



Utilisation de données du Web communautaire à des fins de surveillance de l'usage de médicaments

Redhouane Abdellaoui

► To cite this version:

Redhouane Abdellaoui. Utilisation de données du Web communautaire à des fins de surveillance de l'usage de médicaments. Médecine humaine et pathologie. Sorbonne Université, 2018. Français. NNT : 2018SORUS548 . tel-03006554

HAL Id: tel-03006554

<https://theses.hal.science/tel-03006554>

Submitted on 16 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT EN
SCIENCES DE L'INFORMATION BIOMÉDICALE

ÉCOLE DOCTORALE 393

PIERRE LOUIS DE SANTÉ PUBLIQUE À PARIS
EPIDÉMIOLOGIE ET SCIENCES DE L'INFORMATION BIOMÉDICALE

Utilisation de données du Web communautaire à des fins de surveillance de l'usage de médicaments

Auteur :

REDHOUANE ABDELLAOUI

Dirigée par :

PR. ANITA BURGUN

DR. STÉPHANE SCHÜCK

Soutenue le 14 novembre 2018, Devant le jury composé de :

Mme. BURGUN Anita (Directrice de thèse)

M. SCHÜCK Stéphane (Encadrant en entreprise)

M. PARIENTE Antoine (Rapporteur)

M. FALISSARD Bruno (Rapporteur)

Mme. BRINGAY Sandra (Examineur)

Mme. SEROUSSI Brigitte (Examineur)

M. CUGGIA Marc (Examineur)

*J'adresse mes plus sincères remerciements
à ma directrice de thèse, Anita Burgun,
ainsi qu'à mon directeur scientifique
en entreprise, Stéphane Schück, pour
leurs conseils et pour la méthodologie
de recherche qu'ils m'ont inculquée.*

*Je remercie également mes parents, mes sœurs
ainsi que Charlotte, mon épouse, qui m'ont
toujours poussé et encouragé.*

*Je souhaite également remercier mes amis
de Kappa Santé et de l'UMRS 1138 pour
leurs accueils, leurs conseils et leur accompagnement.*

Résumé

Contexte

La pharmacovigilance souffre d'une sous-notification chronique des effets indésirables de la part des professionnels de santé. En 2014, les agences sanitaires de 44 pays ont intégré la déclaration d'effets indésirables directement par les patients, ce qui représente aujourd'hui 9% des notifications d'effets indésirables de médicaments dans ces pays. Malgré cet apport de données, la sous-notification demeure un problème majeur en pharmacovigilance. La US Food and Drug Administration (FDA), l'European Medicines Agency (EMA), et d'autres agences sanitaires, suggèrent que les réseaux sociaux pourraient constituer une source de données supplémentaire pour la détection de signaux faibles de pharmacovigilance.

Les difficultés rencontrées par les patients durant leurs traitements constituent une information présente dans les échanges sur les médias sociaux. Ces difficultés peuvent entraîner des manquements des patients dans le suivi des recommandations accompagnant leurs traitements (i.e : non-observance au traitement). L'Organisation Mondiale de la Santé (OMS) a publié un rapport en 2003 exposant le problème que pose la non-observance au traitement sur le long terme et son caractère préjudiciable à l'efficacité des systèmes de santé au niveau mondial.

L'extraction d'information de santé depuis les réseaux sociaux constitue un enjeu aujourd'hui. L'avantage des réseaux sociaux réside dans la disponibilité des données presque en temps réel et dans le caractère spontané des déclarations des patients.

Des équipes de recherches ont proposé différentes approches pour l'exploitation de ces données dans un cadre de pharmacovigilance, mais aussi dans des contextes sanitaires infectieux ou encore de mésusage des médicaments. Une large majorité des systèmes développés reste basée sur de l'analyse de textes en anglais.

Objectifs

Nos travaux ont pour objectif la mise au point d'une approche permettant l'extraction et l'exploitation d'informations de santé à partir d'échanges (en français) entre patients sur les forums de santé.

Nous axons notre démarche sur la proposition d'un outil de pharmacovigilance depuis les forums de patients dans un premier temps. Dans un second temps, nous proposons une approche de détection de cas de non-observance au traitement à partir de ces mêmes données.

Méthode

Nous présentons deux approches d'extraction d'information de santé depuis les forums

de santé. Chacune des approches proposées suit une logique de filtrage de l'information non-pertinente. Les données nécessaires à la réalisation de nos objectifs sont obtenues via la base de données Detec't mise à disposition par la société **Kappa Santé**.

La première approche proposée s'inscrit dans un contexte de détection de cas de pharmacovigilance à partir d'échanges entre patients sur des forums de santé. Nous proposons un filtre basé sur le nombre de mots séparant le nom du médicament évoqué dans un message du terme considéré comme un potentiel effet indésirable. Nous nous appuyons sur un modèle de mélange de gaussiennes afin de filtrer des termes qui ne correspondent pas à des effets indésirables. Nous utilisons ensuite une machine à support de vecteurs pour la classification des éléments conservés après filtrage.

Nous proposons une seconde approche basée sur les « topic models » afin de cibler les groupes de messages abordant les thèmes de non-observance. Nous ciblons plus particulièrement les « arrêts de traitement » et les « modifications de la posologie indiquée ».

Résultats

En terme de pharmacovigilance, le filtre gaussien proposé permet d'identifier 50,03% des faux positifs avec une précision de 95,8% et un rappel de 50%. La machine à support de vecteurs utilisée pour l'identification des effets indésirables après filtrage réalise une précision de 19,3% avec un rappel de 87,2%.

L'approche de détection de cas de non-observance permet l'identification de ces derniers avec une précision de 32,6% et un rappel de 98,5%. La revue manuelle des messages analysés montre la présence de cas de non-observance au traitement dans 6,17% des échanges considérés.

Conclusion

Les résultats montrent l'utilité des forums de santé comme source de données dans un cadre de surveillance d'usage médicamenteux en terme d'analyse des effets adverses et de la non-observance au traitement.

