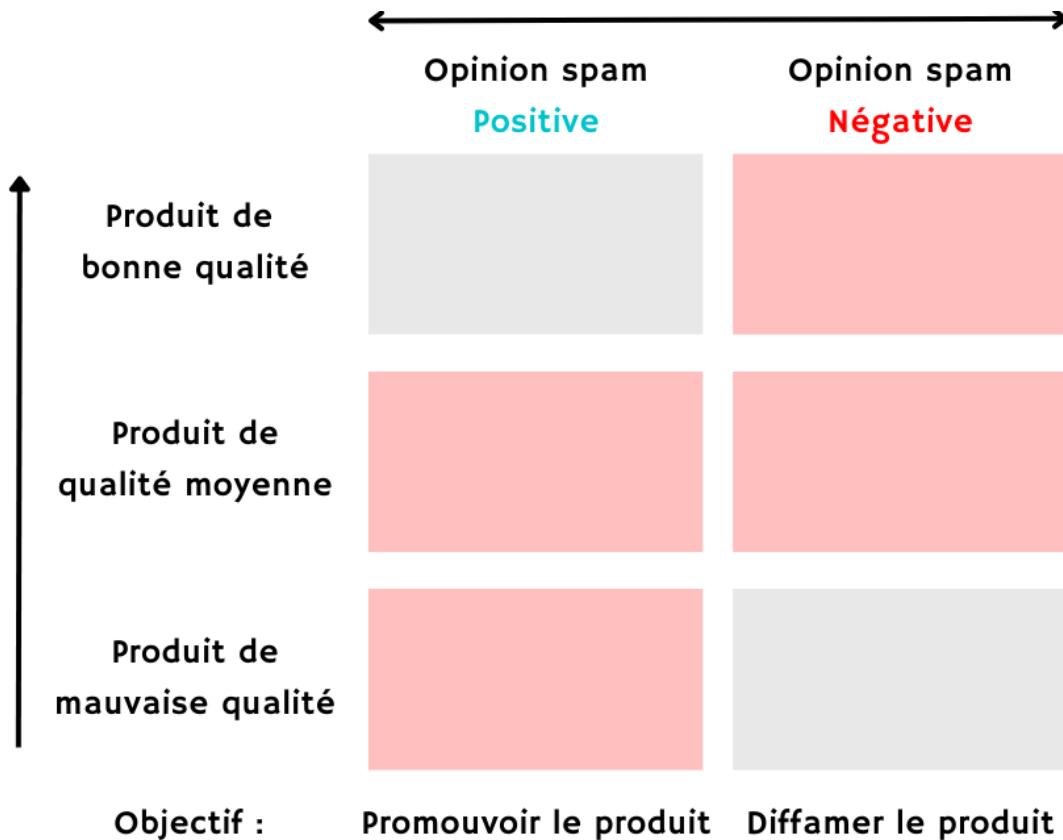


FIG. I.9 : La relation entre la polarité des opinions spam et la qualité des produits selon (JINDAL et LIU 2008)

En revanche, les opinions de la deuxième colonne (i.e. Les opinions trompeuses négatives) sont très probablement écrites par des concurrents et même si la critique peut être vraie (cas de la dernière case où l'opinion spam est négative et le produit est de mauvaise qualité) les intentions des rédacteurs restent malveillantes.

Il est évident que les opinions des cas particuliers cités auparavant (les cases colorées en gris) ne sont pas si nocives, car même si elles ne sont pas fondées sur des expériences réelles, elles correspondent à la véritable qualité du produit ciblé. Ceci contrairement aux opinions spam correspondant aux cases en rouges qui peuvent avoir un impact néfaste sur non seulement les vendeurs, mais aussi les clients. C'est pour cela que (JINDAL et LIU 2008) affirme la nécessité de s'accentuer sur la détection du spam d'opinions dans ces cas précis.

2.3.3 Le spameur

Étant donné que le terme « spam » désigne couramment le spam par e-mail, il est évident que les spameurs seraient souvent définis comme une personne ou une organisation qui envoie des messages non pertinents ou non sollicités par email, généralement à un grand nombre d'utilisateurs, à des fins de publicité, d'hameçonnage, de diffusion de logiciels malveillants, etc.

Cependant, dans notre domaine de recherche concernant le spam d'opinion, nous définissons un spameur comme une entité (personne ou organisation) qui publie des avis spam. D'après la définition de (JINDAL et LIU 2008) un spameur est un rédacteur qui a publié au moins un avis trompeur.

Selon (ANDRESINI et al. 2022), L'étude du comportement d'un rédacteur d'avis pour la détection du spam provient de l'hypothèse qu'un spameur écrit constamment du spam. Cela

nous fournit une autre perspective pour détecter le spam d'avis : « Reconnaître qu'un rédacteur est un spameur permet de marquer toute opinion dont il est l'auteur, qu'il ait déjà publié ou qu'il publiera à l'avenir, comme étant spam. »

Un spameur peut être classifié, suivant le critère de professionnalisme ou de motivation, en deux catégories principales (LIU 2020) :

- **Rédacteurs de spam professionnels** qui sont motivés par un gain financier. Ils sont généralement des indépendants (e.g. la figure I.6 précédente qui présente des exemples d'offres de rédaction d'avis trompeurs réellement publiées sur Fiverr) ou des employés d'entreprises chargés d'écrire des fausses critiques dans le cadre de leurs activités. Les opinions spam professionnelles sont généralement plus faciles à détecter, puisque leurs auteurs en écrivent en grand nombre, ce qui peut laisser des modèles linguistiques et comportementaux facilement détectables par les algorithmes d'exploration de données. Cependant, le problème est qu'au moment où leur style d'écriture et leur comportement anormal sont reconnus et ces spameurs sont finalement identifiés, leurs mensonges pourraient déjà être répandus et leur but pourraient être atteint. Donc, il est primordial de détecter ces modèles le plus tôt possible, mais cela reste difficile et pour empirer les choses, un spameur identifié peut juste abandonner son compte et en créer un nouveau pour propager ses fausses critiques encore une fois.

- **Rédacteurs de spam non professionnels** : qui ne génèrent pas beaucoup de spam et qui ne sont généralement pas motivés par l'argent. Ces personnes écrivent principalement afin d'aider eux-mêmes, leurs connaissances ou leurs entreprises. Leur but est soit de promouvoir leurs produits ou ceux de leurs amis, de détruire l'image de leurs concurrents, ou même de nuire à leurs employeurs (anciens ou actuels) et entreprises. Il existe aussi des spameurs qui rédigent de fausses critiques juste pour le plaisir.

Comme ils n'écrivent pas beaucoup de critiques, ils peuvent ne pas avoir les mêmes habitudes que les professionnels, cependant, cela ne veut pas dire que leurs activités ne laisseront pas de trace. Par exemple, si un utilisateur consulte fréquemment une page d'avis sans en laisser un, puis suite à une critique négative, laisse une critique fortement positive, cet avis est considéré suspect car cette personne peut être le propriétaire du produit/service cible ou quelqu'un d'associé.

De plus, un autre groupe de personnes se situe à la frontière entre les vrais et les faux critiques, un autre groupe qui brise l'hypothèse faite ci-dessus, les Influenceurs et autres, ceux qui ont contribué à de nombreuses critiques authentiques et, ce faisant, ont construit leur réputation, mais ont ensuite été contactés par des marques pour présenter leurs produits comme contenu sponsorisé et le critiquer positivement. Ce type de 'spameurs' offrent un ensemble d'avis qui contient un mélange d'avis authentiques et trompeurs.

D'autre part, nous pouvons catégoriser les spameurs selon deux types d'entité, les spameurs individuels et les spameurs de groupe (MUKHERJEE et al. 2011, 2012, 2013a ; LIU 2020). Ces deux types ont des caractéristiques différentes qui peuvent être exploitées afin de faciliter leur détection.

- **Les spameurs individuels** : sont des personnes qui se chargent de rédiger des opinions spam eux même, sans travailler avec d'autres personnes, avec un seul compte (ou userId).

- **Les groupes de spameurs**, ou de comptes spam, travaillent ensemble, consciemment ou pas, dans le but de promouvoir ou rabaisser certains produits ou services. Généralement professionnels et parfois non, les spameurs de groupe suivent principalement un des deux modèles suivants :

1) Le premier modèle est celui d'un groupe de spameurs (personnes) qui travaillent en collusion pour promouvoir une entité cible et/ou nuire à la réputation d'une autre. Professionnels ou pas, les membres de ce groupe pourraient ne pas connaître l'existence les uns des autres, ou pourraient ne pas connaître leurs activités. Par exemple, un auteur de livre peut demander à un groupe d'amis d'écrire des critiques positives pour l'un de ses nouveaux ouvrages. Ces amis peuvent ne pas connaître l'activité de l'autre, et ils ne sont normalement pas des spameurs professionnels.

2) Le deuxième modèle est la pratique qu'on appelle "sock puppeting" : Une seule personne ou organisation enregistre plusieurs comptes (chacun avec un ID utilisateur différent) et commence à rédiger des spams en utilisant ces comptes. Ces multiples comptes se comportent exactement comme un groupe travaillant en collusion.

Hélas cette classification n'est pas binaire, il y a d'autres scénarios qui pourraient venir compliquer la détection des spameurs. Par exemple, une personne pourrait parfois travailler individuellement et puis en groupes pour certaines cibles, il pourrait aussi arriver qu'elle achète un produit et le critique honnêtement. Avec toutes ces situations qui peuvent arriver, le problème de détection de spameurs se montre complexe à résoudre.

2.4 Processus de la détection de spam d'opinion

Dans la littérature, la détection du spam d'opinion était autrefois divisée en deux tâches, la première étant la classification d'un avis comme vérifique ou spam, la seconde étant l'identification des spameurs. Ce n'est que ces dernières années que les chercheurs ont commencé à envisager le problème sous un autre angle, celui des produits ciblés .i.e découvrir quels produits sont ciblés par les spameurs. De nombreuses études montrent l'intérêt de prendre en considération les trois points de vue (RASTOGI et al. 2020). Cette information supplémentaire s'avérera utile lorsqu'il s'agira d'extraire des caractéristiques plus tard.

Le processus de détection de spam suit les mêmes étapes que le processus générique de fouille d'opinions, comme illustrée dans la figure I.10 :

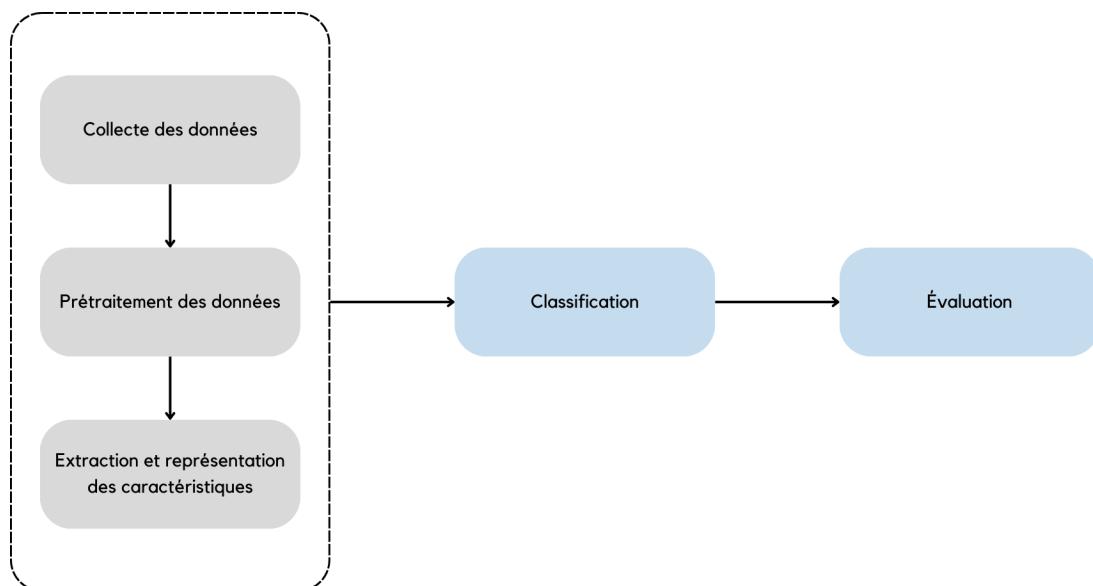


FIG. I.10 : Processus de détection de spam

La première étape est celle de la collecte de données et elle consiste à se procurer des jeux de

données nécessaires pour l’entraînement et l’évaluation des modèles de classification. Après la collecte, les données passent par un processus qui garantit qu’elles sont suffisamment nettoyées, structurées et optimisées pour la suite. Puis, on passe à l’étape d’extraction et représentation des caractéristiques dans le but d’optimiser les données prétraitées et de mettre en valeur les indicateurs pertinents pour fournir un bon ensemble de données pour le modèle de classification. Enfin, les données sont divisées, principalement, sous deux sous-ensembles, le premier servir de données d’entraînement pour le modèle de classification. Le deuxième, quant à lui, permettra le calcul des différentes métriques afin d’évaluer le modèle.

2.5 Jeux de données

Cette section concerne la première étape du processus de détection de spam, à savoir la collecte des données. Nous présenterons donc des jeux de données utilisés pour l’entraînement et l’évaluation des classificateurs implémentés ainsi que les méthodes de construction de ces jeux de données.

2.5.1 Méthodes de construction des ensembles de données

Après avoir recueilli suffisamment d’avis provenant de sources sélectionnées (i.e. des réseaux sociaux comme Facebook et Twitter, des sites Web d’e-commerce comme Amazon et Yelp, ...) à l’aide d’un crawler⁶ ou de toute autre méthode, il est maintenant temps d’annoter les données non seulement afin d’entraîner des modèles de classification supervisés ou semi-supervisés, mais aussi dans le but d’évaluer des modèles après leur apprentissage. Pour cela nous allons expliquer les méthodes de construction des ensembles de données, d’après (REN et JI 2019), les méthodes de construction de datasets sont classées selon 04 catégories :

- Les méthodes basées sur des règles (JINDAL et LIU 2008), (GILBERT et KARAHALIOS 2010)
- Les méthodes manuelles (F. H. LI et al. 2011), (REN et JI 2019), (REN et al. 2014)
- Les méthodes basées sur des algorithmes de filtrage (MUKHERJEE et al. 2013b), (MUKHERJEE et al. 2013b), (RAYANA et AKOGLU 2015), (H. LI et al. 2014)
- Les méthodes basées sur le crowdsourcing (OTT et al. 2013), (Jiwei LI et al. 2014)

2.5.2 Ensembles de données disponibles

Les ensembles de données peuvent être catégorisés selon les différentes méthodes de construction. Les chercheurs et entreprises peuvent construire leur propre datasets selon ces méthodes ou ils peuvent exploiter les jeux de données disponibles publiquement. Dans le tableau I.6, nous présentons certains ensembles de données de référence (Gold-standard) utilisés dans les travaux de la littérature pour la détection de spam d’opinion, ces datasets sont catégorisées selon les méthodes de construction mentionnées auparavant afin de résumer les exemples cités dans le point précédent. Ces corpus ne sont pas vraiment récents, mais ils sont des références sur lesquelles de nombreux chercheurs ont basé leurs travaux.

⁶<https://guersanguillaume.com/ressources/glossaire-marketing-digital/definition-crawler/>

TAB. I.6 : Ensembles de données de référence (Gold-standard) utilisés dans les travaux de la littérature pour la détection de spam d’opinion.

Méthode de construction	Source	Volume	Domaine	Référence
Basée sur des règles	Amazon ⁷	5,8M d’avis, 2,14M d’utilisateurs	Livres, Musique, DVD, Produits industriels	(JINDAL et LIU 2008)
Manuelle	Epinions	6000 avis	Produits	(F. H. LI et al. 2011)
	TripAdvisor ⁸	6000 avis	Hôtels	(REN et al. 2014)
Basée sur des algorithmes de filtrage	Yelp	67395 avis, 38063 utilisateurs	Hôtels, Restaurants	YelpCHI (MUKHERJEE et al. 2013b)
	Yelp	9765 avis	Hôtels	(H. LI et al. 2014)
	Yelp	359052 avis, 160225 utilisateurs	Hôtels	YelpNYC (RAYANA et AKOGLU 2015)
	Dianping	608598 avis, 260277 utilisateurs	Hôtels	YelpZip (RAYANA et AKOGLU 2015)
Basée sur le crowdsourcing	TripAdvisor	800 avis	Hôtels	(OTT et al. 2011)
	TripAdvisor	1600 avis	Hôtels	(OTT et al. 2013)
	TripAdvisor	3032 avis	Hôtels, Médecins, Restaurants	(Jiwei LI et al. 2014)

2.6 Approches de prétraitement

En raison de la grande taille des données collectées, en addition à leurs origines multiples et hétérogènes dans la plupart des cas, ces datasets sont très susceptibles au bruit, aux données manquantes et incohérentes. La qualité des données affecte les performances des classificateurs, pour cela une phase de prétraitement est nécessaire pour assurer une exploration de données efficace. Un processus de prétraitement peut inclure plusieurs étapes :

D’abord, il faut **évaluer la qualité des données collectées**, en les analysant dans le but de trouver des données manquantes, incohérentes ou incompatibles. Une fois ces problèmes détectés, nous devons **nettoyer les données** soit en ignorant les tuples défectueux ou en remplaçant manuellement les données erronées avec des valeurs nulles ou par défaut. Nous devons également mettre toutes les données, telles que les dates, sous un format uniforme, notamment pour les données collectées à partir de sources multiples. Une autre étape couramment utilisée est la suppression des doublons, cependant, cela ne nous convient pas puisque les commentaires en double peuvent être un indicateur de spam d’opinion qui mérite d’être vérifiée.

En ce qui concerne les données textuelles, ici généralement le corps de l’opinion, nous pouvons nettoyer ce que nous jugeons inutile, par exemple les URLs, les mentions, les balises HTML,...

⁷[Amazon.com](https://www.amazon.com)

⁸<https://www.tripadvisor.com/>

ne sont pas pertinents pour notre analyse. Cependant, les émojis, les hashtags et les symboles peuvent contenir des indicateurs utiles pour la détection de spam ou pour l’analyse des sentiments, si nous prenons la polarité des opinions en considération. On peut ajouter à cela la suppression de la ponctuation, l’élimination des mots vides,... Cela est ensuite suivi par la transformation des données à travers plusieurs méthodes telles que la tokénisation, la racinisation, la lemmatisation,...

Comme l’a souligné (SUN et al. 2017), le prétraitement des données a atteint des performances prometteuses pour les documents canoniques (actualités, ...) cependant, il est insatisfaisant lorsqu’il est appliqué à des critiques qui ne sont généralement pas grammaticales. Le problème ne fait qu’empirer lors de l’application de tâches telles que la tokenisation sur des langues plus flexibles comme le chinois.

2.7 Approches de représentation

Comme nous l’avons déjà mentionné, la qualité et la représentation des données alimentant les modèles de classification ont un impact important sur les performances de ces derniers. La phase de représentation concerne la définition des caractéristiques à exploiter durant la classification. Dans la littérature, il arrive que certaines étapes de prétraitement et de représentation se chevauchent, ou qu’elles se confondent complètement, il est donc souvent difficile de trouver une ligne claire qui les sépare. Il existe deux familles d’approches utilisées dans cette étape : l’ingénierie des caractéristiques et l’apprentissage des représentations. Nous allons présenter dans cette section la taxonomie de ces approches comme l’illustre la figure I.11. Cependant, nous devons garder à l’esprit que certaines catégories présentées pourraient se chevaucher, et que le domaine de l’apprentissage automatique est en constante évolution, de nouvelles méthodes et techniques pouvant brouiller les frontières entre ces catégories.

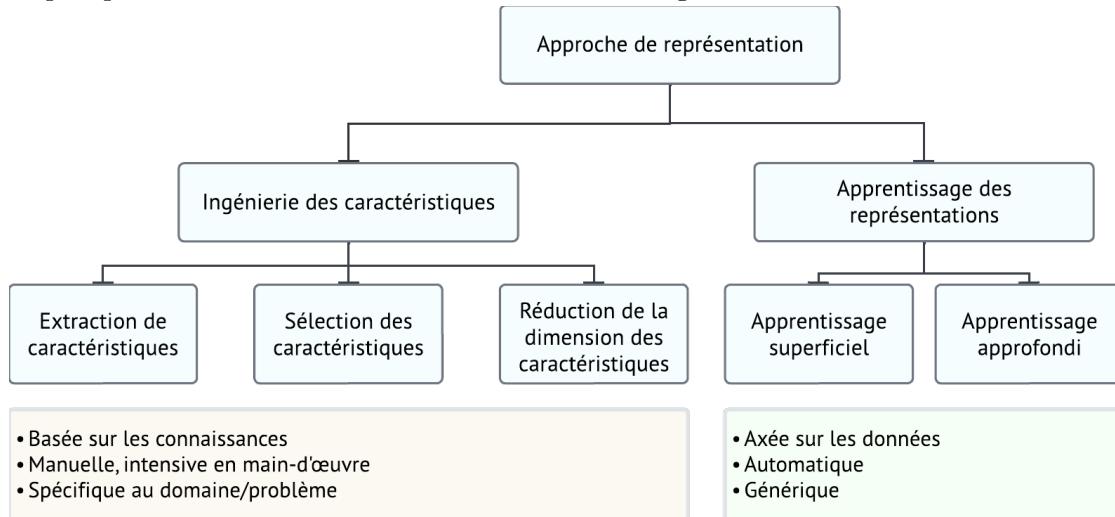


FIG. I.11 : Taxonomie des approches de représentation

2.7.1 Ingénierie des caractéristiques

En parcourant la littérature, il ne semble pas y avoir beaucoup d’attention concentrée sur l’ingénierie des caractéristiques (BROWNLEE 2014 ; YAN et YU 2019), nous trouvons différentes définitions et classifications, mais tout se résume à ceci : l’ingénierie des caractéristiques est le processus de conversion manuelle des données brutes en informations significatives (caractéristiques) en utilisant des connaissances spécifiques au domaine. Ce processus vise à aboutir à un nombre raisonnablement limité de caractéristiques utiles à faible dimensionnalité afin d’améliorer les performances du classifieur par rapport à ne lui fournir que des données brutes. De cette

définition, nous pouvons déduire trois sous-classes de cette catégorie : la sélection de caractéristiques, la réduction de la dimensionnalité des caractéristiques et l'extraction de caractéristiques.

2.7.1.1 Sélection des caractéristiques

La sélection des caractéristiques est le processus d'identification des caractéristiques les plus cohérentes, non redondantes et pertinentes à utiliser dans la construction de modèles. La réduction méthodique de la taille des ensembles de données est importante, car la taille et la variété des ensembles de données continuent de croître. Souvent, il est considéré comme faisant partie de la phase de prétraitement, car il n'introduit aucune nouvelle caractéristique, mais sélectionne plutôt un sous-ensemble des caractéristiques brutes qui sont plus interprétables. Son but est de trouver les caractéristiques qui ont un impact déterminant sur les résultats du modèle prédictif afin de réduire le nombre de caractéristiques et les risques de surajustement sans sacrifier le pouvoir prédictif Ceci revient à supprimer les données superflues et pour permettre au modèle de se concentrer uniquement sur les caractéristiques importantes. Dans la littérature, la sélection de caractéristiques est classifiée en deux catégories, celle des méthodes supervisées et non supervisées. Pour les approches non supervisées, on détermine la corrélation entre les caractéristiques sans prendre en considération la variable cible (ici l'étiquette de crédibilité). En revanche, les techniques supervisées examinent la relation entre chacune des caractéristiques et l'étiquette cible et incluent trois sous-catégories : les Méthodes d'enveloppement, les méthodes de filtrage et les méthodes embarquées (ZHENG et Y. ZHANG 2008).

- Les Méthodes d'enveloppement (ZHENG et Y. ZHANG 2008)
- Les méthodes de filtrage (ZHENG et Y. ZHANG 2008)
- Les méthodes embarquées (ZHENG et Y. ZHANG 2008)

2.7.1.2 Réduction de la dimension des caractéristiques

Parfois, nous ne pouvons pas obtenir une bonne dimensionnalité juste en sélectionnant les caractéristiques. Pour combler cette lacune, nous utilisons des techniques de réduction de dimensionnalité (DR). Bien que les deux soient utilisés pour réduire le nombre d'entités dans un jeu de données, il existe une différence importante : la sélection de caractéristiques consiste simplement à sélectionner et à exclure des caractéristiques données sans les modifier, par contre, la réduction de dimensionnalité transforme les entités en une dimension inférieure. Le but des techniques DR est de réduire les caractéristiques sans perdre beaucoup d'informations authentiques et d'améliorer les performances. Il en existe deux principaux types : les **DR linéaires**, qui incluent l'analyse en composantes principales (ACP)(PEARSON 1901), l'analyse en composantes indépendantes (ICA)(COMON 1994), l'analyse discriminante linéaire (LDA)(MI et al. 2003), la décomposition en valeurs singulières (SVD)(GOLUB et VAN LOAN 1996), l'indexation sémantique latente (LSI)(DEERWESTER et al. 1990), etc, ainsi que les **DR non linéaires** dont l'analyse en composantes principales du noyau (KPCA)(SCHÖLKOPF et al. 1998), l'incorporation de voisins stochastiques distribués en T, l'analyse discriminante linéaire du noyau (KLDA), la mise à l'échelle multidimensionnelle, cartographie isométrique (Isomap)(TENENBAUM et al. 2000), etc.

Le tableau 2.2 de mon rapport de master (SLIMANI 2023) résume les méthodes de réduction de dimensionnalité les plus utilisées dans le domaine de détection de spam.

2.7.1.3 Extraction de caractéristiques

L'extraction de caractéristiques est le processus de création de nouveaux indicateurs qui pourraient ne pas être vus dans les données brutes, et cela, en utilisant des connaissances d'experts. Contrairement à la sélection de caractéristiques qui conserve un sous-ensemble des caractéristiques d'origine, les algorithmes d'extraction de caractéristiques transforment les données

en un nouvel espace de caractéristiques. Parfois, il arrive que l'extraction de caractéristiques et l'ingénierie de caractéristiques soient utilisées de manière interchangeable alors que les deux sous-catégories précédentes sont considérées comme faisant partie de la phase de prétraitement.

En matière de détection de spam d'opinion, la littérature classe les caractéristiques principalement en deux grandes catégories : les caractéristiques textuelles qui concernent le contenu de l'opinion, et les caractéristiques comportementales qui concernent le détenteur de l'opinion. Souvent des caractéristiques concernant le produit ciblé et sa popularité sont ajoutées à la liste.

- **Les caractéristiques liées au texte de l'avis**

Les caractéristiques basées sur le contenu ont été les plus utilisées dans les méthodes de l'état de l'art, ces dernières sont principalement basées sur l'hypothèse qu'un spammeur a peu ou pas de connaissances sur le produit qu'il examine et, à ce titre, les avis authentiques et spam devraient présenter différents styles d'écriture, un indice pour les distinguer. Ces caractéristiques incluent des caractéristiques linguistiques telles que le lexique, la grammaire et les caractéristiques sémantiques du texte de l'opinion ainsi que des métadonnées. (MOHAWESH et al. 2021 ; REN et JI 2019). Nous présentons maintenant certaines caractéristiques textuelles utilisées dans les travaux de recherche, nous posons d'abord les deux documents D₁ et D₂, extraits de TripAdvisor, comme exemples sur lesquels nous expliquerons certains algorithmes présentés :

- **D₁** : « *Had a meal here, food was cold and tasted bad. The smell was weird and nauseating. Our waitress was friendly but the service was very slow.* »

- **D₂** : « *We had an excellent meal, food was freshly cooked and delicious. Our waitress was very friendly. It was enjoyable.* »

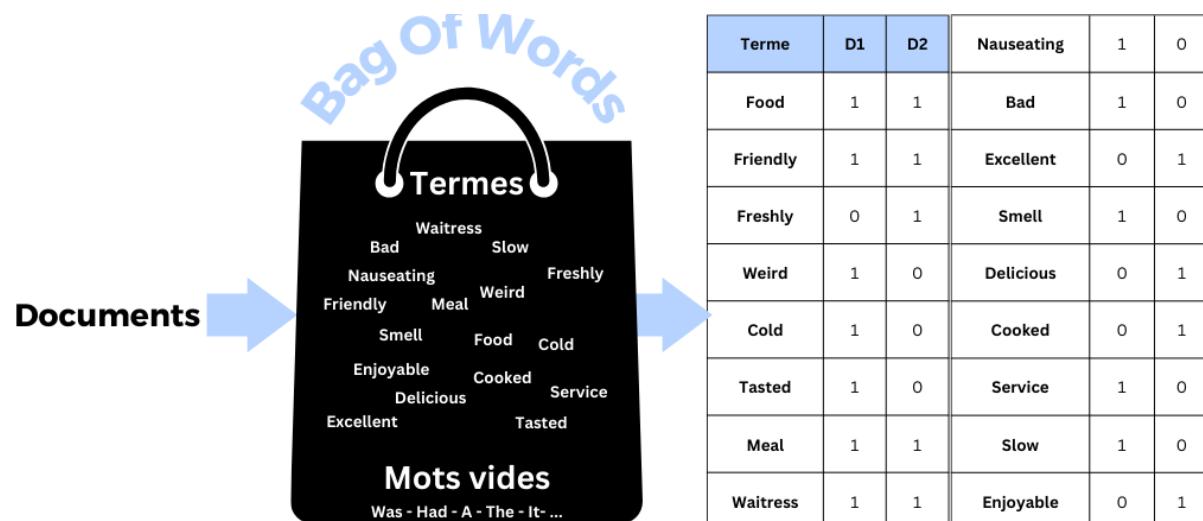


FIG. I.12 : Exemple du fonctionnement de Bag-of-Words (BOW)

1. **Les sacs de mots**, Bag of Words (BOW) en anglais, sont l'une des techniques de vecteurisation les plus utilisées. L'algorithme BOW résulte en une représentation du texte qui décrit l'occurrence des mots dans un document. La figure I.12 présente le résultat de l'application de BOW sur les deux documents D₁ et D₂. Cet algorithme est simple et intuitif, et il ne rencontre pas d'erreurs "En dehors du vocabulaire". Cependant, il ne prend pas en compte les nouveaux mots ni l'ordre de la phrase. Parmi les travaux ayant exploité la technique BOW, on peut citer : (HAJEK et al. 2020 ; KENNEDY et al. 2020 ; PATEL et THAKKAR 2014).

2. **Les n-grammes** sont une chaîne connectée de N éléments apparaissant consécutivement dans un document. Cette méthode est basée sur l'idée que ce type de représentation pourrait

capturer la structure d’une langue du point de vue statistique tout en maintenant l’ordre des mots, contrairement à BOW. La figure I.13 est un exemple simple de n-gramme appliqué sur une phrase extraite de D₁, elle montre le résultat de l’utilisation d’un unigramme (lorsque N=1), d’un bigramme (N= 2) et d’un trigramme (N=3). Même si cet algorithme est également simple et capable de capturer le sens sémantique de la phrase, il ignore toujours les nouveaux mots et à mesure que N augmente, la dimension de la formation de vecteurs augmente et le ralentit. Cet algorithme a été largement utilisé dans de nombreuses recherches (AHMED 2017; FUSILIER et al. 2015; GOEL et al. 2021; HAJEK et al. 2020; Jiwei LI et al. 2014; MUKHERJEE et al. 2012; OTT et al. 2013, 2011; SEDIGHI et al. 2017; T. WANG et ZHU 2014).

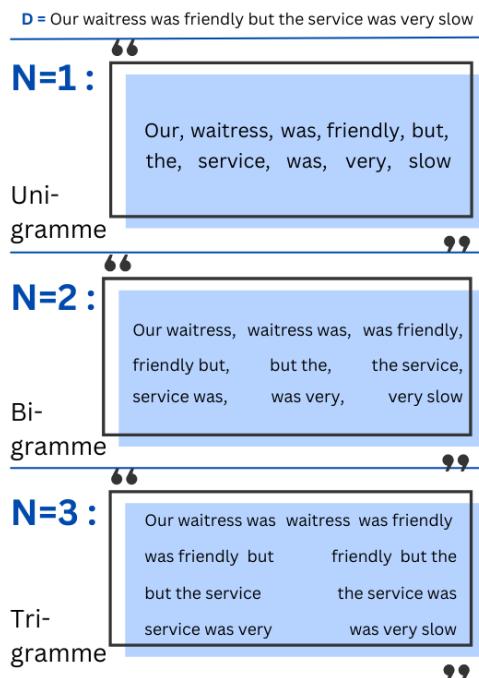


FIG. I.13 : Exemple du fonctionnement des n-grammes

3. POS, Part-of-speech tagging, et également appelé étiquetage grammatical ou désambiguisation des catégories de mots, est le processus consistant à attacher chaque mot dans le texte (le corps de l’opinion) avec une étiquette grammaticale, appelée étiquette POS, en fonction de son emplacement. La figure I.14 représente un exemple simple où nous avons appliqué la technique POS sur une C₁ et nous avons représenté les étiquettes par différentes couleurs. Parmi les études ayant exploité POS on peut citer : (ALSUBARI et al. 2020; GOEL et al. 2021; KENNEDY et al. 2020; Jiwei LI et al. 2014; MUKHERJEE et al. 2012; OTT et al. 2011; SEDIGHI et al. 2017; T. WANG et ZHU 2014). Selon (ALSUBARI et al. 2020), il semblerait que les avis véridiques contiennent plus de noms et d’adjectifs alors que les avis frauduleux ont plus de verbes et d’adverbes.



FIG. I.14 : Exemple du fonctionnement de la technique POS (Part-of-speech)

4. **LIWC**, ou en anglais Linguistic Inquiry and Word Count qui veut dire Recherche linguistique et comptage de mots, est un outil d’analyse de texte populaire⁹ qui peut être utilisé pour analyser, extraire et calculer des caractéristiques linguistiques significatives sous différents aspects. Il s’appuie sur plus de 100 dictionnaires intégrés pour classer les mots cibles en fonction de différentes catégories. Sa dernière version est LIWC-22, c’est un logiciel assez puissant mais il est cher. (ALSUBARI et al. 2020 ; Jiwei LI et al. 2014 ; MUKHERJEE et al. 2013b ; OTT et al. 2011) ont utilisé cet outil dans leur recherche.

5. **Tf-Idf**, de l’anglais Term Frequency-Inverse Document Frequency, est une métrique de pondération souvent utilisée dans la recherche d’informations et le traitement du langage naturel, c’est une version modifiée de l’approche Term-Frequency (TF). TF est une technique qui utilise le nombre de mots apparaissant dans des documents pour déterminer la similitude entre eux. Il représente chaque document d par un vecteur (de même longueur) présentant chaque terme t , appartenant à l’ensemble des documents, avec sa probabilité d’exister dans d . La valeur TF concernant un terme t dans un document d est calculée selon l’équation I.5. Comme TF pondère l’importance d’un terme et que cela augmente avec le nombre de fois qu’il apparaît, cela pourrait faire en sorte que des termes généralement plus courants que d’autres tels que les mots vides (“The”, “Then”,...) dominent le nombre de fréquences sans accorder assez d’importance aux mots plus discriminants, c’est là que TF-IDF est une amélioration : l’utilisation d’IDF (calculé à l’aide de l’équation I.6, où N est le nombre total de documents et $|deD : t\epsilon d|$ est le nombre de documents contenant le terme t) réduit l’impact de tels termes. TF-IDF est calculé selon l’équation I.7. TF-IDF a été utilisé dans (AHMED 2017; HASSAN et ISLAM 2021; H. LI et al. 2014; PATEL et THAKKAR 2014; SEDIGHI et al. 2017; T. WANG et ZHU 2014).

$$TF(t, d) = \frac{f_{t,d}}{\sum_{t'} f_{t',d}} \quad (I.5)$$

$$IDF(t, D) = \log \frac{N}{|deD : t\epsilon d|} \quad (I.6)$$

$$TF-IDF(t, d) = tf \cdot idf_{t,d} = \frac{f_{t,d}}{\sum_{t'} f_{t',d}} \cdot \log \frac{N}{|deD : t\epsilon d|} \quad (I.7)$$

6. **La stylométrie** est le domaine de la linguistique qui exploite des méthodes statistiques afin de définir les propriétés du style d’un texte. Les caractéristiques stylométriques peuvent être divisées en caractéristiques lexicales (telles que la longueur moyenne d’une phrase, le nombre total de termes,...) et syntaxiques (qui incluent la fréquence de ponctuation, etc.) (AHMED 2017). Les auteurs de (SHOJAEET AL. 2013), qui ont extrait 234 caractéristiques stylométriques du

⁹<https://www.liwc.app/>

dataset de (OTT et al. 2011), ont obtenu un F-score de 84% en utilisant à la fois des caractéristiques lexicales et syntaxiques surpassant les résultats qu’ils ont obtenus sans les combiner. Cette méthode a été utilisée dans de nombreux travaux tels que (HARRIS 2022 ; JINDAL et LIU 2008 ; KENNEDY et al. 2020 ; RASTOGI et MEHROTRA 2018 ; RASTOGI et al. 2020 ; SHAN et al. 2021 ; SHOJAEET AL. 2013). Parmi les caractéristiques utilisées, on trouve la longueur du corps de l’opinion (RL), la diversité lexicale (le pourcentage de termes uniques), le ratio de mots négatifs (RNW) et positifs (RPW), le ratio de mots imaginatifs (RImW) et informatifs (RInW), le nombre de mots spatio-temporels etc.

7. Les caractéristiques sémantiques traitent de la signification cachée des mots et de la relation entre eux. Cela aide à identifier les similitudes entre les avis, car même si le spameur modifie certains mots dans l’un des avis afin d’induire les lecteurs en erreur, les caractéristiques sémantiques arrivent à détecter les avis générés à partir d’autres. Ainsi, par exemple, si l’auteur remplace le mot “Delicious” par “Tasty” ou “Horrible” par “Awful”, un modèle basé sur des caractéristiques sémantiques pourrait toujours être en mesure de reconnaître la similarité entre eux. Les expériences de (L. LI et al. 2015) et de (KIM et al. 2015) montrent que l’utilisation de caractéristiques sémantiques a effectivement amélioré les performances de leurs modèles, dans (L. LI et al. 2015) ces caractéristiques ont surpassé LIWC, POS et n-grammes dans les textes inter-domaines. Dans (ZIANI et al. 2021), les auteurs ont travaillé sur les caractéristiques sémantiques de la langue arabe en exploitant les relations de cohérence entre les syntagmes d’une phrase, car deux syntagmes peuvent soit se contredire, soit se renforcer. Parmi les multiples connecteurs présents dans la langue arabe, ils ont en choisi 6 : les connecteurs d’explication, de cause, de condition, de ressemblance, d’opposition et de différence.

- **Caractéristiques liées au comportement de l’auteur**

Dans le but de pallier ces lacunes, les chercheurs se sont tournés vers une autre entité importante en matière de spam d’opinion, le spameur, c’est-à-dire qu’ils ont commencé à utiliser des caractéristiques mettant en valeur le comportement de l’examinateur. Ces caractéristiques ont tendance à concerner l’opinion actuelle (les informations qui l’entourent) et l’histoire de son auteur. Nous allons maintenant présenter certaines des caractéristiques les plus utilisées dans la littérature, nous les avons extraites des travaux suivants (ANDRESINI et al. 2022 ; HARRIS 2022 ; HUSSAIN et al. 2019 ; JINDAL et LIU 2008 ; MUKHERJEE et al. 2013a ; OTT et al. 2011 ; RASTOGI et MEHROTRA 2018). Dans ce qui suit, **r** fait référence à un avis, **a** à un auteur et **p** à un produit.

1) **ARD : Average rating deviation** : L’écart de notation moyen peut être utile pour identifier les spameurs, vu qu’un examinateur impartial devrait évaluer un produit en fonction de la note moyenne attribuée à celui-ci. Cependant, les spameurs tentant de promouvoir, ou de rétrograder un produit lui attribuent généralement une note élevée ou extrêmement basse, ce qui peut nous aider à les identifier. Cette caractéristique peut être calculée selon la formule I.8.

$$ARD(a) = \text{avg} \frac{|r_{ap} - \bar{r}_p|}{4} \quad (\text{I.8})$$

2) **Maximum Review Similarity (MRS)** : Le fait qu’un seul évaluateur publie des avis similaires sur un produit distinct pourrait être une forte indication qu’il est un spameur, vu que la rédaction de nouvelles critiques prend du temps, ces derniers ont tendance à copier leurs avis pour différents produits. C’est pourquoi il peut être utile de calculer la similarité maximale des avis à l’aide de la similarité cosinus afin de détecter cette anomalie. On peut la calculer selon la formule I.9 tandis que le cosine entre ces deux opinions s’obtient par la formule I.10.

$$MRS(a) = \max_{r_i, r_j \in R_a, i < j} \cos(r_i, r_j) \quad (\text{I.9})$$

$$\cos(r_i, r_j) = \frac{r_i \cdot r_j}{\|r_i\| \|r_j\|} \quad (\text{I.10})$$

3) **Average Review Length (ARL)** : Plusieurs recherches ont montré que la plupart des spameurs ne détaillent pas leurs critiques, ce qui semble logique car la plupart d'entre eux ne connaissent pas grand-chose aux produits qu'ils examinent et ne passent pas beaucoup de temps à rechercher leur cible et à rédiger leurs avis. Ainsi, la longueur moyenne de l'opinion d'un évaluateur a été considérée par beaucoup comme une caractéristique qui pourrait aider à distinguer les spameurs des examinateurs honnêtes. Pour calculer cette caractéristique, il suffit de rassembler tous les avis écrits par l'examinateur et de calculer la moyenne de leur longueur, certaines études transforment ensuite cette valeur calculée en un booléen égalant vrai s'il était inférieur à une certaine valeur, ou faux sinon.

4) **Burstiness (BST)** : Des études rapportent que les examinateurs authentiques publient des critiques de temps en temps, tandis que les spameurs d'opinion qui ne sont généralement pas des membres de longue durée d'un site publient plus fréquemment. Cette pratique irrégulière consistant à publier plusieurs avis dans un court laps de temps pourrait être un indicateur qu'un utilisateur est un spameur. L'éclatement des avis est défini à l'aide de la fenêtre d'activité d'un utilisateur (la différence entre les dates de publication du premier et du dernier avis), si les avis sont publiés sur une période raisonnablement longue, cela indique probablement une activité normale, cependant, lorsque la plupart des avis sont publiés dans un délai très court cela pourrait indiquer une infliction de spam. En fixant une valeur seuil τ (28 jours dans certaines études que nous avons examinées), nous définissons la formule de cette caractéristique comme expliqué dans la formule I.11, où $F(a)$ et $L(a)$ désignent respectivement le premier et le dernier commentaire publié par un auteur.

$$BST(a) = \begin{cases} 0, & \text{Si } L(a) - F(a) > \tau \\ 1 - \frac{L(a) - F(a)}{\tau}, & \text{Sinon} \end{cases} \quad (\text{I.11})$$

5) **Early review ratio (ERR)** : Étant donné que les premiers avis ont un impact significatif sur les ventes de produits et services, la plupart des spameurs essaient d'être parmi les premiers examinateurs afin de contrôler le sentiment et d'induire les acheteurs en erreur. Afin de calculer le ratio des premières critiques pour chaque auteur, nous appliquons la formule I.12.

$$ERR(a) = \frac{|\{r \in R_a : r \text{ is first review}\}|}{|R_a|} \quad (\text{I.12})$$

6) **Maximum Number of Reviews per Day (MRD)** :

Dans de nombreuses études, il a été observé que 90% des utilisateurs n'écrivent jamais plus d'un avis en une journée, tandis qu'environ 75% des spameurs publient plus de 5 avis certains jours. Ainsi, ce comportement anormal pourrait être un indicateur utile. Nous définissons cette caractéristique en calculant le nombre maximum d'avis par jour pour un auteur et en le normalisant par la valeur maximale de nos données en suivant la formule I.13.

$$MRD(a) = \frac{MaxRev(a)}{\max_{a \in A}(MaxRev(a))} \quad (\text{I.13})$$

7) Positive Ratio (PR) et Negative Ratio (NR) :

La polarité des critiques rédigées par un utilisateur peut être un indicateur pour savoir s'il est un spameur ou non. Par exemple, si un auteur a un pourcentage élevé d'avis positifs sur certains produits, il pourrait ne pas être digne de confiance, et comme environ 85% des spameurs écrivent plus de 80% de leurs avis positifs, cela pourrait valoir la peine d'être étudié. Le ratio d'avis positifs ou négatifs pour un auteur est calculé en divisant le nombre d'avis de la polarité choisie par le nombre total d'avis qu'il a publié.

8) Early rating deviation (ERD) :

Comme indiqué précédemment, les spameurs essaient de maximiser leur impact en publiant leurs avis le plus tôt possible. Nous utilisons l'écart de notation précoce pour capturer la nature frauduleuse des avis rédigés au début du lancement du produit en utilisant la formule I.14 où $w_{rp} = \frac{1}{(t_{rp})^\alpha}$ est le poids d'un avis sur un produit et t_{rp} est l'heure à laquelle il a été publié.

$$ERD(r) = \frac{|r_p - \bar{r}_p|}{4} \cdot W_{r_p} \quad (\text{I.14})$$

9) Les métadonnées :

Les métadonnées fournissent des informations sur l'avis et son auteur plutôt que sur le contenu, des caractéristiques telles que l'adresse IP d'un utilisateur, son comportement sur le site Web, etc. lorsqu'elles sont combinées avec d'autres caractéristiques peuvent s'avérer bénéfiques pour détecter des avis et des comportements anormaux ou étranges. Cependant, ces données ne sont pas disponibles sur la plupart des sources de données, ce qui limite ainsi leur utilité pour détecter les spams.

Puisque les caractéristiques générées selon cette famille d'approches sont spécifiques et dépendent du contexte et de la tâche à accomplir, il n'existe pas d'ensemble universel de caractéristiques qui fonctionne bien pour toutes les tâches. De plus, l'ingénierie des caractéristiques est manuelle et dépend fortement de la contribution d'experts, ce qui demande beaucoup de travail et de temps.

2.7.2 Apprentissage des représentations

L'apprentissage des représentations ou apprentissage des caractéristiques est un ensemble de techniques permettant la transformation automatique des données d'entrée. Son objectif principal est de fournir des représentations abstraites et utiles pour les tâches de ML telles que la classification, la prédiction, la segmentation d'image, la reconnaissance de la parole et bien d'autres encore (BENGIO et al. 2013 ; YAN et YU 2019).

Par rapport aux caractéristiques conçues à la main décrites ci-dessus, l'apprentissage des caractéristiques est automatique, prend moins de temps et nécessite une connaissance minimale du domaine humain pour produire de meilleurs résultats. De plus, contrairement à l'ingénierie des caractéristiques, ce type d'apprentissage ne nécessite pas d'efforts supplémentaires pour concevoir des caractéristiques pour une nouvelle tâche et a une plus grande capacité de généralisation. Selon la plupart des études (ZHONG et al. 2019), l'apprentissage des caractéristiques peut être divisé en deux catégories selon le niveau de leur abstraction hiérarchique des données : superficiel ou profond.

En raison de l'indisponibilité de données à grande échelle, de l'existence de nombreux paramètres et de la non-convexité élevée des réseaux, les architectures profondes complexes n'ont pas pu être facilement entraînées (ZHONG et al. 2019), ce qui a fait que les méthodes d'**apprentissage superficiel des représentations** ont dominé ce domaine auparavant. Ce type de techniques

ML consiste à extraire des caractéristiques des données sans transformations profondes. Ces techniques impliquent généralement des modèles linéaires et non linéaires ou de simples réseaux de neurones à anticipation avec un petit nombre de couches, et elles ne nécessitent pas de ressources de calcul étendues ni de grandes quantités de données d'entraînement. Ces techniques sont souvent utilisées dans des cas où les données sont relativement simples et peuvent inclure des méthodes supervisées telles que LDA, et non supervisées comme le clustering (K-means, Modélisation des mélanges gaussiens (GMM), etc), la factorisation matricielle (PCA, ICA,...) et certains réseaux de neurones simples tels que les cartes auto-organisées (SOM) et les auto-encodeurs simples. Certains de ces concepts ont été présentés précédemment pour d'autres tâches, cela est dû au fait que plusieurs techniques ML peuvent servir à diverses fins.

Plus récemment, grâce à la disponibilité des données, à l'augmentation de la puissance de calcul et des machines, l'apprentissage profond est devenu plus renommé et les chercheurs ont commencé à l'utiliser davantage pour les tâches de représentation de caractéristiques. Il a été introduit pour la première fois dans le domaine de NLP en 2012 après le succès majeur qu'il a démontré dans la reconnaissance d'objets et de la parole (ZHOU et al. 2020). Contrairement aux algorithmes d'apprentissage superficiel, **les méthodes d'apprentissage profond** des représentations peuvent généralement entraîner de meilleures abstractions des données d'origine pour les tâches de classification et de détection. Dans ce qui suit, nous discuterons de certaines des méthodes les plus intéressantes dans ce type d'apprentissage de représentation :

2.7.2.1 Auto-encodeurs

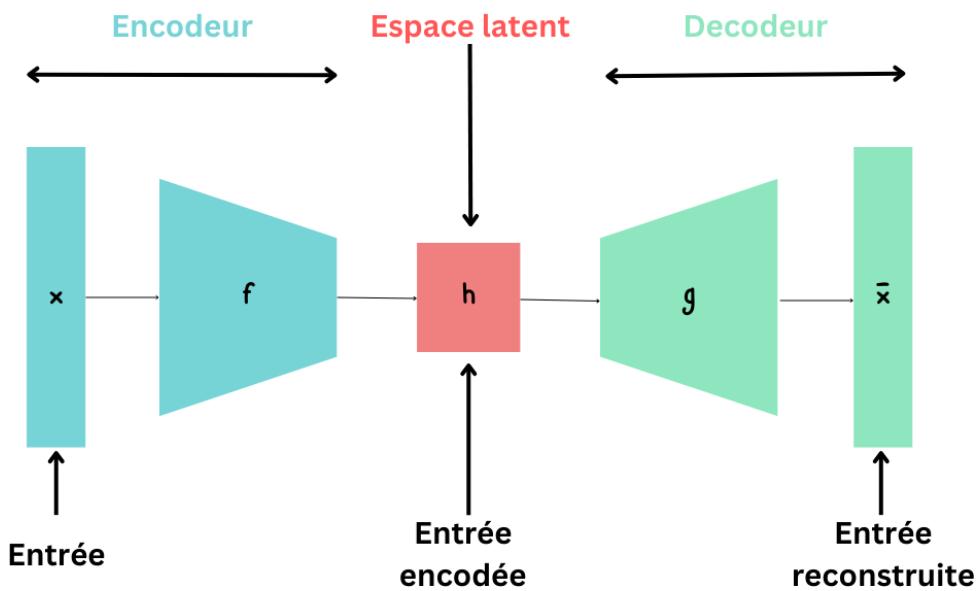


FIG. I.15 : Architecture simplifiée d'un auto-encodeur

Les encodeurs automatiques sont une technique de ML que nous avons vue précédemment, c'est un outil polyvalent qui peut être utilisé pour une variété de tâches. Dans le contexte de la réduction de dimensionnalité, les auto-encodeurs sont entraînés sur des données d'entrée de grande dimension et apprennent une représentation compressée des données dans un espace de dimension inférieure. Alors qu'en termes d'apprentissage de caractéristiques, un auto-encodeur peut être formé sur un ensemble de données pour apprendre une représentation compacte et significative des données d'entrée. Les modèles basés sur l'auto-encodeur sont des réseaux de neurones qui tentent de copier leurs entrées vers leurs sorties, ils sont considérés comme faisant partie des modèles d'apprentissage non supervisés les plus robustes pour extraire des caractéristiques efficaces et discriminantes à partir d'un grand ensemble de données non étiqueté

(BAGHAEI et al. 2022). Leur architecture générale est constituée de deux composants, la fonction **encodeur** $f(x)$ qui vise à transformer l'entrée x en une variable latente h de dimensions inférieures et la fonction **décodeur** g dont le but est de reconstruire l'entrée \hat{x} étant donné la variable précédente h comme le montre la figure I.15. Le processus d'apprentissage consiste à ajuster les poids de ces composantes en fonction de la fonction de perte de la reconstruction, et il existe de nombreuses variables d'auto-encodeurs : les sous-complets, les débruiteurs (DAE), les clairsemés (SAE), les variationnels (VAE) et les contractifs (CAE) (BAGHAEI et al. 2022). (DONG et al. 2020) ont proposé une architecture de forêt décisionnelle d'auto-encodeur neuronal et rapportent une mesure F-score de 95,11%. (SAUMYA et SINGH 2022), de leurs coté, ont utilisé un auto-encodeur LSTM dans le but d'apprendre la structure interne des avis spams tout en conservant les longues séquences existantes dans ces revues, Et ils ont signalé un F-score de 99% en utilisant une représentation One Hot Embedding.

2.7.2.2 Word Embeddings

Les Word Embeddings, ou prolongements de mots, consistent en des représentations des mots comme des vecteurs denses dans un espace de grande dimension, dans le but de capturer les relations sémantiques et syntaxiques entre les mots. Chaque mot est mappé sur un vecteur, et les mots ayant des significations similaires sont mappés sur des vecteurs proches les uns des autres dans l'espace vectoriel. Dans d'autres études, les chercheurs ont également appliqué des intégrations de phrases (Sentence embeddings) via des RNNs, des CNNs et des réseaux d'auto-attention (ZHOU et al. 2020). Dans la littérature, il existe deux familles principales de Word embeddings, celle des embeddings statiques et l'autre dynamique (Y. WANG et al. 2019). Ces prolongements jouent souvent le rôle de la première couche de traitement de données pour les approches d'apprentissage profond.

- **Word Embeddings statiques :**

Les Word Embeddings statiques sont un type de technique de word embedding utilisée dans le traitement du langage naturel pour représenter les mots sous forme de vecteurs denses de dimension fixe dans un espace de grande dimension. Ces incorporations sont « statiques » dans le sens où elles sont pré-formées sur un grand corpus de texte et ne changent pas au cours d'une tâche ou d'un processus d'apprentissage spécifique. Les vecteurs capturent les relations sémantiques et syntaxiques entre les mots en fonction de leurs modèles de co-occurrence dans le corpus d'entraînement, et peuvent être utilisés pour représenter la signification de mots individuels ou de documents entiers. Dans les intégrations de mots statiques, un mot est mappé sur une représentation vectorielle fixe, quel que soit son contexte ou la manière dont il est utilisé dans une phrase ou un document spécifique. Cela signifie qu'un mot aura toujours la même représentation vectorielle, quelle que soit sa signification dans différents contextes. Par exemple, dans la phrase “I rode the train to get to the gym so I could train with my peers” le mot **train** sera mappé sous le sens ‘train’ et donc ces autres significations (e.g. ‘formation’) ne seront pas prises en considération. Les techniques d'incorporation de mots statiques populaires incluent Word2Vec (MIKOLOV et al. 2013) et ses extensions (par exemple Sent2Vec (Q. LE et MIKOLOV 2014) et Doc2Vec (Q. LE et MIKOLOV 2014)), GloVe (PENNINGTON et al. 2014) et fastText (BOJANOWSKI et al. 2017).

- **Word2vec** a été introduit pour la première fois en 2013 (MIKOLOV et al. 2013), c'est l'un des modèles de word embeddings les plus populaires qui est basé sur un réseau neuronal pour générer des intégrations de mots statiques. Le modèle Word2vec apprend à représenter des mots dans un espace vectoriel continu où les mots ayant des significations similaires sont plus proches les uns des autres. L'un des exemples fréquemment donnés est l'équation “ $Word2vec(king) - Word2vec(man) + Word2vec(woman) = Word2vec(queen)$ ” c'est-à-dire que Word2vec mappe les

mots de manière à ce que la valeur vectorielle obtenue par la soustraction et l'addition de ces vecteurs est égale au vecteur correspondant à l'expression "queen". Le modèle est entraîné sur un grand corpus de texte, et l'objectif est soit de prédire la probabilité d'un mot compte tenu de son contexte, soit la probabilité d'un contexte compte tenu d'un mot. Sur cette base, il existe deux algorithmes principaux utilisés dans Word2vec : Sac continu de mots (CBOW : Continuous Bag of Words), dans lequel le modèle est formé pour prédire un mot en fonction de son contexte (l'ensemble de mots qui apparaissent dans une fenêtre fixe du mot cible), et Skip-Gram, qui prédit le contexte étant donné le mot cible (MOHAWESH et al. 2021). La figure I.16 présente une architecture simplifiée de ces deux algorithmes pour une meilleure compréhension. Afin de minimiser la différence entre les probabilités prédites et les probabilités réelles, le modèle est formé en ajustant les poids du réseau de neurones à l'aide de la rétropropagation. Parmi les travaux ayant utilisé Word2vec on peut citer (ANDRESINI et al. 2022 ; SHEHNEPOOR et al. 2021) pour l'algorithme CBOW et (HAJEK et al. 2020 ; KENNEDY et al. 2020) pour celui de Skip-Gram.

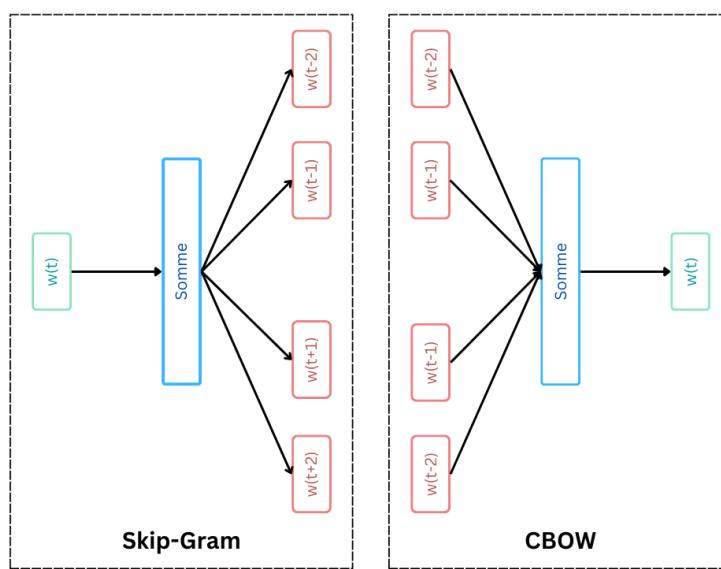


FIG. I.16 : Architecture simplifiée des algorithmes Word2vec

- **GloVe**, abréviation de Global Vectors for Word Representation, a été introduit pour la première fois dans (PENNINGTON et al. 2014) suite au succès de Word2vec. Ce modèle d'intégration est basé sur une matrice globale de co-occurrence de mots qui utilise une approche similaire à Word2Vec, mais au lieu de prédire les mots en s'appuyant uniquement sur les informations locales, il vise à optimiser directement les vecteurs de mots afin qu'ils capturent les propriétés distributionnelles du corpus incorporant à la fois des relations locales et globales entre les mots. Il s'agit d'une méthode basée sur le comptage, ce qui signifie qu'elle utilise le nombre de co-occurrences de mots pour apprendre les vecteurs de mots. L'idée principale derrière GloVe est que le rapport des probabilités de co-occurrence de deux mots doit être égal au rapport de leurs longueurs vectorielles. Cette approche permet à GloVe de capturer les similitudes sémantiques et syntaxiques entre les mots d'une manière plus directe et efficace que d'autres méthodes basées sur le comptage comme LSA par exemple.

Le problème rencontré par Word2vec et Glove est leur incapacité à traiter les mots hors vocabulaire (OOV), donc si le modèle ne voyait pas un terme pendant la formation, il ne serait pas en mesure de le représenter par la suite.

- **FastText** tente de résoudre ce problème. **FastText** est une variante de Word2vec basée sur le modèle Skip-Gram, et il a été introduit dans (BOJANOWSKI et al. 2017). L'une des principales

différences entre eux est que FastText apprend les représentations vectorielles des unités de sous-mots, telles que les n-grammes de caractères, en plus des mots complets. Cela permet au modèle de mieux gérer les mots OOV ainsi que de capturer la morphologie des mots. De plus, les représentations de sous-mots de FastText peuvent capturer la signification de mots qui ne sont pas compositionnels, tels que le préfixe et le suffixe d'un mot. FastText représente chaque mot avec un sac de n-grammes, par exemple le mot ‘Delight’ est représenté pour n=4 par le 4-gramme : ‘<Del’, ‘Deli’, ‘elig’, ‘ligh’, ‘ight’, ‘ght>’ où ‘<’ et ‘>’ indiquent le début et la fin du terme, respectivement. L'un des avantages de FastText est sa vitesse et son évolutivité, ce qui lui permet de s'entraîner efficacement sur de grands ensembles de données contenant des millions ou des milliards de mots. FastText comprend également des modèles pré-formés pour plusieurs langues, ce qui facilite son utilisation pour les langues autres que l'anglais.

- **Word embeddings dynamiques :**

Contrairement aux plongements de mots statiques, dans les plongements de mots dynamiques, la représentation vectorielle d'un mot n'est pas figée et peut varier en fonction du contexte dans lequel il apparaît. Cela signifie qu'un mot peut avoir différentes représentations vectorielles en fonction du contexte dans lequel il apparaît, ce qui peut capturer les différentes significations du mot dans différents contextes. Les Word embeddings dynamiques, également appelées Word embeddings contextualisées, sont générées en formant un réseau de neurones sur un grand corpus de données textuelles à l'aide d'une tâche de modélisation du langage, comme prédire le mot suivant dans une phrase ou remplir un blanc. Il en résulte un modèle qui peut générer une représentation vectorielle pour n'importe quel mot donné, en tenant compte de ses mots environnants dans le contexte donné. Des exemples courants d'incorporations de mots dynamiques incluent ELMo (PETERS et al. 2018), ULMFiT (HOWARD et RUDER 2018) et BERT (DEVLIN et al. 2018b).

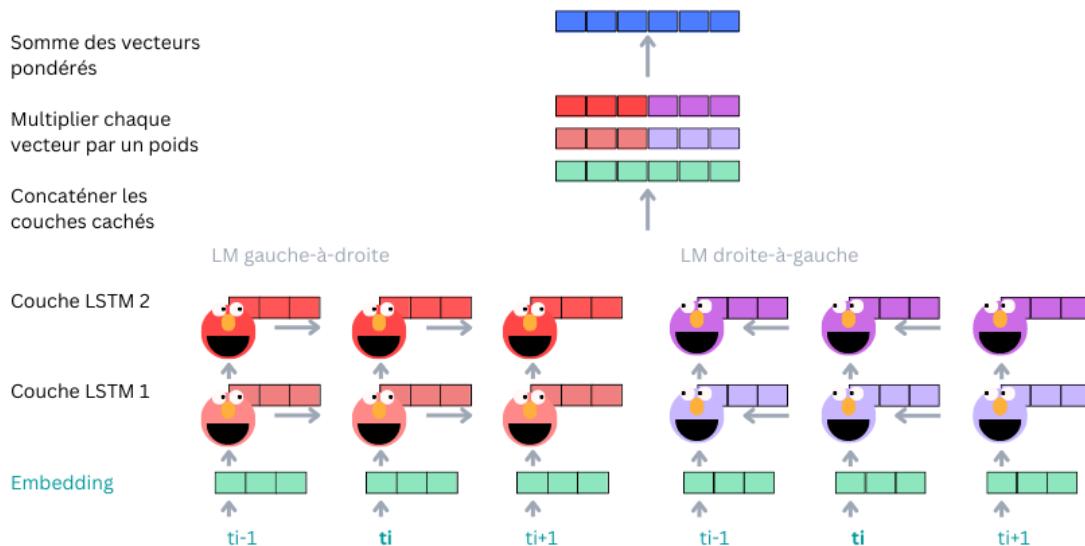


FIG. I.17 : Architecture simplifiée de ELMo

- **ELMo**, qui signifie “Embeddings from Language Models”, est une représentation de mots contextualisée profonde introduite pour la première fois en 2018 (PETERS et al. 2018), elle génère des incorporations de mots en fonction des contextes dans lesquels ils sont utilisés pour enregistrer le sens du mot et récupérer des informations contextuelles supplémentaires. ELMo est basé sur un modèle de langage neuronal qui apprend à prédire le mot suivant dans une séquence de mots et est entraîné sur de grands corpus de texte à l'aide d'un réseau bidirectionnel profond LSTM (mémoire longue à court terme), ce qui lui permet de capturer les dépendances

contextuelles des mots dans une phrase. La sortie de ce réseau est un ensemble de word embeddings contextualisés qui représentent la signification d'un terme dans le contexte des mots environnants dans une phrase, cette architecture est présentée (de manière simplifiée) dans la figure I.17. Grâce à la pré-formation, il peut représenter plus précisément des mots polysémiques dans une variété de contextes et est plus informatif sur la sémantique de niveau supérieur du texte.

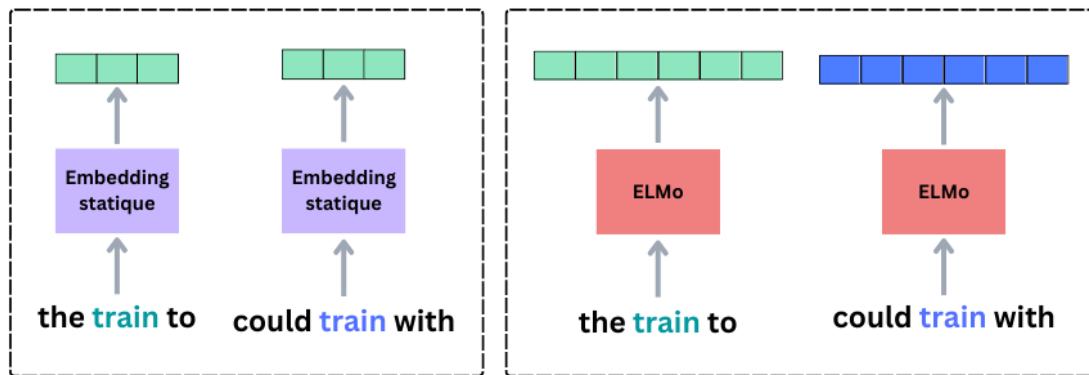


FIG. I.18 : Exemple du fonctionnement de ELMo

Si nous utilisons la phrase “I rode the *train* to get to the gym so I could *train* with my peers” comme exemple une fois de plus, les prolongements présentés précédemment mapperaient le mot “train” avec le même vecteur, ELMo de son côté forme son réseau bidirectionnel afin que son modèle de langage ait un sens à la fois du mot précédent et du suivant par conséquent le mot ‘train’ sera représenté avec un vecteur différent selon le contexte, comme l’illustre la figure I.18.

- **BERT**, ou Bidirectional Encoder Representations from Transformers (Représentations d'encodeurs bidirectionnels à partir de transformateurs), est un réseau neuronal profond qui a été introduit en 2018 pour les tâches de traitement du langage naturel (DEVLIN et al. 2018b), il s'agit d'un modèle de langage pré-entraîné développé par Google qui utilise une architecture de transformateur bidirectionnel combinant l'apprentissage non supervisé et supervisé pour générer des représentations de mots contextualisées de haute qualité. L'architecture de BERT comprend un encodeur avec plusieurs couches d'attention multi-têtes. BERT a été entraîné sur des quantités massives de données textuelles (paragraphes de texte Wikipédia anglais de 2500 millions de mots et corpus de livres de 800 millions de mots), l'une de ses caractéristiques déterminantes est sa stratégie de pré-formation, appelée Masked Language Model (MLM) où il masque de manière aléatoire des mots dans la phrase d'entrée, et l'objectif du modèle est de prédire ces mots masqués en fonction de leur contexte environnant. De plus, il utilise une tâche de prédiction de phrase suivante, où il entraîne le modèle pour déterminer si deux phrases sont consécutives ou non.

Une fois que tout cela est fait, l'encodeur peut être réglé avec précision pour la tâche spécifique à accomplir. Cela implique la formation du modèle à l'aide d'un ensemble de données plus petit et spécifique à la tâche. Au cours de ce processus, les poids du modèle BERT pré-formé sont ajustés pour faire des prédictions sur la tâche spécifique. La figure I.19 est directement extraite de l'article original (DEVLIN et al. 2018b) présentant des exemples de fine-tuning de BERT sur différentes tâches.

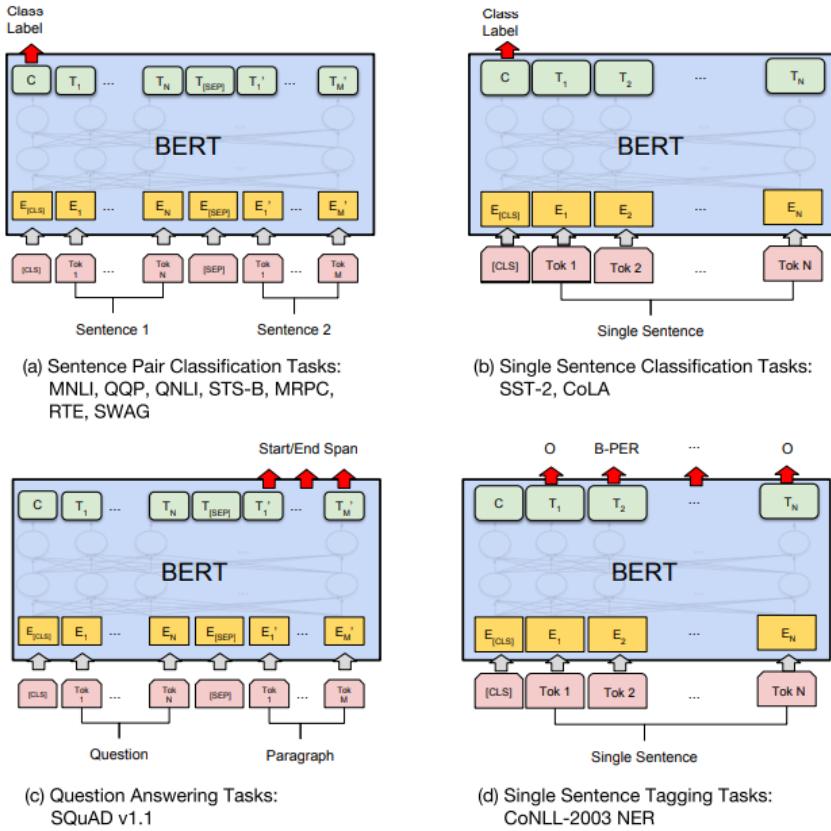


FIG. I.19 : Exemples d’affinage de BERT sur différentes tâches (DEVLIN et al. 2018b)

Dans le même article (DEVLIN et al. 2018b), les auteurs ont introduit deux variantes de BERT, à savoir **BERT-BASE** et **BERT-LARGE**. BERT-BASE se compose d’une pile de 12 blocs de Transformations (une séquence d’encodeurs) ainsi que 12 têtes d’attention. BERT-LARGE, quant à lui, se distingue de BERT-BASE par 24 blocs de transformations et 16 têtes d’attention, mais pas que ! BERT-LARGE a été entraîné sur un corpus de texte beaucoup plus grand que la version BERT-BASE, ce qui lui permet de capturer des relations plus complexes entre les mots, cependant il nécessite plus de mémoire et de puissance de calcul pour l’entraînement et l’inférence que la version BERT-BASE.

D’autres variantes de BERT sont apparues, telles que **DistilBERT** qui en est une version allégée visant à atténuer certaines des limites de BERT telles que la complexité de calcul, la taille de longueur d’entrée fixe et le problème d’intégration de mots, et **RoBERTa** une version étendue de BERT qui peut surpasser les performances de BERT en entraînant le modèle plus longtemps et sur des séquences plus longues (MOHAWESH et al. 2021). Nous pouvons également mentionner **DziriBERT**, le premier modèle de langage basé sur Transformer qui a été pré-entraîné spécifiquement pour le dialecte algérien (ABDAOUI et al. 2021).

L’approche de fine-tuning n’est pas la seule façon d’utiliser BERT. Tout comme ELMo, le modèle BERT pré-entraîné peut être directement utilisé pour extraire des représentations de mots contextualisées riches à partir de données textuelles, sans avoir besoin d’une formation supplémentaire, avant d’alimenter ces prolongements dans un modèle de classification. L’un des avantages de cette approche est qu’elle ne nécessite pas de données étiquetées pour l’affinage, ce qui peut être particulièrement utile dans des contextes où les données étiquetées sont rares ou coûteuses à obtenir. Cependant, l’utilisation de cette méthode peut nécessiter des ressources de calcul importantes, en particulier pour les modèles plus grands tels que **BERT-LARGE**, en raison de la complexité de la tâche de calcul des word embeddings contextualisées.

Parmi les travaux qui ont exploité les performances exceptionnelles de BERT pour la problématique de détection de spam d’opinion, nous citons : (ANDRESINI et al. 2022; KENNEDY et al. 2020; MIR et al. 2023; MOHAWESH et al. 2021; SHAALAN et al. 2021).

Nous avons passé en revue les principales techniques de représentation et leur classification. L’extraction et l’apprentissage de caractéristiques transforment tous deux les données brutes, comme expliqué ci-dessus, afin de trouver des caractéristiques pertinentes, tandis que la sélection de caractéristiques et la réduction de dimensionnalité ont pour objectif de rendre le processus moins complexe sans sacrifier trop de puissance de prédiction. Il peut être intéressant de noter que les techniques d’extraction et d’apprentissage produisent souvent un grand nombre de caractéristiques qui peuvent être redondantes et entraîner des exigences de calcul excessives qui, à leur tour, peuvent rendre leur application en temps réel peu pratique ou inutilement difficile, par conséquent, il est courant d’utiliser la sélection de caractéristiques et la réduction de la dimensionnalité pour remédier à ce problème. La combinaison de différentes caractéristiques et différentes représentations peut améliorer les performances obtenues (ANDRESINI et al. 2022).

2.8 Approches de classification des spams d’opinion

L’objectif principal de la détection de spam d’opinions est d’identifier chaque avis spam ou chaque spameur, les approches adoptées pour réaliser cela sont diverses et peuvent être classifiées selon plusieurs critères : l’entité à détecter (opinion spam, auteur spameur, produit ciblé,...), le type d’approche de représentation avec qui cette classification est utilisée, etc. Dans cette section, nous nous intéresserons aux approches les plus citées dans la littérature pour la tache de classification selon leur architecture, nous les catégoriserons donc en deux familles : celle des modèles ML traditionnels et celle des modèles neuronaux (MOHAWESH et al. 2021 ; REN et JI 2019).

2.8.1 Modèles ML traditionnels

Les modèles de classification traditionnels ont souvent une structure simple et un temps d’apprentissage et de prédiction relativement rapides, ils sont souvent associés à des approches d’ingénierie de caractéristiques (généralement Bag-Of-Words et n-grammes) qui pourraient améliorer les performances des modèles, en particulier dans le cas des données structurées. Cependant, ils pourraient ne pas être en mesure de capturer des relations plus complexes et non linéaires dans les données, sans oublier que l’utilisation d’approches d’ingénierie de caractéristiques nécessite du temps et une expertise du domaine. Du point de vue ML, ces modèles peuvent être divisés en 3 classes, celle de l’apprentissage supervisé, semi-supervisé, et non supervisé (REN et JI 2019).

2.8.1.1 Apprentissage supervisé

Puisque le problème de détection de spam d’opinion a été considéré comme un problème de classification binaire avec deux classes ”Spam” et ”Non Spam” (si l’entité considérée est l’opinion) (JINDAL et LIU 2008), de nombreux chercheurs ont opté pour des méthodes d’apprentissage supervisé. Ces méthodes entraînent un modèle d’apprentissage automatique à l’aide de données étiquetées pour classer les spams d’opinion. L’entrée du modèle se compose de caractéristiques extraites des données brutes telles que les n-grammes, les sacs de mots, les POS tags et d’autres caractéristiques textuelles et comportementales jugées importantes. Les algorithmes couramment utilisés dans l’apprentissage supervisé comprennent les machines à vecteurs de support (SVM), les arbres de décision, Naïve Bayes et la régression logistique.

Le principal avantage de l’utilisation de ces modèles pour la détection des spams d’opinion est qu’ils sont relativement faciles à mettre en œuvre et peuvent atteindre de bonnes performances avec de petits ensembles de données. De plus, ils permettent l’interprétabilité, ce qui signifie qu’il

est possible de comprendre comment le modèle fait des prédictions et quelles caractéristiques sont les plus importantes. Cependant, nous devons garder à l'esprit que les ensembles de données étiquetés de qualité nécessaires pour entraîner ces modèles sont difficiles à acquérir.

Parmi les multiples approches qui utilisent l'apprentissage supervisé, mentionnons d'abord (OTT et al. 2011), dans ce travail, les auteurs ont construit un ensemble de données de référence qui est maintenant largement utilisé dans la détection du spam d'opinion, et est couramment appelé OpSpam¹⁰. Les auteurs de cet article ont testé à la fois les classificateurs **Naïve Bayes** et **SVM** avec des combinaisons de caractéristiques (identification de genre avec POS, psycholinguistiques avec LIWC, textuelles avec n-grammes, ...). Leurs expérimentations ont montré des résultats qui ont surpassé les résultats obtenus par le jugement humain et les méthodes de la littérature, leur approche la plus performante était la combinaison de LIWC et de bi-grammes avec le classificateur SVM, qui a donné une exactitude de 89,8%.

(SHAHARIAR et al. 2019) ont appliqué plusieurs techniques sur les données acquises de Yelp¹¹, après leur pré-traitement. Ils ont appliqué à la fois les techniques ML traditionnelles (SVC, KNN, NB) et des méthodes d'apprentissage approfondi. Ces modèles ont été alimentés avec des caractéristiques sélectionnées via TF-IDF, N-Grammes et Word2vec. La combinaison la plus performante a été de jumeler Uni-Gram et Naïve Bayes, l'exactitude signalée était de 91,73%.

(RASTOGI et al. 2020) est la première étude de son genre qui considère trois perspectives à la fois, celui de l'opinion, de son auteur et du produit qu'elle cible, afin d'analyser l'efficacité de la détection de spam en utilisant des caractéristiques textuelles et comportementales ainsi que 4 classificateurs (SVM, LR, MLP et NB) sur YelpZIP et YelpNYC (RAYANA et AKOGLU 2015). L'architecture de l'approche adoptée dans cette étude est illustrée dans I.20.

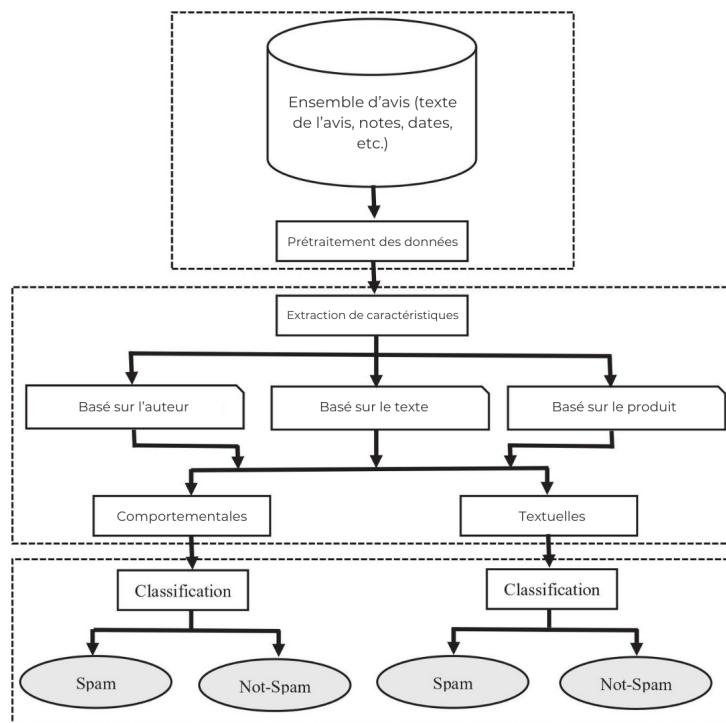


FIG. I.20 : Architecture de l'approche adoptée dans (RASTOGI et al. 2020)

Les résultats indiquent que les caractéristiques comportementales sont plus efficaces que les textuelles pour détecter le spam d'opinion dans les trois configurations avec les 4 classificateurs.

¹⁰<https://myleott.com/op-spam.html>

¹¹<https://www.yelp.com/>

De plus, les modèles formés sur des caractéristiques hybrides surpassent légèrement ceux formés sur des caractéristiques comportementales et il y a des cas où ces derniers font aussi bien voire mieux que les hybrides. Les caractéristiques utilisées dans ce travail apportent une amélioration par rapport aux caractéristiques existantes utilisées dans d’autres travaux connexes. De plus, l’analyse du temps de calcul pour la phase d’extraction des caractéristiques montre une meilleure rentabilité des caractéristiques comportementales par rapport au textuel.

Enfin, (YAO et al. 2021) proposent un modèle d’ensemble combinant 5 classificateurs supervisés (RF, Xgboost, Lightgbm, Catboost, GBDT) et des caractéristiques basées sur l’opinion (Notation (rating), Nombre de mots, TF-IDF) et son auteur (Nombre d’avis, Nombre de premiers avis, note moyenne,...), et les ont testées sur deux sous-ensembles de YelpCHI (MUKHERJEE et al. 2013b) (un pour les restaurants et l’autre pour les hôtels). Les auteurs ont traité le déséquilibre des données en combinant la méthode de recherche de grille et le rééchantillonnage en trouvant le meilleur rapport d’échantillonnage pour chaque classificateur. Ensuite, les caractéristiques extraites sont transmises séparément à chaque classificateur. Enfin, ils ont proposé deux méthodes afin d’améliorer les performances du modèle de classification : une méthode de vote majoritaire et d’empilement. Les résultats expérimentaux ont montré que le modèle proposé fonctionnait assez bien, mais ne surpassait pas les méthodes de pointe. Ils rapportent une exactitude de 72.06% sur le sous-ensemble YelpCHI-Restaurant et de 79.46% sur YelpCHI-Hôtel en suivant la stratégie d’empilement.

2.8.1.2 Apprentissage semi-supervisé

L’apprentissage semi-supervisé est un type d’apprentissage qui combine l’apprentissage supervisé et non supervisé, où un algorithme apprend à partir d’un petit ensemble de données étiquetées et d’un ensemble de données non étiquetées beaucoup plus grand. L’idée est d’exploiter la grande quantité de données non étiquetées pour améliorer les performances du modèle dans l’apprentissage des données étiquetées. Ceci est particulièrement utile lorsque le coût de l’étiquetage des données est élevé ou lorsque les données étiquetées sont rares, c’est donc plus efficace que d’étiqueter toutes les données à la main, mais cela nécessite une sélection minutieuse des exemples à étiqueter et à réapprendre. D’après (LIGTHART et al. 2021a), on peut trouver dans cette catégorie plusieurs méthodes telles que l’apprentissage multi-vues, les modèles génératifs, l’apprentissage basé sur des graphes, ainsi que le **Self-Training** (YAROWSKY 1995), le **Co-Training** (BLUM et MITCHELL 1998) et le (X.-L. LI et LIU 2005 ; LIU et al. 2002) auxquels nous allons nous intéresser.

L’approche **Self-Training** (Auto-apprentissage) consiste à former un classificateur sur une petite quantité de données étiquetées, puis à utiliser le classificateur pour prédire les étiquettes des données non étiquetées. Les prédictions à haute confiance sont ajoutées aux données étiquetées, et le processus est répété de manière itérative jusqu’à ce qu’un critère d’arrêt soit satisfait (toutes les données ont été étiquetées, aucune observation supplémentaire ne satisfait les critères, le nombre maximum d’itérations a été atteint,...). L’approche **Co-Training** (Co-Apprentissage) implique l’entraînement de deux classificateurs sur différents ensembles de caractéristiques et l’utilisation des prédictions de haute confiance de chaque classificateur pour étiqueter les données de l’autre. Cela vient de l’hypothèse que s’il y a deux vues indépendantes du même objet (elles utilisent chacune un ensemble de caractéristiques différent) alors deux classificateurs correctement entraînés sur ces vues doivent étiqueter cet objet de la même manière. Cela peut être efficace lorsque les caractéristiques ont des forces différentes dans des contextes différents. L’approche (Positive-Unlabeled learning), quant à elle, traite d’un ensemble d’échantillons positifs étiquetés (opinions spam) et d’un ensemble d’échantillons non étiquetés (qui peuvent inclure à la fois des avis de spam et non-spam). L’objectif de est d’entraîner un classificateur binaire capable

de classer avec précision les échantillons positifs et de les distinguer des échantillons négatifs, sans utiliser d’exemples étiquetés négatifs. L’idée de base est de différencier entre les instances négatives et les instances positives appartenant à l’ensemble des données non étiquetées, puis utiliser le résultat de cette classification comme un jeu de données annotées pour apprendre un classifieur supervisé. Le est particulièrement utile dans les cas où les échantillons négatifs sont difficiles à identifier ou à collecter, ce qui est souvent le cas dans les tâches de détection de spam ou de fraude. En utilisant cette approche, il est possible d’obtenir des résultats précis en termes de classification des échantillons positifs, sans avoir besoin d’une grande quantité d’exemples négatifs étiquetés.