Mots clés

Pharmacovigilance, Adhérence, Web 2.0, Data Mining, Text Mining, Traitement Automatique du Langage Naturel, Machine learning, Modèles de mélange gaussien, Support Vector Machine, Topic Models

Table des matières

Résumé	iv
Abréviations	xii
1 Contexte	1
1.1 La société Kappa Santé	1
1.2 Le projet Detec't	2
1.3 Nos travaux	2
2 Introduction	3
2.1 Médias sociaux et pharmacovigilance	3
2.2 Médias sociaux et non-observance au traitement	4
3 État de l'art	7
3.1 La pharmacovigilance	7
3.1.1 Pharmacovigilance et effets indésirables : Définitions	7
3.1.2 Notifications spontanées d'effets indésirables de médicaments	8
3.1.3 Médias sociaux et Big Data	8
3.2 Extraction d'information de pharmacovigilance depuis les messages sur les médias sociaux	9
3.2.1 Inventaire des approches	9
3.2.1.1 Choix des sources de données	9
3.2.1.2 Extraction des messages postés par les patients	10
3.2.1.3 Prétraitement des données	11
3.2.1.4 Identification d'entités médicales dans les messages	12
3.2.1.5 Relation « effet indésirable » entre médicament et concept médical	12
3.2.1.6 Évaluation	13
3.2.2 Synthèse des approches observées	16
3.3 Approches par filtrage	20
3.3.1 Les filtres sémantiques	20
3.3.2 Les filtres statistiques	21

3.4	Pharmacovigilance sur les médias sociaux : Synthèse	22
3.5	Non-observance au traitement, Web communautaire et Topic Models	24
4	Objectifs	27
4.1	Pharmacovigilance	27
4.2	Non-observance du traitement	28
I	Détection d'informations de pharmacovigilance sur les médias sociaux	29
5	Matériel	31
5.1	Base de données Detec't	31
5.2	Données	32
5.2.1	Données de conception des algorithmes	33
5.2.2	Données d'évaluation de l'algorithme de prise	34
5.2.3	Annotation manuelle des jeux de données	34
5.2.3.1	Annotation de la prise du traitement	34
5.2.3.2	Annotation des effets indésirables	35
5.3	Dictionnaire Médical	36
5.4	Detec't Annotation Tool	38
6	Méthode	41
6.1	Filtre sémantique : Algorithme de détection de prise du médicament par le déclarant	41
6.1.1	Description de l'algorithme de détection de prise	41
6.1.2	Évaluation de l'algorithme de détection de prise	43
6.2	Reconnaissance d'entités médicales	43
6.2.1	Prétraitement des données	43
6.2.2	Détection des noms de médicaments	45
6.2.3	Détection de concepts médicaux	45
6.2.4	Issue de la détection d'entités médicales	46
6.3	Filtre statistique : Algorithme des distances entre entités médicales	47
6.3.1	Hypothèse	47
6.3.2	Calcul des distances	47
6.3.3	Principe de la classification par mélange gaussien	48
6.4	Caractérisation des effets indésirables : L'algorithme de pertinence	48
6.4.1	Support Vector Machine (SVM)	49
6.4.2	Stratégie de classification	50
6.4.3	Post-traitement des données	50

6.4.3.1	Rétrogradation des SOC's non informatifs	50
6.4.3.2	Dédoublonnage des concepts	51
6.4.3.3	Contextualisation des concepts médicaux polysémiques	51
6.5	Synthèse de notre approche de détection de cas de pharmacovigilance	51
7	Résultats	53
7.1	Package DetectAnalyzer	53
7.2	Algorithme de détection de prise du traitement dans les messages	54
7.2.1	Description des données d'évaluation de l'algorithme de prise	54
7.2.2	Évaluation de l'algorithme de Prise	55
7.3	Algorithme des distances	56
7.3.1	Description des données	56
7.3.2	Distribution des concepts médicaux dans le corpus	58
7.3.3	Classification des concepts médicaux	60
7.3.3.1	Cluster 1	60
7.3.3.2	Cluster 2	61
7.3.3.3	Cluster 3	61
7.3.4	Stratégies de filtrage	61
7.3.4.1	Stratégie 1	62
7.3.4.2	Stratégie 2	62
7.4	Algorithme de pertinence	63
II	Détection de pratiques de non-observance au traitement des patients sur les réseaux sociaux	66
8	Matériel et Méthode	68
8.1	Matériel	68
8.2	Méthode	69
8.2.1	Traitement préliminaire des données	69
8.2.1.1	Prétraitements standards	69
8.2.1.2	Standardisation des mentions de dosage	69
8.2.1.3	Estimation du modèle	70
8.2.1.4	Évaluation	73
9	Résultats	76
9.1	Caractéristiques du jeu de données	76
9.2	Variations de doses et arrêts de traitements	77
9.2.1	Estimation du modèle	77
9.2.2	Interprétation des thèmes identifiés	78

9.2.2.1	Thèmes liés à l'Escitalopram	78
9.2.2.2	Thèmes liés à l'Aripiprazole	80
9.2.3	Évaluation	82
9.2.3.1	Variations de doses	83
9.2.3.2	Arrêts de traitements	84
10	Discussion	86
10.1	Principaux résultats	86
10.2	Comparaison avec la littérature	88
10.3	Limites et perspectives	90
10.4	Discussion générale	93
11	Conclusion	99
	Bibliographie	102

Table des figures

3.1	Synthèse des approches observées pour l'extraction d'informations de pharmacovigilance depuis les réseaux sociaux	17
5.1	Données utilisées	33
5.2	Processus d'annotation des concepts médicaux	35
5.3	Medical Dictionary for Regulatory Activities	36
5.4	Detec't Annotation Tool	38
6.1	Synthèse de l'approche proposée	52
7.1	Description des données d'évaluation de la détection de prise du traitement.	54
7.2	Concepts médicaux automatiquement identifiés (au niveau SOC).	56
7.3	Description du jeu de données utilisé	57
7.4	Densité des distances observée entre concepts médicaux et noms de médicaments.	58
7.5	Q-Q plot normal.	58
7.6	Répartition des effets indésirables et des autres types de concepts médicaux.	59
7.7	Les classes de concepts médicaux obtenues.	60
8.1	Récapitulatif de l'approche proposée	71
9.1	Nombre de thèmes sélectionnés pour Escitalopram en utilisant le log-Bayes Factor.	77