Il existe plusieurs travaux ayant exploité les techniques semi-supervisées, (LIGTHART et al. 2021a) de leur part en ont étudié plusieurs, à savoir l’auto-apprentissage, le co-apprentissage et le SVM transductif sur le dataset OpSpam. Ils ont aussi testé ces modèles avec différentes combinaisons de caractéristiques dont les Uni-grammes, les Bi-grammes, la combinaison de TF-IDF et des Bi-grammes ainsi que la combinaison de POS et les Uni-grammes. Le meilleur résultat obtenu sur le dataset OpSpam était une exactitude de 93% en combinant les Bi-Grammes avec l’algorithme de **Self-training** (avec Naïve Bayes comme classifieur de base), cette méthode a donné les meilleures performances même durant les tests additionnels sur des sous-ensembles de YelpCHI.

(J. WANG et al. 2020), de leur part, ont proposé un modèle combinant plusieurs caractéristiques (Sémantiques, lexicales, basées sur les sentiments, sur des informations externes telles que le comportement anormal de l’auteur, etc), ils ont même proposé une méthode afin de juger si l’émotion exprimée par l’auteur pourrait être une caractéristique significative. Ils ont jugé que la combinaison donnerait une bonne performance et qu’une combinaison optimale de classificateurs n’est sélectionnée qu’à travers plusieurs expérimentations, pour cela, ils ont utilisé 7 modèles ML (SVM, NB, Random Forest, KNN, LDA, et les arbres de décision) afin d’identifier la crédibilité de l’opinion. En expérimentant sur YelpCHI, l’approche de **Co-training** avec fusion des multiples caractéristiques a donné les meilleurs résultats avec une exactitude de 84.45% et une mesure F-score de 81.89%.

En ce qui est de l’approche , nous pouvons citer le travail de (FUSILIER et al. 2015) qui ont proposé une variante conservatrice de la méthode originale (LIU et al. 2002) utilisant le dataset OpSpam. En ce qui est du classifieur utilisé, selon leurs expérimentations, Naïve Bayes a surpassé SVM en utilisant une combinaison d’Uni-grammes et de Bi-grammes. Ils ont aussi essayé d’analyser le rôle de la polarité pour la détection de spam d’opinion dans un dataset de polarité mixte avec deux configurations : soit en utilisant un seul classificateur pour les spams négatifs et positifs, soit en considérant les spams positifs et négatifs comme deux problèmes différents (donc 2 classificateurs). Les résultats de cette expérimentation montrent que le traitement de la détection de spam positif et négatif comme un seul problème donne de meilleurs résultats, on constate donc que ces deux types de spam ont des points en commun qui aident à améliorer la classification. L’approche adoptée a donné une mesure F-score de 81.1% pour la détection de spams positifs et de 72.3% pour les spams négatifs, ces résultats montrent que cette approche modifiée surpassé le travail original et que la détection du spam négatif est plus difficile que la détection du spam positif, mais qu’il reste préférable d’avoir un seul classificateur pour analyser les deux types d’opinions que d’utiliser deux classificateurs distincts.

Plus récemment, (TIAN et al. 2020) ont proposé une approche robuste non-convexe semi-supervisée sous le nom de “Ramp One-Class SVM”. Elle est nommée comme cela car elle combine SVM à une classe et la fonction Ramp. **One-Class SVM** est une technique d’apprentissage non supervisée pour apprendre à différencier les échantillons de test d’une classe particulière d’autres classes, ils l’ont appliqué afin de gérer le manque de données étiquetées pour les opinions spams.

En ce qui concerne la fonction de perte **Ramp**, il s’agit d’une fonction de perte non convexe, et ils exploitent cela afin d’éliminer les effets des valeurs aberrantes et des non-avis. Ils ont obtenu une exactitude de 74.37% sur Yelp et de 92.13% sur OpSpam.

Nous avons déjà mentionné le travail de (ZIANI et al. 2021) en ce qui est des caractéristiques sémantiques qu’ils ont proposées pour la langue arabe. En combinant ces caractéristiques avec d’autres statiques, ils ont appliqué un modèle de classification SVM semi-supervisé, ou S3VM (BENNETT et DEMIRIZ 1998), sur plusieurs datasets dont une version d’OpSpam traduite en arabe, pour lequel ils ont obtenu une exactitude de 93%.

2.8.1.3 Apprentissage non-supervisé

Comme nous pouvons le conclure de ce qui a été présenté jusqu’à présent, l’un des problèmes majeurs que rencontrent les méthodes supervisées et semi-supervisées est le fait qu’elles dépendent des datasets annotés pendant la phase d’apprentissage, et que la qualité desdits datasets peut affecter leur performance. Ainsi, de nombreux chercheurs ont décidé d’aborder le problème de détection de spam d’opinion sous l’angle de classification non supervisée. Ces modèles peuvent être utiles lorsque les données étiquetées sont rares ou coûteuses à obtenir, mais ils peuvent être moins efficaces que les méthodes supervisées, car ils n’ont pas accès à des données étiquetées sur lesquelles s’entraîner et ils nécessitent souvent plus de connaissances et d’affinage pour obtenir de bonnes performances.

Parmi les approches non supervisées, nous pouvons citer ‘**RLOSD**’, acronyme de “Representation Learning based Opinion Spam Detection” (SEDIGHI et al. 2017) qui propose une approche basée sur les arbres de décision ainsi qu’un apprentissage de représentation non supervisé. Pour la sélection des caractéristiques, ils utilisent des méthodes traditionnelles dont POS, TF-IDF et N-grammes, PCA et MMI. Le modèle a obtenu de bonnes performances durant les expérimentations sur Yelp et le dataset construit par (Jiwei LI et al. 2014) avec une mesure F-score toujours supérieure ou égale à 75%. Et selon (MOHAWESH et al. 2021), ce modèle proposé peut être amélioré en prenant en compte la corrélation des données dans le choix des caractéristiques appropriées.

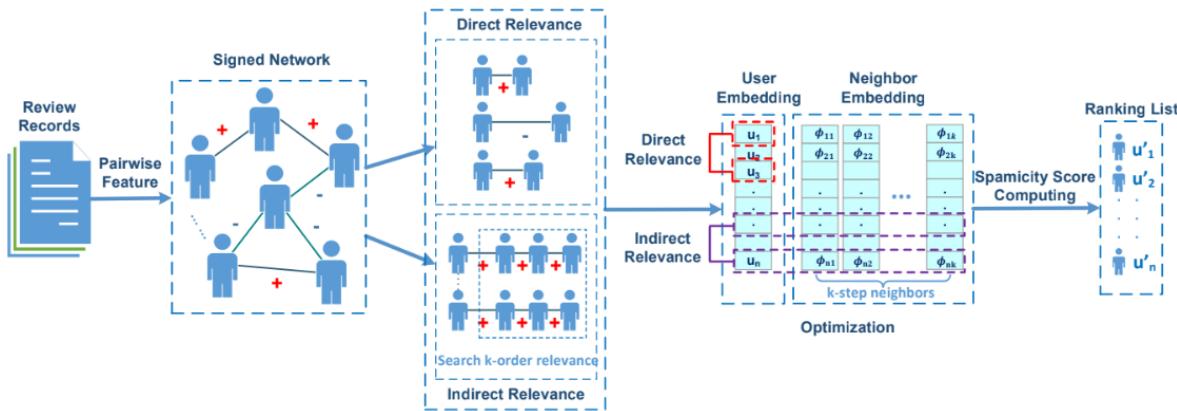


FIG. I.21 : Architecture de l’approche adoptée dans (Z. WANG et al. 2020)

L’étude de (Z. WANG et al. 2020) présente une approche basée sur l’intégration de réseau non supervisée pour apprendre les intégrations d’utilisateurs en exploitant conjointement les relations directes et indirectes entre les utilisateurs par paires. La figure I.21 montre l’approche proposée au sein de l’architecture unifiée proposée nommée COSD, son objectif est de combiner conjointement l’exploration de voisinage directe et indirecte pour apprendre la représentation intégrée de chaque utilisateur pour identifier plus précisément les spameurs. L’intégration de pertinence directe contrôle l’apprentissage des intégrations d’utilisateurs vers les utilisateurs

par paires avec une forte intensité des caractéristiques collusives. L’intégration indirecte de la pertinence, de son côté, a tendance à rapprocher les utilisateurs par paires partageant plus fréquemment les voisins de co-évaluation. Les auteurs de ce papier ont opté pour l’utilisation des deux parce que ces deux types se renforcent mutuellement pour rapprocher les utilisateurs concernés des relations directes et indirectes des utilisateurs. Les auteurs de ce papier combinent 4 types différents de caractéristiques hétérogènes par paire (la proximité de la notation du produit, la proximité temporelle du produit, la catégorie cote proximité, la catégorie proximité horaire). Les résultats des expérimentations, évalués sur deux jeux de données AmazonCn et YelpHotel, montrent la supériorité de COSD comparée à d’autres méthodes de la littérature. On peut remarquer une amélioration de l’AUC de 12.04% sur AmazonCn et de 12.78% sur YelpHotel.

Le tableau 2.3 de mon rapport de master (SLIMANI 2023) résume quelques travaux présents dans la littérature en mettant en valeur, en addition à l’approche de classification utilisée, les caractéristiques représentées, le dataset exploité et certains résultats obtenus.

2.8.2 Modèles de Deep Learning

Les modèles d’apprentissage en profondeur sont principalement conçus pour apprendre automatiquement les caractéristiques les plus informatives directement à partir des données brutes, éliminant ainsi le besoin d’ingénierie manuelle des caractéristiques. Ceci est réalisé grâce à l’utilisation de plusieurs couches de transformations non linéaires, la première consistant généralement en une technique d’apprentissage de caractéristiques comme mentionné précédemment, ce qui permet au modèle de capturer des relations complexes dans les données plus efficacement que les modèles traditionnels. Cependant, il convient de mentionner que ces techniques nécessitent généralement de grandes quantités de données et des ressources importantes. Des exemples de tels modèles incluent CNN, RNN, LSTM, et GAN (MOHAWESH et al. 2021 ; REN et JI 2019). Ces modèles ont atteint de bonnes performances sur une variété de tâches de traitement du langage naturel, y compris l’analyse des sentiments et la détection des spams d’opinion.

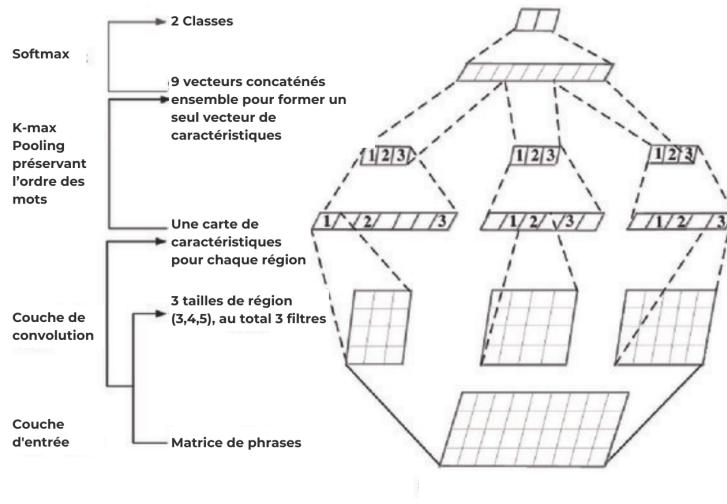


FIG. I.22 : Architecture générale de OpCNN (ZHAO et al. 2018)

De nombreux travaux ont utilisé CNN, citons par exemple le modèle **OpCNN** introduit par (ZHAO et al. 2018), qui est un modèle qui prend en considération les caractéristiques de l’ordre des mots pendant le processus d’analyse. Ce modèle CNN est composé de quatre couches, une

couche d'entrée, une couche de convolution, une couche de regroupement et une couche de sortie comme l'illustre la figure I.22. Le modèle prend en entrée les avis avec un certain ordre de mots, puis dans les couches de convolution et de regroupement, l'ordre consécutif des mots est conservé en appliquant la méthode de pooling “word order-preserving k-max” qu'ils ont conçue. Afin d'évaluer leur modèle, ils ont construit un corpus contenant 24,166 avis sur des hôtels collectés du site chinois dianping.com puis les ont annoté en suivant la méthode de (F. H. LI et al. 2011), 4,132 de ces avis ont été annotés comme Spam. Les auteurs ont déclaré une exactitude de 70.02% surpassant les performances d'autres méthodes de l'état de l'art (TF-IDF + SVM, Bigrammes + SVM, CNN). Ils ont aussi testé sa capacité de généralisation sur le dataset OpSpam, OpCNN a obtenu de meilleurs résultats que CNN avec une exactitude de 84.50% et une mesure F-score de 82.84%.

Les **Réseaux de neurones récurrents** (RNN) peuvent utiliser les informations dans des séquences de longueurs arbitraires, mais c'est surtout en théorie. En pratique, un RNN standard se limite à ne regarder que quelques étapes en arrière en raison du problème de disparition du gradient ou de gradient explosif. Les chercheurs ont développé des types de RNN plus sophistiqués pour éviter ces défauts tels que la mémoire à long court terme (LSTM), l'unité récurrente fermée (GRU)(CHO et al. 2014), le LSTM bidirectionnel, le LSTM empilé et le LSTM avec méthode d'attention (MOHAWESH et al. 2021). **LSTM**, ou Long Short-Term Memory (HOCHREITER et SCHMIDHUBER 1997), est un type de RNN qui peut apprendre des dépendances à long terme et, contrairement à un RNN traditionnel, il dispose d'un mécanisme de déclenchement qui lui permet d'oublier ou de conserver de manière sélective les informations des étapes de temps précédentes, il a été introduit afin de résoudre le problème de disparition du gradient que nous avons mentionné ci-dessus. L'architecture de base d'un LSTM comprend une cellule mémoire (qui stocke les informations de l'instant précédent), une porte d'entrée (qui détermine quelles informations doivent être stockées dans la cellule mémoire à l'instant actuel), une porte de sortie (qui décide quelles informations de la cellule mémoire doivent être transmises à la couche ou l'instant suivant) et la porte d'oubli (qui décide quelles informations doivent être supprimées de la cellule mémoire).**Bi-LSTM**, d'autre part, est l'abréviation de LSTM bidirectionnel, et c'est un type de LSTM qui traite les entrées dans les sens avant et arrière (SCHUSTER et PALIWAL 1997). Il se compose de deux LSTM, l'un traitant l'entrée vers l'avant et l'autre vers l'arrière. La sortie des deux LSTM est concaténée pour former la sortie finale.

(ZENG et al. 2019) ont proposé une approche qui applique un modèle Bi-LSTM afin de détecter les faux avis en fonction de la structure de l'avis. La méthode d'ensemble proposée, **RSBE**, est composée de quatre encodeurs LSTM bidirectionnels distincts, qui encodent la première phrase de l'examen, le contexte intermédiaire, la dernière phrase et le texte entier en quatre représentations de document, ainsi que deux couches de mécanismes d'attention qui intègrent les 4 représentations précédentes en une représentation finale qui est ensuite transmise à une fonction softmax comme illustré dans la figure I.23. L'intérêt pour la structure des critiques est venu de leurs conclusions selon lesquelles les fausses critiques exprimaient des émotions plus fortes que les véritables critiques, que ces émotions étaient principalement présentes dans les premières et dernières phrases et que les fausses critiques commençaient ou se terminaient par des phrases similaires. Ils ont évalué leur modèle à l'aide du corpus construit par (Jiwei LI et al. 2014) et de Word embeddings pré-entraînés sur le corpus Wikipédia. Le modèle proposé a obtenu de meilleurs résultats que d'autres méthodes de l'état de l'art dans un domaine (hôtel, médecin et restaurant) avec une précision de 85,7%, 84,7% et 85,5%, respectivement. De plus, dans le domaine mixte, il atteint une précision de 83,4%. Cependant, le modèle proposé n'a pas réussi à obtenir de bons résultats dans les domaines croisés avec seulement 71,6%.

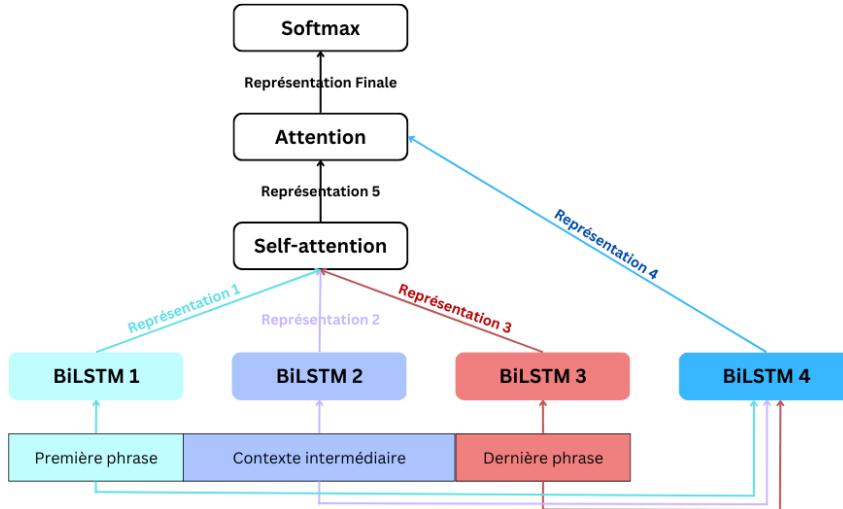


FIG. I.23 : Architecture générale de l'approche RSBE de (ZENG et al. 2019)

En ce qui est des simples LSTMs, (SHAHARIAR et al. 2019), dont nous avons présenté les méthodes traditionnelles utilisées précédemment, ont aussi expérimenté avec des classificateurs neuronaux, à savoir CNN et LSTM, et cela, avec des embeddings Word2vec. Les meilleurs résultats ont été obtenus avec LSTM avec une exactitude de 94.565% sur le corpus OpSpam et 96.75% sur Yelp. Plus récemment et toujours concernant les LSTMs, les auteurs de (HARRIS 2022) ont proposé une approche combinant les caractéristiques linguistiques et comportementales avec une architecture exploitant LSTM. Ils entraînent d'abord trois modèles LSTM : *LSTM1* utilisant 12 caractéristiques linguistiques (dont le nombre de mots, de verbes, et de mots spatio-temporels, etc), *LSTM2* utilisant 6 caractéristiques comportementales (MRD, BST,...) et *LSTM3* utilisant l'approche K-L Divergence, qui est une mesure de l'information perdue lorsqu'une distribution est estimée par une autre. Cette séparation leur a permis de configurer et d'évaluer chaque modèle LSTM séparément. Ces modèles sont ensuite combinés dans un modèle d'ensemble, comme le montre la figure I.24, avec deux couches d'auto-attention présentes, l'auto-attention étant un type de mécanisme d'attention qui peut extraire les dépendances positionnelles (sa sortie est la moyenne pondérée des différentes positions dans la séquence d'entrée) tout en nécessitant moins de paramètres et à une complexité de calcul inférieure. Afin d'évaluer leur approche, ils ont choisi les avis de YelpZIP portants sur les restaurants. Et étant donné que les données authentiques sont plus fréquentes que les fausses (l'ensemble de données est déséquilibré) ils ont utilisé une technique largement appliquée pour suréchantillonner la classe minoritaire, appelée SMOTE (CHAWLA et al. 2002), séparément pour les modèles linguistiques et comportementaux.

En ce qui est des résultats de l'étude (HARRIS 2022), l'évaluation des 12 caractéristiques linguistiques montre que les avis signalés comme spam ont moins de diversité lexicale et de contenu, utilisent moins de mots spatio-temporels, ont une longueur de mot/terme légèrement plus courte et utilisent globalement moins de clauses. L'évaluation des 6 caractéristiques comportementales, quant à elle, montre que les fenêtres d'activité des spameurs ont tendance à être beaucoup plus courtes, utilisent des contenus plus similaires, sont plus extrêmes et utilisent davantage les majuscules. Ils ont trouvé que les caractéristiques comportementales ont de meilleurs résultats que les linguistiques et que le meilleur résultat a été observé lors de la combinaison de modèles, ce modèle a surpassé chacune des 3 approches indépendantes et d'autres approches de l'état de l'art basées sur la divergence K-L appliquées sur le même ensemble de données avec une AUC de 89.30%.

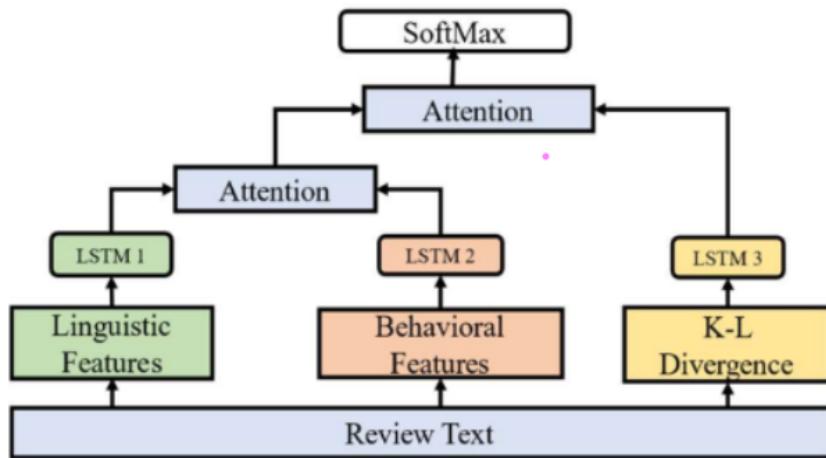


FIG. I.24 : Architecture générale de l'approche de (HARRIS 2022)

Passons maintenant à **GAN**, qui est l'abréviation de "Generative Adversarial Networks". Il s'agit d'un type de modèle d'apprentissage profond qui se compose de deux réseaux de neurones, un générateur et un discriminateur, qui sont entraînés ensemble pour générer de nouvelles données similaires aux données d'entraînement. Le générateur prend un bruit d'entrée aléatoire et produit un échantillon qui ressemble aux données d'apprentissage, tandis que le discriminateur prend à la fois des échantillons réels et générés et essaie de les distinguer. Le générateur est formé pour produire des échantillons qui peuvent tromper le discriminateur, tandis que le discriminateur est formé pour identifier correctement les vrais échantillons à partir de ceux générés. Au fur et à mesure que les deux réseaux sont entraînés ensemble, ils se font concurrence et s'améliorent dans leurs tâches respectives. Finalement, le générateur est capable de produire des échantillons très similaires aux données d'apprentissage, et le discriminateur n'est pas capable de les distinguer des données réelles.

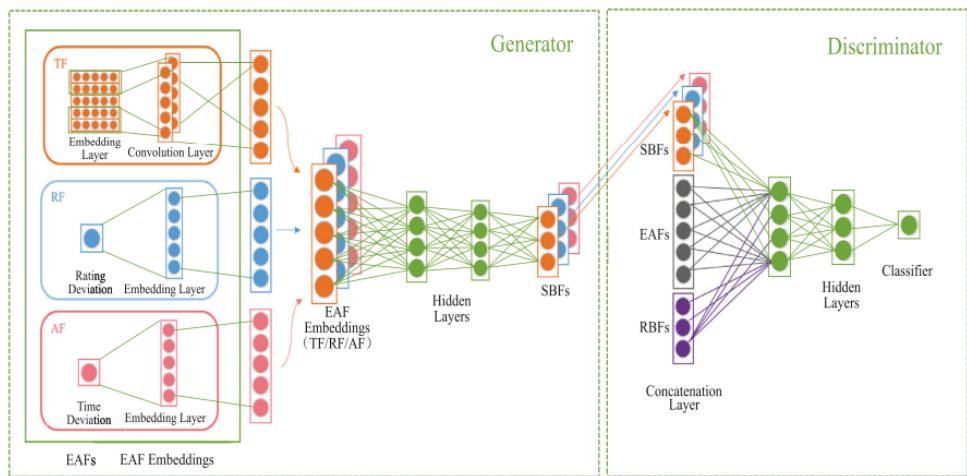


FIG. I.25 : Architecture générale de l'approche de (TANG et al. 2020)

(TANG et al. 2020) se sont intéressés au manque de caractéristiques comportementales efficaces pour les nouveaux utilisateurs qui ne publient qu'un seul avis. Afin de gérer le problème des démarrages à froid (ou des spammeurs singleton), ils ont exploité le réseau GAN, l'idée de base est de générer des caractéristiques de comportement synthétique (SBFs) pour les nouveaux utilisateurs à partir de leurs caractéristiques facilement accessibles (EAFs). D'abord, 6 caractéristiques réelles (RBFs) ont été extraites pour les utilisateurs réguliers existants (Activity Window - Maximum Number of Reviews - Percentage of Positive Reviews - Review Count - Reviewer Deviation - Maximum Content Similarity) ainsi que 3 caractéristiques facilement

accessibles pour les nouveaux utilisateurs : des caractéristiques textuelles (TF) via des emmbeddings à 100 dimensions pré-formés à l'aide de CNN (puisque l'évaluateur doit publier au moins un avis), des caractéristiques d'évaluation (RF) via un vecteur à 100 dimensions représentant l'écart de note (puisque l'évaluateur doit évaluer au moins un produit) et des caractéristiques d'aspect (AF) via un vecteur à 100 dimensions représentant la déviation temporelle (puisque l'utilisateur doit avoir des horodatages d'inscription et de publication). En utilisant ces caractéristiques, ils entraînent le modèle GAN, y compris un *générateur* utilisé pour apprendre le mappage des EAFs d'entrée aux SBFs sous la contrainte de maintenir les SBFs proches des RBFs et un *discriminateur* qui consomme les SBFs du générateur, ainsi que les EAFs et les RBFs à partir des données de formation en entrée et vise à distinguer les caractéristiques de comportement synthétiques et réelles. L'architecture globale de cette approche peut être visualisée dans la figure I.25. Le modèle proposé a été évalué sur deux sous-ensembles de YelpCHI, Hôtel et Restaurant, et a surpassé certaines méthodes de pointe avec une exactitude de 83% sur le domaine hôtelier et 75,7% sur le domaine restaurant. Les caractéristiques combinées ont amélioré les performances du modèle de classification.

À la suite de cela, (KENNEDY et al. 2020) ont analysé plusieurs méthodes neuronales et traditionnelles afin de distinguer entre les avis véridiques et spam ainsi que différentes caractéristiques. Ils ont expérimenté sur Yelp et OpSpam, et ont proposé les caractéristiques suivantes : Bag-Of-Words, Word2vec pré-entraîné avec une dimensionnalité de 300 sur un dataset de Google Actualités¹² ainsi que des caractéristiques structurelles (Longueur de l'avis, longueur moyenne des mots et des phrases, pourcentage de mots en majuscules et pourcentage de chiffres), des caractéristiques comportementales (nombre maximal d'avis en une journée, durée moyenne des avis, écart type des notes et pourcentage de notes positives et négatives), des pourcentages Part-Of-Speech et des pourcentages de polarité pour chaque mot. Cependant, après une sélection de caractéristiques à l'aide de la régression logistique, il s'est avéré que ces deux dernières caractéristiques n'étaient pas prédictives et que de meilleures performances peuvent être obtenues en combinant les caractéristiques comportementales et BOW. À noter que les caractéristiques comportementales ne peuvent être exploitées qu'avec le dataset Yelp, et que pour la combinaison de ces caractéristiques avec BOW, elles sont directement concaténées. En ce qui est des modèles utilisés, on trouve FFNNs (Réseaux de neurones Feed-Forward), CNNs, LSTMs ainsi qu'une version affinée de BERT (bert-base-uncased). Les auteurs précisent dans leur article les spécifications techniques de chacun de ces modèles pour chacun des datasets et pour chaque ensemble de caractéristiques. En comparant les performances des classificateurs FFNN, CNN et LSTM, les meilleures performances ont été données par FFNN et BOW sur OpSpam (une exactitude de 88,8%), suivis de LSTM et BOW (exactitude = 87,6%). En revanche, sur Yelp, LSTM+BOW et CNN+Word2vec ont donné les meilleurs résultats avec 73,1%. Les meilleures performances sur l'ensemble de données OpSpam, qui sont compétitifs par rapport aux techniques de l'état de l'art, sont obtenues en affinant BERT avec une précision de 90,5% (avec l'implémentation TensorFlow).

La dernière étude que nous allons présenter est l'approche appelée **EUPHORIA**, “nEural mUlti-view aPproach fOr RevIew spAm” (ANDRESINI et al. 2022) qui est une approche qui utilise l'apprentissage multi-vues en exploitant plusieurs vues (chevauchées) des données afin que le modèle de classification ne manque aucun aspect important. Dans cette méthode, chaque vue de données est représentée par un vecteur de caractéristiques distinct qui est ensuite traité par le réseau neuronal à entrées multiples. Actuellement, le modèle prend en charge trois vues de données différentes et en accepte d'autres. Deux de ces vues sont liées aux caractéristiques textuelles et utilisent Word2Vec (CBOW) pour une vue spécifique au domaine du texte de

¹²<https://code.google.com/archive/p/word2vec/>

révision, et BERT pour une vue plus générale. La troisième vue, quant à elle, comporte 6 caractéristiques comportementales qui représentent le profil de l'examinateur au fil du temps (nombre maximal d'avis par jour (MRD), rapport positif (PR), durée moyenne de l'examen (ARL), écart de l'examinateur (RD), similarité moyenne de l'examen (ARS), similarité maximale des avis (MRS)).

Comme l'illustre la figure I.26, après l'extraction des caractéristiques, les vecteurs d'intégration et les caractéristiques comportementales sont transmis à un réseau de neurones à entrées multiples. Comme indiqué, la première couche du réseau est constituée de 3 couches distinctes entièrement connectées avec dropout, et crée une vue compressée et spécifique à la tâche de chaque vecteur de caractéristiques. Les trois vues sont concaténées ensemble, puis transmises par la couche de classification. L'utilisation d'un réseau de neurones à entrées multiples s'est avérée, après les expérimentations présentées, aider réellement à bénéficier de la richesse des données multi-vues par rapport au réseau de neurones à entrée unique dont les performances sont dégradées par les effets de la malédiction de la dimensionnalité.

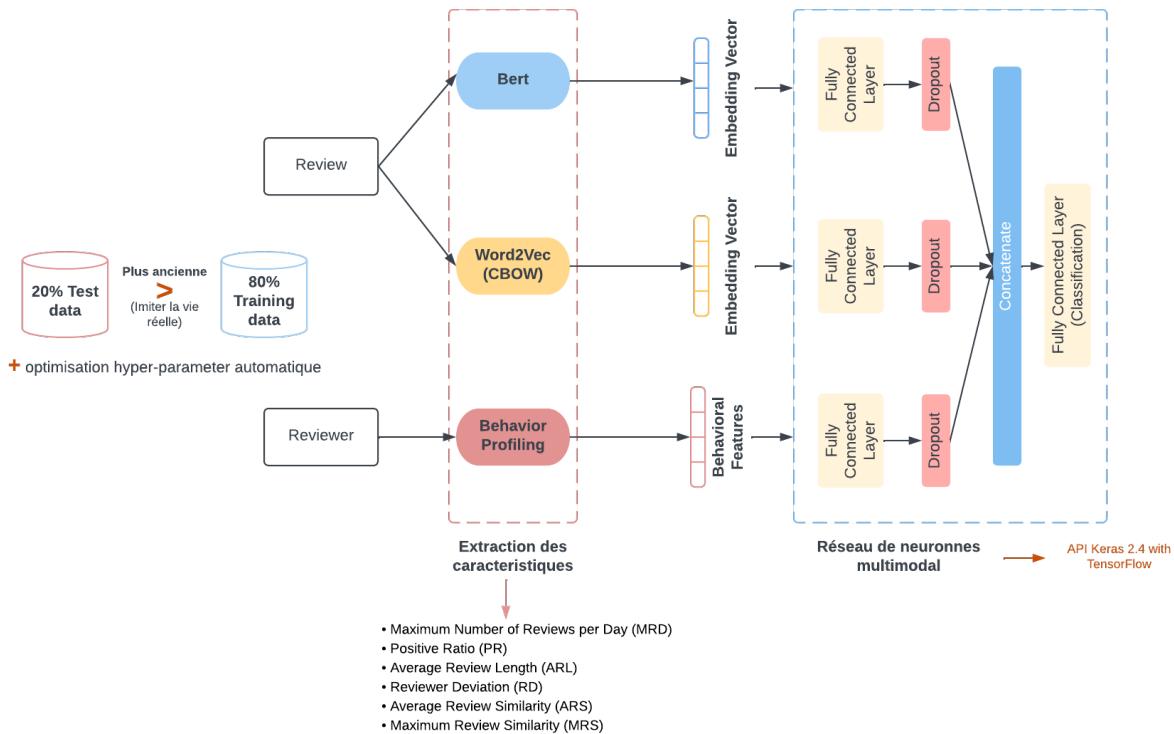


FIG. I.26 : Architecture générale de l'approche de (ANDRESINI et al. 2022)

Les données utilisées dans cette étude ont été extraites de deux sous ensembles de YelpCHI, à savoir Hôtel and Restaurant, contenant des avis sur 85 hôtels et 130 restaurants, respectivement, dans la région de Chicago. Le cadre expérimental adopté dans cette étude est différent de celui couramment adopté dans la littérature, la détection de spam, où les divisions des datasets Entraînement-Test sont générées de manière aléatoire, négligeant ainsi la date/heure de publication des avis. Afin d'adapter les expériences à un scénario réaliste où le modèle ne serait pas en mesure d'utiliser les informations produites après la rédaction de l'avis, dans les ensembles de validation et de test, les valeurs des caractéristiques comportementales pour chaque opinion sont mises à jour en fonction de l'ordre dans lequel ils ont été écrits. Une autre particularité de la méthode proposée est que même si les deux ensembles de données sont classifiés séparément, car tout domaine a ses caractéristiques spécifiques, le classifieur considère également les connaissances hors domaine en tenant compte des critiques écrites par le même auteur concernant d'autres domaines, plus tard leurs expériences ont montré que c'était un pas dans la bonne

voie car l'acquisition de plus de connaissances a reflété un gain de performances.

Afin de savoir dans quelle mesure chaque point de vue individuel influence la précision du modèle de classification, les auteurs ont analysé comment les connaissances contenues dans les vues textuelles et de comportement peuvent influencer les performances d'EUPHORIA. Les résultats ont montré que les caractéristiques comportementales transmettent les informations les plus pertinentes nécessaires pour détecter les faux avis surpassant à la fois Word2Vec et Bert et leur combinaison également. D'autre part, il a été démontré que les caractéristiques de contenu extraites conjointement par Word2Vec et BERT donnaient de meilleurs résultats que de les traiter séparément. Le principal résultat de cette analyse a montré que la combinaison des trois vues surpassait toutes les vues précédentes et, par conséquent, Euphoria a obtenu la meilleure performance avec une AUC de 70,8% sur le corpus de restaurants et de 81,3% sur le corpus d'hôtels. Euphoria a aussi eu de meilleurs résultats par rapport à deux algorithmes de l'état de l'art basés sur SVM, dans le premier, les vecteurs de caractéristiques ont d'abord été concaténés en un seul vecteur d'entrée, puis traités comme par le classificateur SVM tandis que dans le second (noté Ens-SVM) un ensemble de trois SVM distincts sont formés séparément des trois vecteurs de caractéristiques, puis la règle de majorité d'ensemble est utilisée pour la classification finale. Les résultats ont montré que l'apprentissage de classificateurs séparés pour chaque vue est moins performant que l'apprentissage d'un seul classificateur en concaténant toutes les vues ensemble. Cela ne pose pas de problème en ce qui concerne Euphoria en raison de la rétro-propagation dans le réseau neuronal à entrées multiples, contrairement à Ens-SVM où les informations ne peuvent pas être partagées dans ce cadre.

Le tableau 2.4 de mon rapport de master (SLIMANI 2023) résume certains travaux exploitant les modèles Deep Learning pour le problème de détection de spam d'opinion, dont certaines que nous avons expliquées précédemment et d'autres non.

2.9 Défis et discussions

Comme présenté précédemment, les études antérieures ont construit des bases solides pour résoudre le problème de détection d'opinions frauduleuses. Cependant, il reste de nombreux défis à surmonter dans ce domaine et comme l'a souligné (MEEL et VISHWAKARMA 2020), ces challenges peuvent être considérés comme de nouvelles opportunités de recherche. Dans ce qui suit, nous présenterons certains de ces défis selon la phase du processus affectée :

2.9.1 L'acquisition des données :

D'après la littérature et ce que nous avons discuté, on constate que la plupart des datasets utilisés dans les travaux dans ce domaine sont créés de manière synthétique puisque l'acquisition d'une large quantité de données non étiquetées est facile, cependant la collecte de données étiquetées est difficile et coûteuse (MA et F. LI 2012). Cependant, ces données ne sont pas nécessairement représentatives du spam d'opinion du monde réel. D'où, l'établissement de datasets de référence de qualité est une tache de haute importance afin d'assurer la qualité des modèles de détection de spam entraînés sur ces données (MEEL et VISHWAKARMA 2020).

En plus de cela, il peut s'avérer intéressant de comparer les opinions et les auteurs sur plusieurs sites. Comme l'indique (LIU 2020), cette comparaison pourrait nous aider à découvrir des anomalies telles que des avis similaires publiés sur plusieurs plates-formes, à peu près au même moment, avec des adresses IP et des ID utilisateur similaires, un avis comme celui-ci est peu susceptible d'être authentique par exemple, pourquoi un utilisateur prendrait-il la peine de publier le même avis positif sur un produit sur plusieurs sites ? ne serait-il pas embêtant de s'inscrire sur plusieurs plateformes pour publier cette critique si ce n'est pour un certain

gain ? Cependant, la collecte de toutes ces données à travers plusieurs sites, leur analyse et leur comparaison peut être difficile et coûteux, surtout avec la quantité immense de données présentes sur les réseaux qui ne cesse d'augmenter de jour en jour.

Le dernier point que nous voudrions mentionner concernant les défis en matière de données spam serait leur détection. Comme les modèles mettent un certain temps à se former dans les comportements des spameurs et les spams textuels, il est difficile de détecter les avis trompeurs rapidement, si nous attendons trop longtemps, il pourrait être trop tard, car le mal serait déjà fait (LIU 2020). C'est pourquoi que les algorithmes de détection doivent être conçus pour réduire le temps nécessaire pour identifier efficacement le spam.

2.9.2 Les caractéristiques :

D'autant plus, du point de vue des caractéristiques (features), déterminer la meilleure combinaison de caractéristiques décrivant au mieux un contenu spam ou l'attitude d'un spameur est aussi un défi majeur dans la détection d'opinions frauduleuses. La difficulté de choisir ces caractéristiques est qu'il est également difficile pour les humains de distinguer si un avis est un spam ou non vu que les spameurs fournissent des efforts considérables pour que leur style de rédaction ressemble le plus possible à celles d'un être humain qui a vécu l'expérience avec le produit. Nous pouvons aussi remarquer que la plupart des travaux réalisés se basent sur les caractéristiques textuelles et non pas comportementales, or certaines études montrent qu'une combinaison des deux donnent de meilleurs résultats (ANDRESINI et al. 2022).

Un autre défi critique qui pourrait entraver la pertinence des caractéristiques, en particulier les caractéristiques comportementales, est l'existence des opinions spam singletons (ALIARAB et FOULADI 2022). Une opinion spam singleton est un commentaire dont l'auteur n'en a publié qu'un seul, il n'a donc pas laissé beaucoup d'informations utiles pour modéliser. Dans la plupart des études, ce type de commentaire est supprimé de l'ensemble de données, seuls quelques-uns d'entre eux ont essayé de résoudre ce problème, comme (SHAALAN et al. 2021) par exemple qui ont utilisé des schémas d'aspect-sentiment temporels anormaux profonds et des auto-encodeurs. Sans oublier le problème de dimensionnalité qui, comme l'affirme (ANDRESINI et al. 2022), détériore les performances des modèles d'apprentissage avec chaque caractéristique considérée.

2.9.3 Les modèles de classification :

D'après (MEEL et VISHWAKARMA 2020), les travaux présents dans la littérature sont principalement concentrés sur les techniques d'apprentissage supervisé, or ces dernières sont dépendantes des ensembles de données étiquetées et, comme nous l'avons discuté dans les points précédents, peuvent être difficile, voire impossible, à acquérir. Pour cela, il serait plus intéressant de développer des approches d'apprentissage non supervisé.

N'oublions pas que puisque la polarité d'une opinion peut être une caractéristique utile à exploiter dans le processus de détection de spam, tout défi rencontré durant l'analyse des sentiments (Section 1.8) est un défi pour la détection de spam d'opinion.

2.10 Conclusion

Au cours de ce chapitre, nous avons exploré les différents travaux de l'état de l'art traitants le problème de l'identification de la crédibilité d'une opinion. Nous avons d'abord commencé par définir quelques généralités sur le domaine pour une meilleure compréhension. Puis, nous avons présenté le processus de détection de spam d'opinion et ses différentes étapes en passant par les différents jeux de données disponibles et les métriques d'évaluation utilisées. Enfin, nous avons évoqué les différents défis présents et que les prochaines contributions doivent relever afin d'assurer la crédibilité de l'opinion.

Partie II

Contribution

Chapitre 1

État des lieux

1.1 Introduction

Au début de notre projet, nous avons proposé à la Direction Marketing Opérationnel de Brandt, une filiale de Cevital, l'idée d'un tableau de bord d'e-réputation capable de surveiller et d'analyser les avis, les commentaires et les opinions des clients en ligne. Nous avons également voulu donner une dimension de recherche à notre projet en travaillant sous l'encadrement de madame Said EL Hadj Lynda, maître de conférence au Laboratoire de la Communication dans les Systèmes Informatiques (LCSI) de l'ESI. Ainsi, notre projet de fin d'études a été conçu comme une combinaison d'application pratique pour répondre aux besoins spécifiques de Brandt et de recherche pour étudier les défis et les avancées dans les domaines de l'e-réputation, de l'analyse de sentiments et de la détection de spam.

Dans cette partie, nous mettrons en avant les deux organisations dans lesquelles nous avons effectué notre stage : la Direction Marketing Opérationnel de Brandt et le Laboratoire de la Communication dans les Systèmes Informatiques (LCSI) de notre école.

1.2 Présentation de l'organisme d'accueil

1.2.1 Laboratoire LCSI

Le Laboratoire de Communication dans les Systèmes Informatiques (LCSI)¹ est un laboratoire de recherche affilié à l'Ecole Nationale Supérieure d'Informatique (ESI) à Alger, en Algérie. Agréé depuis 2010 par l'Arrêté ministériel n° 93 du 25 mars 2010, le laboratoire se situe à l'adresse suivante : E.S.I (ex.INI), BP 68 M Oued-Smar, 16309, Alger, Algérie.

Le LCSI se concentre sur divers domaines de recherche, notamment la conception de structures de données adaptées aux environnements distribués et mobiles, les architectures avancées des Systèmes de Gestion de Base de Données (SGBD) de nouvelle génération, les processus intelligents, l'intelligence artificielle, la robotique et la perception. Le laboratoire s'intéresse également aux milieux poreux et composites, au traitement d'images, à l'apprentissage automatique (machine learning), à l'analyse de données, aux processus aléatoires et à la théorie des nombres. Grâce à ses recherches innovantes et à ses avancées dans ces différents domaines, le LCSI contribue activement au développement des technologies de l'information et de la communication en Algérie et au-delà.

¹<https://lcsi.esi.dz/>

1.2.2 Groupe Cevital

1.2.2.1 a. Histoire du Groupe Cevital

Le Groupe Cevital² est un conglomérat algérien de premier plan, créé en 1998 par l'entrepreneur Issad Rebrab, et opérant dans les domaines de l'industrie agroalimentaire, de la grande distribution, de l'industrie et des services. Sous la forme juridique d'une Société par Actions (SPA), ce groupe familial est actuellement dirigé par Malik Rebrab, qui occupe le poste de Président-Directeur Général (PDG). Le siège social est situé à Îlot D, N° 6 Zhun Garidi II, Kouba 16005 Alger. La figure II.1 illustre le logo du groupe.



FIG. II.1 : Logo du groupe Cevital.

Avec une dynamique de croissance soutenue depuis sa création, Cevital compte actuellement **26** filiales telles que Mediterranean Float Glass, Brandt et Evcon, réparties en **4** pôles d'activités majeurs : l'Agro-Industrie et la Distribution, l'Automotive, l'Immobilier et les Services, ainsi que l'Industrie.

Fort d'un effectif de **18 000** employés répartis sur **3** continents, le Groupe a enregistré un chiffre d'affaires de **4 milliards** de dollars en 2015, confirmant ainsi sa position de **premier** groupe privé algérien et sa présence à l'international. En effet, Cevital se distingue comme la **troisième** entreprise du pays en termes de chiffre d'affaires et demeure le leader du secteur agroalimentaire en Afrique.

1.2.2.2 b. Les Activités et la Vision de Cevital

Le Groupe Cevital s'est construit autour d'une vision ambitieuse de son fondateur, Issad Rebrab : bâtir un ensemble économique d'envergure mondiale, tourné vers l'exportation et l'international. Au fil des investissements, Cevital s'est diversifié dans divers secteurs, allant de l'industrie agroalimentaire à la grande distribution, en passant par l'électronique, l'électroménager, la sidérurgie, l'industrie du verre plat, la construction industrielle, l'automobile, les services, et les médias. Le Groupe possède des unités de production de taille mondiale, équipées des technologies les plus évoluées, et s'appuie sur des compétences en Recherche et Développement, innovation et talent de ses collaborateurs.

Selon monsieur Rebrab, le succès de Cevital repose sur sept points clés : le réinvestissement systématique des gains dans des secteurs porteurs à forte valeur ajoutée, l'utilisation des technologies les plus évoluées, l'attention portée au choix et à la formation des collaborateurs, l'esprit d'entreprise, la recherche de l'excellence, le sens de l'innovation et la fierté de servir l'économie nationale.

1.2.2.3 c. Structure du Groupe Cevital

Le Groupe Cevital est organisé en quatre grandes divisions :

- **Automotive, Immobilier et Services** : Fondée en 2012, cette division regroupe quatre filiales actives dans les domaines de l'automobile, de l'immobilier, de la location de voitures et de l'affichage publicitaire. (Hyundai, Cevital-MTP, Futur Media,...)

²<https://www.cevital.com/>

- **Agro-industrie et Distribution :** Crée en 2015 et employant plus de 6000 personnes, cette division se consacre à la transformation et la distribution alimentaires à travers la chaîne de magasins UNO, ainsi qu'à la distribution non alimentaire, la production agricole, la gestion de centres commerciaux et le commerce de détail spécialisé. (Cevital, numidis, CeviAgro,...)
- **Industrie :** Fondée en 2014, cette division regroupe douze filiales actives dans la production de verre et dérivés, de fenêtres et volets, la construction, la logistique et les services, ainsi que les appareils électriques. (Brandt, Baticompos, Cevital Minerals,...)
- **Cevital Group Projects :** Cette division est dédiée à la gestion de projets spécifiques développés par le Groupe Cevital.

1.2.3 Brandt

1.2.3.1 Présentation du groupe Brandt

Le Groupe Brandt³, marque historique, est un acteur de référence dans l'équipement domestique. Présente sur l'ensemble des familles de produits en Lavage, Cuisson, Froid, Petit Électroménager, Télévision et Climatisation, la marque bénéficie d'une forte notoriété auprès des consommateurs. Depuis son lancement, la mission dévolue à Brandt est de faciliter le quotidien des utilisateurs en leur proposant des produits innovants et pertinents, conçus pour accompagner les consommateurs tout au long de leur vie, d'où son slogan "Chaque jour, le meilleur" (illustré dans le logo du groupe Brandt dans la figure II.2).



FIG. II.2 : Logo du groupe Brandt.

Le groupe Brandt est constitué de plusieurs marques emblématiques, dont De Dietrich, Brandt, Sauter et Vedette, qui occupent une place centrale dans le quotidien de millions de foyers à travers le monde. Avec 6 filiales en Europe, en Asie et en Afrique du Nord, le Groupe exporte ses produits vers de nombreux pays, dont les États-Unis, la Chine, l'Afrique et le Moyen-Orient.

En 2014, le Groupe Cevital a repris le Groupe Brandt, renforçant ainsi sa position dans le domaine de l'équipement domestique. Suite à cette acquisition, Cevital a créé une filiale de Brandt en Algérie⁴, avec une usine de production à Sétif, deux sites de recherche et développement et un site de service après-vente. Cette implantation en Algérie a permis au Groupe Brandt de renforcer sa présence dans le marché local et de soutenir ses exportations. Depuis, Brandt fait partie de la division industrielle de Cevital, plus précisément dans le secteur d'activité électroménager.

1.2.3.2 Organisation de la filiale Brandt Algérie

La filiale Brandt Algérie comprend plusieurs directions, notamment la direction générale, la direction Finance et Comptabilité, la direction Contrôle de Gestion, la direction commerciale, ainsi que la direction Marketing Opérationnel où nous avons effectué notre stage.

³<http://www.groupebrandt.com/en/>

⁴<https://brandt.dz/>

Un aperçu de l'organisation de la filiale est présenté dans la figure II.3, en mettant l'accent sur la direction générale, qui compte un effectif de 22 employés sur les 27 postes disponibles. Cette structure organisationnelle permet une répartition claire des responsabilités et une gestion efficace des activités au sein de la filiale Brandt Algérie.

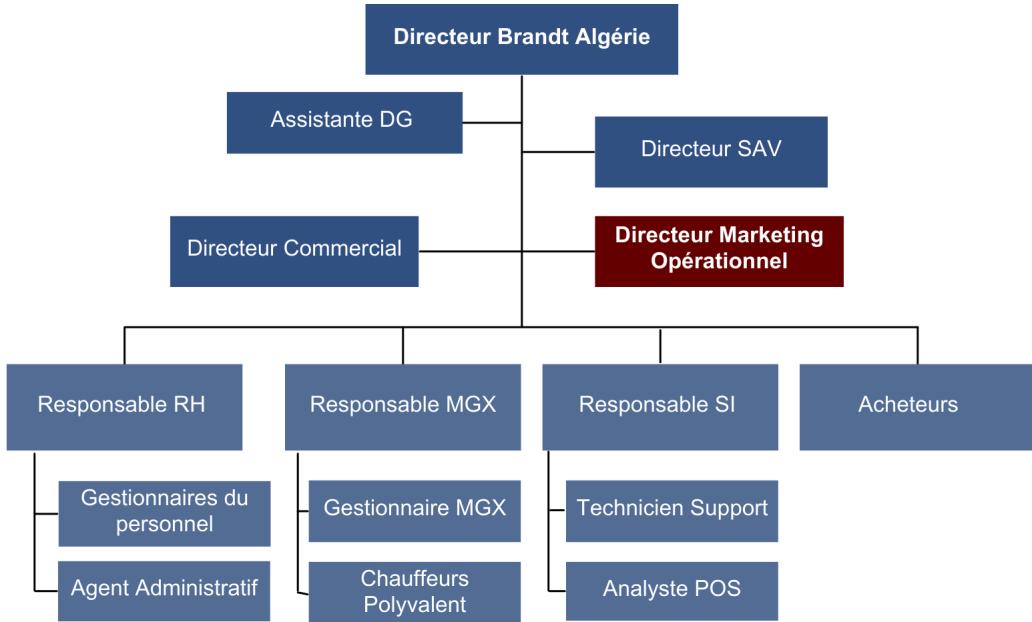


FIG. II.3 : Directions de la filiale Brandt Algérie.

1.2.3.3 Direction Marketing Opérationnel

Nous allons maintenant nous intéresser à la Direction Marketing Opérationnel qui occupe une place essentielle au sein de la filiale Brandt Algérie, étant responsable de la planification et de la mise en œuvre des stratégies de communication et digitales. Notre solution, conçue dans le cadre de notre stage, est destinée à répondre aux besoins spécifiques de cette direction.

La figure II.4 représente l'organisation de la direction marketing opérationnel. Cependant, les rôles qui nous concernent (nos acteurs principaux pour le dashboard) sont : le responsable de communication, le responsable digital, le chargé du digitale, ainsi que l'agence digitale.

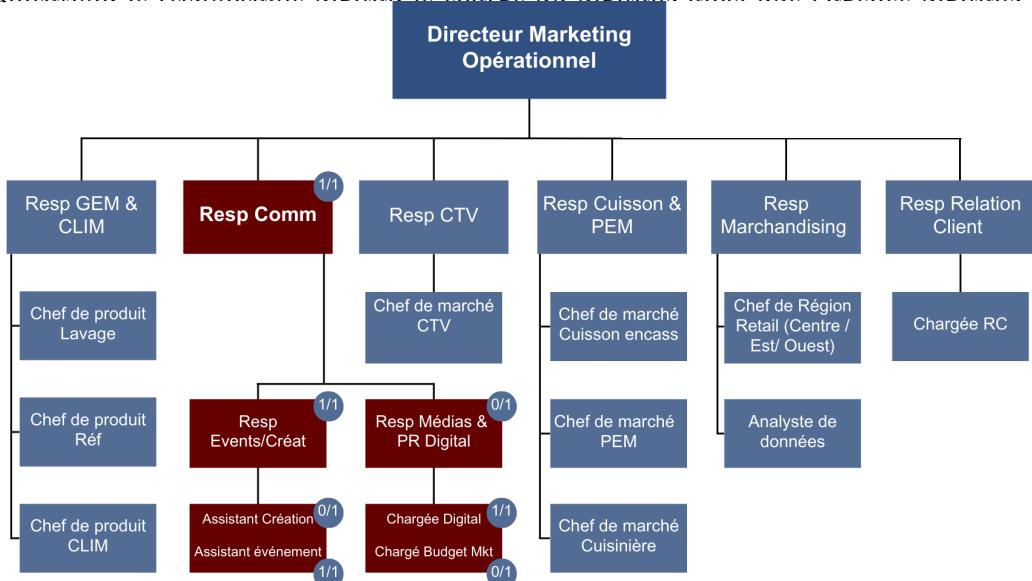


FIG. II.4 : Organisation de la Direction Marketing Opérationnel.

- **Le responsable de communication** est chargé de planifier, d'organiser et de mettre en œuvre les stratégies de communication. Ses rôles clés incluent le développement de la stratégie de communication, la gestion de la marque, des relations publiques, et de l'équipe de communication, la planification des campagnes publicitaires et la coordination des événements.
- **Le responsable digital** est chargé de planifier, de mettre en œuvre et de gérer les stratégies et les initiatives digitales. Parmi ses rôles clés, on retrouve l'élaboration de la stratégie digitale, la gestion des canaux de communication digitale, l'optimisation de la présence digitale, la gestion de la relation client digitale, l'analyse des données digitales et la coordination des projets digitaux.
- **Le Chargé du Digital** est impliqué dans la planification des activités digitales, la création de contenu en collaboration avec l'agence, la gestion des réseaux sociaux en collaboration avec l'agence, l'optimisation de la présence digitale, l'analyse des données digitales, la coordination des projets digitaux et la veille concurrentielle.
- Et enfin, nous devons mentionner **l'agence digitale**, dont la collaboration avec Brandt constitue un partenariat étroit visant à atteindre les objectifs en utilisant des stratégies digitales innovantes et efficaces. Cette agence joue un rôle crucial dans l'élaboration de la stratégie digitale en déterminant les objectifs, les publics cibles, les canaux de communication, les messages clés, les KPIs, etc. De plus, elle est responsable de la création de contenu, de la planification et de la gestion des campagnes digitales, ainsi que du suivi et de l'analyse des performances. L'optimisation de la présence digitale et la modération font également partie des responsabilités de cette agence.

1.2.3.4 Processus de suivi d'e-réputation existant

le processus actuel de suivi de la réputation chez Brandt Algérie consiste à ce que les collaborateurs de la direction marketing mentionnés auparavant surveillent certaines sources pour détecter les avis négatifs et y réagir (en répondant aux commentaires, par exemple) ou prendre note de ces commentaires pour améliorer leurs stratégies.

Ils essaient ensuite de générer des rapports hebdomadaires simples avec une synthèse de la réputation sur le marché et une veille concurrentielle, ainsi que des rapports mensuels plus détaillés contenant les indicateurs clés de performance comme le nombre de publications de la page,...

Nous constatons donc que le processus actuel est principalement manuel et nécessite beaucoup de temps et d'efforts. Aussi, cette approche manuelle ne permet pas d'avoir une vue globale optimale sur la réputation en ligne de la marque, et en addition à cela, le fait de vérifier chaque avis individuellement risque d'engendrer la perte d'informations pertinentes, par exemple les avis non authentiques.

Par ailleurs, cette approche devient encore plus couteuse durant l'été où les pages officielles de Brandt et les groupes de médias sociaux sont inondés de personnes à la recherche de recommandations pour des climatiseurs et autres produits.

1.3 Conclusion

Dans ce chapitre, nous avons étudié le contexte général de l'organisme d'accueil. Dans le prochain, nous nous intéresserons à l'étude des solutions et processus existants pour la résolution de notre problématique.

Chapitre 2

Solution proposée

2.1 Introduction

Dans ce qui suit, nous allons présenter notre solution proposée qui vise à répondre à la problématique essentielle que nous avons identifiée, à savoir l'amélioration du processus de veille sur la réputation de l'entreprise. Comme nous l'avons présenté dans une section précédente, ce processus est actuellement géré de manière manuelle, ce qui peut entraîner des retards, des oubli et une perte d'efficacité globale.

2.2 Présentation de la solution

Pour relever le défi posé, notre vision de départ fut de mettre en place un système de veille périodique sur des sources spécifiques, accompagné d'un système de notification. Ce premier niveau d'amélioration permettrait de suivre de près les avis concernant l'entreprise (ou la marque) sur différentes plateformes en ligne spécifiées, telles que les pages Facebook, les blogs, et autres sources pertinentes.

Conscients que la simple surveillance de la réputation ne serait pas suffisante pour aider à élaborer des stratégies efficaces, nous avons proposé la création d'un dashboard dédié au suivi de l'é-réputation. Ce tableau de bord fournirait aux décideurs et aux professionnels du marketing des statistiques et des informations clés (le ratio d'avis positifs, le nombre d'avis concernant la marque,...) pour prendre des décisions éclairées, mieux comprendre les clients cibles et améliorer l'image de la marque sur le marché.

Nous avons estimé que fournir une analyse sans véritable solution pratique, une manière d'impacter directement les clients, n'était pas suffisant. Ainsi, nous avons pris la décision de développer un moyen de communiquer avec les clients et de les impliquer davantage. Pour ce faire, nous avons opté pour la création d'une page d'accueil, mettant à disposition des informations, des moyens de contact, et même des avis positifs soigneusement sélectionnés par l'entreprise pour représenter les points forts de la marque.

Afin de créer une connexion plus dynamique avec eux, nous avons décidé de transformer la page d'accueil statique en une interface conversationnelle, permettant aux clients d'interagir directement avec un chatbot, de lui poser des questions et de recevoir des réponses en temps réel, renforçant ainsi leur sentiment de connexion avec l'entreprise. Les requêtes que l'agent conversationnel n'arrive pas à résoudre sont enregistrées dans une base de données permettant aux experts de l'entreprise de non seulement mieux comprendre les besoins de leurs clients, mais aussi d'enrichir les bases de connaissances de l'agent afin d'améliorer ses performances.

2.3 Vue conceptuelle de la solution

Jusqu'à présent, nous avons décrit notre système de manière abstraite sans exposer ses composants. Dans ce qui suit, nous décrirons avec plus de détail son architecture comme illustrée ci-dessous dans la figure II.5.

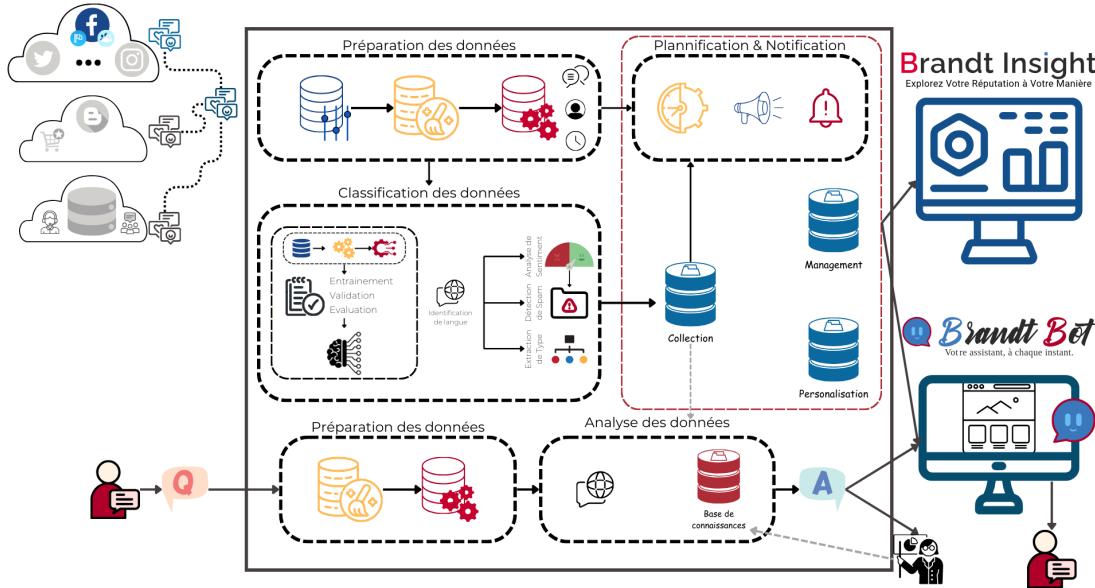


FIG. II.5 : Vue conceptuelle de la solution

2.3.1 Entrées du système

D'abord, nous introduisons les entrées de notre système, à savoir initialement les sources de collecte d'avis, puis les conversations des clients avec le chatbot intégré après le déploiement de la page d'accueil.

2.3.1.1 Les sources de collecte

Les sources depuis lesquelles nous pouvons collecter les avis de clients peuvent inclure :

- **Les réseaux sociaux :** Tels que les pages Facebook (exclusives à l'entreprise ou pas), les groupes Facebook, les pages Instagram,...
- **Les pages Web :** Comme les blogs, les micro-blogs, les sites de vente en ligne,...
- **Les données d'entreprise :** Qui peuvent inclure les historiques de transactions, les feedbacks des clients, les conversations avec le service client,...

Actuellement, seules les sources de médias sociaux telles que les pages Facebook ou les groupes Facebook sont prises en charge par notre système, cependant il est assez facile de mettre en place d'autres sources et d'autres méthodes de collecte étant donné que notre système a été développé en prenant en considération un attribut de qualité très important : l'extensibilité.

2.3.1.2 Entrées des utilisateurs

Il s'agit des conversations des clients avec le chatbot de l'entreprise (après le déploiement de la page d'accueil). Les entrées des clients sont traitées par le système (le chatbot en particulier) afin de déterminer la réponse idéale et la transmettre à ces derniers. Certaines de ces entrées, telles que les questions, les retours clients ou les plaintes, peuvent être enregistrées dans la base de données afin d'être ultérieurement analysés par un expert. Celui-ci pourrait ensuite utiliser les connaissances acquises soit pour améliorer la marque et ses produits, soit pour enrichir l'apprentissage du chatbot.

2.3.2 Le cœur du système

Le système est conçu sous forme de deux processus. Le premier processus traite les avis après les avoir collectés depuis les sources de collectes spécifiées par l'administrateur, et le second analyse les conversations des clients avec l'agent conversationnel.

2.3.2.1 Système de collecte et traitement

Ce système concerne la partie du processus qui collecte les avis depuis les sources de collecte puis les traite afin d'extraire les informations pertinentes qui alimenteront le dashboard d'e-réputation ainsi que les rapports périodiques que recevront les collaborateurs (acteurs principaux du système). Dans ce qui suit, nous présenterons divers composants de ce système.

a. Le collecteur d'avis

La collection d'avis est la brique de notre système, le collecteur prend en entrée la source définie par l'administrateur et collecte les avis y présents en utilisant un des composants les plus importants qu'il comporte : **le scraper**. Si la source est une page ou un groupe de réseaux sociaux, le scraper recueille les publications, les commentaires et les réponses associés. De même, si la source est un blog, le scraper récupère les articles et les commentaires des lecteurs.

En parlant de réseaux sociaux, nous limitons notre collecte aux avis émis sur les pages et les groupes auxquels nous avons accès grâce à des comptes spécifiques. De plus, il est important de souligner que, dans le respect des lois protégeant les données personnelles des citoyens en ligne, nous ne collectons que les avis des clients, sans recueillir leurs informations personnelles telles que les adresses e-mail ou les numéros de téléphone, par exemple. De plus, nous prenons soin d'anonymiser les données collectées, garantissant ainsi la confidentialité et la protection des utilisateurs.

Lors de la première collecte, après que l'administrateur ait défini la source, le scraper récupère les avis jusqu'à une date préétablie. Si l'administrateur a configuré l'outil pour collecter les données avec un intervalle d'un mois, par exemple, le scraper arrête sa collecte lorsqu'il atteint un avis dont la date dépasse la limite définie. Après cette première collecte, des collectes ultérieures sont planifiées périodiquement, afin de garantir une veille continue sur la réputation de l'entreprise. Lors de ces collectes ultérieures, le système vérifie d'abord si nous avons toujours accès à la source avant de lancer la recherche pour récupérer uniquement le nouveau contenu.

Les données collectées sont ensuite soumises à un processus de prétraitement afin de déterminer les avis qui concernent la marque. Prenons l'exemple d'un réseau social, parmi les règles appliquées dans ce processus :

- Si l'avis est une publication, son contenu est analysé pour déterminer s'il concerne la marque. Dans ce cas, la publication ainsi que tous ses commentaires et leurs réponses sont considérés comme pertinents pour la marque.

- Sinon, s'il s'agit d'un commentaire sur une publication qui ne concerne pas la marque, le contenu du commentaire est analysé. S'il est relatif à la marque, ses réponses sont également considérées comme pertinentes pour la marque. En revanche, la publication d'origine est considérée comme concernant la marque, mais non informative.

- De même, si l'avis est une réponse à un commentaire qui ne concerne pas la marque, son contenu est analysé. S'il concerne la marque, les mêmes règles que précédemment s'appliquent, mais cette fois-ci pour la publication d'origine et le commentaire parent.

Une fois les avis concernant la marque sont déterminés, ils passent par un processus de prétraitement comprenant plusieurs étapes. Au cours de cette dernière étape, certains avis sont

identifiés comme non informatifs en termes de polarité ou de type d'avis. Parmi les exemples, on peut citer les mentions, les URLs, et d'autres informations qui ne contribuent pas à une analyse significative de la réputation de la marque. Ces avis sont donc écartés des futurs processus de détections de langue, analyse de sentiments ou d'identification de type, mais pas du processus de détection de spam qui, comme nous le verrons, exploitent les métadonnées autant que les données cachées dans le texte de l'avis.

b. L'extracteur d'information

Dans cette étape, nous cherchons à extraire, pour chaque avis collecté, les données pertinentes qui alimenteront le dashboard de suivi ainsi que les rapports périodiques. Pour ce faire, un avis passe par plusieurs modules, comme illustré dans la figure II.6.

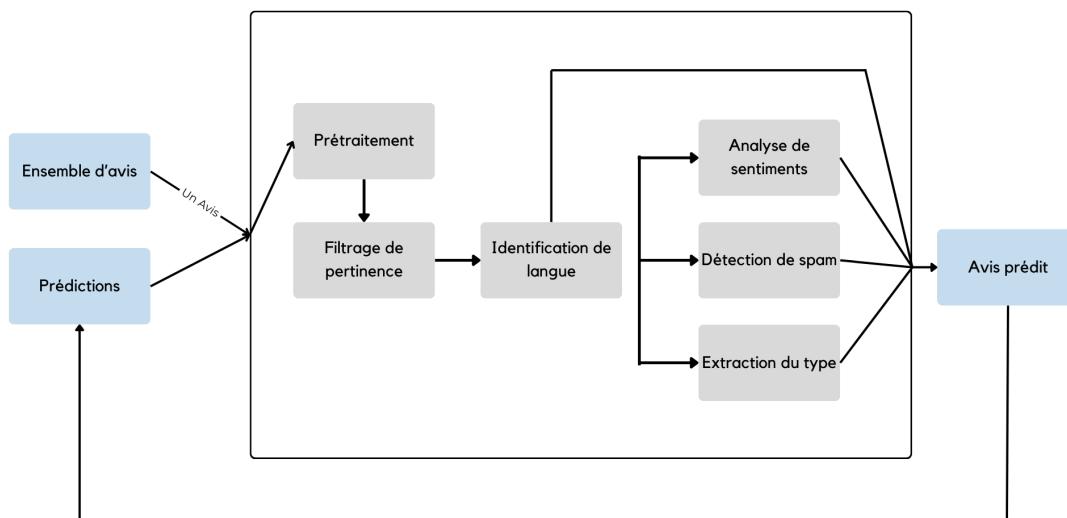


FIG. II.6 : Pipeline de prediction.

En premier lieu, les avis (à l'exception de ceux qui ont été identifiés comme non informatifs) sont soumis à un modèle d'**identification de langue** comme illustré dans la figure II.7. Ce modèle prend en entrée l'avis et détermine sa langue parmi les suivantes : anglais (en), français (fr), arabe (ar), le dialecte algérien exprimé en utilisant des lettres romaines (dz_roman), le dialecte algérien exprimé en utilisant les lettres arabes (dz_ar), ainsi que les autres langues qui ne sont pas prises en charge (not_supported). Les avis non pris en charge ne sont pas considérés pour le reste des étapes du processus.

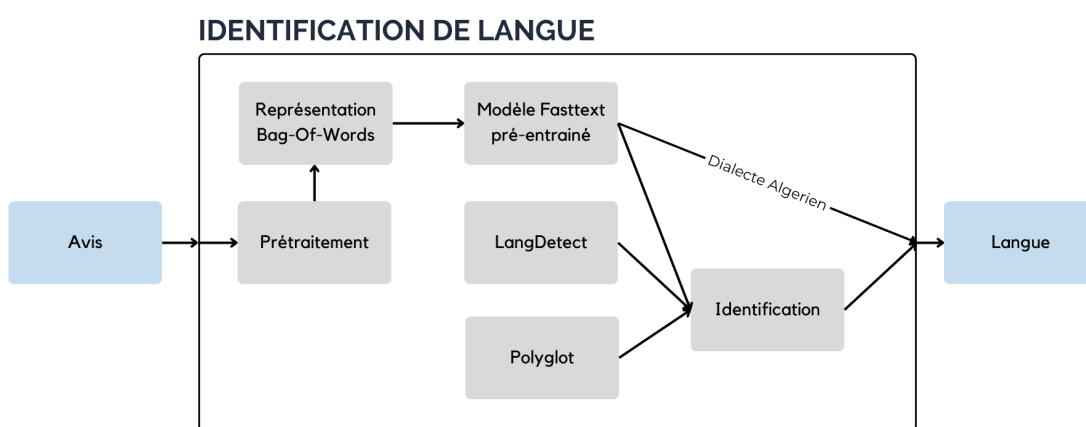


FIG. II.7 : Module d'identification de langue

Ce module extrait une information qui peut être pertinente et permet aux collaborateurs de mieux connaître leur public cible ainsi que leurs tendances d'expression. Mais ce n'est pas tout ! Étant donné que notre système doit prendre en compte la diversité linguistique du public visé, l'utilisation du module d'identification de langue pour chacun des modules qui suivent s'est avérée essentielle. En effet, afin de développer des modules multilingues performants, cette approche fut cruciale. Dans le prochain chapitre, nous discuterons en détail de nos expérimentations ainsi que des résultats obtenus, démontrant ainsi l'efficacité de cette stratégie.

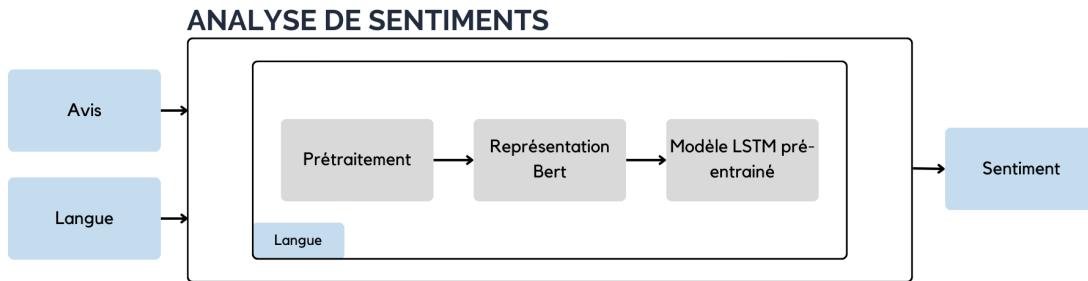


FIG. II.8 : Module d'analyse de sentiments

Après avoir déterminé sa langue, l'avis est traité par le module d'analyse de sentiments, comme le montre la figure II.8, qui permet de déterminer sa polarité. Pour ce faire, nous avons mené des expérimentations avec plusieurs architectures et diverses représentations. Au cours de ces expérimentations, nous avons décidé d'exploiter la puissance de la représentation BERT en utilisant un modèle spécifique à chaque langue. Cependant, cette étape présente plusieurs défis, par exemple, lorsque l'avis est une réponse à un avis parent, il peut être difficile de déterminer sa polarité sans prendre le contexte en compte. Prenons un exemple simple :

- Avis parent (publication ou commentaire) P : "Je déteste cette marque ! Leurs produits sont de qualité médiocre et leur service client est terrible. À éviter à tout prix!"
- Reponse 1 R1: "Tu as totalement raison !"
- Reponse 2 R2: "Ce n'est pas vrai. N'induisez pas les gens en erreur."
- Reponse 3 R3: "Franchement, je trouve que cette marque a un bon rapport qualité-prix, c'est pas mal du tout!"

TAB. II.1 : Exemple de détection de polarité avec et sans prise en compte du contexte

Avis	Polarité par simple modèle de détection	Polarité avec prise en compte du contexte	Polarité de l'avis parent
R1	Positif	Négatif	Négatif
R2	Négatif	Positif	Négatif
R3	Positif	Positif	Négatif

Comme nous pouvons le voir dans le tableau II.1, déterminer la polarité de l'avis sans prendre le contexte en compte peut donner de faux résultats si l'avis est dépendant du contexte (de l'avis parent). Cependant, s'il est indépendant (un avis en lui-même et pas une réponse), l'avis parent n'affecte pas sa polarité.

Pour résoudre ce problème, nous avons décidé de mettre en place une approche simple :

- Tout d'abord, l'avis enfant passe par un modèle qui détermine s'il est dépendant ou indépendant.

- S'il est indépendant, il est traité comme un avis parent et passe directement par le modèle d'analyse de sentiments.

- Sinon, il passe par un autre modèle qui détermine s'il est en accord ou en désaccord avec l'avis auquel il est lié.

- S'il est en accord, il prend la même polarité que son parent (soit positive, soit négative) ; sinon, il prend la polarité opposée.

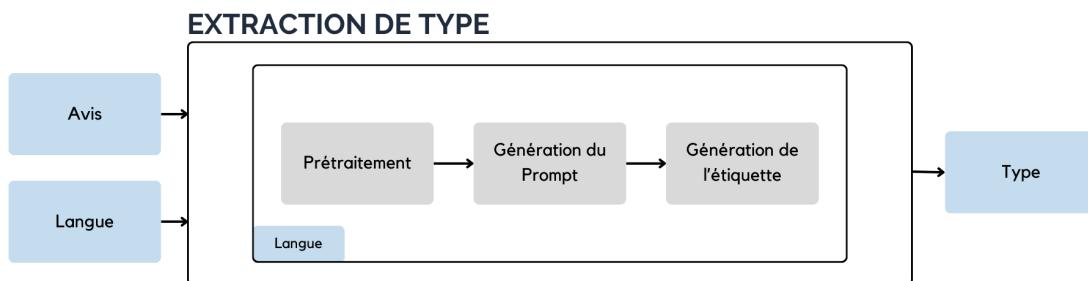


FIG. II.9 : Module d'extraction de type

Une autre information intéressante que l'on peut extraire est le type d'avis exprimé par les clients, qu'il s'agisse d'une description (d'un produit, de la marque, de l'expérience avec le service après-vente, etc.), d'une comparaison (entre deux produits, deux marques, etc.) ou d'une simple question. En raison du manque de données, ce module ne prend en considération que les avis rédigés en anglais, en français ou en arabe. Pour ce modèle, illustré dans la figure II.9, nous avons appliqué la même méthode que pour les deux modèles précédemment présentés (le modèle de détection de dépendance et celui de détection d'accord).

Le module de détection de spam d'opinion, illustré dans la figure II.10, ne prend en charge que les avis informatifs rédigés en anglais, en français ou en arabe, en plus des avis non informatifs. Dans notre approche, nous avons pris en considération des caractéristiques textuelles et comportementales, à savoir : la représentation BERT spécifique à chaque langue, la polarité, la longueur de l'avis (RL), les bursts (BST), le ratio positif (PR), le nombre maximal d'avis par jour (MRD), la longueur moyenne des avis (ARL), la similarité maximale entre les avis (MRS) ainsi que la similarité moyenne entre les avis (ARS).

L'utilisation de ces caractéristiques nous permet d'identifier et de filtrer les avis potentiellement suspects ou non pertinents, afin de garantir la qualité et la fiabilité des analyses effectuées. En intégrant des paramètres textuels et comportementaux dans notre module de détection de spam d'opinion, nous visons à fournir des résultats plus précis.

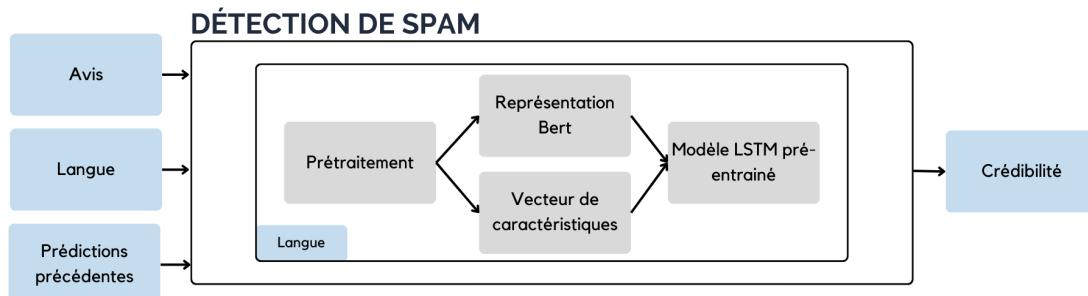


FIG. II.10 : Module de détection de spam

TAB. II.2 : Exemples d'entrées sorties du système

Entrée	Sortie
"I really loved this! I buy this serum every summer and I feel like I just can't live without it anymore !!!"	"I really loved this! I buy this serum every summer and I feel like I just can't live without it anymore !!!" 28/07/2023 Auteur01 publication01 facebook_page_01 Informatif En Positif Description Non Spam
"Pour le restaurant ABC, je n'ai pas trop apprécié le personnel, ils étaient méchants. Et la nourriture était infecte... Je préfère le restaurant DFE, il est beaucoup moins cher et c'est trop bon !"	"Pour le restaurant ABC, je n'ai pas trop apprécié le personnel, ils étaient méchants. Et la nourriture était infecte... Je préfère le restaurant DFE, il est beaucoup moins cher et c'est trop bon !" 09/09/2023 Auteur02 publication04 facebook_groupe_04 Informatif Fr Négatif Comparaison Spam
"لم يرق هذا الفندق إلى توقعاتنا. الغرف قديمة و الضوضاء المحيطة جعلت من الصعب الاسترخاء بإقامتنا"	"لم يرق هذا الفندق إلى توقعاتنا. الغرف قديمة و الضوضاء المحيطة جعلت من الصعب الاسترخاء بإقامتنا" 23/8/2023 Auteur03 commentaire06 facebook_page_02 Informatif Ar Negatif Description Non spam
"3jbni bzf had telephone, rapide wla batterie ta3 tshad bzf. hayel!"	"3jbni bzf had telephone, rapide wla batterie ta3 tshad bzf. hayel!" date Auteur04 commentaire07 facebook_page_05 Informatif Dz_roman Positif Description Non Spam
"ماشي بنينة لماكلة تاعكم و غالية تانيت"	"ماشي بنينة لماكلة تاعكم و غالية تانيت" date Auteur05 commentaire16 facebook_page_06 Informatif Dz_ar Négatif Description Spam
"La marca XYZ ofrece productos de belleza de alta calidad que han transformado mi piel. Sus ingredientes naturales y resultados visibles la convierten en una marca confiable para el cuidado de la piel."	"La marca XYZ ofrece productos de belleza de alta calidad que han transformado mi piel. convierten en una marca confiable para el cuidado de la piel." 25/06/2023 Auteur06 / facebook_page_02 Informatif Non pris en charge / / /
"@ami"	"@ami" 02/05/2023 Auteur07 commentaire02 facebook_groupe_02 Non informatif / / / Spam

Le tableau II.2 illustre des exemples d'entrées sorties de ce système. La sortie du processus qui prend en entrée un avis (son texte, sa date, l'id de son auteur,...) est du format suivant : **Texte | Date | Id_auteur | Id_parent | Id_source | Pertinence | Langue | Polarité | Type | Crédibilité** Où :

- **Texte** : Le contenu de l'avis exprimé par l'utilisateur.

- **date** : La date à laquelle l'avis a été émis.
- **Id_auteur** : L'identifiant unique de l'auteur de l'avis.
- **Id_parent** : L'identifiant unique de l'avis parent (dans le cas où l'avis est une réponse ou un commentaire).
- **Id_source** : L'identifiant unique de la source d'où provient l'avis.
- **Pertinence** : Indication sur la pertinence de l'avis (informatif ou non).
- **Langue** : La langue dans laquelle l'avis a été exprimé (anglais, français, arabe, etc.).
- **Polarité** : L'orientation sentimentale de l'avis (positif ou négatif).
- **Type** : Le type d'avis exprimé par l'utilisateur (description, comparaison, question).
- **Crédibilité** : Le niveau de crédibilité attribué à l'avis en fonction de diverses caractéristiques (Spam ou non spam).

Les avis collectés et traités sont ensuite sauvegardés dans la base de données. À partir de ces données, nous sommes en mesure de créer de nombreuses statistiques intéressantes qui alimenteront le tableau de bord. Par exemple, nous pourrons obtenir le nombre d'avis par source, le ratio d'avis positifs, ainsi que les pourcentages respectifs des différents langages utilisés.

c. Le planificateur

Le planificateur est le cœur de l'outil de veille qui compose notre système. Une fois les avis collectés lors de la première exécution (après que la source ait été définie dans le système), il programme la prochaine exécution en fonction d'un paramètre défini X par l'administrateur. Par exemple, toutes les 3 heures, le système lance une nouvelle collecte à partir de la source spécifiée en utilisant un compte d'accès spécifié. Il récupère les nouvelles données, les enregistre dans la base de données et planifie la prochaine exécution. Ce processus assure une veille périodique et régulière sur la source désignée, garantissant que les données restent à jour et que le système continue de fournir des informations pertinentes.

d. Le notificateur

En utilisant les données nouvellement collectées à partir des exécutions planifiées après la première exécution, le système génère un rapport présentant les nouvelles données acquises ainsi que des statistiques à leur sujet. Grâce au notificateur, ce rapport est envoyé à chaque personne devant être informée. En cas de problème survenant dans le système, tel que la perte d'accès à une source, une erreur pendant une exécution, etc., le notificateur avertit immédiatement l'administrateur. Cette fonctionnalité assure une surveillance proactive du système, permettant à l'administrateur de réagir rapidement en cas de problème, et garantit que les parties concernées restent informées périodiquement des de la réputation en ligne de la marque.