Liste des tableaux

3.1	Prétraitements répertoriés dans les études retenues	11
3.2	Métriques d'évaluations utilisées dans les 26 études répertoriées	14
3.3	Systèmes en réponse aux problématiques 1 et 2	16
3.4	Caractéristiques des études sélectionnées (avec évaluation métrologique) .	18
3.5	Caractéristiques des études sélectionnées (avec évaluation qualitative) . .	19
3.6	Résultats de l'évaluation de Sarker & Gonzalez.	22
7.1	Résultats de l'annotation manuelle des messages de prise.	55
7.2	Répartition des concepts médicaux après classification.	61
7.3	Filtrage par union des classes 1 et 3.	62
7.4	Filtrage par union des classes 1 et 2.	63
7.5	Données de conception du modèle SVM	63
7.6	Calibrage du modèle SVM	64
9.1	Description des corpus.	76
9.2	Thèmes liés au corpus Escitalopram.	79
9.3	Thèmes liés au corpus Aripiprazole.	81
9.4	Thèmes associés aux deux corpus.	82
9.5	Annotations du corpus Escitalopram.	83
9.6	Thèmes associés aux deux corpus.	83

Abréviations

AFNOR	A ssociation F rançaise de N ormalisation
AMM	A utorisation de M ise sur le M arché
ANSM	A gence N ationale de S écurité du M édicament et des P roduits de S anté
API	A pplication P rogramming I nterfaces
BCPNN	B ayesian C onfidence P ropagation N eural N etwork
CHD	C lassification H iéarchique D escendante
DAT	D etec't A nnotation T ool
DTM	D ocument- T erm M atrix
EBGM	E mpirical B ayes G eometric M ean
EM	E spérance- M aximisation
FAERS	F DA A dverse D rug E vent R eporting S ystem
FAIR	F indable A ccessible I nteroperable R eusable
FDA	U S F ood and D rug A dministration
GMM	G aussian M ixture M odel
HLGT	H igh- L evel G roup T erms
HLT	H igh- L evel T erms
LDA	L atent D irichlet A llocation
LLT	L owest L evel T erms
MedDRA	M edical D ictionary for R egulatory A ctivities
NLP	N atural L anguage P rocessing
OMS	O rganisation M ondiale de la S anté

PT	P referred T erms
PV	P harmaco V igilance
RCP	R ésumé des C aractéristiques du P roduit
SOC	S ystem O rgan C lass
SVM	S upport V ector M achine
TF-IDF	T erm- F requency I nverse- D ocument- F requency
TIC	T echnologies de l' I nformation et de la C ommunication
UMLS	U nified M edical L anguage S ystem
VIH	V irus de l' I mmunodéficience H umaine
VPN	V aleur P rédictive N égative
VPP	V aleur P rédictive P ositive

Chapitre 1

Contexte

1.1 La société Kappa Santé

Kappa Santé est une société de recherche indépendante, de droit privé, créée en 2003. Spécialisée dans la pharmacoépidémiologie et dans les interventions en santé publique et numérique, elle réunit aujourd’hui une équipe de 32 personnes. L’entreprise produit des études épidémiologiques sur l’efficacité et la sécurité d’emploi des médicaments. **Kappa Santé** est constituée de professionnels de la santé et de la recherche (médecins, pharmaciens), de l’informatique (développeurs, gestionnaires de bases de données) de la statistique (statisticiens, data managers).

Les activités de **Kappa Santé** se déclinent selon deux axes : l’épidémiologie et les technologies de l’information et de la communication (TIC). En tant que prestataire de service pour l’industrie pharmaceutique et les institutions de santé, **Kappa Santé** intervient sur de multiples aires thérapeutiques telles que la Pneumologie, l’Oncologie, la Psychiatrie, l’Hématologie ou encore l’Allergologie. Ainsi, l’entreprise est amenée à travailler pour plus d’une dizaine de laboratoires pharmaceutiques (GSK, Astra Zeneca, Boiron, Sandoz, Sanofi, ...) ainsi que pour des sociétés savantes ou des institutions, telles que la Communauté européenne, le ministère de la Santé, l’Agence Nationale de Sécurité du Médicament et des Produits de Santé (ANSM).

Kappa Santé est également membre de l’Alliance Big Data et du groupe de travail de l’Association Française de Normalisation (AFNOR) sur la normalisation du Big Data.

1.2 Le projet Detec't

Depuis 2012, **Kappa Santé** entreprend une diversification vers la santé digitale. L'entreprise propose des solutions basées sur les technologies Big Data et les méthodes d'apprentissage automatique. C'est dans ce contexte que le projet Detec't est initié par **Kappa Santé**.

Detec't est un projet de mise en place d'un système de surveillance des effets indésirables des médicaments basé sur l'exploitation et l'analyse de données issues du web communautaire. Ce projet a été initié en 2012 par **Stéphane Schück**, président et directeur scientifique de **Kappa Santé** dans le cadre de cette diversification.

1.3 Nos travaux

Cette thèse s'est déroulée dans le cadre d'un projet CIFRE proposé par la société **Kappa Santé** pour la conception et le développement de la solution Detec't. Notre apport dans ce projet a consisté en la conception d'un prototype d'outil de détection d'informations dans les messages rédigés par les patients sur les forums de santé français. La série d'algorithmes constituant ce prototype sont détaillés dans ce manuscrit.

Chapitre 2

Introduction

2.1 Médias sociaux et pharmacovigilance

Les effets indésirables liés aux médicaments constituent un important problème de santé publique. La pharmacovigilance a pour objectif d’analyser ces effets indésirables, de les expliquer et de mettre en place des mesures pour les prévenir. Son organisation repose principalement sur la notification spontanée des effets indésirables par les professionnels de santé. Des directives européennes et françaises ont été mises en place afin de renforcer la pharmacovigilance et ont donné aux patients la possibilité de déclarer eux-mêmes des effets indésirables sur les sites des autorités de santé [1, 2].

L’avènement des nouvelles technologies de l’information a profondément modifié la relation médecin-malade. Les patients cherchent des informations sur internet et partagent leurs expériences notamment sur les médicaments prescrits et les effets indésirables ressentis. Il existe ainsi une importante source de données, encore peu exploitée à ce jour, pour la pharmacovigilance sur les forums de discussions.

La nouvelle réglementation européenne en pharmacovigilance et les bonnes pratiques qui en résultent, applicables depuis 2012, s’intéressent au traitement des données de pharmacovigilance identifiées sur internet et les médias numériques. Cette ouverture vers d’autres sources d’informations non structurées, produites en grand volume (Big Data), s’accompagne d’une évolution des procédures et méthodes de détection des cas d’effets indésirables et d’identification de signaux.