2.3.2.2 Le système conversationnel

Le système conversationnel concerne à son tour la partie du processus qui a pour objectif de renforcer le lien avec le client. Les interactions entre les clients et le chatbot sont essentielles pour établir une communication fluide et efficace. Lorsque les clients interagissent avec le chatbot sur la page d'accueil déployée, leurs entrées sont envoyées depuis le frontend vers notre backend pour traitement. Le chatbot utilise alors un algorithme de recherche dans sa base de connaissance pour trouver la correspondance la plus pertinente parmi les réponses enregistrées. Ainsi, il est en mesure de fournir la réponse la plus appropriée aux clients en fonction de leurs requêtes spécifiques.

Dans le cas où le chatbot ne dispose pas d'une réponse appropriée ou si la langue utilisée n'est pas prise en charge (étant donné qu'il ne prend en charge que le français, l'anglais et l'arabe), il est programmé pour réagir de manière adéquate. Grâce au module de détection de langue mentionné précédemment, le chatbot détermine s'il ne possède pas l'information requise

ou si le client converse dans une langue non supportée. Si tel est le cas, le chatbot présente ses excuses en utilisant le langage qui sera le plus susceptible d'être compréhensible par le client. Par exemple, s'il reçoit une requête en dialecte algérien exprimé en utilisant les lettres romaines (dz_roman), il répondra en français, tandis que pour un dialecte algérien exprimé en utilisant les lettres arabes (dz_ar), la réponse sera en arabe. Dans le cas de toute autre langue, le chatbot répondra en anglais.

2.4 Conclusion

En conclusion de ce chapitre, nous avons expliqué notre solution, son architecture, ainsi que les divers éléments qui la composent. Dans les chapitres à suivre, nous aborderons les expérimentations que nous avons menées pour chaque modèle du système, mettant en évidence les résultats obtenus ainsi que les défis que nous avons rencontrés tout au long de notre démarche.

Par la suite, nous présenterons le prototype que nous avons développé. Ce prototype serait ainsi une démonstration concrète de notre système illustrant la faisabilité et l'intérêt de notre approche. De plus, nous discuterons en profondeur de l'image de marque et des différents choix conceptuels que nous avons dû entreprendre pour répondre au mieux à leurs besoins spécifiques.

Chapitre 3

Expérimentations et résultats

3.1 Introduction

Comme mentionné précédemment, notre système est composé de plusieurs modèles, chacun jouant un rôle clé dans l'accomplissement de notre objectif. Dans les sections à venir, nous décrirons en détail les expérimentations que nous avons menées pour chaque modèle, nous avons expérimenter avec différents modèles de ML et de DL dans le but d'obtenir les meilleurs résultats et performances possibles.

3.2 Identification de la langue

Le modèle d'identification de langue prend le texte d'un avis en entrée et renvoie sa langue : qu'il s'agisse d'anglais, de français, d'arabe, du dialecte algérien (exprimé avec des lettres arabes ou romaines) ou s'il n'est actuellement pas pris en charge par le système.

Ce modèle fournit non seulement à notre tableau de bord des informations importantes et de nouvelles perspectives, mais il est également crucial pour l'exécution correcte des autres modèles du système de réputation ainsi que pour le chatbot.

Au début, nous avons envisagé d'utiliser des solutions existantes bien connues capables d'identifier la langue d'un texte telles que Langdetect¹, Polyglot² et PyclD2³, ainsi qu'un modèle Fasttext préentraîné : lid.176.bin⁴. Cependant, le problème est rapidement devenu évident : ces solutions, bien que puissantes, ne sont pas capables d'identifier le dialecte algérien, ce qui nous a poussé à créer notre propre modèle d'identification de langue.

Nous avons expérimenté avec plusieurs approches et avons finalement opté pour l'utilisation d'un modèle Fasttext, que nous avons entraîné sur des données que nous avons collectées ainsi que sur des données créées manuellement. Le Tableau II.3 présente notre ensemble de données ainsi que les différentes sources à partir desquelles nous avons collecté nos données.

Comme nous pouvons le constater, l'ensemble de données est très déséquilibré, car il y a très peu de données pour le dialecte algérien, étant donné qu'il n'y a pas beaucoup de ressources disponibles à ce sujet. Toutefois, en collectant davantage de données à l'aide du collecteur de notre système, nous pourrions les utiliser pour améliorer davantage les performances de notre modèle.

¹<https://pypi.org/project/langdetect/>

²<https://polyglot.readthedocs.io/en/latest/index.html>

³<https://pypi.org/project/pyclD2/>

⁴<https://fasttext.cc/docs/en/language-identification.html>

Chapitre 3. Expérimentations et résultats

TAB. II.3 : Répartition des langues dans le jeu de données d'entraînement pour le modèle d'identification de langues.

Langue	Nombre de lignes	Sources
En	244339	Acquis manuellement, Tatoeba ⁵ , Papluca ⁶ , oea_algdi ⁷ , YelpChi
Fr	276674	Acquis manuellement, Tatoeba, Papluca, oea_algdi, Allocine ⁸
Ar	185024	Acquis manuellement, Tatoeba, Papluca, oea_algdi, NArabizi ⁹ , saurabhshahane ¹⁰
Dz_ar	22628	Acquis manuellement, msda ¹¹ , NArabizi
Dz_roman	6597	Acquis manuellement, NArabizi, (BOUAMOR et al. 2018), (MATAOUI et al. 2016), (CHADER et al. 2019)
Autres	76500	Acquis manuellement, Papluca, (BOUAMOR et al. 2018), (MATAOUI et al. 2016), (CHADER et al. 2019)
Total	811762	

Étant donné que les modèles de détection de langues bien connus sont très efficaces pour détecter les langues autres que le dialecte algérien, nous avons décidé de tirer parti de cela en créant une méthode d'ensemble. Cette méthode effectue une première prédiction à l'aide de notre modèle basé sur Fasttext, puis, si la prédiction n'est pas le dialecte algérien, elle passe par un processus utilisant Polyglot et Langdetect, en plus de certaines règles logiques, afin de faire une prédiction plus précise, améliorant ainsi nos performances. Le Tableau II.4 présente les résultats de nos expériences ainsi que des comparaisons avec les solutions existantes.

TAB. II.4 : Résultats des expérimentations des modèles d'identification de langue.

Modèles	Précision	Rappel	Accuracy	F-score
Langdetect	0.9138	0.9002	0.9002	0.9045
Polyglot	0.9069	0.9235	0.9235	0.9136
Pycll2	0.9226	0.8826	0.8826	0.9001
Modèle Fasttext pré-entraîné : lid.176.bin	0.9185	0.9434	0.9434	0.9299
Notre modèle Fasttext	0.9851	0.9853	0.9853	0.9851
Notre méthode d'ensemble	0.9852	0.9854	0.9854	0.9852

3.3 Analyse des sentiments

Nous poursuivons avec la polarité, l'un des aspects les plus importants de notre système de suivi de l'e-réputation. La polarité d'un avis révèle l'opinion exprimée par le client, qu'elle soit positive ou négative à l'égard de la marque, du produit ou du service consommé ou fourni.

⁵tatoeba.org

⁶<https://huggingface.co/datasets/papluca/language-identification>

⁷https://github.com/kinmokusu/oea_algdi/blob/master/data/dataset/data.csv

⁸<https://www.kaggle.com/djilax/allocine-french-movie-reviews>

⁹<https://github.com/SamiaTouileb/NArabizi/blob/main/README.md>

¹⁰<https://www.kaggle.com/datasets/saurabhshahane/arabic-classification>

¹¹https://msda.um6p.ma/msda_datasets

Cependant, la détermination précise de la polarité peut être un défi, car les avis peuvent être rédigés de manière complexe avec des nuances subtiles. Dans ce qui suit, nous ne considérons que les avis positifs et les avis négatifs. Pour représenter les avis dans notre système, nous utilisons des embeddings BERT (DEVLIN et al. 2018b) qui nous permettent de capturer les informations sémantiques et contextuelles essentielles des avis pour une analyse plus approfondie de la polarité.

Pour appuyer notre choix, nous avons d'abord réalisé des expérimentations en anglais en utilisant le jeu de données Yelpchi, qui n'est pas annoté avec des scores de sentiment, mais qui contient la note de chaque avis. Nous avons suivi la pratique conventionnelle en considérant les avis avec une note inférieure à trois comme négatifs et les autres comme positifs. Le jeu de données est décrit plus en détail dans le tableau II.6. Nous avons d'abord expérimenté avec des techniques d'apprentissage machine traditionnelles (SVM, NB, etc), puis avec des réseaux neuronaux simples, mais les résultats étaient insatisfaisants. Nous avons ensuite exploré des modèles LSTM plus complexes et avons limité nos expériences à trois modèles LSTM avec différentes architectures. Dans la suite des expérimentations, nous avons considéré plusieurs configurations en prenant en compte les paramètres suivants :

- Fonctions d'optimisation : Nous avons expérimenté avec différentes fonctions d'optimisation, telles que l'optimiseur Adam, RMSprop et SGD, pour déterminer celle qui offre les meilleures performances en termes de convergence et de précision du modèle.
- Nombre d'unités : Nous avons varié le nombre d'unités dans les couches LSTM pour évaluer son impact sur la capacité du modèle à capturer les motifs et les dépendances dans les avis textuels.
- Taux d'apprentissage : Le taux d'apprentissage a été ajusté pour contrôler la vitesse à laquelle le modèle apprend à partir des données, afin de trouver une valeur optimale qui favorise une convergence rapide sans surapprentissage.
- Taux d'abandon (Dropout Rate) : Nous avons utilisé différentes valeurs de taux d'abandon pour les couches LSTM afin de régulariser le modèle et éviter le surapprentissage en désactivant aléatoirement certains neurones pendant l'entraînement.

Certains de nos meilleurs résultats sur le jeu de données en anglais sont présentés dans le tableau II.5. Les expériences sont nommées comme suit : Model_numero-model_optimiser_units_learning-rate_dropout-rate.

TAB. II.5 : Meilleurs résultats pour les expériences d'analyse de sentiment pour l'anglais.

Cross Validation Expériences	Exactitude	Précision	Rappel	F1-Score	AUC-ROC
Model_2_adam _64_1e-4	0.9295	0.9491	0.9646	0.9560	0.8829
Model_3_adam _64_1e-4_0.2	0.9011	0.9242	0.9527	0.9381	0.8329
Model_2_rmsprop_32 _1e-4	0.9182	0.9463	0.9518	0.9484	0.8735
Model_4_adam _64_1e-4_0.2	0.9097	0.9344	0.9518	0.9430	0.8538
Test Expériences	Exactitude	Précision	Rappel	F1-Score	AUC-ROC
Model_2_adam _64_1e-4	0,8693	0,9342	0,8966	0,915	0,8332

Model_3_adam _64_1e-4_0.2	0,8215	0,9471	0,8183	0,878	0,8258
Model_2_rmsprop_32 _1e-4_	0,8173	0,9572	0,803	0,8734	0,836
Model_4_adam _64_1e-4_0.2	0,8232	0,9472	0,8205	0,8793	0,8269

Compte tenu de la diversité de notre public et du rôle central de la polarité dans notre système, nous avons essayé de rendre notre modèle multilingue et capable de prendre en charge les principales langues parlées par notre public cible (anglais, français, arabe et le dialecte algérien, à la fois en utilisant les lettres arabes et les lettres romaines). Pour y parvenir, nous avons rassemblé des ensembles de données pour chacune de ces langues, qui sont présentés dans le tableau II.6.

TAB. II.6 : Descriptions des jeux de données pour le modèle d'analyse de sentiments.

Langue	Nombre d'avis	Taux d'avis positifs	Source
Anglais	5854	78%	YelpChi
Français	19999	50%	Allociné
Arabe	11358	50%	Arabic Sentiment Twitter Corpus ¹²
Dialecte algérien avec lettres arabes	3438	50%	(BOUAMOR et al. 2018), (MATAOUI et al. 2016), (CHADER et al. 2019)
Dialecte algérien avec lettres romaines	3448	67%	(BOUAMOR et al. 2018), (MATAOUI et al. 2016), (CHADER et al. 2019)

Dans notre quête pour rendre ce modèle compatible avec nos langues cibles, nous avons d'abord effectué des expérimentations pour chaque langue individuellement, dans le but de développer un modèle performant pour chaque langue séparément. Pour chaque langue, nous avons utilisé différentes techniques de prétraitement, différentes listes de mots vides, ainsi que des lemmatiseurs, des tokenizers et des représentations BERT spécifiques. Pour le traitement des emojis, nous avons utilisé un dictionnaire d'emojis (et d'émoticons) disponible en ligne¹³ pour l'anglais, et nous l'avons traduit manuellement pour les autres langues. Le tableau II.7 présente les lemmatiseurs spécifiques et les modèles/tokenizers BERT que nous avons utilisés pour chaque langue, tandis que le tableau II.8 rapporte les résultats du meilleur modèle pour chaque langue.

TAB. II.7 : Lemmatiseurs et Modèles BERT utilisés pour chaque langue.

Langue	Lemmatiseurs	Modèles BERT
Anglais	en_core_web_sm de Spacy ¹⁴	(DEVLIN et al. 2018a)

¹²<https://www.kaggle.com/datasets/mksaad/arabic-sentiment-twitter-corpus>

¹³<https://www.kaggle.com/datasets/divyansh22/emoji-dictionary-1>

¹⁴<https://spacy.io/>

Français	fr_core_news_sm de Spacy	(H. LE et al. 2020b), (H. LE et al. 2020a)
Arabe	(ZERROUKI 2020)	(ANTOUN et al. s. d.)
Dialecte algérien avec lettres arabes	/	(ABDAOUI et al. 2021)
Dialecte algérien avec lettres romaines	/	(ABDAOUI et al. 2021)

TAB. II.8 : Meilleurs résultats du modèle d'analyse de sentiments pour chaque langue.

Expériences Cross Validation	Exactitude	Précision	Rappel	F1-Score	AUC-ROC
EN	0,96	0,98	0,97	0,97	0,94
FR	0,93	0,92	0,94	0,93	0,93
AR	0,93	0,94	0,91	0,93	0,93
DZ_ar	0,96	0,96	0,97	0,97	0,95
DZ_roman	0,95	0,96	0,94	0,95	0,95
Expériences Test	Exactitude	Précision	Rappel	F1-Score	AUC-ROC
EN	0,86	0,92	0,89	0,91	0,81
FR	0,65	0,64	0,70	0,67	0,65
AR	0,72	0,75	0,66	0,70	0,72
DZ_ar	0,81	0,85	0,87	0,86	0,77
DZ_roman	0,80	0,84	0,73	0,78	0,80

Par la suite, nous avons commencé à expérimenter afin de trouver la meilleure méthode pour créer un modèle prenant en charge toutes les langues en une seule fois.

Dans notre première expérience (Expérience 01), nous avons essayé de mélanger toutes les données collectées pour former un ensemble de données multilingue, puis d'appliquer un prétraitement de base (celui appliqué à l'anglais) et d'utiliser la représentation BERT la plus couramment utilisée, tout comme nous l'avions fait pour l'anglais (DEVLIN et al. 2018a). Cependant, les résultats n'étaient pas satisfaisants, car la représentation BERT utilisée ne reconnaissait pas les jetons d'autres langues, donc ni la tokenisation ni la représentation ne donnaient de résultats corrects.

Puisque la représentation BERT(DEVLIN et al. 2018a) n'a pas donné de résultats satisfaisants sur notre jeu de données, nous avons décidé d'utiliser une représentation BERT multilingue (DEVLIN et al. 2018a) à la place (Expérience 02). Les résultats étaient toujours peu satisfaisants, car le modèle BERT ne prenait pas en charge le dialecte algérien.

Ainsi, traiter l'ensemble des données comme un seul ensemble semblait impossible, car il n'y avait aucun moyen de représenter toutes les langues en même temps. C'est là que le modèle d'identification de langue entre en jeu ! En l'utilisant, nous pouvons ignorer les langues non prises en charge, puis diviser l'ensemble de données en sous-ensembles en fonction de la langue, et appliquer à chaque sous-ensemble un processus de prétraitement et de représentation différent.

Suivant cette ligne de pensée, c'est précisément ce que nous avons fait : nous avons divisé l'ensemble de données en fonction de la langue et acheminé chaque sous-ensemble vers un processus de prétraitement et de représentation différent, le même processus que nous avons utilisé

précédemment pour les modèles individuels. Ensuite, nous avons fusionné les données traitées et les avons divisées en lots d'entraînement et de test, formant ainsi un modèle unique sur l'ensemble des données multilingues (Expérience 03). Cependant, nous avons remarqué un biais dans cette approche : le fait que les avis de chaque langue étaient représentés à l'aide d'un modèle différent signifiait qu'il pouvait exister un jeton X_i dans la langue i, et un autre jeton X_j dans la langue j, et que, en utilisant deux modèles différents M_1 et M_2 la représentation pouvait être la même même si les jetons n'étaient pas les mêmes ($X_i \neq X_j$ mais $M_1(X_i) \equiv M_2(X_j)$), ce qui explique pourquoi les résultats n'étaient toujours pas satisfaisants.

Ce constat nous a conduits à notre approche proposée : comme précédemment, nous avons acheminé les données vers un processus de prétraitement et de représentation en fonction de la langue des avis, mais cette fois-ci, nous n'avons pas combiné les données une fois cela terminé. Pour chaque langue, nous avons utilisé les meilleurs modèles expérimentés précédemment pour chaque langue individuelle, créant ainsi une méthode d'ensemble qui, en fonction de la langue, achemine les avis vers le processus de prétraitement et de représentation spécifique à la langue, ainsi que vers le meilleur modèle de classification pour la langue (Expérience 04). Le tableau II.9 résume les résultats des expériences que nous avons décrites et montre la différence entre les résultats obtenus.

TAB. II.9 : Résultats des différentes expériences d'analyse de sentiments multilingues

Expériences Cross-Validation	Exactitude	Précision	Rappel	F1-Score	AUC-ROC
Expérience 01	0.78	0.81	0.76	0.80	0.78
Expérience 02	0.73	0.85	0.66	0.74	0.75
Expérience 03	0.87	0.91	0.89	0.90	0.86
Expérience 04	0.93	0.95	0.94	0.95	0.93
Expériences Test	Exactitude	Précision	Rappel	F1-Score	AUC-ROC
Expérience 01	0.69	0.75	0.72	0.73	0.67
Expérience 02	0.65	0.78	0.58	0.66	0.68
Expérience 03	0.72	0.76	0.74	0.77	0.70
Expérience 04	0.78	0.83	0.81	0.83	0.74

Dans le chapitre précédent, nous avons présenté un problème que nous rencontrons dans le processus d'analyse des sentiments, non pas dans le domaine théorique, mais dans le domaine pratique, à savoir le manque de contexte résultant de la nature des plateformes de médias sociaux, ce qui entraîne que certaines critiques sont des réponses à des commentaires ou des publications. Nous avons proposé deux solutions pour résoudre ce problème : l'un pour déterminer si un commentaire/réponse est dépendant ou indépendant, et l'autre pour déterminer si la critique dépendante est d'accord ou en désaccord avec ce à quoi elle est liée. Compte tenu du manque de ressources, nous avons décidé de prendre en charge uniquement l'anglais, le français et l'arabe pour cette partie du système, tandis que nous transmettons directement les critiques en dialecte algérien à la méthode d'analyse des sentiments. Nous avons également décidé d'utiliser un modèle GPT pré-entraîné (Generative Pre-trained Transformers) pour cette tâche, tout en l'ajustant à cette tâche en utilisant ce que l'on appelle l'apprentissage à faible coût (Few-Shot Learning, FSL), qui est un cadre d'apprentissage automatique permettant à un modèle pré-entraîné d'apprendre efficacement à partir d'un nombre limité d'échantillons étiquetés par classe, même pour de nouvelles catégories de données qui n'ont pas été vues pendant le processus d'entraînement.

TAB. II.10 : Exemple de classification de dépendance/indépendance et d'accord/désaccord.

Avis	Classe dépendance	Classe Position
"You are correct!"	Dépendance	Accord
"I couldn't agree more!"	Dépendance	Accord
"If that's what you want to believe..."	Dépendance	Désaccord
"I've been going to that store for as long as I can remember and it's still my favorite!"	Indépendance	/
"The first movie was much better than the second!"	Indépendance	/
"True!"	Dépendance	Accord
"you couldn't be more wrong..."	Dépendance	Désaccord
"I go there every now and then. and i love what they do"	Indépendance	/

Pour cette tâche, nous aurions souhaité profiter de grands modèles pré-entraînés tels que GPT-4 ou GPT-3.5 d'OpenAI¹⁵, ou l'alternative open source GPT-j (B. WANG et KOMATSUZAKI 2021), etc. Cependant, en raison de la non-disponibilité de certains et de notre manque de ressources et de puissance de calcul, nous n'avons pas pu le faire. Nous avons donc décidé d'utiliser un modèle plus petit, qui est un modèle GPT-j ajusté¹⁶. L'idée générale de notre approche est la suivante : si l'avis est en arabe ou en français, nous le traduisons en anglais, les avis en anglais sont ensuite passés au modèle GPT et le modèle génère une étiquette en fonction des étiquettes possibles et des exemples précédents qui lui ont été donnés. Le tableau II.10 présente quelques exemples d'entrées et les étiquettes retournées par nos modèles.

3.4 Classification du type

La classification d'une critique en tant que description, comparaison ou question était une préoccupation secondaire de l'entreprise. Nous aurions aimé fournir une classification plus informative et complexe, prenant en charge plus que seulement l'anglais, l'arabe et le français. Cependant, étant donné qu'il n'y avait pas de données disponibles et le fait que nous aurions dû créer et annoter à nouveau des données manuellement, nous avons décidé d'essayer à nouveau l'apprentissage à faible coût en utilisant le même modèle GPT pré-entraîné que celui que nous avons utilisé pour le modèle de classification de dépendance et le modèle de classification de position, et de suivre le même processus. Nous lui avons fourni quelques exemples et lui avons ensuite transmis l'entrée souhaitée en lui demandant de générer son étiquette. Le tableau II.11 présente quelques exemples d'entrées et les étiquettes renvoyées par ce modèle.

TAB. II.11 : Exemple de classification de type de texte.

Avis	Classe Type
"I brought this product last week, and so far it has been amazing!"	Description
"The second movie is so much better than the first one"	Comparaison
"I tried everything I could and still couldn't get this to work. Can someone please tell me how to fix this?"	Question
"I like Brand1 but Brand2 is so much better!"	Comparaison

¹⁵<https://openai.com/>

¹⁶<https://huggingface.co/togethercomputer/GPT-JT-Moderation-6B>

"The customer service in this restaurant is just awful..."	Description
"How can I get to your store from the station ?"	Question

3.5 Detection de spam

Lors de nos recherches et de la lecture d'articles scientifiques, nous avons rapidement réalisé que la rareté des données était un problème majeur, en particulier pour les langues autres que l'anglais. Cependant, nous ne pensions jamais que cela serait aussi difficile, en particulier lorsqu'il s'agit de données comportementales. Ils existaient quelques articles scientifiques mentionnaient l'existence de jeux de données pour les langues ciblées, mais ils ne les rendaient pas publics, les réservant souvent à des fins de recherche internes ou à des collaborations restreintes. Cela signifiait que, malgré nos efforts de recherche, nous n'avons pas pu trouver de jeux de données appropriés pour ces langues, ce qui a limité nos possibilités d'expérimentation et d'analyse.

Cela nous a conduits à envisager de traduire les avis du jeu de données yelpchi (qui contient 13% d'avis spam), qui était destiné à être utilisé pour le modèle de détection de spam en anglais, vers les autres langues. Nous avons réussi à le faire pour le français et l'arabe en utilisant le modèle de traduction GoogleTranslator de Deep Translator¹⁷. Cependant, nous avons rapidement découvert qu'il n'existe pas de traducteur pour le dialecte algérien, et nous avons également constaté que la traduction manuelle de chaque avis ou la création d'un traducteur personnalisé demandait beaucoup de temps et d'efforts. Nous avons donc abandonné cette idée, ce qui signifie que ce modèle prend en charge uniquement l'anglais, le français et l'arabe pour le moment. Cette limitation est un exemple de la complexité que peut représenter la prise en charge de plusieurs langues dans un système, et cela souligne l'importance de l'accès à des ressources linguistiques appropriées pour des tâches de traitement du langage naturel multilingues.

Concernant les caractéristiques exploitées, comme mentionné précédemment, nous avons utilisé : la représentation BERT spécifique à chaque langue, la polarité, la longueur de l'avis (RL), les Bursts (BST), le ratio positif (PR), le nombre maximal d'avis par jour (MRD), la longueur moyenne des avis (ARL), la similarité maximale entre les avis (MRS) ainsi que la similarité moyenne entre les avis (ARS). En ce qui concerne l'architecture du modèle de classification, nous avons mené plusieurs expérimentations, en utilisant des réseaux de neurones simples, des LSTM, des architectures à entrée unique ou à entrée multiple, etc. À la fin, nous avons réussi à les réduire à quatre réseaux LSTM à entrées multiples que nous avons testés avec de multiples combinaisons. Le tableau suivant (Tableau II.12) montre certains des meilleurs résultats que nous avons obtenus au cours de nos expériences, le format suivi dans la dénomination des expériences étant le même que celui utilisé précédemment pour l'analyse de sentiment.

TAB. II.12 : Meilleurs résultats obtenus durant les Expériences de détection de spam pour l'anglais.

Expériences Cross-Validation	Exactitude	Précision	Rappel	F1-Score	AUC-ROC
Model_4_adam _32_1e-4_0.2	0.7758	0.3326	0.6833	0.4474	0.7366
Model_3_adam _32_1e-4_0.2	0.7865	0.3528	0.7283	0.4753	0.7618

¹⁷ <https://pypi.org/project/deep-translator/>

Model_4_adam _64_1e-4_0.2	0.7971	0.3640	0.7058	0.4803	0.7585
Model_1_adam _64_1e-4_	0.8003	0.3822	0.8167	0.5208	0.8073
Expériences Test	Exactitude	Précision	Rappel	F1-Score	AUC-ROC
Model_4_adam _32_1e-4_0.2	0.7549	0.2921	0.5897	0.3907	0.6850
Model_3_adam _32_1e-4_0.2	0.7566	0.2784	0.5192	0.3624	0.6562
Model_4_adam _64_1e-4_0.2	0.7635	0.2949	0.5577	0.3858	0.6764
Model_1_adam _64_1e-4_	0.7284	0.2500	0.5192	0.3375	0.6399

Pour la détection des spams, nous avons suivi la même ligne de réflexion que pour l'analyse de sentiment, étant donné que nous exploitons la polarité et les embeddings BERT, nous devions également prendre en compte la langue de l'avis. Ainsi, notre approche pour la détection des spams est également une méthode d'ensemble qui achemine l'avis en fonction de sa langue vers le modèle le plus approprié. Dans le tableau II.13, nous présentons les meilleurs résultats pour chaque modèle individuel (chaque langue), ainsi que le résultat final obtenu par la méthode d'ensemble.

TAB. II.13 : Meilleurs résultats obtenus avec le modèle de détection de spam pour chaque langue et pour la méthode d'ensemble.

Expériences Cross-Validation	Exactitude	Précision	Rappel	F1-Score	AUC-ROC
EN	0.80	0.38	0.83	0.82	0.81
FR	0.63	0.21	0.66	0.32	0.64
AR	0.83	0.42	0.70	0.52	0.77
Méthode d'ensemble	0.70	0.31	0.72	0.53	0.71
Expériences Test	Exactitude	Précision	Rappel	F1-Score	AUC-ROC
EN	0.73	0.25	0.52	0.34	0.6
FR	0.65	0.24	0.75	0.36	0.69
AR	0.76	0.27	0.51	0.33	0.63
Méthode d'ensemble	0.66	0.25	0.65	0.30	0.62

3.6 Agent conversationnel

En ce qui concerne l'agent conversationnel, nous avons d'abord testé des modèles génératifs. Cependant, ces modèles se sont avérés très complexes, lourds et souffraient d'un problème bien connu (un problème souvent fortement critiqué chez ChatGPT), à savoir que s'il ne connaît pas la réponse à quelque chose, il peut simplement dériver et générer des réponses incohérentes et sans signification. De plus, l'acquisition des données nécessaires pour affiner efficacement de tels modèles s'est avérée difficile.

C'est pourquoi nous avons décidé d'adopter une approche plus simple. Étant donné que notre tâche est assez spécifique, nous avons choisi de concevoir des flux de travail pour notre chatbot

et de le rendre basé sur la récupération d'informations. Le chatbot prend l'entrée de l'utilisateur et renvoie la réponse la plus adéquate après avoir vérifié sa base de données en utilisant un algorithme complet.

La base de données du chatbot se compose essentiellement de cinq tables (sans compter celle contenant les noms et codes des Wilayas algériennes). La première table contient des expressions régulières décrivant l'intention de l'utilisateur, que ce soit une question sur l'emplacement des magasins ou les produits proposés par notre marque, par exemple. Les deux tables suivantes contiennent des informations sur les magasins et les produits de la marque, respectivement. La quatrième table permet de capturer d'autres questions et déclarations faites par le client. Et enfin, la dernière table fait correspondre les questions aux réponses appropriées. Le Tableau II.14 présente des exemples de questions et les réponses renvoyées par le chatbot.

TAB. II.14 : Exemples de réponses du chatbot.

Entrée	Sortie
Hey Bbot, who's CEVITAL ?	The Cevital Group is an Algerian conglomerate of the food industry, mass distribution, industry and services. Created by the entrepreneur Issad Rebrab in 1998, Cevital is the first Algerian private group, also present internationally and the third Algerian company by turnover. It employs around 18,000 people. The Cevital group is the leader in the agri-food sector in Algeria.
Wshno homa llhwant li ybi3o les produits ta3 Brandt ?	Toutes nos excuses, la langue dans laquelle vous conversez n'est pas encore prise en charge par notre agent conversationnel. Pour le moment, notre agent ne peut converser avec vous qu'en français, arabe et anglais. Si vous avez fait une demande dans une langue prise en charge et que vous voyez cette réponse, veuillez formuler votre demande différemment. Toutes nos excuses pour le dérangement.
What type of products does your brand sell ?	Our brand specifies in appliances, we can find : televisions, washing machines, ovens, cook tops, hoods, cookers, fridges, air conditioners as well as other small appliances. Feel Free to visit our official website for more information, or ask and I'll do my best to assist you !
براندت هي علامتنا التجارية! يمكنك أن تسألني عنها إذا كنت ترغب في معرفة المزيد!	

3.7 Conclusion

En conclusion, nous avons mené de nombreuses expérimentations tout au long de ce projet. Certains de nos résultats auraient pu être améliorés si nous avions disposé de plus de ressources et de données. Cependant, malgré ces limitations, nos résultats sont prometteurs et constituent un excellent point de départ pour des recherches futures. Nous sommes convaincus que cette solution a un grand potentiel et qu'elle peut être encore affinée et améliorée avec davantage de ressources et d'efforts.

Chapitre 4

Prototype de la solution

4.1 Introduction

Dans ce chapitre, nous aborderons en détail le prototype de la solution que nous avons élaborée pour répondre aux besoins spécifiques de Brandt Algérie en matière de suivi et d'amélioration de la réputation en ligne. Nous nous concentrerons sur la mise en place de la plateforme "Brandt Insight", en examinant à la fois les perspectives de l'administrateur et des collaborateurs. De plus, nous explorerons la page d'accueil ainsi que la fonctionnalité du chatbot, tous deux conçus dans le but de faciliter une expérience utilisateur plus fluide et interactive. La charte graphique est présentée dans l'annexe 4.5

4.2 Architecture de la solution

Dans cette section, nous aborderons l'architecture de la solution, tant du point de vue de son architecture interne que des relations entre ses différents composants.

4.2.1 Architecture au niveau général

Dans cette section, nous allons aborder l'architecture générale de la solution, plus précisément au niveau physique. Cela concerne les serveurs, les bases de données ainsi que la manière dont les interactions entre eux sont modélisées.

Dans l'ensemble, nous avons adopté une architecture client-serveur. L'architecture client-serveur est un modèle dans lequel les fonctions sont divisées entre les fournisseurs de ressources (les serveurs) et les demandeurs de ressources (les clients). Les clients envoient des demandes aux serveurs pour obtenir des ressources ou des services, et les serveurs répondent en fournissant ces ressources ou services. Cela permet une répartition des tâches et une gestion plus efficace des ressources.

Notre solution est présentée aux utilisateurs sous la forme de deux plateformes web : une pour les employés de l'entreprise (avec une interface pour les collaborateurs et une pour l'administrateur, accessible via une page de connexion sécurisée avec des jetons d'authentification pour garantir l'aspect sécurité, qui est d'une importance primordiale), et une autre pour les clients de l'entreprise. Ces plateformes sont développées en ReactJS¹, une librairie JavaScript réputée pour sa flexibilité et sa réactivité. Les bonnes pratiques de ReactJS sont également mises en œuvre pour garantir l'attribut de qualité de réutilisabilité des différents composants.

Ces éléments concernent la partie frontend de notre solution. En ce qui concerne le backend, il est hébergé sur des serveurs spécifiques et optimisé pour garantir la disponibilité et la performance aux utilisateurs. Les serveurs sont configurés en utilisant deux ports distincts : un pour

¹<https://react.dev/>

le serveur de Brandt Insight et un pour le serveur de Brandt Bot. Ces deux serveurs partagent l'accès à des ressources, des fonctions de machine learning et à des modèles pré-entraînés (non accessibles depuis l'extérieur), avec le serveur de Brandt Bot n'accédant qu'à la fonction d'identification de la langue, tandis que le serveur de Brandt Insight utilise l'intégralité du pipeline de prédiction. Le backend a également accès à des bases de données PostgreSQL² spécialisées, un système de gestion de bases de données relationnelles, open source, reconnu pour sa stabilité, sa sécurité et ses performances.

La communication entre le frontend et le backend se fait via une API REST (Representational State Transfer)³. Une API REST est un ensemble de règles et de conventions qui permettent aux systèmes de communiquer entre eux de manière cohérente et efficace. Elle repose sur le protocole HTTP, ce qui la rend adaptée aux communications web. L'utilisation d'une API REST assure des interactions fluides et évolutives entre les différentes parties de notre solution, tout en maintenant une architecture modulaire et flexible.

4.2.2 Architecture au niveau logique

Dans cette section, nous allons approfondir les détails de l'architecture au niveau logique de l'architecture client-serveur. Nous aborderons d'abord les clients, puis les serveurs, et enfin, nous présenterons les bases de données.

4.2.2.1 Clients

Comme mentionné précédemment, nous avons deux clients : **BrandtInsight**, qui offre une interface à l'administrateur et une autre aux collaborateurs, accessibles via la même plateforme, mais avec des interfaces de connexion différentes (l'authentification se fait via Firebase, avec des fonctions supplémentaires accessibles via l'API du serveur BrandtInsight) ainsi que **la page d'accueil de Brandt** qui intègre **BBot**.

L'architecture des deux plateformes repose en fait sur un patron courant et largement utilisé avec ReactJS, souvent appelé modèle de composant-conteneur. Cela permet de séparer les préoccupations et d'optimiser la réutilisabilité du code. La figure II.11 présente l'arborescence du répertoire du projet résultant de l'application de ce modèle.

- **Assets** contient les différentes ressources à utiliser dans le projet (polices, images, logos, animations, etc.).
- **Components** contient des modules entièrement réutilisables et stylisés qui peuvent être utilisés dans l'ensemble du projet, tels que les boutons, les champs de saisie, les bannières, les messages pour le chatbot, les formulaires, etc.
- **Config** contient les configurations générales, notamment celles liées aux requêtes Axios⁴, aux configurations Firebase⁵, etc.
- **Containers** contient les pages et les structures importantes du projet.
- **Services** contient les fonctions d'appels API communes.
- **Utils** contient les fonctions communes partagées dans toute l'application, telles que les fonctions de validation, etc.

²<https://www.postgresql.org/>

³<https://restfulapi.net/>

⁴<https://axios-http.com/docs/intro>

⁵<https://firebase.google.com/>

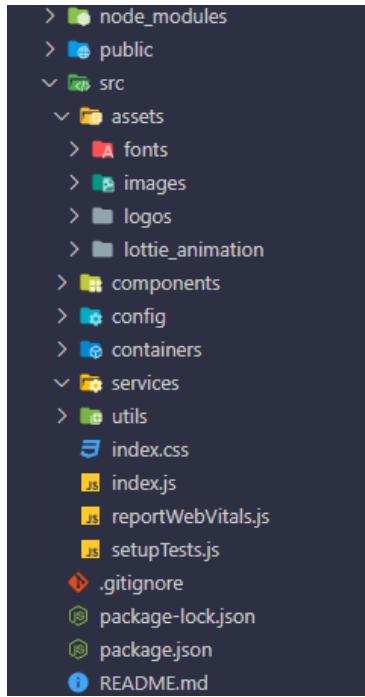


FIG. II.11 : Arborescence du répertoire d'un projet React

Cette organisation garantit une séparation claire des responsabilités et facilite la maintenance et l'extension de la solution tout en favorisant la réutilisabilité du code.

4.2.2.2 Serveurs

Nous disposons de deux serveurs accessibles via des appels API : BrandtInsight et BrandtBot. Ces serveurs partagent tous deux l'accès à nos fonctions de modèles ML singleton ainsi qu'à d'autres fichiers et ressources partagés (comme les dictionnaires d'emoji, les fonctions de prétraitement, etc.), comme illustré dans la figure II.12a ci-dessous.

- Serveur BrandtInsight

BrandtInsight suit une architecture de microservices, un concept qui privilégie la création d'applications par assemblage de petits services indépendants, chaque service ayant sa propre responsabilité et pouvant être développé et déployé de manière autonome. Cela permet une meilleure maintenabilité, évolutivité et réutilisabilité. Comme le montre la figure II.12b, nous avons actuellement trois microservices : **AccountManagement** (pour gérer les comptes, les connexions, etc.), **OpinionCollection** (responsable de la collecte, du filtrage et de la classification des opinions, ainsi que de la planification des exécutions et de la notification des utilisateurs) et **ProductCustomisation** (offrant des fonctions de personnalisation pour l'administrateur du produit).

Pour chaque microservice, nous avons adapté une architecture MVC (Modèle-Vue-Contrôleur), qui sépare la logique en trois composants principaux, le **modèle** qui gère la logique métier et la manipulation des données, la vue qui, dans notre contexte, est l'interface API exposée par le microservice et le contrôleur agit comme un intermédiaire entre le modèle et la vue.

Comme le montre la figure II.12c, nous avons une structure bien définie comprenant des routes, des contrôleurs, des services et des modèles. En plus de cela, nous incluons des fonctions d'infrastructure, telles que la connexion à la base de données, l'utilisation de Firebase pour la gestion des connexions et la vérification des jetons JWT, la planification des tâches avec un ordonnanceur, ainsi que la gestion des exceptions personnalisées et des fonctions utilitaires.

- Serveur BrandtBot

Le serveur de BrandtBot suit une architecture MVC plus simple que celle des microservices de BrandtInsight, car il s'agit d'un modèle relativement direct. bb_route définit les routes de la plateforme ainsi que les contrôleurs appropriés, bb_controllers appelle les services nécessaires et bb_service appelle la logique métier (les fonctions du modèle présentées dans bb_chatbot), comme le montre la figure II.12c.



FIG. II.12 : Architecture logique des serveurs de la solution

4.2.2.3 Base de données

La dernière sous-section de cette section concerne les différentes bases de données contenant les données nécessaires à la solution, qui sont accessibles via le backend comme précédemment présenté. Ces bases de données sont divisées en fonction de leur objectif respectif.

- Base de données BrandtInsight : Cette base de données contient les définitions des sources, les opinions collectées, les planifications, ainsi que d'autres données nécessaires à la solution BrandtInsight.
- Base de données BrandtBot : Cette base de données contient les différentes tables utilisées pour déterminer l'intention du client et la réponse la plus adéquate à son entrée.
- Base de données Gestion : Cette base de données contient les informations concernant les différents utilisateurs de la solution (l'administrateur et les collaborateurs).
- Base de données Personnalisation : Cette base de données contient différentes informations concernant l'outil pour adapter la solution au contenu de l'utilisateur de manière personnalisée.

4.3 Tableau de bord

Le tableau de bord est la partie principale de l'interface destinée à nos clients, et son rôle principal est la surveillance de l'e-réputation de la marque. Dans les sections suivantes, nous

allons présenter en détail cet outil, ses utilisateurs et les différentes interfaces qui le composent.

4.3.1 Acteurs

L'outil comprend deux plateformes principales, chacune destinée à un groupe d'utilisateurs spécifique : l'administrateur et le collaborateur.

4.3.1.1 Administrateur

L'administrateur représente les individus qui détiennent le site et sont responsables de sa maintenance. Ses préoccupations principales englobent la gestion des utilisateurs (ajout, suppression et modification des informations), la gestion des sources (ajout, activation, suppression, etc.), ainsi que la personnalisation de l'outil pour l'adapter aux besoins de la marque.

Ce type d'utilisateurs possède généralement des compétences informatiques avancées, ce qui nous donne une certaine latitude dans la création de fonctionnalités techniques pour l'outil. Cependant, il est essentiel de maintenir une interface simple et accessible pour répondre à ses besoins tout en facilitant son travail.

4.3.1.2 Collaborateur

Les collaborateurs sont les employés de l'entreprise et représentent les acteurs principaux de cet outil. Ils possèdent une compréhension solide de la gestion de la réputation de l'entreprise, souvent affiliés au service marketing de l'organisation.

Leurs objectifs incluent la gestion et la compréhension de la réputation de la marque, en utilisant les différents indicateurs de performance et les statistiques à leur disposition. Ils assurent également une veille proactive de la réputation grâce aux notifications et aux rapports périodiques méticuleusement conçus pour transmettre les informations les plus pertinentes.

4.3.2 Collecteur d'avis

Le collecteur d'avis constitue un élément essentiel de notre système, ayant pour mission de collecter les avis concernant la marque à partir de diverses sources spécifiées par l'administrateur de l'outil. Pour l'instant, notre solution est configurée pour recueillir des données uniquement à partir des réseaux sociaux, en se focalisant plus particulièrement sur les groupes et les pages Facebook. Toutefois, il est important de noter que notre code est conçu de manière extensible, ce qui signifie que nous avons anticipé la possibilité d'intégrer facilement d'autres types de sources de collecte à l'avenir.