À ce jour, l'élaboration et la validation de ces méthodes ont fait l'objet de travaux dans le cadre de démarches exploratoires sur des périodes et des périmètres de surveillance limités [3–5]. La mise au point de méthodes de détection de signaux dans un contexte plus large de surveillance continue dans le temps et sur de nombreux sites est à consolider. Si des méthodes de détection de signaux en pharmacovigilance existent à partir de bases de données telles que les bases FDA Adverse Event Reporting System (FAERS) aux USA [6, 7] ou Eudravigilance en Europe, ces méthodes doivent être maintenant adaptées à la particularité des données issues du web communautaire.

2.2 Médias sociaux et non-observance au traitement

Un rapport publié par l'Organisation Mondiale de la Santé (OMS) en 2003 a souligné que la non-observance (ou la non-adhérence) au traitement de longue durée était un problème mondial préjudiciable à l'efficacité globale du système de santé [8].

L'observance est définie dans ce rapport comme le degré de correspondance entre le comportement d'un patient (prendre des médicaments, suivre les règles d'hygiène et son régime alimentaire) et les recommandations faites par un professionnel de santé. Le non-respect de ces recommandations a un impact sur la qualité de vie des patients, les résultats du traitement et les coûts du système de soins.

L'OMS a identifié plusieurs causes de non-adhérence aux traitements, notamment les caractéristiques du système de santé, la maladie du patient et le déroulement du traitement. Pour les patients souffrant de dépression, l'observance est liée à la fréquence d'administration d'un médicament et à un traitement concomitant. Pour les patients atteints de cancer, la crainte des effets indésirables liés au traitement a un impact négatif sur l'observance. Pour les patients diabétiques, l'observance peut varier avec l'âge, le sexe et la qualité de la relation avec le médecin. Plusieurs méta-analyses ont montré que les méthodes actuelles d'amélioration de l'observance des médicaments pour les maladies chroniques étaient pour la plupart complexes et peu efficaces [9, 10].

Le groupe Cochrane [9] a conclu qu'une mesure plus systématique et objective de l'observance ainsi que les innovations pour aider les patients à suivre les prescriptions de médicaments pour les maladies chroniques étaient des points importants à considérer

dans ce domaine. Considérant les médias sociaux comme des plateformes où les patients peuvent discuter de leurs traitements et partager leurs témoignages, ces médias pourraient constituer une nouvelle source de données pour mesurer l'observance aux traitements de manière globale et régulière.

L'utilisation des médias sociaux permet à de vastes groupes d'individus de créer et de partager des informations, des expériences et des opinions sur leurs problèmes de santé et médicamenteux grâce à des discussions en ligne [11]. Les médias sociaux fournissent aux experts en pharmacovigilance une source d'information pertinente [12].

Les médias sociaux sont un moyen prometteur pour améliorer la communication et l'engagement des patients [13]. Horvath et al. [14] et Taggart et al. [15] ont montré que le partage d'informations et la socialisation étaient les critères les plus souvent cités lorsque des patients atteints du Virus de l'Immunodéficience Humaine (VIH) décrivent un réseau social idéal. Wang et al [16] ont modélisé les discussions et les intérêts des utilisateurs d'un forum pour femmes enceintes en utilisant les topic models et ont montré que ces patientes partageaient leurs expériences, leurs peurs et leurs inquiétudes au sujet des médicaments. Stellefson et al. [17] ont passé en revue les initiatives sur les médias sociaux proposant un programme d'autogestion aux patients de plus de 50 ans pour leur maladie chronique. Les patients ont souligné l'avantage d'interagir avec d'autres patients. Par exemple, le partage d'informations via les réseaux sociaux a permis aux patients de mieux communiquer avec les professionnels de la santé.

Les patients utilisent les médias sociaux pour discuter de l'observance des traitements. Mao et al. [18] ont étudié les messages de patientes atteintes de cancer du sein traitées par des inhibiteurs d'aromatase. Près de 12,8% des patientes arrêtent les inhibiteurs d'aromatase. Chary et al. [19] ont étudié les corrélations entre la distribution géographique de l'abus d'opioïdes sur ordonnance estimée à partir des médias sociaux et les données de la National Survey on Drug Usage and Health (NSDUH). Ils ont conclu que les mentions d'abus de drogues sur Twitter étaient fortement corrélées avec les estimations du NSDUH sur l'abus d'opioïdes.

Les médias sociaux peuvent avoir une incidence sur l'observance du traitement. Dans l'étude de Horvath et al. [14], les résultats d'une enquête sur le Web pour les patients atteints du VIH ont montré que 52,6% des participants étaient considérés comme non-observants. La méta-analyse publiée par Taggart et al. [15] a identifié 2 études sur les

populations de VIH qui ont démontré un lien entre l'utilisation des médias sociaux et l'amélioration de la conformité au traitement chez les utilisateurs. Mao et al. [18] ont montré que les patientes atteintes de cancer du sein partagent des astuces pour faire face aux effets secondaires des médicaments et se soutenir mutuellement. Par exemple, 28,1% des messages ont mentionné une méthode pour traiter leurs arthralgies liées à la prise d'inhibiteur d'aromatase. Ces conseils, partagés par les patients, peuvent inclure de l'exercice et/ou des produits pharmaceutiques, qu'ils soient prescrits ou en vente libre.

L'analyse d'un grand nombre de récits nécessite des techniques d'exploration de texte automatisées [12]. Ces techniques sont déjà utilisées pour extraire des informations à partir des dossiers médicaux électroniques. Par exemple, Topaz et al. [20] ont utilisé ces méthodes pour identifier des cas de patients insuffisants cardiaques qui ne respectaient pas leur traitement à partir de comptes-rendus médicaux. La détection des comportements de non-observance dans les médias sociaux nécessite également des techniques d'exploration de texte. Parmi ces techniques, figurent les topic models. Les topic models pourraient être utilisés pour découvrir des structures sémantiques cachées dans de grands ensembles de messages provenant des médias sociaux. Ils pourraient fournir une exploration plus approfondie des comportements de non-observance.

Chapitre 3

État de l'art

3.1 La pharmacovigilance

3.1.1 Pharmacovigilance et effets indésirables : Définitions

L'OMS définit la pharmacovigilance comme « *la science et les activités relatives à la détection, à l'évaluation, à la compréhension et à la prévention des effets indésirables ou de tout autre problème lié aux médicaments* » [21]. La pharmacovigilance comprend la surveillance de l'évolution du rapport bénéfice-risque d'un médicament ainsi que l'identification de potentiels signaux d'effets indésirables.

Toujours selon l'OMS, un effet indésirable est défini comme une « *réaction nocive et non voulue à un médicament, se produisant aux posologies normalement utilisées chez l'homme pour la prophylaxie, le diagnostic ou le traitement d'une maladie ou pour le rétablissement, la rectification ou la modification d'une fonction physiologique* » [22]. L'exhaustivité des potentiels effets indésirables d'un traitement ne peut être garantie lors d'un essai clinique en raison du nombre de sujets inclus dans les essais qui ne reflètent pas la variabilité de la population cible. Il est nécessaire de maintenir une surveillance après l'Autorisation de Mise sur le Marché (AMM) d'un médicament [23] afin d'éviter de potentielles hospitalisations [24] ou encore des décès [25].

3.1.2 Notifications spontanées d'effets indésirables de médicaments

Les déclarations spontanées d'effets indésirables de médicaments constituent une source de données importante en pharmacovigilance. Dans un premier temps, les professionnels de santé ont été les seuls habilités à rapporter un effet indésirable. C'est dans un second temps que des études [26–30] ont prouvé l'intérêt de relever les déclarations d'effets indésirables directement auprès des patients. Le constat suivant a été fait : la remontée d'effets indésirables médicamenteux par les patients constitue une source de données complémentaire en pharmacovigilance [31]. En 2014, les agences sanitaires de 44 pays ont régulé la remontée d'effets indésirables par les patients [32]. Cette source de données représentait 9% des déclarations d'effets indésirables de médicaments dans ces pays.

Malgré cet apport de données, la sous-notification d'effets indésirables est restée un problème majeur en pharmacovigilance [33–35]. Aujourd'hui, de nouvelles sources de données sont donc considérées [36–43].

3.1.3 Médias sociaux et Big Data

L'avènement du Web 2.0, ou encore *web social*, a été accompagné de l'expansion de plateformes d'échanges entre patients autour des médicaments. Du fait du grand volume de données, ces médias représentent une nouvelle source d'informations pour l'identification de signaux de pharmacovigilance [44]. Leur exploitation nécessite des méthodes pour extraire, structurer et analyser ces données. L'hétérogénéité des données combinée à l'utilisation de plusieurs sources implique l'utilisation de traitements automatiques afin d'obtenir des corpus de données exploitables.

De nombreuses questions se posent sur la qualité de l'information échangée par les patients sur les médias sociaux. Les notifications d'effets indésirables par les professionnels de santé bénéficient d'un mode de remontée de données structurées contrairement aux déclarations spontanées de patients sur les réseaux sociaux et les forums de santé.

Dans les paragraphes suivants, nous proposons un inventaire des approches répertoriées extrayant de l'information pertinente en pharmacovigilance depuis les médias sociaux.

3.2 Extraction d'information de pharmacovigilance depuis les messages sur les médias sociaux

Nous nous appuyons sur notre revue de la littérature [12] pour présenter 13 études publiées entre 2010 et 2015 retenues pour cette revue ainsi que 13 publications supplémentaires [5, 45–56] identifiées dans le même champ de recherche publiées entre 2015 et 2018. Nous nous appuyons donc sur un total de 26 études traitant de la problématique de pharmacovigilance sur les médias sociaux.

3.2.1 Inventaire des approches

L'étude des 26 articles cités dans ce chapitre nous a permis de distinguer un processus général d'analyse des médias sociaux dans un cadre de pharmacovigilance. Ce processus est composé des six étapes suivantes :

1. Choix des sources de données ;
2. Extraction des messages postés par les patients ;
3. Prétraitement des posts ;
4. Identification d'entités médicales ;
5. Caractérisation d'une relation « effet indésirable » ;
6. Évaluation de l'approche proposée ;

3.2.1.1 Choix des sources de données

Dix-sept études parmi les 26 considérées (65, 4%) [1, 3, 5, 49, 50, 52, 54, 57–66] ont utilisé des forums de santé comme source de données. Neuf études parmi les 26 (34, 6%) ont utilisé des réseaux sociaux (type Twitter, Facebook, etc.) [45–48, 53–56, 67].

La majorité des études (84, 6%), 22 parmi les 26 sélectionnées, ont analysé des messages rédigés en anglais. Deux études (7, 7%) se sont intéressés à des messages rédigés en français [5, 50], une étude (3, 9%) considérait des messages rédigés en japonais [51], et une étude (3, 9%) portait sur l'analyse de messages rédigés en serbe [60]. Les différentes équipes ont extrait des volumes de données variant de l'ordre du million de messages [61]

ou du milliard de tweets [67] à des volumes moins importants, comme les 1 290 messages inclus dans l'étude de Hadzi-Puric & Grmusa [60].

Les listes de médicaments étudiés par ces équipes étaient hétérogènes. Parmi ces 26 études, 24 (92,3%) se sont concentrées sur un nombre restreint de médicaments. Les autres études avaient pour objectif la détection de signaux de pharmacovigilance basée sur un grand nombre de médicaments. C'était le cas pour Liu & Chen [63] qui ont considéré l'ensemble des médicaments répertoriés dans le Unified Medical Language System (UMLS) [68] ou dans la base FAERS de la FDA.

3.2.1.2 Extraction des messages postés par les patients

Les méthodes utilisées par les différentes équipes pour extraire des messages depuis les réseaux sociaux et les forums de santé dépendaient de la nature de la source de données. Pour les forums de santé, 13 des 26 études retenues (soit 50%) ont fait état de l'utilisation d'un web crawler (ou web scraper) permettant la collecte de pages web [1, 3, 5, 49–52, 62–66, 69]. L'étape d'extraction de données a été réalisée suivant 2 approches :

1. Dans la première approche, l'extracteur procédait en deux temps, l'extraction des codes sources des pages visées puis le nettoyage de ces derniers en éliminant les balises HTML et autres éléments non-pertinents pour l'analyse (les publicités par exemple).
2. La seconde approche consistait à s'appuyer sur la structure de la page web (balises HTML) afin d'en extraire les messages uniquement.

Le premier type d'extracteur a été utilisé par Benton et al. [3]. Dans cette étude, seulement 48% du contenu des pages extraites correspondait aux messages de patients. Abbe & Falissard [5] ont également utilisé ce système en deux étapes. La première étape consistait en l'extraction de pages web incluant des liens d'accès aux discussions du site *Doctissimo*. La seconde étape était le nettoyage du contenu des pages extraites des éléments non-informatifs.