Afin de garantir une conformité avec les directives et les règlements des différentes plateformes de collecte, nous avons mis en place des mesures spécifiques. Tout d'abord, nous avons mis l'accent sur la protection des données personnelles des utilisateurs. Cela se traduit par le fait que nous ne collectons que des informations non identifiables sur les auteurs des avis. En d'autres termes, nous conservons uniquement des données anonymes permettant d'analyser les caractéristiques comportementales à des fins de détection de spam, sans aucune collecte d'informations personnelles sensibles.

De plus, nous accordons une grande importance à ne pas surcharger les serveurs sources lors de la collecte des données. Pour atteindre cet objectif, nous avons mis en place des mécanismes d'espacement entre les requêtes de collecte. Cette approche permet de garantir que nos activités de collecte n'impactent pas négativement les performances ou la disponibilité des serveurs sources.

En résumé, notre collecteur d'avis repose sur des principes de conformité rigoureuse, de protection des données personnelles et d'efficacité opérationnelle, garantissant ainsi la qualité et la

légitimité des informations collectées tout en préservant de bonnes relations avec les plateformes sources.

4.3.3 Indicateurs clés de performance

Dans cette partie, nous définirons les indicateurs de performance et statistiques que nous mettons à disposition des utilisateurs afin de surveiller et de gérer l'e-réputation de l'entreprise. Nous discuterons d'abord de celles présentées dans le tableau de bord puis celles du rapport périodique.

4.3.3.1 Tableau de bord

Le tableau de bord a comme objectif principal fournir une vue détaillée et à jour de la réputation en ligne de la marque. Pour ce faire, nous avons élaboré différents indicateurs de performance spécifiques à chaque aspect que nous souhaitons analyser. Parmi ces indicateurs, nous mentionnons :

- Indicateurs spécifiques aux sources de collecte :

- Nombre de sources : Ceci fait référence au nombre total de sources de collecte surveillées. Une augmentation du nombre de sources peut indiquer une plus grande exposition en ligne de la marque.

- Distribution des avis par source : Il s'agit de répartir les avis collectés en fonction de leur source d'origine, ce qui peut révéler quels canaux sont les plus influents.

- Répartition des sources par type de source : Il s'agit de classer les sources en catégories, selon le type de la source (page Facebook, groupe Facebook, etc). Cela aide à identifier les canaux les plus pertinents pour l'engagement de la marque.

- Polarité par source : Il s'agit d'analyser la tendance des avis (positifs, négatifs ou neutres) générés par chaque source. Ce qui permet de déterminer quels canaux ont tendance à présenter une opinion positive ou négative.

- Authenticité par source : L'authenticité des avis est cruciale pour évaluer leur crédibilité. Cet indicateur permet d'examiner si certaines sources ont tendance à générer des avis plus authentiques que d'autres.

- Indicateurs spécifiques au sentiment :

- Distribution des avis par sentiment : Cet indicateur permet de répartir les avis en fonction de leur sentiment. Il offre un aperçu rapide de l'opinion globale des clients à l'égard de la marque.

- Aperçu de la Polarité des avis : Il examine la polarité globale des avis exprimés envers la marque.

- Tendance du sentiment au fil du temps : Cet indicateur suit les fluctuations du sentiment des avis au fil du temps, permettant d'identifier des tendances à court et à long terme.

- Distribution du score de sentiment par source : Cette visualisation permet d'identifier rapidement les sources de collecte générant le plus d'opinions positives, ainsi que celles engendrant le plus d'opinions négatives.

- Indicateurs spécifiques à l'authenticité :

- Distribution des avis par authenticité : Cet indicateur permet de répartir les avis en fonction de leur authenticité, identifiant les avis authentiques et mettant en évidence tout contenu suspect ou indésirable.

- Aperçu de l'authenticité des avis : Il évalue l'authenticité globale des avis collectés.
 - Tendance des avis spam au fil du temps : Cet indicateur suit la fréquence des avis considérés comme du spam au fil du temps, permettant de réagir rapidement aux activités non authentiques.
 - Corrélation entre le spam et la polarité : Il identifie si le spam est de nature positive ou négative, ce qui peut donner des informations sur les tentatives de manipulation de la réputation de la marque.
 - Distribution du score de crédibilité par source : Cette représentation visuelle facilite la distinction entre les sources de collecte les plus fiables et les moins fiables.
- Indicateurs spécifiques à la langue :
 - Distribution des avis par langue : Cet indicateur permet de répartir les avis en fonction de la langue dans laquelle ils ont été exprimés, offrant des informations sur la diversité linguistique des opinions sur la marque.
 - Classement des langues les plus utilisées pour les avis : Il identifie les langues les plus fréquemment utilisées par les auteurs d'avis, ce qui peut être essentiel pour cibler des audiences spécifiques.
 - Autres statistiques linguistiques montrant les relations entre la langue de l'opinion et différentes mesures comme sa polarité, authenticité et pertinence, ce qui peut fournir d'intéressantes informations sur les audiences cibles.
 - Indicateurs divers :
 - Nombre d'Avis Collectés : Cet indicateur fait référence au nombre total d'avis recueillis, offrant une vue globale de l'activité en ligne liée à la marque.
 - Distribution des avis par Type de Publication : Il classe les avis en fonction de leur type de publication, tels que les publications, les commentaires ou les réponses. Cela peut révéler où les clients sont les plus actifs et où leurs opinions sont le plus susceptible d'être exprimées.
 - Distribution des avis par Type : Cet indicateur permet de répartir les avis en fonction de leur type, comme les avis de comparaison, les questions ou les descriptions. Il permet de comprendre les différents aspects abordés par les clients.

Ces indicateurs sont essentiels pour une analyse approfondie de la réputation en ligne de la marque, permettant de surveiller les tendances, d'identifier les problèmes potentiels et de prendre des mesures pour améliorer l'image de la marque. Le tableau de bord est conçu pour offrir une vue d'ensemble complète de ces données à travers des statistiques et des diagrammes, facilitant ainsi la prise de décisions en matière de gestion de la réputation en ligne.

4.3.3.2 Rapport périodique

Le rapport périodique est un élément essentiel de notre système d'analyse de la réputation en ligne. Il s'agit d'un rapport envoyé régulièrement, par défaut toutes les trois heures environ, à la boîte de réception de l'utilisateur sur la plateforme, ainsi qu'à son adresse e-mail. L'objectif de ce rapport est de tenir l'utilisateur informé des nouvelles données concernant la marque au cours d'une période spécifique, s'il y en a. Les données contenues dans ce rapport sont spécifiques à une seule source de collecte et visent à résumer les changements survenus au cours de cette période.

Parmi les informations clés incluses dans ce rapport, nous avons choisi de fournir : des informations sur la Source, le nombre d'avis collectés, un aperçu de la Polarité (si les opinions collectées sont majoritairement positives, négatives ou neutres), un aperçu de l'authenticité des

Chapitre 4. Prototype de la solution

opinions collectées, ainsi qu'un aperçu sur leur pertinence.

Le rapport périodique comprend également plusieurs graphiques qui visent à présenter visuellement ces données pour une meilleure compréhension :

- Diagramme circulaire pour la distribution des sentiments : Ce graphique permet de visualiser la répartition des sentiments parmi les avis collectés. Il indique si la majorité des avis sont positifs, négatifs ou neutres, offrant ainsi un aperçu rapide de l'opinion générale des clients à l'égard de la marque.
- Diagramme en donut pour la distribution des spams : Ce graphique présente la répartition des avis qui peuvent être considérés comme du spam. Il est essentiel pour identifier et gérer tout contenu non authentique ou indésirable.
- Graphiques à barres pour les types d'avis, les Types de Publications et la distribution des langues : Ces graphiques offrent des informations détaillées sur la diversité des avis collectés, notamment les types d'avis (comparaisons, questions, descriptions, etc.), les types de publications (publications, commentaires, réponses), et la répartition des langues dans les avis. Ils aident à comprendre la nature variée des interactions en ligne concernant la marque.

En somme, le rapport périodique est conçu pour tenir l'utilisateur au courant des évolutions significatives concernant la réputation de la marque. Il combine des données descriptives et visuelles pour fournir une vue d'ensemble complète de la perception de la marque par les clients au fil du temps, facilitant ainsi la prise de décisions éclairées. La figure II.13 illustre une portion d'un exemple de rapport périodique.

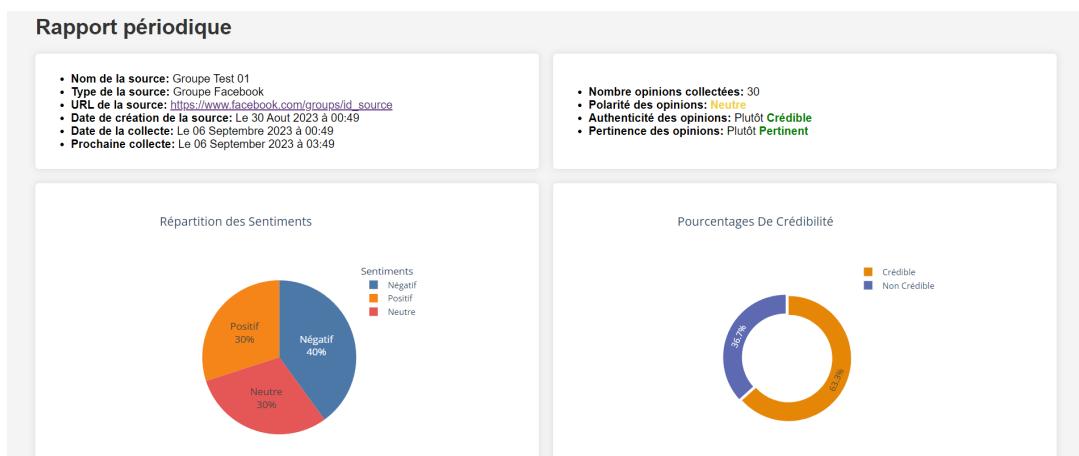


FIG. II.13 : Aperçu d'un rapport périodique exemple

4.3.4 Plateforme collaborateur

La plateforme destinée aux collaborateurs est la partie principale de notre outil permettant au collaborateur de suivre l'e-réputation de la marque à travers divers statistiques et diagrammes. Dans ce qui suit, nous présenterons les différentes interfaces de cette plateforme.

4.3.4.1 Interface de connexion

L'interface de connexion, illustrée dans la figure II.14 est la première interface de la plateforme permettant aux utilisateurs ayant un compte (créé par l'administrateur) de se connecter en utilisant son email et son mot de passe et d'accéder aux différentes fonctionnalités. Si l'utilisateur tentant de se connecter est un administrateur souhaitant accéder à la plateforme d'administration, il peut naviguer vers son interface de connexion avec un simple clique.

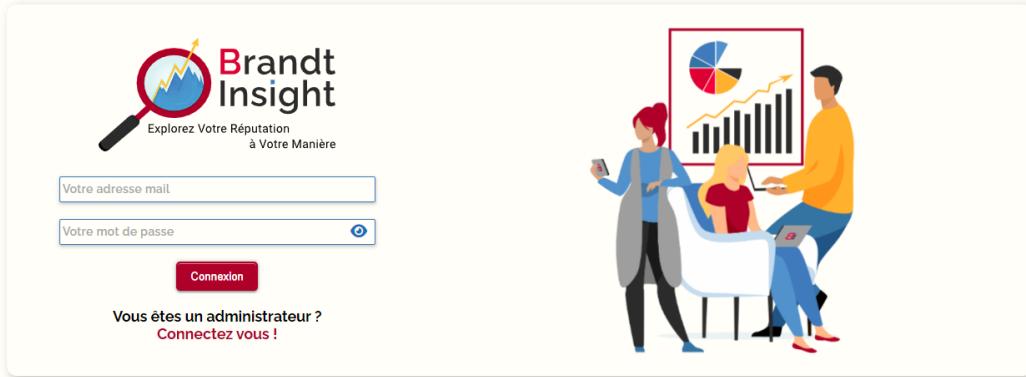


FIG. II.14 : Interface de connexion de la plateforme collaborateur

4.3.4.2 Interface de suivi d'e-réputation

L'interface de suivi de l'e-réputation est la première interface que l'utilisateur découvre après s'être connecté à la plateforme. De manière générale, elle est composée d'un en-tête, d'un menu et d'un pied de page, en plus de l'espace réservé au contenu, comme illustré dans la figure II.15.

L'en-tête permet à l'utilisateur de naviguer dans le tableau de bord, d'accéder à son compte ou de se déconnecter, et de consulter ses notifications. Les notifications peuvent concerner de nouvelles sources ajoutées ou des notifications de rapports périodiques. L'utilisateur a la possibilité de marquer une notification comme lue et de télécharger un rapport périodique s'il y en a un de disponible. Le pied de page, quant à lui, est simple et affiche la date et l'heure, ainsi que différents liens vers les réseaux sociaux de la marque.



FIG. II.15 : Interface du tableau de bord de la plateforme collaborateur

Le menu, de sa part, permet à l'utilisateur de naviguer entre les différents tableaux de bord de la plateforme. Actuellement, l'utilisateur a accès à six tableaux de bord, chacun étant conçu pour offrir une expérience efficace dans l'analyse de la réputation en ligne de la marque :

- Vue d'Ensemble

Chapitre 4. Prototype de la solution

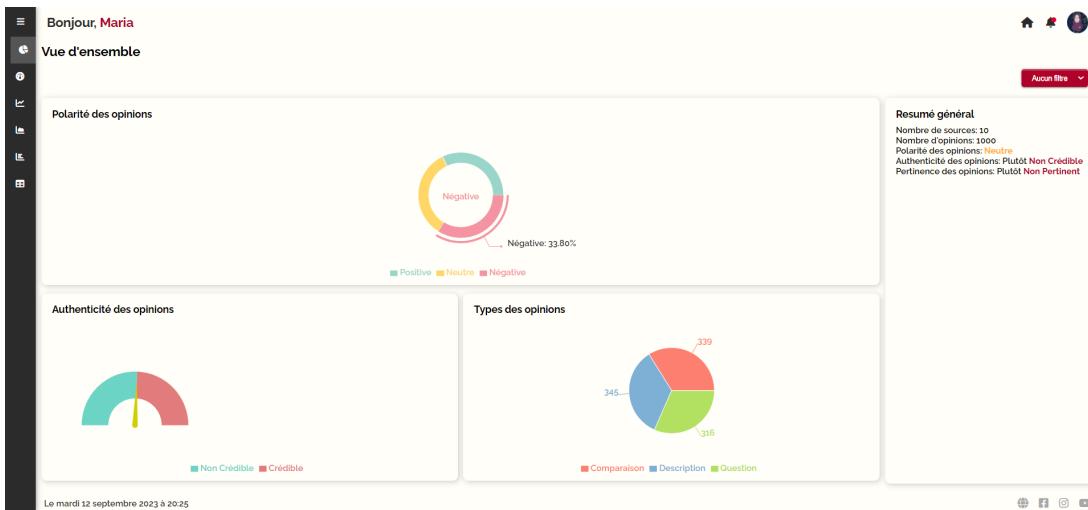


FIG. II.16 : Interface de la vue d'ensemble de la plateforme collaborateur

Le premier tableau de bord fournit au collaborateur une vue d'ensemble des données concernant la réputation, comme illustré dans la figure II.16. Il contient un résumé des différentes sources, du nombre d'avis, etc. Il présente également différents diagrammes sur la répartition des avis par polarité, authenticité et types.

- Analyse de Sentiments

La deuxième vue, illustrée dans la figure II.17, offre divers aperçus sur l'aspect de la polarité, avec un diagramme pour la distribution des avis par sentiment, un pour la distribution du score de sentiment par source et un autre pour la tendance des sentiments au fil du temps. Permettant aux collaborateurs de suivre l'évolution de la polarité des opinions et d'identifier des tendances à court et à long terme.



FIG. II.17 : Interface de la vue d'analyse de sentiments de la plateforme collaborateur

- Analyse de Crédibilité

La troisième perspective, comme illustré dans la figure II.18, présente divers aperçus sur l'aspect de l'authenticité en utilisant plusieurs représentations visuelles :

- Un diagramme présentant la répartition des avis en fonction de leur crédibilité, offrant ainsi aux collaborateurs une vision globale de la crédibilité des avis collectés.
- Un diagramme montrant l'évolution des avis de spam au fil du temps, permettant aux

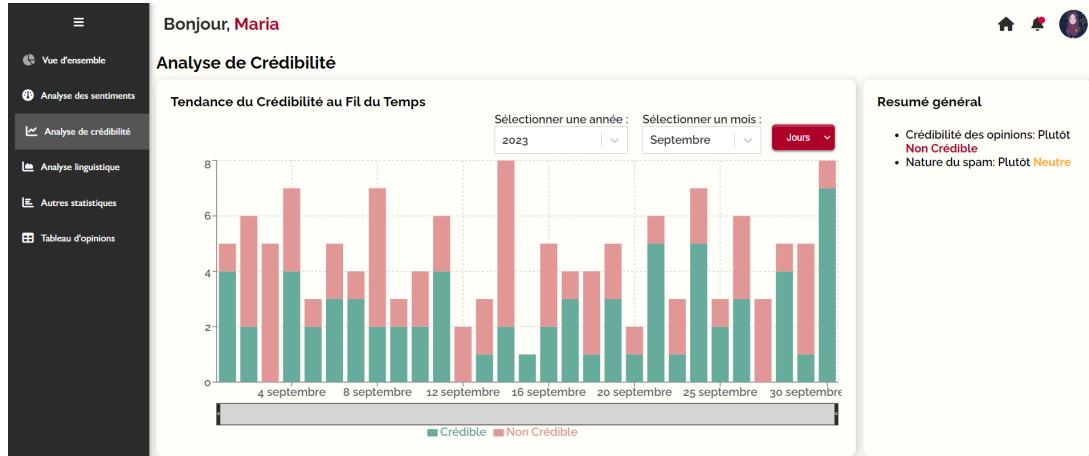
Chapitre 4. Prototype de la solution

collaborateurs de détecter d'éventuelles tendances ou variations dans la présence de spam.

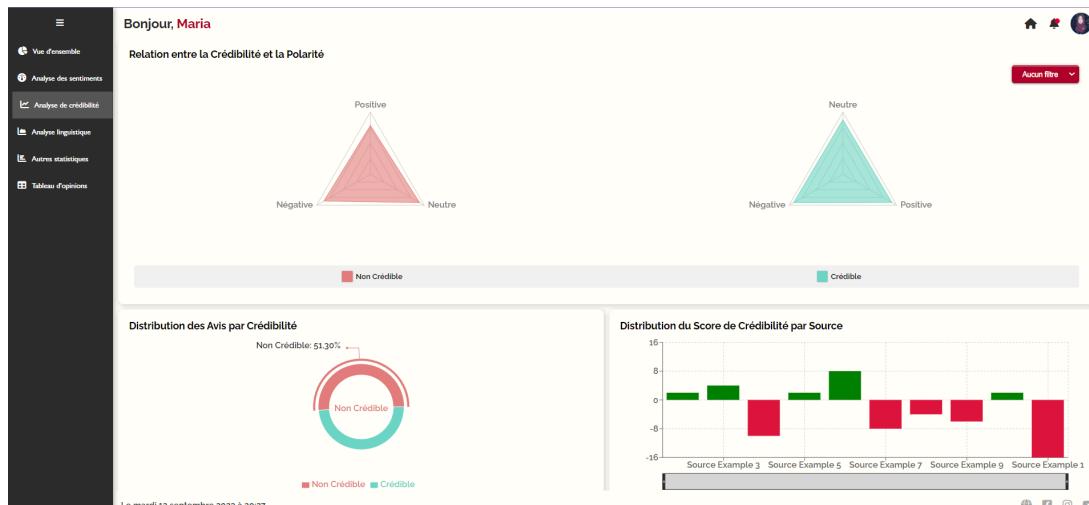
- Un diagramme présentant la distribution des scores de crédibilité par source, fournissant un aperçu des sources les plus fiables en termes d'authenticité.

- Un diagramme illustrant la corrélation entre l'authenticité et la polarité, aidant les collaborateurs à comprendre la relation entre le sentiment exprimé dans un avis et son niveau d'authenticité.

Cette vue détaillée permet aux collaborateurs de détecter d'éventuelles tendances et attaques de spam qui pourraient affecter la réputation de leur marque.



(a) Partie 01



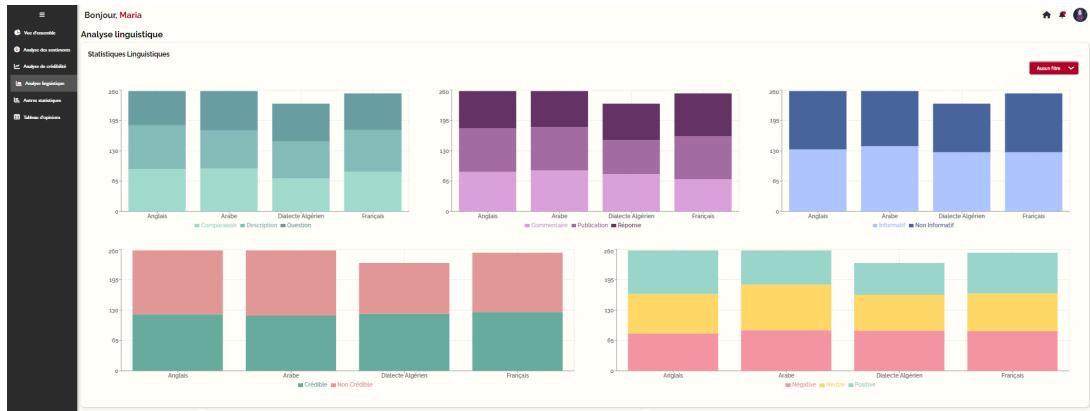
(b) Partie 02

FIG. II.18 : Interface de la vue d'analyse de crédibilité de la plateforme collaborateur

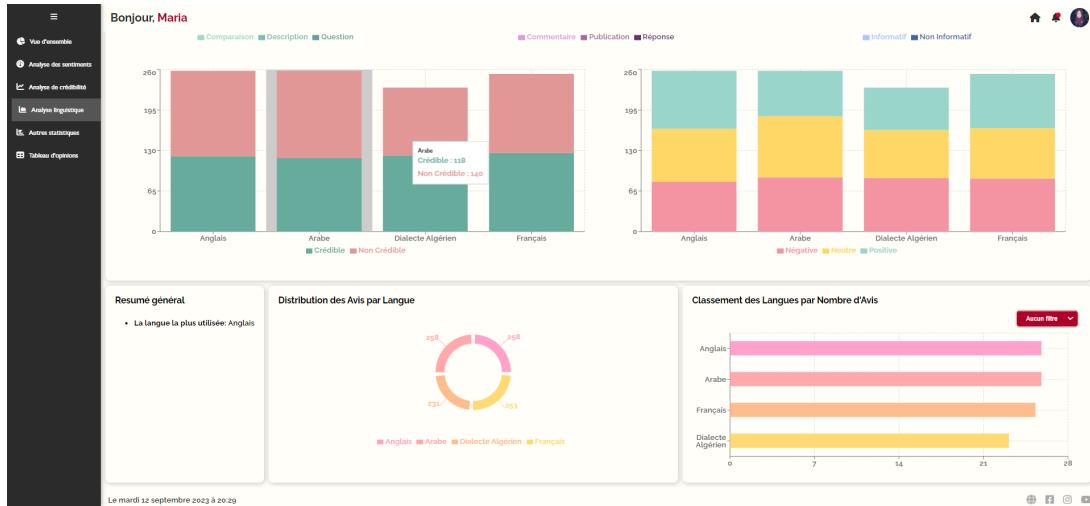
- Analyse de Langues

La quatrième vue, illustrée dans la figure II.19, se concentre sur l'aspect linguistique. Elle présente apporte des informations concernant la distribution des avis par langue ainsi qu'un classement des langues selon leurs taux d'utilisation dans les opinions collectées. Cette vue présente aussi des diagrammes illustrant la relation entre la langue et différentes autres mesures telles que le type de l'opinion, sa crédibilité ou sa polarité. Ceci peut offrir des informations intéressantes concernant la diversité linguistique du public de la marque.

Chapitre 4. Prototype de la solution



(a) Partie 01



(b) Partie 02

FIG. II.19 : Interface de la vue d'analyse de langues de la plateforme collaborateur

- Statistiques Diverses



FIG. II.20 : Interface de la vue de statistiques divers de la plateforme collaborateur

La cinquième vue, montrée dans la figure II.20, présente diverses autres mesures. Certaines concernent le nombre des avis, telles que le nombre total d'avis collectés et l'évolution du nombre des opinions au fil du temps. D'autres concernent la source, telle que le nombre de sources, la distribution des avis par source, la répartition des sources par type de source, la polarité par

Chapitre 4. Prototype de la solution

source, et l'authenticité par source. De plus, elle offre des informations sur la distribution des avis par type de publication, et la distribution des avis par type.

La dernière vue, visible dans la figure II.21, est un tableau présentant les avis collectés. Elle permet à l'utilisateur de trier et de filtrer les avis, ainsi que de télécharger le tableau affiché.

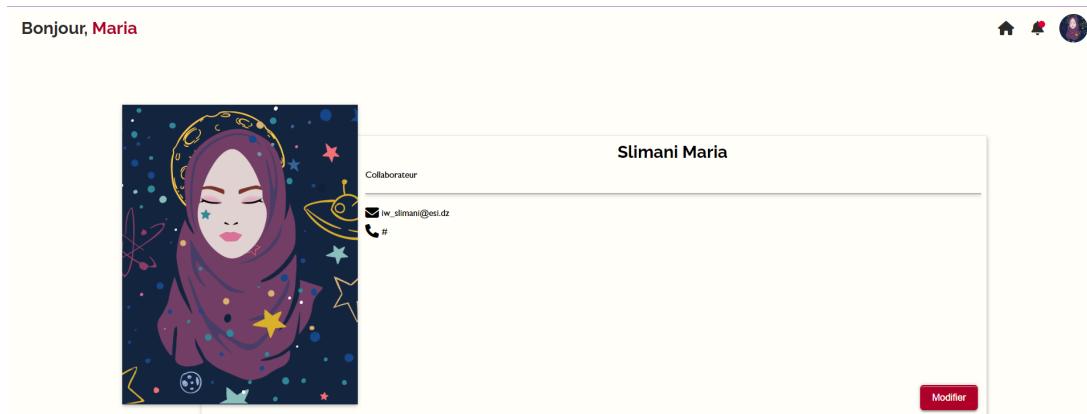
The screenshot shows a user interface titled "Bonjour, Maria". On the left is a sidebar with icons for navigation. The main area displays a table of 16 rows, each representing a review example. The columns are labeled: #, Source, Opinion, Type de publication, Sentiment, Pertinence, Authenticité, Langue, Type, Date, and Lien. The table includes a header row and footer links for "Aucun filtre", "Télécharger", and "Réinitialiser les filtres".

#	Source	Opinion	Type de publication	Sentiment	Pertinence	Authenticité	Langue	Type	Date	Lien
1	Source Example 2	This is review 1.	Commentaire	Positive	Non Informatif	Non Crédible	Anglais	Comparaison	2023-05-16	
2	Source Example 3	This is review 2.	Commentaire	Neutre	Non Informatif	Crédible	Arabe	Comparaison	2023-05-27	
3	Source Example 4	This is review 3.	Commentaire	Positive	Informatif	Crédible	Anglais	Description	2023-09-06	
4	Source Example 5	This is review 4.	Réponse	Négative	Informatif	Crédible	Arabe	Comparaison	2023-06-01	
5	Source Example 6	This is review 5.	Réponse	Neutre	Informatif	Crédible	Dialecte Algérien	Question	2023-02-19	
6	Source Example 7	This is review 6.	Publication	Négative	Informatif	Crédible	Français	Description	2023-07-10	
7	Source Example 8	This is review 7.	Commentaire	Négative	Informatif	Crédible	Anglais	Description	2023-02-15	
8	Source Example 9	This is review 8.	Publication	Neutre	Informatif	Non Crédible	Français	Comparaison	2023-04-29	
9	Source Example 10	This is review 9.	Publication	Neutre	Informatif	Non Crédible	Arabe	Description	2023-05-26	
10	Source Example 1	This is review 10.	Réponse	Neutre	Non Informatif	Crédible	Anglais	Description	2023-09-19	
11	Source Example 2	This is review 11.	Commentaire	Positive	Non Informatif	Non Crédible	Arabe	Question	2023-05-08	
12	Source Example 3	This is review 12.	Réponse	Positive	Informatif	Crédible	Anglais	Comparaison	2023-06-12	
13	Source Example 4	This is review 13.	Réponse	Neutre	Informatif	Non Crédible	Arabe	Comparaison	2023-03-27	
14	Source Example 5	This is review 14.	Réponse	Positive	Non Informatif	Non Crédible	Anglais	Question	2023-09-06	
15	Source Example 6	This is review 15.	Publication	Négative	Informatif	Crédible	Français	Description	2023-04-12	
16	Source Example 7	This is review 16.	Publication	Neutre	Informatif	Crédible	Arabe	Description	2023-06-30	

FIG. II.21 : Interface de la vue du tableau d'opinions de la plateforme collaborateur

4.3.4.3 Interface de gestion de compte

L'interface de gestion de compte est simple et intuitive permettant au collaborateur de visionner ses informations ainsi que de les modifier. La figure II.22 illustre l'interface d'un utilisateur souhaitant modifier ces informations.



(a) Section 01 : Visualisation des informations personnelles

Bonjour, Maria

Votre nom: Slimani

Votre prénom: Maria

Votre adresse mail: iw_simmani@esi.dz

Votre numéro de téléphone: #

Importer la photo de profil Modifier la photo de profil Modifier

Votre mot de passe Confirmez votre mot de passe

Modifier le mot de passe Annuler

Le mardi 12 septembre 2023 à 21:11



(b) Partie 02 : Modification des informations personnelles

FIG. II.22 : Interface de gestion de compte de la plateforme collaborateur

4.3.5 Plateforme administrateur

La plateforme destinée aux administrateurs leur permet de gérer l'outil et d'assurer son bon fonctionnement. Dans ce qui suit, nous présenterons les différentes interfaces de cette plateforme.

4.3.5.1 Interface de connexion

L'interface de connexion, illustrée dans la figure II.23 est,似ilairement à la plateforme collaborateur, la première interface de la plateforme permettant à l'administrateur de se connecter en utilisant son email et son mot de passe. Si l'utilisateur tentant de se connecter est un collaborateur souhaitant accéder à la plateforme de suivi, il peut naviguer vers son interface de connexion avec un simple clique.

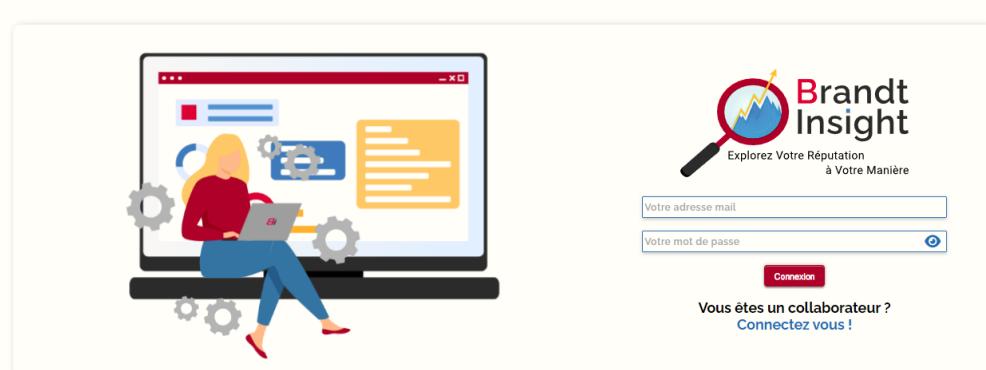


FIG. II.23 : Interface de connexion de la plateforme administrateur

4.3.5.2 Interface de gestion de l'outil

#	Plateforme	Source	Date de création	Nombre D'opinions	Exclusive?	Besoin de compte?	Active?	Lien	Actions
1	Platform E	Source Example 1	2023-09-06	100	Non	Oui	Non		
2	Platform D	Source Example 2	2023-09-07	100	Non	Non	Oui		
3	Platform D	Source Example 3	2023-09-08	100	Oui	Non	Oui		
4	Platform B	Source Example 4	2023-09-09	100	Non	Oui	Oui		
5	Platform A	Source Example 5	2023-09-10	100	Non	Oui	Non		
6	Platform D	Source Example 6	2023-09-11	100	Oui	Non	Non		
7	Platform C	Source Example 7	2023-09-12	100	Non	Non	Oui		
8	Platform C	Source Example 8	2023-09-13	100	Non	Oui	Oui		
9	Platform D	Source Example 9	2023-09-14	100	Non	Oui	Oui		
10	Platform G	Source Example 10	2023-09-15	100	Oui	Non	Non		

FIG. II.24 : Interface de gestion de l'outil de la plateforme administrateur

L'interface de gestion de l'outil, représentée dans la figure II.24, est spécifiquement conçue pour permettre à l'administrateur de gérer divers aspects de l'outil. Cela inclut la gestion des sources, avec des fonctionnalités telles que l'ajout, la suppression ou la modification des informations d'une source. De plus, l'administrateur peut activer ou désactiver une source de collecte selon les besoins.

4.3.5.3 Interface de gestion des employés

La figure II.25, offre à l'administrateur la possibilité de gérer les utilisateurs de l'outil. Il peut créer de nouveaux comptes, supprimer des comptes existants et modifier les informations de ces comptes, ce qui est essentiel pour garantir un accès sécurisé et approprié à l'outil.

#	Nom	Prénom	Email	Téléphone	Administrateur?	Actions	
1	nom_1	prenom_1	user1@example.com	tfn_1	Oui		
2	nom_2	prenom_2	user2@example.com	tfn_2	Non		
3	nom_3	prenom_3	user3@example.com	tfn_3	Non		
4	nom_4	prenom_4	user4@example.com	tfn_4	Oui		
5	nom_5	prenom_5	user5@example.com	tfn_5	Oui		
6	nom_6	prenom_6	user6@example.com	tfn_6	Oui		
7	nom_7	prenom_7	user7@example.com	tfn_7	Oui		
8	nom_8	prenom_8	user8@example.com	tfn_8	Oui		
9	nom_9	prenom_9	user9@example.com	tfn_9	Non		
10	nom_10	prenom_10	user10@example.com	tfn_10	Non		

FIG. II.25 : Interface du tableau de bord de la plateforme administrateur

4.3.5.4 Interface de statistiques et de personnalisation

L'interface de statistiques et de personnalisation, visible dans la figure II.26, fournit à l'administrateur des informations sur l'outil, tel que les performances des différents modèles. De plus, cette interface permet à l'administrateur de personnaliser certains aspects de l'outil, y compris la page conversationnelle, en sélectionnant les avis qui y seront affichés.

The screenshot shows a dashboard interface with a sidebar on the left containing links for 'Gestion du tableau de bord', 'Gestion des utilisateurs', and 'Statistiques et Personnalisation'. The main area displays a table titled 'Performances des modèles de prédition' with columns: #, Model Name, Accuracy, Precision, Recall, F1 Score, and AUC-ROC. The table contains five rows of data. At the bottom, there are navigation buttons for 'Précédent', '1', 'Suivant', and download/reload buttons.

#	Model Name	Accuracy	Precision	Recall	F1 Score	AUC-ROC
1	Model B	35.17	43.61	47.19	23.75	62.44
2	Model D	9.19	82.23	4.60	49.04	19.53
3	Model A	18.63	23.91	63.08	25.92	25.06
4	Model D	94.86	61.99	75.60	16.09	68.11
5	Model B	90.48	32.30	40.06	88.11	50.46

(a) Partie 01

The screenshot shows a dashboard interface with a sidebar on the left. The main area displays a table with columns: #, Source, Opinion, Type de publication, Langue, Type, Date, Lien, and Actions. The table contains 10 rows of data. At the top right, there are buttons for 'Aucun filtre', 'Télécharger', and 'Réinitialiser les filtres'. Below the table, there are navigation buttons for 'Précédent', '1', '2', '...', '5', 'Suivant', and a date stamp 'Le mardi 12 septembre 2023 à 20:51'. Social media sharing icons are at the bottom right.

#	Source	Opinion	Type de publication	Langue	Type	Date	Lien	Actions
1	Source Example 6	This is review 145.	<input checked="" type="button"/> Ordonner	Anglais	Comparaison	2023-05-04	<input checked="" type="button"/>	<input type="button"/>
2	Source Example 9	This is review 148.	<input checked="" type="button"/> Par ordre croissant	Dialecte Algérien	Description	2023-05-27	<input checked="" type="button"/>	<input type="button"/>
3	Source Example 10	This is review 169.	<input checked="" type="button"/> Par ordre décroissant	Arabe	Comparaison	2023-04-22	<input checked="" type="button"/>	<input type="button"/>
4	Source Example 8	This is review 177.	<input checked="" type="button"/> Filtrer	Arabe	Comparaison	2023-08-20	<input checked="" type="button"/>	<input type="button"/>
5	Source Example 3	This is review 202.	<input checked="" type="button"/> Sélectionner tout	Anglais	Comparaison	2023-05-19	<input checked="" type="button"/>	<input type="button"/>
6	Source Example 9	This is review 248.	<input checked="" type="button"/> Commentaire	Dialecte Algérien	Comparaison	2023-09-17	<input checked="" type="button"/>	<input type="button"/>
7	Source Example 3	This is review 252.	<input checked="" type="button"/> Publication	Arabe	Description	2023-05-15	<input checked="" type="button"/>	<input type="button"/>
8	Source Example 6	This is review 255.	<input checked="" type="button"/> Réponse	Anglais	Description	2023-02-11	<input checked="" type="button"/>	<input type="button"/>
9	Source Example 1	This is review 270.	<input checked="" type="button"/> Commentaire	Anglais	Comparaison	2023-09-05	<input checked="" type="button"/>	<input type="button"/>
10	Source Example 5	This is review 294.	<input checked="" type="button"/> Publication	Dialecte Algérien	Description	2023-04-05	<input checked="" type="button"/>	<input type="button"/>

(b) Partie 02

FIG. II.26 : Interface du tableau de bord de la plateforme administrateur

4.3.5.5 Interface de gestion de compte

L'interface de gestion de compte pour l'administrateur est similaire à celle des collaborateurs. Elle offre des fonctionnalités de base pour gérer les informations de compte, y compris le visionnage des informations personnelles et leur modification.

4.4 Page d'accueil conversationnelle

Dans ce qui suit, nous allons présenter la page d'accueil conversationnelle, ses acteurs principaux, son interface ainsi que le chatbot intégré.

4.4.1 Acteurs

Les acteurs principaux de la page d'accueil conversationnelle sont les clients de la marque, issus de divers milieux, qui utilisent l'interface à des fins variées telles que les demandes d'information, les retours d'expérience et la résolution de problèmes. Ces clients représentent une variété de compétences en informatique, elles peuvent aller de basique à intermédiaire. Ils ont généralement la capacité d'accéder à un site web, d'effectuer des recherches et d'interagir efficacement avec les fonctionnalités proposées par les interfaces web.

L'interface conversationnelle agit comme un lien direct et dynamique entre la marque et ces

clients. C'est pourquoi l'interface doit demeurer simple et intuitive afin de répondre aux besoins variés de ces utilisateurs aux niveaux d'expérience divers. L'objectif consiste à rendre l'interaction fluide et accessible, même pour ceux qui ont une connaissance limitée des technologies en ligne.

4.4.2 Interface de la page d'accueil

Abordons maintenant l'interface de la page d'accueil, la figure II.27 présente de manière concise le zonage de cette interface en mettant en évidence les différents éléments qui la constituent ainsi que les différentes sections qui la structurent. Dans le but de garantir une expérience utilisateur optimale, nous avons opté pour un design minimaliste et à une seule page. En utilisant des éléments visuels soignés et une disposition stratégique, nous avons cherché à offrir une expérience fluide, en mettant en avant les informations essentielles de manière attrayante et facilement accessible. Dans ce qui suit, nous allons examiner chaque composant de l'interface individuellement, en suivant l'ordre de numérotation présenté dans la figure. Cela nous permettra de bien présenter les composants tout en expliquant leurs fonctionnalités et la logique derrière nos choix de conception.

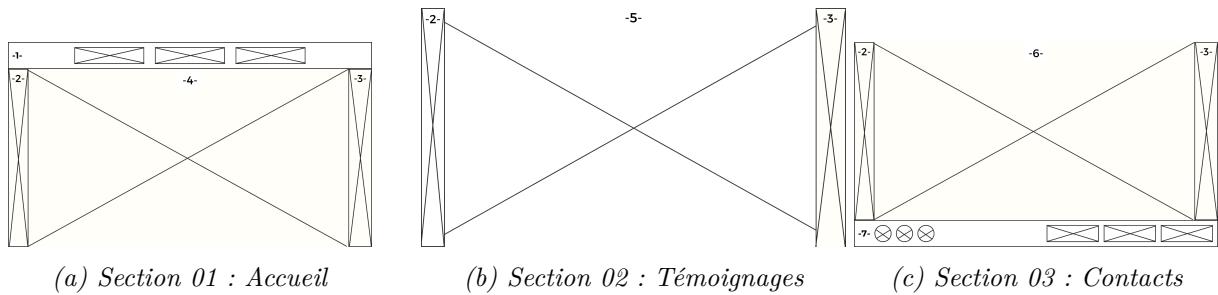


FIG. II.27 : Schéma de Zonage de la Page d'Accueil Conversationnelle

4.4.2.1 En-tête

Le composant identifié par le numéro -1- dans la figure II.27 représente un en-tête minimaliste qui occupe la partie supérieure de la page. Ce composant est conçu pour une navigation aisée vers les différentes sections du site, à savoir la section d'accueil, la section des témoignages et celle dédiée aux contacts, comme illustré dans la figure II.28. L'objectif principal de cet en-tête est de fournir aux utilisateurs un moyen simple et direct d'accéder rapidement aux informations clés. Conçu dans une optique minimaliste, cet en-tête adopte un fond couleur charbon, en cohérence avec la palette visuelle fréquemment employée par Brandt. Cette approche renforce l'alignement avec l'identité visuelle de la marque et garantit une expérience utilisateur fluide et cohérente.