Lorsque la source était un réseau social, type Twitter [45–47, 53, 54, 56, 67], des *Application Programming Interface* (API) ont été utilisées pour la récolte de données. À l'exception d'Adrover et al. [55] qui ont utilisé un ensemble de tweets achetés via *Gnip Inc* un

revendeur de tweets affilié à Twitter, et de Comfort et al. [48] qui ont utilisé un ensemble de messages provenant de Twitter, Tumblr, Facebook et de Blogs fourni par une équipe des laboratoires Roche participant à l'étude.

L'utilisation d'APIs a permis aux équipes d'extraire des métadonnées structurées comme la date du tweet ou encore le pseudonyme associé au compte Twitter, la géolocalisation de l'utilisateur lors de l'envoi du tweet (si autorisée sur l'appareil du patient), si le message était un tweet ou un retweet ou encore le nombre de retweets du message.

L'utilisation d'approches de *Natural Language Processing* (NLP) fut nécessaire pour l'extraction d'information depuis les messages quelle que soit la source de données.

3.2.1.3 Prétraitement des données

Le prétraitement des données se basait sur des approches permettant de considérer des corrections orthographiques ou encore des méthodes pour traiter les abréviations. Les prétraitements répertoriés sont listés en table 3.1 :

Prétraitement	Description
Anonymisation	Suppression des informations personnelles des déclarants
Corrections orthographiques	Correction des fautes d'orthographe dans les messages
Nettoyage des données	Suppression de contenus non-pertinents des pages Web extraites
Racinisation des messages	Suppression des préfixes et suffixes des mots
Segmentation des posts	Découpage du message en segments (mots, phrases ou paragraphes)

TABLE 3.1: Prétraitements répertoriés dans les études retenues

Parmi les publications retenues :

Avec l'objectif d'optimiser la détection d'informations dans les messages de patients, Li [62] a appliqué une méthode de correction des fautes d'orthographe et de frappe. Benton et al. [3], Liu & Chen [63] ainsi que Abbe & Falissard [5] ont appliqué une étape de nettoyage de balises HTML des pages Web extraites. Benton et al. [3], Leaman et al. [1] et Abbe & Falissard [5] ont eu recours à une étape de racinisation des mots des textes analysés. Benton et al. [3] et Leaman et al. [1] ont mentionné l'utilisation de l'algorithme de Porter « Porter Stemmer » [70] pour cette étape de racination. Liu & Chen [63] mettent en place une segmentation en phrases de leur corpus. Benton et al. [3]

et Leaman et al. [1] se sont basés sur des séries respectives de 20 et 5 segments où il y a co-occurrence d'un médicament et d'un potentiel effet. Nikfarjam & Gonzalez [71] ainsi que Yeleswarapu et al. [72] ont aussi eu recours à des approches de segmentation.

3.2.1.4 Identification d'entités médicales dans les messages

Cette étape consiste en la reconnaissance des entités médicales parmi l'ensemble du vocabulaire composant les messages analysés. Par entité médicale, nous entendons :

1. Tout terme médical évoqué dans le message et constituant un potentiel effet indésirable (par exemple, *nausées*, *mal de tête*, *acouphène*, *etc.*);
2. Les noms de médicaments dans les messages;

Afin de distinguer ces deux types d'entités médicales, nous désignons, dans la suite de ce travail, le premier type d'entités par l'appellation « *concept médical* ».

Parmi les études sélectionnées, 18 (69,2%) ont fait état de l'utilisation d'une terminologie médicale (ou dictionnaire médical). Les différentes terminologies répertoriées sont :

- UMLS® (Unified Medical Language System) [68];
- SIDER (Side Effect Resource) [73];
- COSTART (Coding Symbols for Thesaurus of Adverse Reaction Terms) [74];
- MedDRA® (Medical Dictionary for Regulatory Activities) [75];

Au sein des 26 études conservées, 11 (42,3%) ont mentionné une prise en charge du vocabulaire « profane » utilisé par les patients sur les médias sociaux. Une équipe [66] a utilisé MedSyn [76]. Quatre études parmi les 26 (soit 19,2%) [1, 47, 60, 62] ont utilisé une terminologie conçue spécifiquement pour leurs analyses à partir de dictionnaires existants. Cinq études [46, 53, 60, 63, 67] ont procédé à cette étape d'identification de concepts médicaux dans les messages de patients à l'aide de MetaMap [77].

3.2.1.5 Relation « effet indésirable » entre médicament et concept médical

L'étape suivante consiste en la caractérisation sémantique de la relation entre les deux entités détectées (*nom du médicament* et *concept médical*). Il s'agit d'identifier si le

concept médical correspond bien, selon le déclarant, à un effet indésirable dû à la prise du médicament évoqué dans le message.

Cette étape a été implémentée par certaines équipes dans un souci de pertinence de l'information extraite. Ces deux entités (nom de médicament et concept médical) peuvent apparaître dans un message dans le cadre d'une déclaration d'effet indésirable, mais aussi d'une demande d'information, d'une description de symptômes, lors d'un témoignage ou de tout autre type de partage d'expériences, etc.

Les études retenues sont classées en deux catégories :

1. La première catégorie regroupe les méthodes qui avaient pour objectif d'établir une relation entre les concepts médicaux identifiés et les noms de médicaments cités. Ces méthodes ont privilégié l'utilisation d'approches type *machine learning* et/ou de règles sémantiques. Quinze études parmi les 26 sélectionnées (soit 57,7%) ont utilisé ce type d'approches. Parmi ces 15 études, 12 se sont basées sur une approche type *machine learning*.
2. La seconde catégorie observée est constituée des approches adoptant des méthodes exploratoires sans catégoriser la relation entre entités. Ces études sont basées sur la recherche de co-occurrences statistiquement significatives [3, 5, 45, 60].

À titre d'exemple, Comfort et al. [48] ont proposé un système permettant la caractérisation de ces relations. Ce système était basé sur des règles sémantiques et un Support Vector Machine (SVM) pour la classification des effets indésirables depuis les médias sociaux. Ils ont analysé 311 189 messages rédigés en anglais et en provenance de Twitter, Tumblr, Facebook et d'un ensemble de blogs. Leur approche leur a permis l'identification de cas d'effets indésirables avec une sensibilité de 92,1% et une spécificité de 82%.

3.2.1.6 Évaluation

La table 3.2 présente les métriques qui ont été utilisées dans les 26 études retenues pour évaluer les méthodes mises en place.