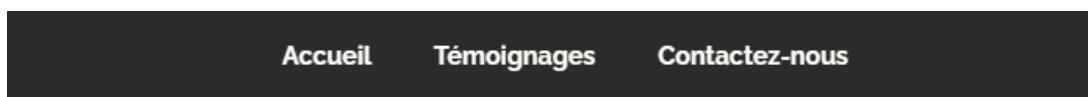


FIG. II.28 : En-tête de la Page d'Accueil Conversationnelle

4.4.2.2 Menu de navigation

Le composant identifié par le numéro -2- dans la figure II.27 concerne l'espace où flotte le menu de navigation, une zone considérée particulièrement accessible. Comme le montre l'illustration II.29, le menu est suffisamment compact pour ne pas perturber le design de la page lorsque l'utilisateur fait défiler le contenu, préservant ainsi l'expérience utilisateur. Ce menu permet à l'utilisateur d'accéder aisément aux différentes sections de la page, même lorsqu'il se trouve sur d'autres parties de la page. Nous avons envisagé simplement maintenir l'en-tête en haut de la page lorsque l'utilisateur fait défiler, mais avons finalement opté contre cette option, puisque cette méthode consommerait une quantité considérable d'espace, or l'approche entreprise évite justement cet inconvénient et nous permet de profiter de la totalité de l'écran disponible. Ainsi,

le design de la page conserve un aspect aéré, évitant toute sensation d'encombrement tout en offrant une méthode de navigation intuitive.

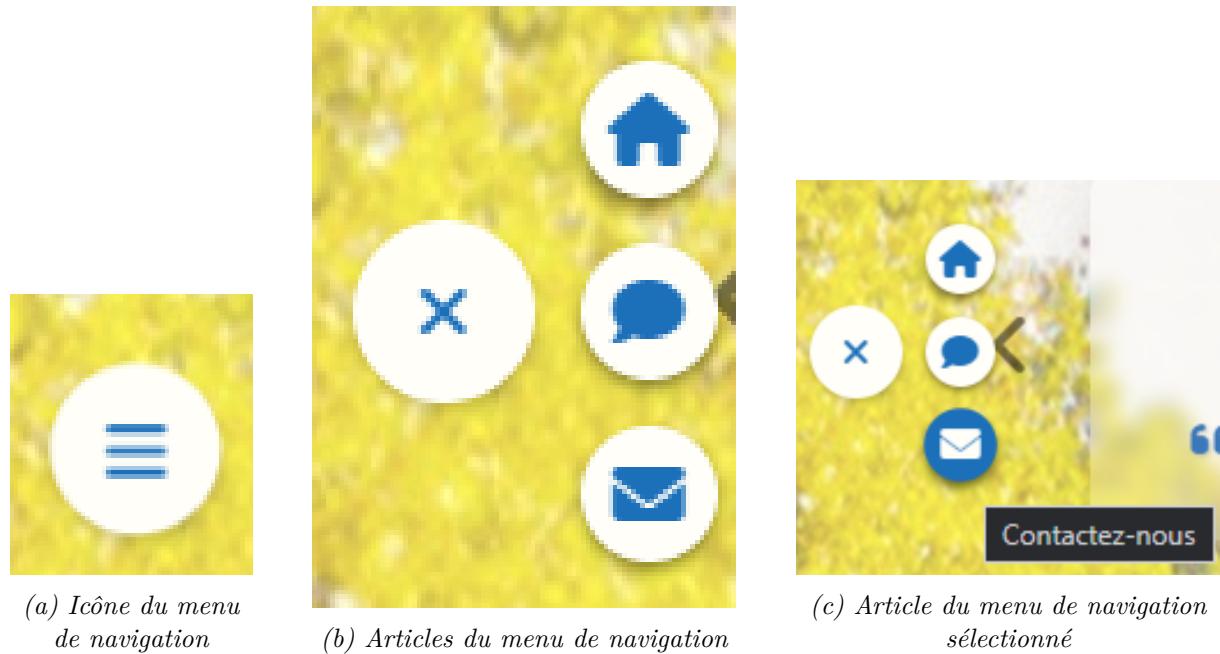


FIG. II.29 : Menu de navigation de la Page d'Accueil Conversationnelle

4.4.2.3 Interface du chatbot

L'un des principaux éléments de notre page d'accueil conversationnelle est le chatbot intégré. Il est doté d'une icône distinctive et ludique, positionnée en bas à droite de l'écran (dans la zone annotée -3- dans la figure II.27). Comme illustré dans la figure II.30a, l'icône est conçue de manière simple, mais amusante, représentant une conversation avec un chatbot. L'expression de l'icône est animée de manière à attirer l'attention du client. Lorsqu'on clique dessus, l'icône est remplacée par une interface de messagerie, telle qu'illustrée dans la figure II.30b, permettant à l'utilisateur d'engager une conversation avec l'agent conversationnel. Nous aborderons en détail le processus de conversation avec le chatbot dans la section suivante.



FIG. II.30 : Chatbot de la Page d'Accueil Conversationnelle

4.4.2.4 Section : Accueil

La section d'accueil a pour rôle d'être le point d'entrée des utilisateurs, leur offrant une première impression attrayante et engageante. Elle exploite la bannière précédemment conçue, illustrée dans la figure II.47, en l'animant et en la plaçant stratégiquement dans une zone hautement visible, captant ainsi l'attention du visiteur (annotée par -4- dans la figure II.27). Cette bannière est également réactive et adaptée à la fois à la résolution web et à la version mobile. Dans la version web, nous utilisons le design original, tel qu'illustré dans la figure II.31a, tandis que la version mobile est présentée dans la figure II.31b.

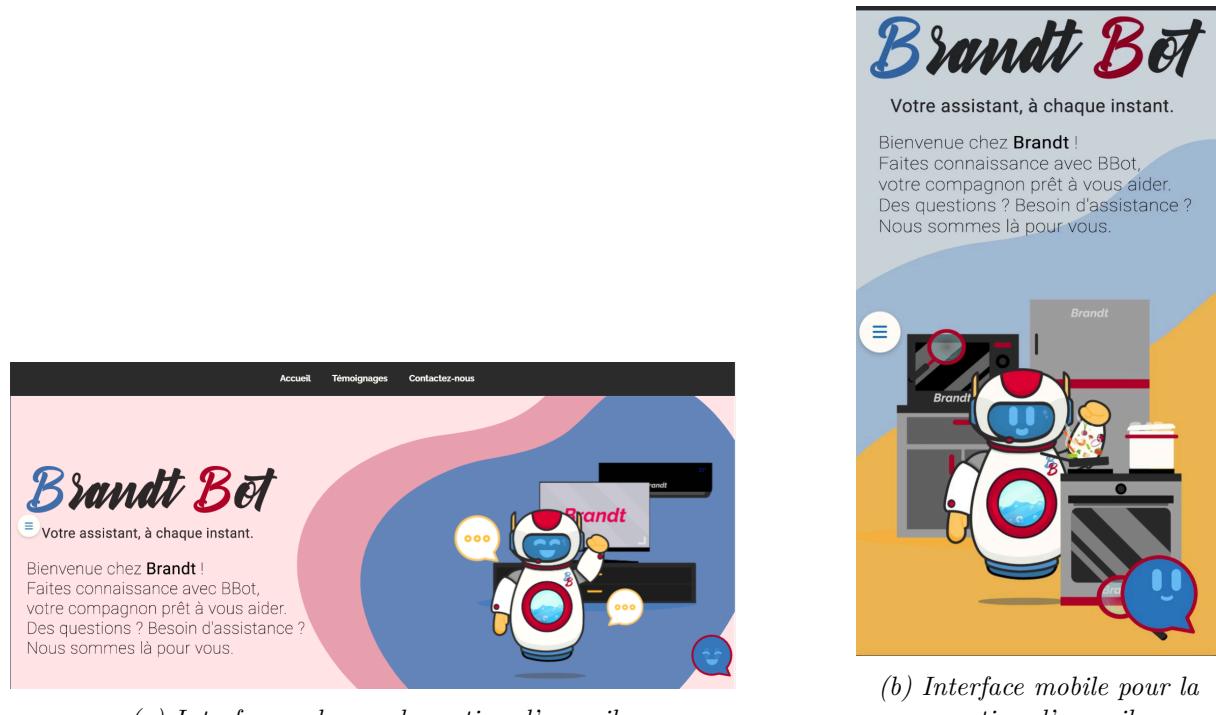


FIG. II.31 : Section d'accueil de la Page d'Accueil Conversationnelle

Le but de cette bannière est d'immerger l'utilisateur dans le contexte de la page. Elle capture à la fois l'essence de la marque (Brandt) en mettant en scène différents appareils électroménagers dans le décor des bannières, et l'aspect conversationnel de la page grâce au chatbot conçu. L'animation est fluide et dynamique, ce qui reflète un aspect souvent présent dans les designs et les publicités créés par Brandt. Cette approche vise à créer une cohérence entre la page d'accueil et l'image de marque de l'entreprise, tout en mettant en évidence les fonctionnalités clés de la page, notamment la conversation avec le chatbot.

4.4.2.5 Section : Témoignages

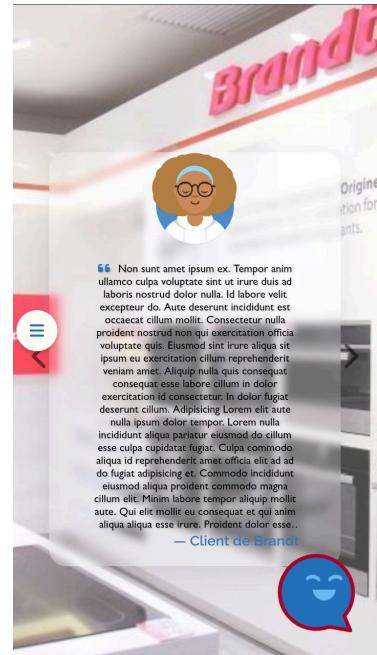
the testimonials section (annotated by -5- in the figure II.27) showcases positive opinions about the brand collected by the scraper and selected thru an algorithm then thru the admin's choices. the opinions are presented in a beautifully made carousel that is responsive as shown in the figure II.32 where we present both a web and mobile version

La section des témoignages (identifiée par -5- dans la figure II.27) met en avant un nombre d'avis authentiques positifs concernant la marque, qui ont été collectés à travers le scraper, puis sélectionnés au moyen d'un algorithme, suivi des choix de l'administrateur. Ces avis sont présentés dans un diaporama qui s'adapte aux différents écrans, comme illustré dans la figure II.32, où nous présentons à la fois la version web et la version mobile. Cette section revêt une importance particulière dans le cadre de la construction d'une réputation en ligne solide. Elle

vise à montrer aux visiteurs l'expérience positive de la clientèle avec la marque Brandt.



(a) Interface web pour la section de témoignages



(b) Interface mobile pour la section de témoignages

FIG. II.32 : Section de témoignages de la Page d'Accueil Conversationnelle

4.4.2.6 Section : Contacts

La dernière section principale de la page, annotée par le numéro -6- dans la figure II.27, concerne la section des contacts.

Comme illustré dans la figure II.33, cette section présente un design simple et épuré, mettant en avant l'ergonomie du formulaire pour une utilisation aisée. De plus, le design est adaptatif, ce qui signifie qu'il s'ajuste de manière fluide en fonction de la taille de l'écran. Il est à noter que dans la version mobile (figure II.33b), nous devons faire défiler la page pour accéder à la deuxième partie du formulaire. Dans le présent rapport, nous les présentons côté à côté pour faciliter la visualisation.

Il est important de souligner que chaque icône présente sur cette section a une fonctionnalité précise. En cliquant sur l'une d'entre elles, l'utilisateur peut effectuer une action spécifique, comme afficher l'adresse sur une carte ou ouvrir l'application de messagerie pour envoyer un e-mail à la marque. Cette approche interactive vise à offrir à l'utilisateur une expérience fluide et sans effort, tout en facilitant et en accélérant le processus de prise de contact avec la marque.

Pour renforcer le lien entre la marque et les clients, nous avons mis en place des méthodes de contact direct. Les coordonnées de l'entreprise sont facilement accessibles, et un formulaire de contact permet aux utilisateurs de soumettre des demandes rapidement. Cela crée une impression de disponibilité et de proximité, renforçant ainsi la confiance des utilisateurs dans la marque. Durant la phase de développement, nous avons utilisé une adresse e-mail fictive. Cependant, en production, cette adresse sera remplacée par celle de la marque⁶.

⁶ Contactez.nous@brandt.dz

CONTACTEZ-NOUS

Vous avez des questions, des réclamations ou des commentaires ? Contactez-nous !

Contactez-nous via ce formulaire, ou en utilisant une des méthodes spécifiées.

Votre Nom
Votre Email
Votre Sujet
Votre Message

Envoyer
– Brandt, toujours à votre écoute !

CONTACTEZ-NOUS

Vous avez des questions, des réclamations ou des commentaires ? Contactez-nous !

Contactez-nous via ce formulaire, ou en utilisant une des méthodes spécifiées.

Votre Nom
Votre Email
Votre Sujet
Votre Message

Envoyer
– Brandt, toujours à votre écoute !

(a) Interface web pour la section de contacts

(b) Interface mobile pour la section de contacts

FIG. II.33 : Section de contacts de la Page d'Accueil Conversationnelle

4.4.2.7 Pied de page



FIG. II.34 : Pied de page de la Page d'accueil Conversationnelle

Le dernier composant, annoté par le numéro -7- à la fin de la page II.27, est le pied de page qui regroupe divers liens sociaux (tels que leur site web officiel⁷, Facebook⁸,...) ainsi que des liens utiles pour aider les clients à trouver des réponses si le chatbot ne peut pas les aider à ce moment-là. Ces liens incluent des informations telles que "Qui est Brandt?"⁹ et "Où acheter?"¹⁰, et servent à orienter les clients vers des ressources qui répondent à leurs questions spécifiques...

Dans l'ensemble, tout au long de la conception de cette interface, nous avons mis en œuvre des principes d'ergonomie cognitive pour assurer une utilisation efficace et conviviale de la page d'accueil. Les critères d'évaluation, tels que les 9 critères de (BASTIEN et SCAPIN 1993), les heuristiques de (NIELSEN et MOLICH 1990) et les 7 règles d'or de (COUTAZ et al. 1990), ont guidé nos choix de conception pour garantir une expérience utilisateur optimale.

L'un des aspects essentiels de l'ergonomie est l'homogénéité, que nous avons maintenue tout au long de l'interface. Un design homogène, comme décrit précédemment dans la charte visuelle, assure une cohérence visuelle qui facilite la compréhension et l'utilisation de la page. De plus, nous avons appliqué le principe de la visibilité du système en utilisant des éléments tels que les fenêtres contextuelles (popups) et les indicateurs de chargement pour informer les utilisateurs sur les actions en cours. De plus, les boutons sont désactivés lorsqu'ils ne sont pas accessibles, ce qui renforce la correspondance entre le système et le monde réel en évitant les actions inutiles ou erronées.

Dans la même optique, nous avons veillé à ce que l'interface soit intuitive pour les utilisateurs, nous avons donc utilisé des icônes communes et adéquates lorsque cela est nécessaire, accompagnées de texte significatif pour assurer une compréhension claire des fonctionnalités.

4.4.3 Agent conversationnel

L'interface de messagerie, permettant d'interagir avec le chatbot, propose une expérience claire et fonctionnelle. Les échanges avec le chatbot sont structurés de manière à faciliter la

⁷ <https://www.brandt.dz/>

⁸ <https://www.facebook.com/Brandt.dz>

⁹ <https://www.brandt.dz/la-marque-page>

¹⁰ <https://www.brandt.dz/stores>

Chapitre 4. Prototype de la solution

communication et à répondre efficacement aux requêtes des utilisateurs. La diversité des types de messages présents dans l'interface reflète la variété des demandes potentielles. Certains messages sont concis, contenant des réponses directes, tandis que d'autres peuvent comporter des éléments plus complexes tels que des tableaux ou des informations détaillées, en fonction de la nature des questions.

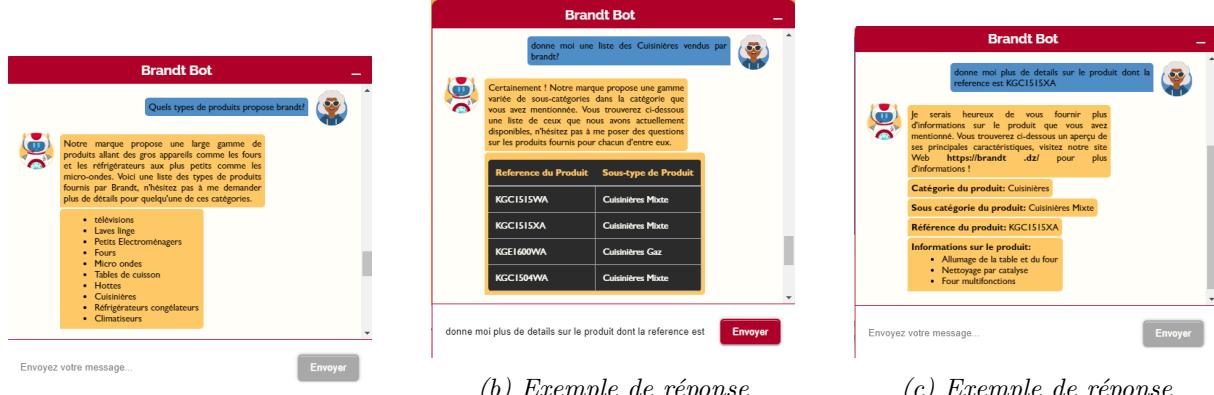


FIG. II.35 : Exemples de Réponses du Chatbot

Pour mieux illustrer ces interactions, examinons quelques exemples concrets. Dans la figure II.35a, le client demande des informations sur les types de produits vendus par Brandt. Le chatbot répond de manière claire en énumérant les différentes catégories de produits disponibles. Par la suite, le client s'informe sur les cuisinières vendues, comme visible dans la figure II.35b. Le chatbot fournit une réponse sous forme de tableau présentant les différents sous-types de cuisinières ainsi que les références des produits correspondants. Enfin, dans la situation représentée dans la figure II.35c, l'utilisateur demande plus de détails sur un produit en précisant sa référence. Le chatbot répond en fournissant des informations détaillées sur le produit en question.

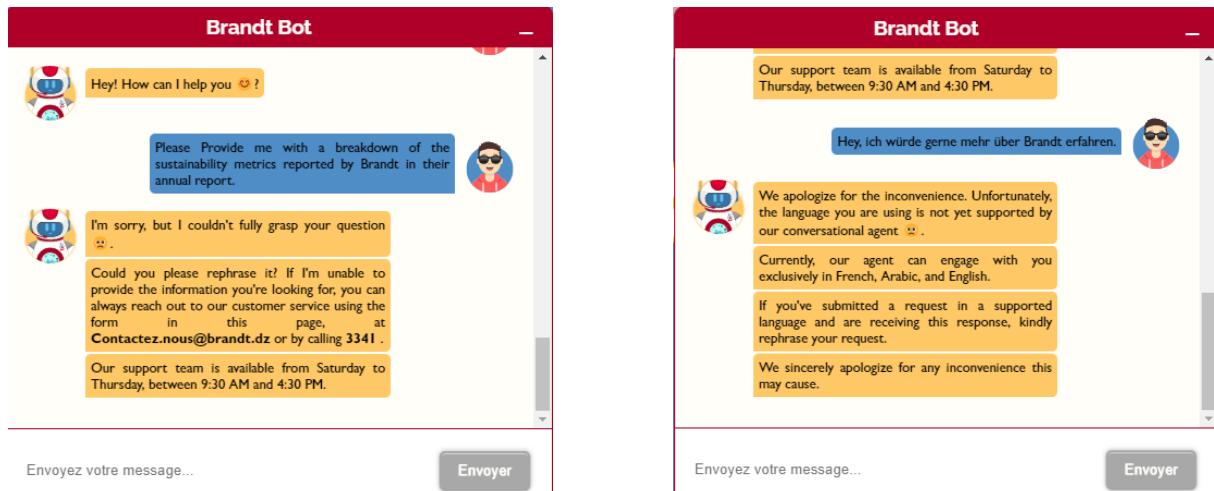


FIG. II.36 : Exemples de réponses Chatbot - Scénarios Non Résolus

Actuellement, le chatbot dispose d'une base de connaissances limitée, principalement axée sur des informations générales et des requêtes plutôt courantes. La figure II.36a illustre comment le chatbot gère une demande non prise en charge ou une demande qu'il ne parvient pas à interpréter. Il convient de souligner que le chatbot est actuellement opérationnel en français, en anglais et en arabe uniquement, comme évoqué précédemment. Cette limitation linguistique est représentée dans la figure II.36b, où l'on voit comment le chatbot gère une demande formulée

dans une langue non prise en charge.

Il est important de mentionner que cette interface est conçue de manière à gérer diverses erreurs, telles que les problèmes de connexion à Internet et autres. Les requêtes non reconnues ou non prises en charge par le chatbot sont enregistrées dans une base de données. Cette fonctionnalité revêt un intérêt stratégique, car les données ainsi recueillies peuvent être exploitées par des experts pour analyser les demandes des clients afin de prendre de meilleures décisions, mais aussi en vue d'enrichir la base de connaissances du chatbot. Cela contribue à l'amélioration continue du chatbot, en l'enrichissant avec de nouvelles informations pertinentes et en permettant des interactions plus efficaces et pertinentes avec les utilisateurs.

Comme illustré dans les figures précédentes, le chatbot adopte une approche conviviale, polie et professionnelle, tout en se montrant particulièrement utile pour les utilisateurs. Cette convivialité se reflète dans l'image de la marque Brandt, qui transmet un sentiment d'amabilité et de dynamisme à travers ses divers contenus en ligne. L'attitude du chatbot s'aligne donc sur ces valeurs de la marque, assurant une expérience utilisateur en harmonie avec l'identité de l'entreprise.

Le chatbot repose sur une approche de récupération basée sur des règles. Pour obtenir une réponse en fonction de l'entrée de l'utilisateur, un processus bien défini, tel qu'illustré dans la figure II.37, est suivi. Dans un premier temps, nous déterminons l'intention de l'utilisateur (qu'il s'agisse d'une demande concernant un produit, un type de produit, un magasin, les magasins dans une wilaya spécifique, ou d'autres types de demandes simples), en utilisant le premier tableau. Ensuite, nous utilisons le deuxième et le troisième tableau pour collecter des informations sur les magasins et les produits respectivement. Le quatrième tableau nous permet de déterminer le type de demande que l'utilisateur formule. Enfin, la cinquième table contient les différentes réponses possibles, allant des réponses directes aux réponses par défaut, en passant par les messages d'erreur et les salutations de base. En fonction de l'intention identifiée et des informations collectées, le chatbot sélectionne la réponse appropriée.

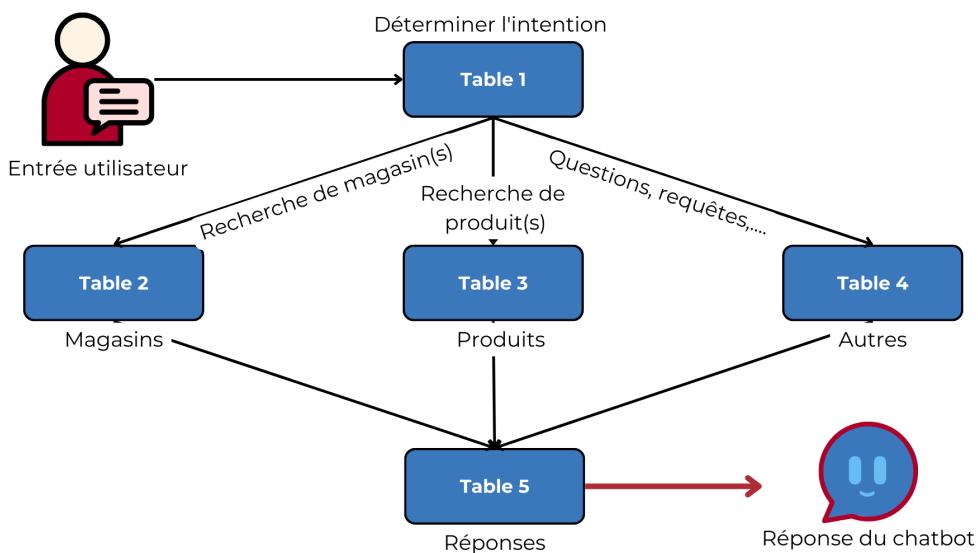


FIG. II.37 : Processus de Réponse du Chatbot

Une fois toutes les informations nécessaires rassemblées, le backend renvoie la réponse au frontend, où elle est affichée pour l'utilisateur sous forme de message dans l'interface. Lorsque le chatbot ne parvient pas à reconnaître une demande, il met en œuvre une approche de dernier recours en tentant de détecter tout jeton identifiable, comme un nom de magasin ou une référence de produit. Dans de tels cas, il est nécessaire d'utiliser le modèle d'identification de la langue pour répondre correctement au client.

Étant donné que l'historique complet de la session de chat est déjà enregistré dans l'interface utilisateur, une amélioration future du chatbot pour prendre en compte davantage de contexte ne devrait pas représenter une difficulté majeure. Cette infrastructure préexistante permettra de développer de nouvelles fonctionnalités tout en garantissant une continuité fluide de l'expérience utilisateur.

4.5 Conclusion

En somme, ce chapitre a présenté en détail le prototype que nous avons conçu pour répondre aux défis de suivi de réputation en ligne auxquels Brandt Algérie est confrontée. En explorant les aspects de la charte graphique, du design, de la plateforme "Brandt Insight", de la page d'accueil et du chatbot, nous avons pu démontrer comment notre solution vise à simplifier et à améliorer la gestion de la réputation en ligne de la marque. Grâce à des choix précis et à une conception adaptée aux besoins de l'entreprise, nous cherchons à offrir une expérience utilisateur optimisée pour les collaborateurs et les administrateurs, tout en permettant une interaction plus efficace avec les clients.

Conclusion et perspectives

Conclusion générale

L'objectif de ce projet de fin d'études était de créer une solution innovante permettant à une marque de suivre efficacement sa réputation en ligne et même de tenter de l'améliorer, étant donné la préoccupation croissante des entreprises concernant leur popularité et la nécessité d'identifier les véritables sentiments des clients au milieu de la présence croissante d'avis spam.

Tout au long du projet, nous avons exploré l'état de l'art en matière d'analyse des sentiments et de détection des spams, en explorant diverses définitions, approches, ensembles de données, méthodes d'évaluation et métriques, ainsi que les défis associés à ce domaine. Cette connaissance fondamentale nous a permis de construire une base théorique solide pour notre contribution.

Nous avons ensuite examiné le processus et les solutions existantes utilisées par notre entreprise d'accueil, Brandt, et avons identifié la nécessité d'une méthode améliorée pour surveiller et améliorer leur réputation en ligne. Avec une compréhension approfondie du domaine du problème, nous avons conçu une architecture pour notre solution, visant à déterminer les véritables sentiments envers la marque en incorporant différentes méthodes pour surmonter le défi de la diversité culturelle et linguistique du public cible.

En menant de nombreuses expériences, nous avons affiné nos approches et développé un prototype illustrant l'efficacité et la faisabilité de la solution proposée, capable d'identifier la langue utilisée dans chaque avis, de distinguer les avis positifs des avis négatifs, les avis spam des avis authentiques, et même de classer un avis comme étant une description (d'un produit, d'une marque, ...), une comparaison ou une question, fournissant ainsi des informations précieuses à Brandt pour gérer plus efficacement leur réputation en ligne.

De plus, notre contribution ne s'est pas arrêtée à la simple identification des sentiments ou à la détection des spams. Nous avons introduit une dimension pratique en mettant en œuvre une page d'accueil conversationnelle, permettant à Brandt d'établir une connexion plus forte avec ses clients et d'améliorer la satisfaction globale des clients.

Les contributions que notre projet a pu apporter peuvent se résumer dans les points suivants :

- **Revue de la Littérature :** Nous avons effectué une revue de l'état de l'art en analyse des sentiments et détection des spams, acquérant ainsi des connaissances sur les différentes techniques et méthodologies utilisées dans ce domaine.
- **Solution Innovante :** Nous avons proposé une solution innovante qui répond efficacement au défi de la polarité des opinions et des spams, contribuant ainsi à une surveillance plus précise de la réputation en ligne des entreprises comme Brandt.
- **Public cible polyglotte :** Notre solution vise à extraire efficacement des informations pertinentes des opinions du public malgré les nombreuses différences culturelles et linguis-

tiques présentées par l'audience de la marque.

- **Application prête à être mise en production :** Notre solution n'était pas limitée à la recherche théorique, mais a démontré des applications pratiques dans le monde réel, mettant en évidence ses avantages potentiels pour Brandt et des entreprises similaires.
- **Gestion Améliorée de la Réputation en Ligne :** En distinguant les avis spam des avis authentiques, notre solution permet aux entreprises de mieux comprendre les sentiments des clients et de prendre des décisions basées sur les données pour améliorer leur identité en ligne.
- **Amélioration de l'Expérience Client :** L'intégration d'une page d'accueil conversationnelle permet aux entreprises d'établir des interactions plus significatives avec les clients, conduisant à une augmentation de la satisfaction et de la fidélité des clients.

Perspectives

Bien que notre projet ait réussi à résoudre les problèmes identifiés dans le contexte de Brandt, il existe plusieurs pistes d'exploration et d'amélioration supplémentaires. Voici quelques perspectives potentielles :

1. **Adaptation Multilingue et Multiplateforme :** Élargir la solution pour traiter les avis et les sentiments dans plusieurs langues et sur différentes plateformes sociales améliorera son utilité et son applicabilité à un public plus large.
2. **Analyse en Temps Réel :** Incorporer des capacités d'analyse en temps réel, au lieu d'analyses périodiques, permettra aux entreprises de répondre rapidement aux commentaires des clients et de gérer leur réputation de manière proactive.
3. **Considérations Éthiques :** Bien que nous ayons fait de notre mieux pour anonymiser les données collectées conformément aux lois sur la protection de la vie privée, il est crucial de prendre en compte des aspects éthiques tels que la confidentialité des clients et les biais potentiels dans les modèles, à mesure que l'analyse des sentiments et la gestion de la réputation deviennent de plus en plus courantes.
4. **Collaboration avec l'Industrie :** Des études supplémentaires peuvent être menées pour mieux comprendre les biais et les limites des techniques d'analyse des sentiments et de détection de spam, en particulier en ce qui concerne les problèmes de représentativité des données et les biais culturels. La collaboration avec des experts du domaine peut aider à mieux comprendre ces questions, à accéder à des ressources nécessaires et à développer des solutions plus adaptées.

Appréciation personnelle

Tout au long de ce projet, j'ai eu l'opportunité inestimable de faire un stage chez Brandt/Cevital et de travailler sur un problème de pointe dans le domaine de l'analyse des sentiments et de la détection de spam. Je suis reconnaissante pour les connaissances acquises lors de la revue de littérature, les idées récoltées pendant la phase d'expérimentation, et l'expérience pratique du développement du prototype. Ce projet a renforcé mes compétences techniques et mes capacités de recherche, et je me sens plus confiante pour appliquer les concepts théoriques à des scénarios réels.

En y réfléchissant, je dois admettre que j'ai ressenti une certaine tristesse et une certaine déception à propos des aspects que je n'ai pas pu accomplir avec la solution en raison du manque de ressources, comme l'absence de jeux de données dans le dialecte algérien, l'absence de traducteurs pour le dialecte algérien, ainsi que l'absence de jeux de données pour la détection de spams d'opinion qui exploitent à la fois les caractéristiques textuelles et comportementales,

Conclusion et perspectives

dans des langues autres que l'anglais. Cela m'a incité à réfléchir à des moyens de fournir de telles données publiquement afin de faire progresser la recherche dans ces domaines.

En conclusion, ce projet a été un voyage d'apprentissage remarquable, et je suis enthousiaste à l'idée de continuer à explorer les domaines du traitement du langage naturel et de l'intelligence artificielle qui sont en constante évolution. Je suis convaincue que les connaissances acquises grâce à ce projet continueront à influencer positivement mes projets futurs.

Bibliographie

- ABDAOUI, Amine et al. (2021). « Dziribert : a pre-trained language model for the algerian dialect ». In : *arXiv preprint arXiv :2109.12346*.
- AHMED, Hadeer (2017). « Detecting opinion spam and fake news using n-gram analysis and semantic similarity ». Thesis.
- ALIARAB, Mahmoud et Kazim FOULADI (2022). « A survey on review spam detection methods using deep learning approach ». In : *International Journal of Web Research* 5.1, p. 19-24.
- ALSUBARI, Saleh Nagi, Mahesh B SHELKE et Sachin N DESHMUKH (2020). « Fake reviews identification based on deep computational linguistic ». In : *International Journal of Advanced Science and Technology* 29.8s, p. 3846-3856.
- ALY, Mohamed et Amir ATIYA (2013). « Labr : A large scale arabic book reviews dataset ». In : *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, p. 494-498.
- AMINUDDIN, Raihah et al. (2021). « Sentiment analysis of online product reviews using Lexical Semantic Corpus-Based technique ». In : *2021 IEEE 11th IEEE Symposium on Computer Applications and Industrial Electronics (ISCAIE)*. IEEE, p. 233-238.
- ANDRESINI, Giuseppina et al. (2022). « EUPHORIA : A neural multi-view approach to combine content and behavioral features in review spam detection ». In : *Journal of Computational Mathematics and Data Science* 3, p. 100036.
- ANTOUN, Wissam, Fady BALY et Hazem HAJJ (s. d.). « AraBERT : Transformer-based Model for Arabic Language Understanding ». In : *LREC 2020 Workshop Language Resources and Evaluation Conference 11–16 May 2020*, p. 9.
- BACCIANELLA, Stefano, Andrea ESULI et Fabrizio SEBASTIANI (mai 2010). « SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining ». In : *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, Malta : European Language Resources Association (ELRA).
- BAGHAEI, Kourosh T et al. (2022). « Deep representation learning : Fundamentals, Perspectives, Applications, and Open Challenges ». In : *arXiv preprint arXiv :2211.14732*.
- BASTIEN, Christian et Dominique SCAPIN (1993). « Ergonomic criteria for the evaluation of human-computer interfaces ». Thèse de doct. Inria.
- BECCHETTI, Luca et al. (2008). « Link analysis for web spam detection ». In : *ACM Transactions on the Web (TWEB)* 2.1, p. 1-42.
- BENGIO, Yoshua, Aaron COURVILLE et Pascal VINCENT (2013). « Representation learning : A review and new perspectives ». In : *IEEE transactions on pattern analysis and machine intelligence* 35.8, p. 1798-1828.
- BENNETT, Kristin et Ayhan DEMIRIZ (1998). « Semi-supervised support vector machines ». In : *Advances in Neural Information processing systems* 11.
- BLUM, Avrim et Tom MITCHELL (1998). « Combining labeled and unlabeled data with co-training ». In : *Proceedings of the eleventh annual conference on Computational learning theory*, p. 92-100.

- BOJANOWSKI, Piotr et al. (2017). « Enriching word vectors with subword information ». In : *Transactions of the association for computational linguistics* 5, p. 135-146.
- BOUAMOR, Houda et al. (2018). « The madar arabic dialect corpus and lexicon ». In : *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*.
- BROWNLEE, Jason (2014). « Discover feature engineering, how to engineer features and how to get good at it ». In : *Machine Learning Process*.
- CASTILLO, Carlos, Marcelo MENDOZA et Barbara POBLETE (2011). « Information credibility on twitter ». In : *Proceedings of the 20th international conference on World wide web*, p. 675-684.
- CHADER, Asma et al. (2019). « Sentiment Analysis for Arabizi : Application to Algerian Dialect ». In : *KDIR*, p. 475-482.
- CHATURVEDI, Iti et al. (2018). « Distinguishing between facts and opinions for sentiment analysis : Survey and challenges ». In : *Information Fusion* 44, p. 65-77.
- CHAWLA, Nitesh V et al. (2002). « SMOTE : synthetic minority over-sampling technique ». In : *Journal of artificial intelligence research* 16, p. 321-357.
- CHO, Kyunghyun et al. (2014). « Learning phrase representations using RNN encoder-decoder for statistical machine translation ». In : *arXiv preprint arXiv :1406.1078*.
- COMON, Pierre (1994). « Independent component analysis, a new concept ? » In : *Signal processing* 36.3, p. 287-314.
- COUTAZ, J, L NIGAY et D SALBER (1990). « Software architecture modelling for interactive systems ». In : *A paraître dans encyclopedia of HCI*, Wiley&son, J. Nielsen eds.
- DAI, Yong et al. (2021). « Unsupervised sentiment analysis by transferring multi-source knowledge ». In : *Cognitive Computation* 13, p. 1185-1197.
- DAVE, Kushal, Steve LAWRENCE et David M PENNOCK (2003). « Mining the peanut gallery : Opinion extraction and semantic classification of product reviews ». In : *Proceedings of the 12th international conference on World Wide Web*, p. 519-528.
- DEERWESTER, Scott et al. (1990). « Indexing by latent semantic analysis ». In : *Journal of the American society for information science* 41.6, p. 391-407.
- DEVLIN, Jacob et al. (2018a). « BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding ». In : *CoRR* abs/1810.04805.
- (2018b). « Bert : Pre-training of deep bidirectional transformers for language understanding ». In : *arXiv preprint arXiv :1810.04805*.
- DING, Xiaowen et Bing LIU (s. d.). « The utility of linguistic rules in opinion mining ». In : *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, p. 811-812.
- DONG, Manqing et al. (2020). « Opinion fraud detection via neural autoencoder decision forest ». In : *Pattern Recognition Letters* 132, p. 21-29.
- FUSILIER, Donato Hernández et al. (2015). « Detecting positive and negative deceptive opinions using PU-learning ». In : *Information processing and management* 51.4, p. 433-443.
- GHELANI, Pooja H et Tosal M BHALODIA (2017). « Opinion mining and opinion spam detection ». In : *International Research Journal of Engineering and Technology (IRJET)* 4, p. 11.
- GILBERT, Eric et Karrie KARAHALIOS (2010). « Understanding deja reviewers ». In : *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, p. 225-228.
- GO, Alec, Richa BHAYANI et Lei HUANG (2009). « Twitter sentiment classification using distant supervision ». In : *CS224N project report, Stanford* 1.12, p. 2009.
- GOEL, Anurag et al. (2021). « Classification of Positive and Negative Fake Online Reviews using Machine Learning Techniques ». In : *International Journal of Advanced Networking and Applications* 12.6, p. 4746-4749.
- GOLUB, GH et CF VAN LOAN (1996). « Matrix computations 3rd edition the john hopkins university press ». In : *Baltimore, MD*.

- HAJEK, Petr, Aliaksandr BARUSHKA et Michal MUNK (2020). « Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining ». In : *Neural Computing and Applications* 32, p. 17259-17274.
- HARRIS, Christopher G (2022). « Combining Linguistic and Behavioral Clues to Detect Spam in Online Reviews ». In : *2022 IEEE International Conference on e-Business Engineering (ICEBE)*. IEEE, p. 38-44.
- HASSAN, Rakibul et Md Rabiul ISLAM (2021). « Impact of sentiment analysis in fake online review detection ». In : *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*. IEEE, p. 21-24.
- HERCIG, Tomás et Ladislav LENC (2017). « The Impact of Figurative Language on Sentiment Analysis. » In : *RANLP*, p. 301-308.
- HOCHREITER, Sepp et Jürgen SCHMIDHUBER (1997). « Long short-term memory ». In : *Neural computation* 9.8, p. 1735-1780.
- HOSSEN, Md Sharif et Niloy Ranjan DEV (2021). « An improved lexicon based model for efficient sentiment analysis on movie review data ». In : *Wireless Personal Communications* 120, p. 535-544.
- HOWARD, Jeremy et Sebastian RUDER (2018). « Universal language model fine-tuning for text classification ». In : *arXiv preprint arXiv :1801.06146*.
- HUSSAIN, Naveed et al. (2019). « Spam review detection techniques : A systematic literature review ». In : *Applied Sciences* 9.5, p. 987.
- HUSSEIN, Doaa Mohey El-Din Mohamed (2018). « A survey on sentiment analysis challenges ». In : *Journal of King Saud University-Engineering Sciences* 30.4, p. 330-338.
- JINDAL, Nitin et Bing LIU (2006a). « Identifying comparative sentences in text documents ». In : *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, p. 244-251.
- (2006b). « Mining comparative sentences and relations ». In : *Aaai*. T. 22. 13311336, p. 9.
- (2008). « Opinion spam and analysis ». In : *Proceedings of the 2008 international conference on web search and data mining*, p. 219-230.
- KENNEDY, Stefan et al. (2020). « Fact or factitious ? Contextualized opinion spam detection ». In : *arXiv preprint arXiv :2010.15296*.
- KIM, Seongsoon et al. (2015). « Deep semantic frame-based deceptive opinion spam analysis ». In : *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, p. 1131-1140.
- LE, Hang et al. (2020a). « FlauBERT : des modèles de langue contextualisés pré-entraînés pour le français ». In : *Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 31e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 2: Traitement Automatique des Langues Naturelles*. ATALA, p. 268-278.
- LE, Hang et al. (mai 2020b). « FlauBERT : Unsupervised Language Model Pre-training for French ». In : *Proceedings of The 12th Language Resources and Evaluation Conference*. Marseille, France : European Language Resources Association, p. 2479-2490.
- LE, Quoc et Tomas MIKOLOV (2014). « Distributed representations of sentences and documents ». In : *International conference on machine learning*. PMLR, p. 1188-1196.
- LI, Fangtao Huang et al. (2011). « Learning to identify review spam ». In : *Twenty-second international joint conference on artificial intelligence*.
- LI, Huayi et al. (2014). « Spotting fake reviews via collective positive-unlabeled learning ». In : *2014 IEEE international conference on data mining*. IEEE, p. 899-904.