Métrique	Nombre d'études (parmi les 26)	Pourcentage d'utilisation
Precision	13	50%
Recall	13	50%
F-score	10	38,5%
Accuracy	6	23,1%
Spécificité	1	3,9%
VPP & VPN	1	3,9%
Log-likelihood ratios	1	3,9%
Support	1	3,9%
Confidence	1	3,9%
Leverage	1	3,9%

TABLE 3.2: Métriques d'évaluations utilisées dans les 26 études répertoriées

Les évaluations des méthodes répertoriées ont été réalisées sur un sous-ensemble des données extraites par les différentes équipes (ensemble de validation ou de tests). Cet ensemble de validation a nécessité une évaluation manuelle pour la construction d'un *gold standard*. Par exemple, Hadzi-Puric & Grmusa [60] ainsi que Li [62] ont divisé la totalité des données extraites en ensembles d'apprentissage et d'évaluation alors que l'équipe de Yates [66] a utilisé 480 messages sur les 400 000 posts qu'ils ont extraits.

Cinq études parmi les 26 (19,2%) sélectionnées ont eu recours à une comparaison avec des bases de données institutionnelles comme la base FAERS [78] pour évaluer leurs approches. Le principe de cette évaluation consistait à retrouver des résultats connus des autorités de santé.

D'autres modes opératoires ont été utilisés pour l'évaluation des méthodes. Benton et al. [3] se sont référés aux éléments du Résumé des Caractéristiques du Produit (RCP) du médicament étudié.

Sept études parmi 26 (26,9%), ne se concentraient pas seulement sur les effets indésirables attendus, mais aussi sur la découverte de couples (*médicament, évènement indésirable*) non répertoriés dans les notices de médicaments, la littérature ou dans les bases de données institutionnelles. Pierce et al. [47] avaient pour objectif d'identifier les couples (*médicament, évènement indésirable*) apparaissant sur Twitter et Facebook avant d'être répertoriés dans la base FAERS. Ils ont extrait 935 246 posts et ont automatiquement

identifié 98 252 concepts médicaux. Parmi ces messages, les déclarants ont fait état d'un lien entre médicaments et concepts médicaux dans 13 tweets. Une seule paire (*médicament, effet indésirable*) est apparue sur ces réseaux sociaux avant d'apparaître dans la base FAERS.

Une seule équipe (Coloma et al. [45]) a utilisé les informations de géolocalisation de posts provenant de Twitter, Facebook et Google+ via leurs APIs respectives. Selon le médicament étudié, entre 85% et 98% des messages évoquant ces produits provenaient de Twitter. Ces métadonnées ont été utilisées pour la contextualisation des concepts médicaux identifiés dans les posts extraits.

Sarker, Nikfarjam & Gonzalez [46] sont à l'origine d'un travail collaboratif sur la détection d'informations de santé depuis les réseaux sociaux. Trois problématiques ont été traitées :

- *Problématique 1* : Détection automatique de déclarations d'effets indésirables ;
- *Problématique 2* : Détection automatique d'entités médicales ;
- *Problématique 3* : Normalisation des entités médicales utilisées par les patients pour standardiser l'extraction d'informations ;

Un corpus de tweets annotés a été mis à disposition des équipes à l'adresse suivante : <http://diego.asu.edu/downloads>.

Onze équipes ont participé à ce travail et ont proposé 24 systèmes pour répondre aux problématiques 1 et 2 (aucune proposition pour le problématique 3). L'ensemble des systèmes proposés utilisaient des approches de classification supervisée. Les deux meilleurs systèmes ont utilisé une approche type *Ensemble learning* [79]. Les caractéristiques des meilleurs systèmes proposés sont exposées en table 3.3.

Problématique	Système	Méthode	Performance
Problématique 1	<i>Mayo-NLP</i> [80]	Ensemble learning	F-score = 41,9%
		Cooccurrences	
		Règles de négation	
		Analyse de sentiments	
	<i>TJZZF</i> [81]	Ensemble learning	F-score = 41,8%
		Principe d'entropie maximale	
		Pondération <i>TF-IDF</i>	
	<i>ReadBioMed</i> [82]	Analyse de sentiments	F-score = 35,8%
		Support Vector Machine	
Problématique 2	<i>NTTUMUNSW</i> [83]	Support Vector Machine	F-score = 33%
		Topic Models	
	<i>SwissChocolate</i> [84]	Analyse de sentiments	F-score = 31,8%
Problématique 2	<i>DLIR</i> [16]	Étiquetage morpho-syntaxique	F-score = 61,1%
	<i>NTTUMUNSW</i> [85]	Word Embedding	F-score = 54%

TABLE 3.3: Systèmes en réponse aux problématiques 1 et 2

3.2.2 Synthèse des approches observées

La figure 3.1 propose un récapitulatif de l'ensemble des étapes observées au cours de cette revue de la littérature. Nous y présentons les étapes ainsi que les traitements, méthodes utilisées dans les études retenues.

Les tables 3.4 et 3.5 rendent compte des caractéristiques des approches observées. Nous présentons dans la table 3.4 les approches avec une évaluation basée sur des indicateurs métrologiques. La table 3.5 présente les études ayant utilisé une évaluation qualitative des résultats obtenus.