- LI, Jiwei et al. (2014). « Towards a general rule for identifying deceptive opinion spam ». In : *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, p. 1566-1576.
- LI, Luyang et al. (2015). « Learning document representation for deceptive opinion spam detection ». In : *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data : 14th China National Conference, CCL 2015 and Third International Symposium, NLP-NABD 2015, Guangzhou, China, November 13-14, 2015, Proceedings 14*. Springer, p. 393-404.
- LI, Ning, Chi-Yin CHOW et Jia-Dong ZHANG (2020). « SEML : A semi-supervised multi-task learning framework for aspect-based sentiment analysis ». In : *IEEE Access* 8, p. 189287-189297.
- LI, Ruohan et Ayoung SUH (2015). « Factors influencing information credibility on social media platforms : Evidence from Facebook pages ». In : *Procedia computer science* 72, p. 314-328.
- LI, Xiao-Li et Bing LIU (2005). « Learning from positive and unlabeled examples with different data distributions ». In : *Machine Learning : ECML 2005: 16th European Conference on Machine Learning, Porto, Portugal, October 3-7, 2005. Proceedings 16*. Springer, p. 218-229.
- LIGTHART, Alexander, Cagatay CATAL et Bedir TEKINERDOGAN (2021a). « Analyzing the effectiveness of semi-supervised learning approaches for opinion spam classification ». In : *Applied Soft Computing* 101, p. 107023.
- (2021b). « Systematic reviews in sentiment analysis : a tertiary study ». In : *Artificial Intelligence Review* 54.7, p. 4997-5053.
- LIU, Bing (2006). *Web data mining : exploring hyperlinks, contents, and usage data*. Springer.
- (2011). *Web data mining : exploring hyperlinks, contents, and usage data*. T. 1. Springer.
- (2012). « Sentiment analysis and opinion mining ». In : *Synthesis lectures on human language technologies* 5.1, p. 1-167.
- (2020). « Sentiment Analysis ». In : *Sentiment Analysis : Mining Opinions, Sentiments, and Emotions*. 2^e éd. Studies in Natural Language Processing. Cambridge University Press, p. i-i.
- LIU, Bing et al. (2002). « Partially supervised classification of text documents ». In : *ICML*. T. 2. 485. Sydney, NSW, p. 387-394.
- MA, Yingying et Fengjun LI (2012). « Detecting review spam : Challenges and opportunities ». In : *8th International Conference on Collaborative Computing : Networking, Applications and Worksharing (CollaborateCom)*. IEEE, p. 651-654.
- MAAFI, Hanene (2020). « Prise en compte de la crédibilité de l'opinion dans le système d'analyse de sentiments ». Ecole nationale Supérieure d'Informatique, ESI.
- MAAS, Andrew et al. (2011). « Learning word vectors for sentiment analysis ». In : *Proceedings of the 49th annual meeting of the association for computational linguistics : Human language technologies*, p. 142-150.
- MATAOUI, M'hamed, Omar ZELMATI et Madiha BOUMECHACHE (2016). « A proposed lexicon-based sentiment analysis approach for the vernacular Algerian Arabic ». In : *Research in Computing Science* 110.1, p. 55-70.
- MAYNARD, Diana G et Mark A GREENWOOD (2014). « Who cares about sarcastic tweets ? investigating the impact of sarcasm on sentiment analysis ». In : *Lrec 2014 proceedings*. ELRA.
- MEEL, Priyanka et Dinesh Kumar VISHWAKARMA (2020). « Fake news, rumor, information pollution in social media and web : A contemporary survey of state-of-the-arts, challenges and opportunities ». In : *Expert Systems with Applications* 153, p. 112986.
- MEHTA, Pooja et Sharnil PANDYA (2020). « A review on sentiment analysis methodologies, practices and applications ». In : *International Journal of Scientific and Technology Research* 9.2, p. 601-609.

- MI, Blei DM Ng AY Jordan et al. (2003). « Latent dirichlet allocation ». In : *J. Mach. Learn. Res* 3.993, p. 1022.
- MIKOLOV, Tomas et al. (2013). « Efficient estimation of word representations in vector space ». In : *arXiv preprint arXiv :1301.3781*.
- MIR, Abrar Qadir, Furqan Yaqub KHAN et Mohammad Ahsan CHISHTI (2023). « Online Fake Review Detection Using Supervised Machine Learning And BERT Model ». In : *arXiv preprint arXiv :2301.03225*.
- MOHAWESH, Rami et al. (2021). « Fake reviews detection : A survey ». In : *IEEE Access* 9, p. 65771-65802.
- MUKHERJEE, Arjun, Bing LIU et Natalie GLANCE (2012). « Spotting fake reviewer groups in consumer reviews ». In : *Proceedings of the 21st international conference on World Wide Web*, p. 191-200.
- MUKHERJEE, Arjun et al. (2011). « Detecting group review spam ». In : *Proceedings of the 20th international conference companion on World wide web*, p. 93-94.
- MUKHERJEE, Arjun et al. (2013a). « Spotting opinion spammers using behavioral footprints ». In : *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, p. 632-640.
- MUKHERJEE, Arjun et al. (2013b). « What yelp fake review filter might be doing ? » In : *Proceedings of the international AAAI conference on web and social media*. T. 7. 1, p. 409-418.
- NANDWANI, Pansy et Rupali VERMA (2021). « A review on sentiment analysis and emotion detection from text ». In : *Social Network Analysis and Mining* 11.1, p. 81.
- NASUKAWA, Tetsuya et Jeonghee YI (2003). « Sentiment analysis : Capturing favorability using natural language processing ». In : *Proceedings of the 2nd international conference on Knowledge capture*, p. 70-77.
- NI, Jianmo, Jiacheng LI et Julian MCAULEY (2019). « Justifying recommendations using distantly-labeled reviews and fine-grained aspects ». In : *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, p. 188-197.
- NIELSEN, Jakob et Rolf MOLICH (1990). « Heuristic evaluation of user interfaces ». In : *Proceedings of the SIGCHI conference on Human factors in computing systems*, p. 249-256.
- OTT, Myle, Claire CARDIE et Jeffrey T HANCOCK (2013). « Negative deceptive opinion spam ». In : *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics : human language technologies*, p. 497-501.
- OTT, Myle et al. (2011). « Finding deceptive opinion spam by any stretch of the imagination ». In : *arXiv preprint arXiv :1107.4557*.
- OUESLATI, Oumayma, Ahmed Ibrahim S KHALIL et Habib OUNELLI (2018). « Sentiment analysis for helpful reviews prediction ». In : *Int. J* 7, p. 34-40.
- ÖZÇELIK, Merve et al. (2021). « HisNet : a polarity lexicon based on wordnet for emotion analysis ». In : *Proceedings of the 11th Global Wordnet Conference*, p. 157-165.
- PADMAJA, S et S Sameen FATIMA (2013). « Opinion mining and sentiment analysis-an assessment of peoples' belief : A survey ». In : *International Journal of Ad hoc, Sensor and Ubiquitous Computing* 4.1, p. 21.
- PATEL, Rinki et Priyank THAKKAR (2014). « Opinion spam detection using feature selection ». In : *2014 International Conference on Computational Intelligence and Communication Networks*. IEEE, p. 560-564.
- PEARSON, Karl (1901). « LIII. On lines and planes of closest fit to systems of points in space ». In : *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2.11, p. 559-572.

- PENNINGTON, Jeffrey, Richard SOCHER et Christopher D MANNING (2014). « Glove : Global vectors for word representation ». In : *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, p. 1532-1543.
- PETERS, ME et al. (2018). « Deep contextualized word representations. arXiv 2018 ». In : *arXiv preprint arXiv :1802.05365* 12.
- PUTRA, Muhammad Mukhtar Dwi, Wikky Fawwaz AL MAKI et Ade ROMADHONY (2021). « Sentiment Analysis on Marketplace Review using Hybrid Lexicon and SVM Method ». In : *2021 9th International Conference on Information and Communication Technology (ICoICT)*. IEEE, p. 66-70.
- RAMESH, Bhavana et Charles M WEBER (s. d.). « State-of-art methods used in sentiment analysis : A literature review ». In : *2022 Portland International Conference on Management of Engineering and Technology (PICMET)*. IEEE, p. 1-13.
- RASTOGI, Ajay et Monica MEHROTRA (2018). « Impact of behavioral and textual features on opinion spam detection ». In : *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, p. 852-857.
- RASTOGI, Ajay, Monica MEHROTRA et Syed Shafat ALI (2020). « Effective opinion spam detection : A study on review metadata versus content ». In : *Journal of Data and Information Science* 5.2, p. 76-110.
- RAVI, Kumar et Vadlamani RAVI (2015). « A survey on opinion mining and sentiment analysis : tasks, approaches and applications ». In : *Knowledge-based systems* 89, p. 14-46.
- RAVI KUMAR, G, K VENKATA SHESHANNA et G ANJAN BABU (2021). « Sentiment analysis for airline tweets utilizing machine learning techniques ». In : *International Conference on Mobile Computing and Sustainable Informatics : ICMCSI 2020*. Springer, p. 791-799.
- RAYANA, Shebuti et Leman AKOGLU (2015). « Collective opinion spam detection : Bridging review networks and metadata ». In : *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining*, p. 985-994.
- REN, Yafeng et Donghong JI (2019). « Learning to Detect Deceptive Opinion Spam : A Survey ». In : *IEEE Access* 7, p. 42934-42945.
- REN, Yafeng, Donghong JI et Liang YIN (2014). « Deceptive reviews detection based on semi-supervised learning algorithm ». In : *J. Sichuan Univ.(Eng. Sci. Ed.)* 46.3, p. 62-69.
- SAUMYA, Sunil et Jyoti Prakash SINGH (2022). « Spam review detection using LSTM autoencoder : an unsupervised approach ». In : *Electronic Commerce Research* 22.1, p. 113-133.
- SCHÖLKOPF, Bernhard, Alexander SMOLA et Klaus-Robert MÜLLER (1998). « Nonlinear component analysis as a kernel eigenvalue problem ». In : *Neural computation* 10.5, p. 1299-1319.
- SCHUSTER, Mike et Kuldip K PALIWAL (1997). « Bidirectional recurrent neural networks ». In : *IEEE transactions on Signal Processing* 45.11, p. 2673-2681.
- SEDIGHI, Zeinab, Hossein EBRAHIMPOUR-KOMLEH et Ayoub BAGHERI (2017). « RLOSD : Representation learning based opinion spam detection ». In : *2017 3rd Iranian Conference on Intelligent Systems and Signal Processing (ICSPIS)*, p. 74-80.
- SHAALAN, Yassien et al. (2021). « Detecting singleton spams in reviews via learning deep anomalous temporal aspect-sentiment patterns ». In : *Data Mining and Knowledge Discovery* 35.2, p. 450-504.
- SHAH, Arkesha (2021). « Sentiment analysis of product reviews using supervised learning ». In : *Reliability : Theory and Applications* 16.SI 1 (60), p. 243-253.
- SHAHARIAR, GM et al. (2019). « Spam review detection using deep learning ». In : *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, p. 0027-0033.

- SHAN, Guohou, Lina ZHOU et Dongsong ZHANG (2021). « From conflicts and confusion to doubts : Examining review inconsistency for fake review detection ». In : *Decision Support Systems* 144, p. 113513.
- SHEHNEPOOR, Saeedreza et al. (2021). « HIN-RNN : A graph representation learning neural network for fraudster group detection with no handcrafted features ». In : *IEEE Transactions on Neural Networks and Learning Systems*.
- SHOJAEE, Somayeh et al. (2013). « Detecting deceptive reviews using lexical and syntactic features ». In : *2013 13th International Conference on Intelligent Systems Design and Applications*. IEEE, p. 53-58.
- SLIMANI, Wassila Maria (2023). « Etat de l'art sur l'intégration de la crédibilité dans la tâche de fouille d'opinion ». Mém. de mast. Ecole nationale Supérieure d'Informatique ESI, Oued Smar Alger, Algérie.
- SOCHER, Richard et al. (2013). « Recursive deep models for semantic compositionality over a sentiment treebank ». In : *Proceedings of the 2013 conference on empirical methods in natural language processing*, p. 1631-1642.
- SUN, Shiliang, Chen LUO et Junyu CHEN (2017). « A review of natural language processing techniques for opinion mining systems ». In : *Information fusion* 36, p. 10-25.
- TANG, Xiaoya, Tieyun QIAN et Zhenni YOU (2020). « Generating behavior features for cold-start spam review detection with adversarial learning ». In : *Information Sciences* 526, p. 274-288.
- TENENBAUM, Joshua B, Vin de SILVA et John C LANGFORD (2000). « A global geometric framework for nonlinear dimensionality reduction ». In : *science* 290.5500, p. 2319-2323.
- TIAN, Yingjie et al. (2020). « A non-convex semi-supervised approach to opinion spam detection by ramp-one class SVM ». In : *Information Processing and Management* 57.6, p. 102381.
- TIFFANI, Ilham Esa (2020). « Optimization of naïve bayes classifier by implemented unigram, bigram, trigram for sentiment analysis of hotel review ». In : *Journal of Soft Computing Exploration* 1.1, p. 1-7.
- VAN ENGELEN, Jesper E et Holger H HOOS (2020). « A survey on semi-supervised learning ». In : *Machine learning* 109.2, p. 373-440.
- WANG, Jingdong et al. (2020). « Fake review detection based on multiple feature fusion and rolling collaborative training ». In : *IEEE Access* 8, p. 182625-182639.
- WANG, Tao et Hua ZHU (2014). « Voting for deceptive opinion spam detection ». In : *arXiv preprint arXiv :1409.4504*.
- WANG, Yile, Leyang CUI et Yue ZHANG (2019). « Using Dynamic Embeddings to Improve Static Embeddings ». In : *arXiv Preprint*.
- WANG, Ziyang et al. (2020). « User-based Network Embedding for Collective Opinion Spammer Detection ». In : *arXiv preprint arXiv :2011.07783*.
- WASHHA, Mahdi et al. (2017). « A topic-based hidden Markov model for real-time spam tweets filtering ». In : *Procedia Computer Science* 112, p. 833-843.
- YADOLLAHI, Ali, Ameneh Gholipour SHAHRAKI et Osmar R ZAIANE (2017). « Current state of text sentiment analysis from opinion to emotion mining ». In : *ACM Computing Surveys (CSUR)* 50.2, p. 1-33.
- YAN, Weizhong et Lijie YU (2019). « On accurate and reliable anomaly detection for gas turbine combustors : A deep learning approach ». In : *arXiv preprint arXiv :1908.09238*.
- YAO, Jianrong, Yuan ZHENG et Hui JIANG (2021). « An ensemble model for fake online review detection based on data resampling, feature pruning, and parameter optimization ». In : *IEEE Access* 9, p. 16914-16927.
- YAROWSKY, David (1995). « Unsupervised word sense disambiguation rivaling supervised methods ». In : *33rd annual meeting of the association for computational linguistics*, p. 189-196.

Bibliographie

- ZENG, Zhi-Yuan et al. (2019). « A review structure based ensemble model for deceptive review spam ». In : *Information* 10.7, p. 243.
- ZERROUKI, Taha (2020). « Towards An Open Platform For Arabic Language Processing ».
- ZHANG, Wenping, Mengna XU et Qiqi JIANG (2018). « Opinion Mining and Sentiment Analysis in Social Media : Challenges and Applications ». In : *STACS 98*. STACS 98, p. 536-548.
- ZHAO, Siyuan et al. (2018). « Towards accurate deceptive opinions detection based on word order-preserving CNN ». In : *Mathematical Problems in Engineering* 2018.
- ZHENG, Hongwen et Yanxia ZHANG (2008). « Feature selection for high-dimensional data in astronomy ». In : *Advances in Space Research* 41.12, p. 1960-1964.
- ZHONG, Guoqiang, Xiao LING et Li-Na WANG (2019). « From shallow feature learning to deep learning : Benefits from the width and depth of deep architectures ». In : *Wiley Interdisciplinary Reviews : Data Mining and Knowledge Discovery* 9.1, e1255.
- ZHOU, Ming et al. (2020). « Progress in neural NLP : modeling, learning, and reasoning ». In : *Engineering* 6.3, p. 275-290.
- ZIANI, Amel et al. (2021). « Deceptive Opinions Detection Using New Proposed Arabic Semantic Features ». In : *Procedia Computer Science* 189, p. 29-36.

Webographie

- KAGGLE (2015). *Twitter US Airline Sentiment*. URL : <https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment>.
- LI, Jessica (2018). *UCI ML Drug Review Dataset*. URL : <https://www.kaggle.com/datasets/jessicali9530/kuc-hackathon-winter-2018>.
- VIMAL et TARUN (2019). *Amazon Reviews : Unlocked Mobile Phones*. URL : <https://www.kaggle.com/datasets/PromptCloudHQ/amazon-reviews-unlocked-mobile-phones>.
- WANG, Ben et Aran KOMATSUZAKI (mai 2021). *GPT-J-6B : A 6 Billion Parameter Autoregressive Language Model*. <https://github.com/kingoflolz/mesh-transformer-jax>.

Annexes

Annexe A

Charte graphique

La définition de la charte graphique est un élément essentiel dans la conception de l'identité visuelle d'une marque. Elle se matérialise sous la forme d'un cahier des normes graphiques qui établit les directives précises pour l'utilisation cohérente des éléments graphiques qui composent l'identité visuelle de la marque. Ce cahier des normes graphiques joue un rôle crucial en garantissant la cohérence et l'harmonie visuelle de tous les supports de communication de la marque.

A.1 Image de l'entreprise

Nous avons analysé l'image de l'entreprise, en commençant par celle de Cevital. Comme nous l'avons précédemment montré dans la figure II.1, le logo de Cevital utilise comme couleurs principales le bleu et un jaune chaleureux. Le bleu transmet une impression de fiabilité et de confiance, tandis que le jaune véhicule la joie et l'énergie. Cette palette de couleurs s'aligne parfaitement avec le ton utilisé dans leur communication et leurs illustrations, contribuant à créer une image chaleureuse et accueillante pour Cevital.

En revanche, pour Brandt, le rouge est omniprésent, comme illustré dans le logo du Groupe Brandt présenté précédemment dans la figure II.2. Il transmet une sensation de puissance et de dynamisme. De plus, Brandt utilise une nuance de gris pour renvoyer une image moderne. Les lignes, le design et l'ensemble du concept visuel évoquent fluidité et dynamisme.

A.2 Dénominations et Slogans

A.2.1 Brandt Insight

Le choix du nom de la solution, "Brandt Insight", a été soigneusement réfléchi pour capturer l'essence même de ce que cet outil représente. En combinant "Brandt" et "Insight", nous avons créé un nom qui évoque à la fois la reconnaissance de la marque et la nouvelle perspective que cette solution offre.

"Brandt" est un élément crucial pour reconnaître la marque et établir la continuité avec l'entreprise. Il garantit que l'outil reste immédiatement identifiable en tant que produit de la marque. D'autre part, "Insight" a été choisi pour souligner l'aspect novateur de la solution. Cette plateforme offre une vue d'ensemble sur la réputation de la marque, offrant ainsi une perspective inédite sur la perception en ligne de l'entreprise.

Nous avons aussi choisi le slogan "Explorez votre réputation à votre manière" reflète la capacité de l'outil à permettre aux utilisateurs d'explorer et d'analyser la réputation de manière personnalisée et adaptée à leurs besoins spécifiques.

A.2.2 Brandt Bot

En ce qui est du chatbot, nous avons opté pour le nom "Brandt Bot", ou simplement "Bbot". Cette dénomination simple permet d'associer immédiatement le chatbot à la marque tout en restant facile à mémoriser. Pour le slogan, nous avons opté pour "Votre assistant, à chaque instant" qui souligne le rôle du chatbot en tant qu'assistant personnel de gestion de réputation, prêt à répondre aux besoins des utilisateurs en temps réel.

A.3 Typographie

Pour la typographie de notre solution, nous avons soigneusement sélectionné des polices qui reflètent l'image dynamique, moderne et puissante de la marque, tout en restant fidèles à une esthétique minimaliste. Pour mettre en valeur ces caractéristiques, nous avons opté pour une police linéale (sans Serif)¹, dont les lignes nettes et épurées contribuent à transmettre une sensation de clarté et de modernité. Cette approche visuelle est également en adéquation avec la typographie utilisée par Brandt sur son site Web, notamment avec la police "Kiro".

Les choix typographiques spécifiques que nous avons faits sont présentés dans la figure II.38. Pour les titres principaux, nous avons utilisé la police "Raleway" en graisse moyenne (semibold). Cette police apporte un aspect élégant et distinctif tout en étant parfaitement lisible. En ce qui concerne le texte principal, nous avons opté pour "Gill Sans" en version légère (light). Cette police est reconnue pour sa lisibilité et sa clarté, ce qui en fait un choix idéal pour le corps du texte. Enfin, pour les autres textes et les éléments d'accentuation, nous avons choisi "Roboto" en graisse fine (thin). Cette police présente une dualité intéressante, alliant un squelette mécanique et des formes essentiellement géométriques à des courbes ouvertes et amicales. Ce mélange crée un rythme de lecture naturel, rappelant les caractéristiques des polices humanistes et à empattements.

En somme, la typographie choisie pour notre solution renforce l'identité visuelle de la marque Brandt en véhiculant ses valeurs de modernité, de dynamisme et de clarté. Cette sélection soignée de polices permet une expérience visuelle cohérente et engageante pour les utilisateurs, qu'ils interagissent avec les interfaces administratives, les tableaux de bord ou le chatbot de la solution.

¹<https://fr.wikipedia.org/wiki/Lin%C3%A9ale>

Typographie

Aa
Raleway
SemiBold

Aa
Gill Sans
Light

Aa
Roboto
Thin

Lorem ipsum

orem ipsum dolor sit amet
orem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit vel eum iriure

Polices Sans Serif

Esthétique Contemporaine Captivante

Simplicité

Dynamisme

Modernité

FIG. II.38 : Typographie de la solution

A.4 Palette de couleurs

Durant la création de la palette de couleurs pour notre solution, nous avons cherché à atteindre un équilibre entre la reconnaissance de la marque et l'intégration d'un design créatif et esthétiquement plaisant. Pour cette raison, nous avons choisi d'utiliser le cramoisi (crimson), qui est la couleur principale associée à Brandt, comme notre couleur dominante. Cette sélection nous permet de maintenir une cohérence visuelle forte tout en intégrant des éléments de design novateurs.

Évidemment, dans la conception web, le noir et le blanc jouent un rôle essentiel pour améliorer la lisibilité du contenu. Cependant, au lieu d'opter pour le blanc pur ou le noir profond, nous avons utilisé des teintes plus douces pour créer une harmonie visuelle. Nous avons choisi la nuance "poudre de bébé" (#FFF8E8) pour le blanc et le "charbon" (#2B2B2B) pour le noir. Cette décision est guidée par les directives d'accessibilité Web ⁽²⁾. En effet, un contraste trop

²<https://www.w3.org/TR/WCAG20/>

élevé pourrait devenir inconfortable pour l'utilisateur en créant une tension entre le contenu et l'arrière-plan.



FIG. II.39 : Palette de couleurs de la solution

Outre le noir et le blanc, notre schéma de couleurs est triadique : il incorpore trois couleurs réparties uniformément sur le cercle chromatique. Nous avons opté pour les trois teintes de couleurs primaires : le rouge, le jaune et le bleu. Le rouge cramoisi #DD0033 (utilisé également par Brandt) transmet dynamisme, énergie et mouvement. Le jaune doré #FFBE38 évoque la créativité et l'énergie. Enfin, le bleu céruleen #3B78BB véhicule la fiabilité, la sécurité et la confiance.

L'ensemble de notre palette de couleurs est présenté dans la figure II.39. Cette palette a été soigneusement élaborée pour garantir une expérience visuelle agréable et engageante pour les utilisateurs, tout en renforçant la cohérence visuelle avec l'identité de la marque Brandt.

A.5 Logos

Dans la conception des logos pour notre solution, nous avons mis en avant notre créativité tout en conservant l'essence de la marque principale. Nous avons créé un logotype pour "Brandt Insight" ainsi qu'un autre pour "Brandt Bot".

A.5.1 Brandt Insight

Ce logo est spécifiquement conçu pour les employés de l'entreprise. Dans la figure II.40, nous présentons les différentes variations de notre logo. Tout d'abord, le logo complet dans la figure II.40a, comprenant le nom de la solution ("Brandt Insight"), le slogan et une icône représentant son utilisation. Notre logo horizontal dans la figure II.40b affiche uniquement le nom de la solution et son slogan. Ensuite, dans la figure II.40c, nous proposons la version monogramme du logo où nous avons choisi d'utiliser les initiales B.I en inversant le B.

Nous avons mis en évidence le B en rouge dans le nom de la solution, car il s'agit d'un élément distinctif de groupe Brandt. Quant à l'icône, nous avons opté pour une loupe pour illustrer l'aspect de détail, de vision et d'analyse qu'offre notre solution. En ce qui est de la représentation montagne/graphique, elle symbolise le suivi des évolutions de la réputation ainsi que les conséquences évidentes d'un suivi et d'une gestion efficaces de la réputation de la marque.



FIG. II.40 : Variations de logo de Brandt Insight

A.5.2 Brandt Bot

Pour le logo du chatbot, qui vise les clients de la marque, nous avons adopté un design amusant et énergique pour refléter le dynamisme de la marque tout en montrant la convivialité du chatbot afin de créer une connexion plus étroite avec le client. Dans la figure II.41, nous présentons le personnage du chatbot de la marque. Il reprend les couleurs de notre palette et adopte un design ludique et dynamique. De plus, il est conçu pour ressembler à un appareil électroménager (comme un lave-linge).

Nous avons également conçu un logo à utiliser sur la page d'accueil de conversation, comme illustré dans la figure II.42. La police utilisée pour le logo complet dans la figure II.42a ainsi que pour la version monogramme dans la figure II.42b est à la fois dynamique et fluide. Le logo complet affiche également le slogan et une icône représentant le chatbot.

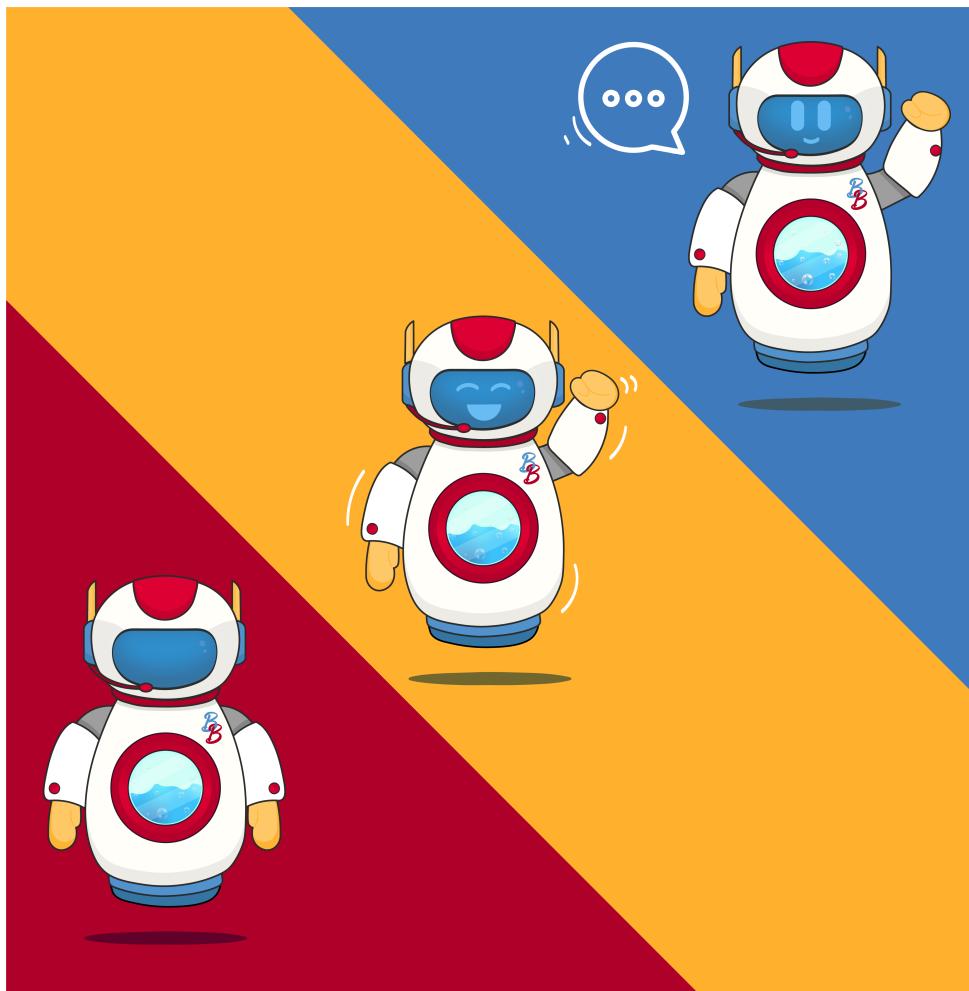


FIG. II.41 : Bbot : Personnage de Brandt Bot



(a) Logo Complet



(b) Logo Monogramme

FIG. II.42 : Variations de logo de Brandt Insight

A.6 Icônes

Lors de notre analyse de l'image de la marque, nous avons remarqué l'utilisation d'icônes simples aux contours légers. Nous avons donc choisi d'utiliser des icônes similaires provenant de Google Material Symbols³. Comme le montre la figure II.43, cela confère une sensation de simplicité et de modernité. Les icônes légèrement dessinées contribuent à maintenir une harmonie visuelle avec la marque et offrent une esthétique épurée et contemporaine.

³<https://fonts.google.com/icons>

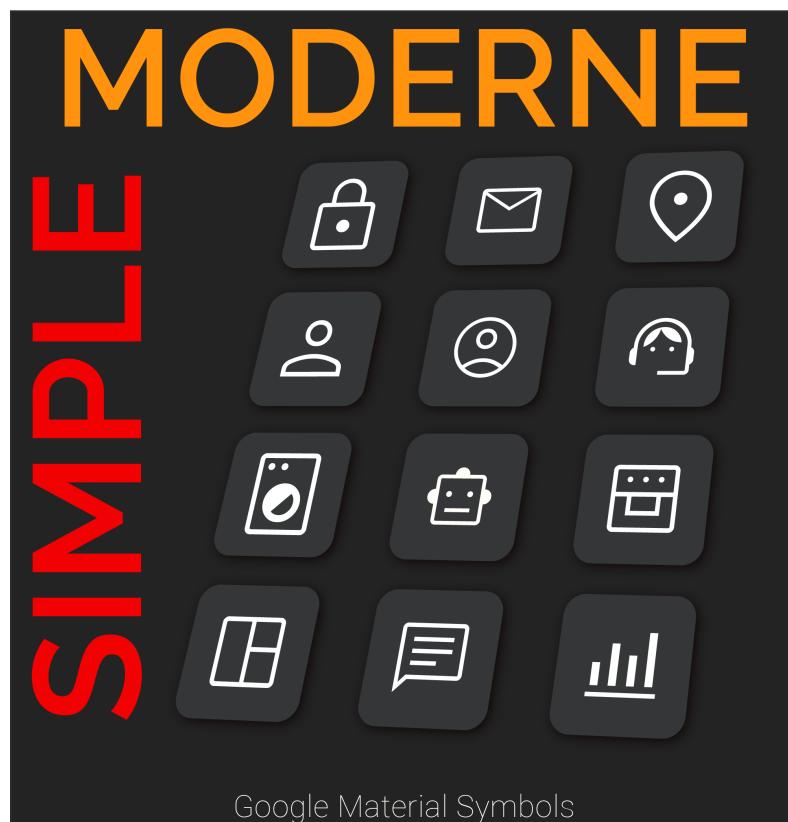


FIG. II.43 : Icônes utilisées pour la solution

A.7 Éléments d'Interface

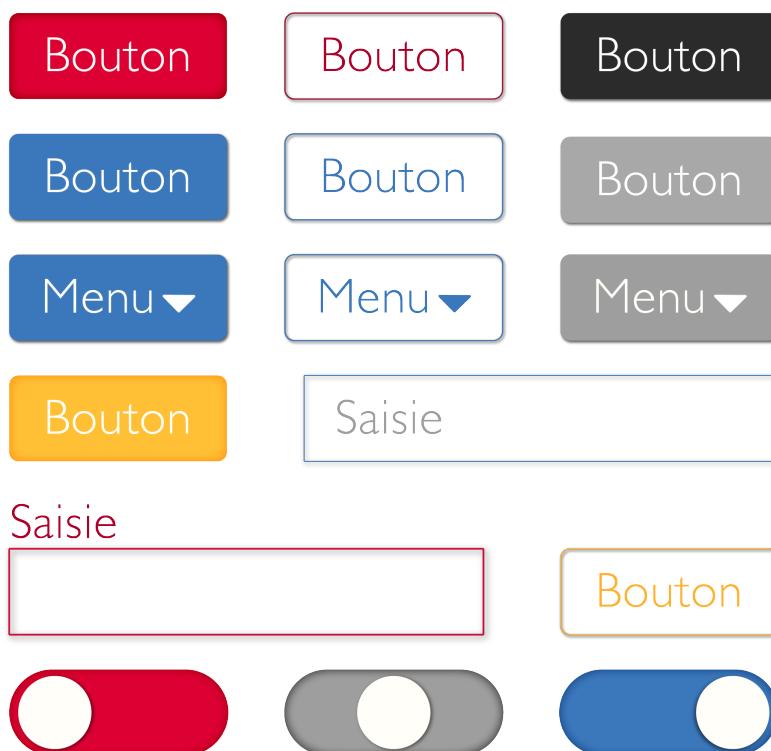


FIG. II.44 : Éléments d'Interface de la solution

Dans cette section, nous définissons les éléments de l'interface tels que les boutons, les menus déroulants et les champs de saisie. Pour les boutons principaux, nous avons opté pour la

couleur rouge cramoisi, tandis que les boutons secondaires sont en bleu céruleen et les boutons d'accentuation sont en jaune doré. Lorsque les boutons sont survolés, ils s'encadrent d'une bordure de leur couleur respective, comme illustré dans la figure II.44. Cette interaction visuelle est également présente sur le site Web de Brandt⁴. Cette cohérence d'interaction renforce l'identité visuelle de la marque tout en offrant une expérience utilisateur familière et intuitive.

A.8 Illustrations

Nous avons également élaboré des illustrations pour chaque solution proposée. Par exemple, pour Brandt Insight, nous avons conçu une illustration pour l'interface de connexion des collaborateurs, comme illustré dans la figure II.45, et une autre pour l'interface de connexion de l'administrateur, visible dans la figure II.46.



FIG. II.45 : Illustration pour la plateforme collaborateur de Brandt Insight

⁴<https://brandt.dz/>

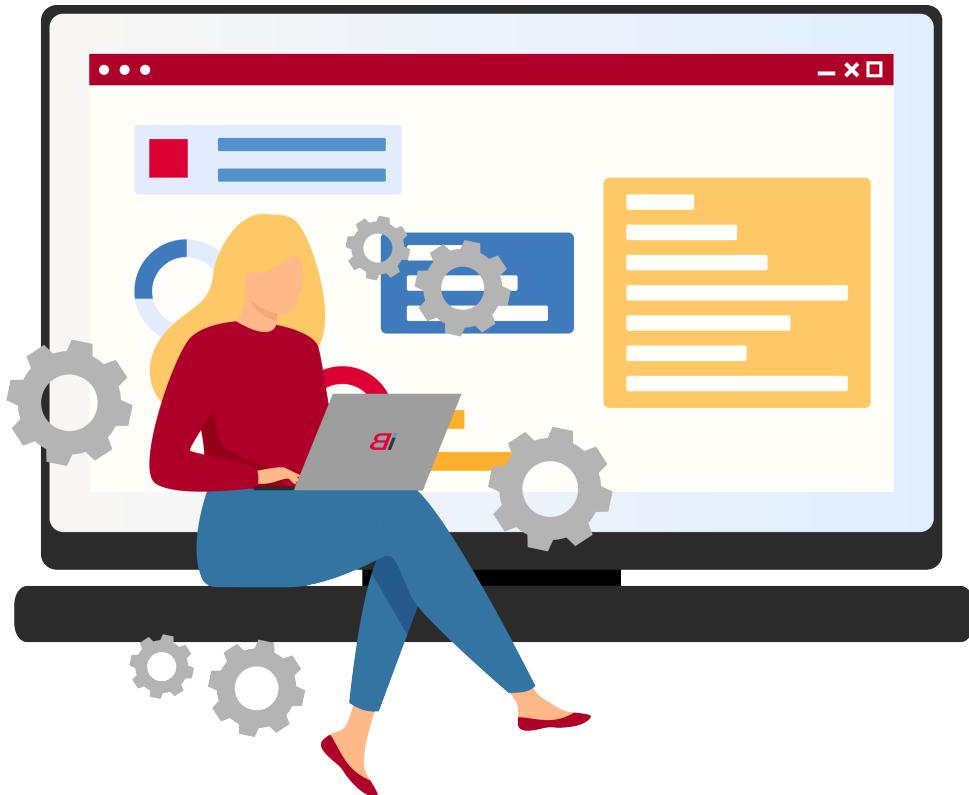
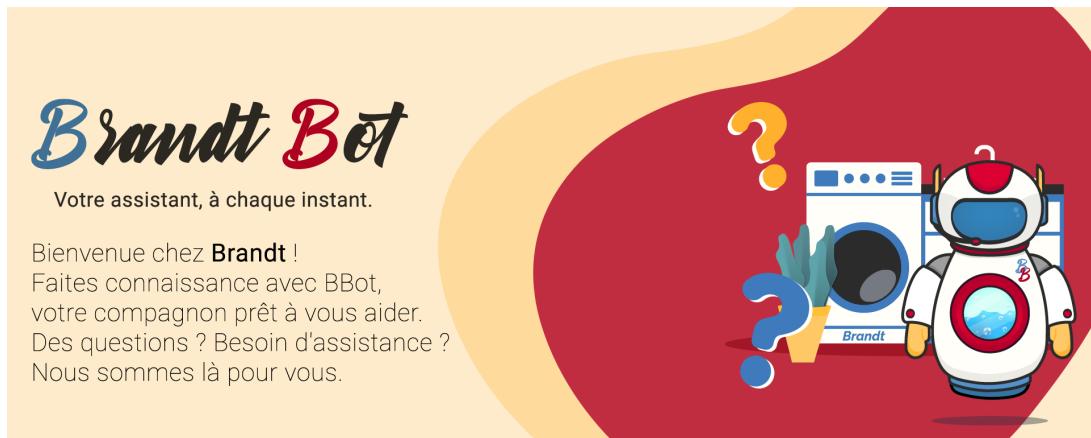
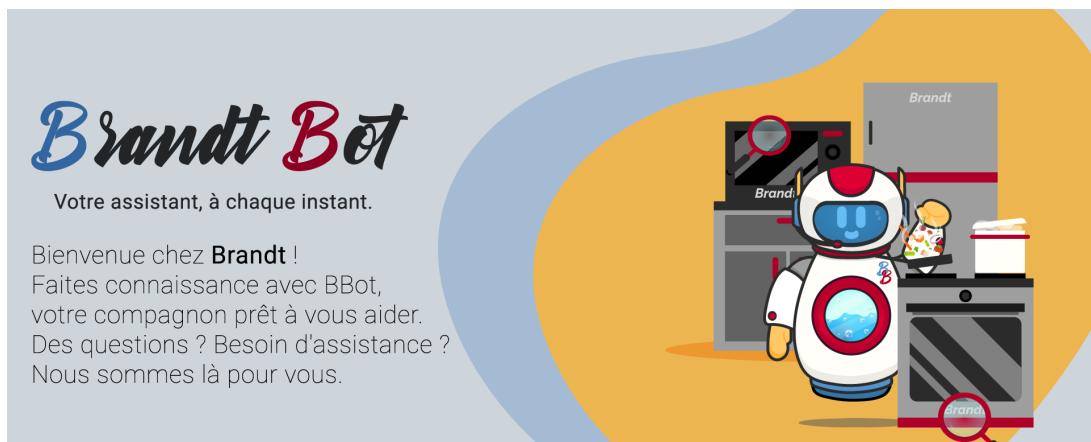


FIG. II.46 : Illustration pour la plateforme administrateur de Brandt Insight

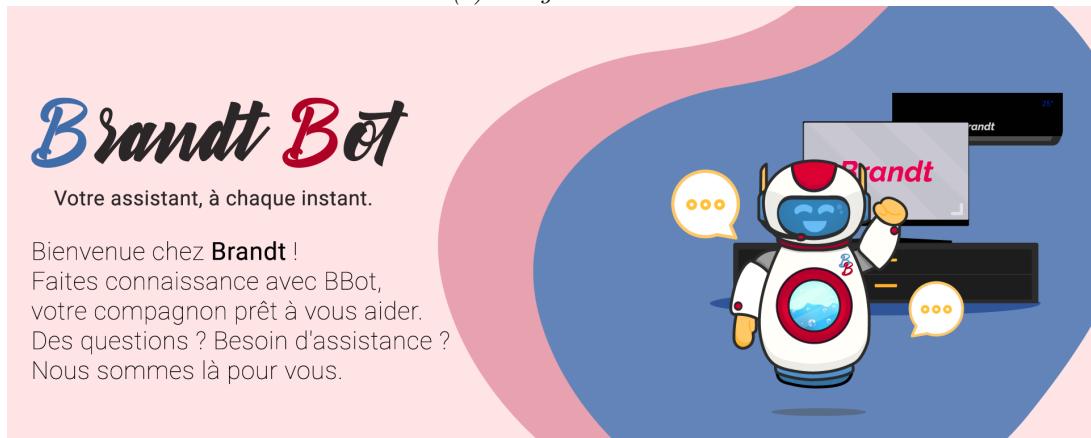
En ce qui concerne la page d'accueil conversationnelle, nous avons conçu une bannière mettant en avant Bbot (le personnage du chatbot). Pour rendre cette illustration encore plus dynamique, nous avons créé une animation avec différentes expressions affichées par le bot. Cette animation se déroule dans des décors variés, présentant différents types de produits de Brandt. Les trois principales images clés de l'animation sont présentées dans la figure II.47. Cette approche visuelle et animée ajoute une touche ludique et dynamique à la page d'accueil.



(a) Images clé 01



(b) Images clé 02



(c) Images clé 03

FIG. II.47 : Illustrations pour la page d'accueil conversationnelle