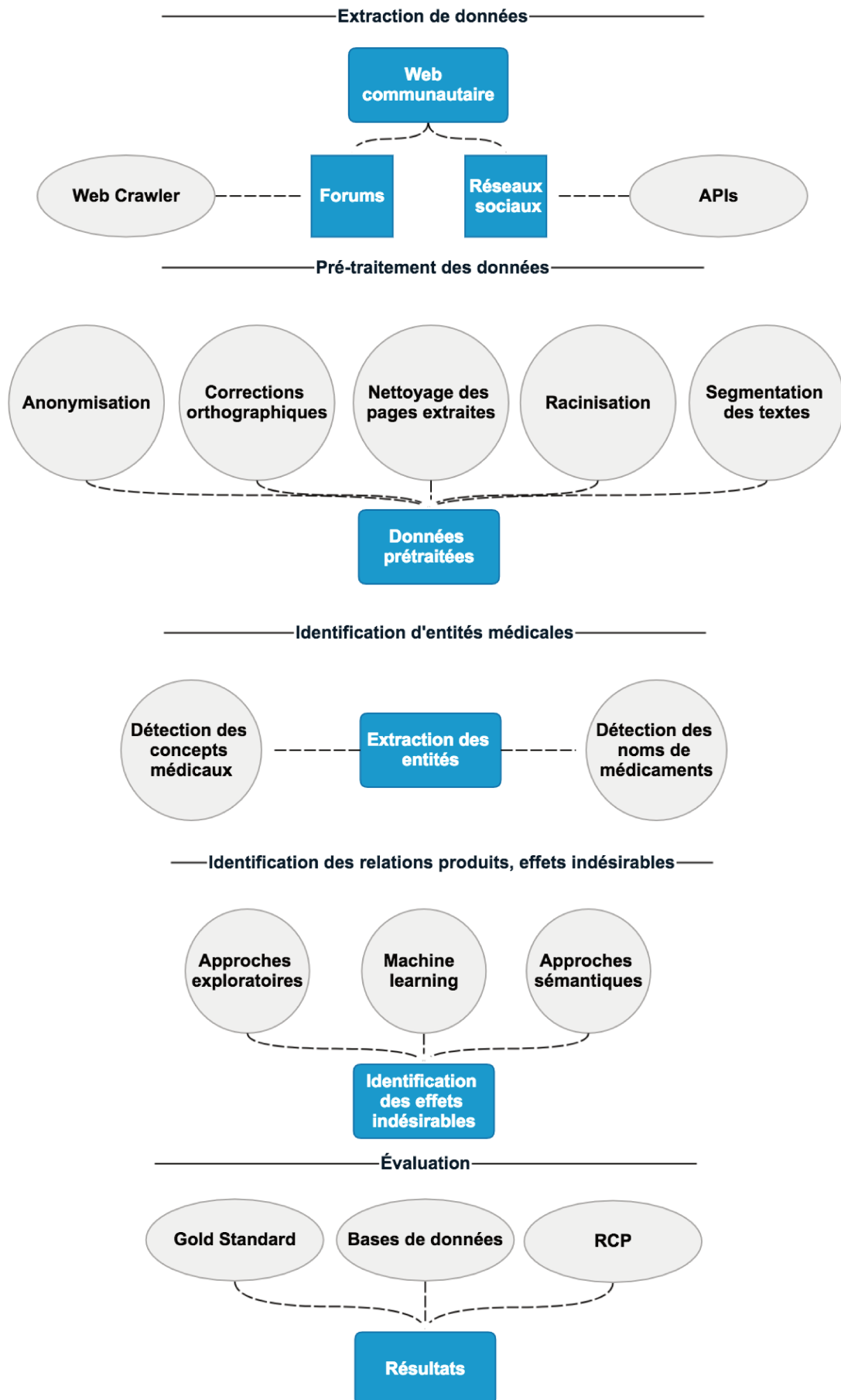


FIGURE 3.1: Synthèse des approches observées pour l'extraction d'informations de pharmacovigilance depuis les réseaux sociaux

Auteurs	Source	Volumétrie	Évaluation
<i>Benton et al.</i> [3]	Forums	1 100 000 messages	Precision = 35,1% Recall = 77,0%
<i>Bian et al.</i> [67]	Twitter	2 000 000 000 tweets	Accuracy = 74% AUC = 82%
<i>Hadzi-Puric & Grmusa</i> [60]	Forums	1 290 messages	Precision = 75,3% Recall = 64,7% F score = 69,6%
<i>Jiang et al.</i> [61]	Forums	12 500 000 messages	Precision = 69,3% Recall = 64,6% F score = 67% Accuracy = 64,7%
<i>Leaman et al.</i> [1]	Forums	6 890 messages	Precision = 78,3% Recall = 69,9% F score = 73,9%
<i>Li</i> [62]	Forums	12 118 (drug reviews)	log-likelihood ratio $\in [5, 98, 14, 61]$
<i>Liu & Chen</i> [63]	Forums	185 874 messages	Precision = 82% Recall = 56,5% F score = 66,9%
<i>Nikfarjam & Gonzalez</i> [71]	Forums	6 890 messages	Precision = 70% Recall = 66,3% F score = 68%
<i>Sampathkumar et al.</i> [69]	Forums	7 961 messages	F score = 86,4%
<i>Wu et al.</i> [64]	Forums	178 871 (discussions)	Precision = 70% Recall = 69%
<i>Yang et al.</i> [65]	Forums	3 149 (discussions)	Support $\in [9, 53 \times 10^{-4}, 1, 06 \times 10^{-1}]$ Confidence $\in [4, 37 \times 10^{-3}, 1, 00 \times 10^{-1}]$ Leverage $\in [-8, 68 \times 10^{-4}, 5, 85 \times 10^{-2}]$
<i>Yates et al.</i> [66]	Forums	400 000 messages	Precision = 63% Recall = 36%
<i>Yeleswarapu et al.</i> [72]	Forums	13 500 messages	BCPNN Variance $\in [6, 76, 1, 83]$
<i>Comfort et al.</i> [48]	Twitter Trumblr Facebook Blogs	311 189 posts	Sensibilité = 92,1% Spécificité = 82,3% Accuracy = 78%
<i>Liu et al.</i> [49]	Forums	261 464 messages	AUC $\in [77, 88\%, 81, 48\%]$
<i>Chen et al.</i> [50]	Forums	325 435 messages	F score = 70%
<i>Eshleman et al.</i> [53]	Twitter	157 735 tweets	F score = 81%
<i>Korkontzelos et al.</i> [54]	Twitter Forum	8 061 posts	F score $\in [72, 14\%, 73, 22\%]$
<i>Sarker, Nikfarjam & Gonzalez</i> [46]	Twitter	15 717 tweets	F score $\in [31, 8\%, 61, 1\%]$
<i>Cocos et al.</i> [56]	Twitter	81 médicaments	F score = 75,5%

TABLE 3.4: Caractéristiques des études sélectionnées (avec évaluation métrologique)

Auteurs	Source	volumétrie
<i>Abbe & Falissard</i> [5]	Forums	2 415 (titres de discussions)
	Twitter	
<i>Coloma et al.</i> [45]	Facebook	4 773 posts
	Google+	
<i>Pierce et al.</i> [47]	Twitter	935 246 posts
	Facebook	
<i>Matsuda et al.</i> [51]	Blogs	54 010 blogs
<i>Duh et al.</i> [52]	Forums	1 268 messages
<i>Adrover et al.</i> [55]	Twitter	39 988 306 tweets

TABLE 3.5: Caractéristiques des études sélectionnées (avec évaluation qualitative)

Suite à l'étude des articles présentés dans cette section, nous proposons la synthèse suivante :

En termes de disponibilité d'informations sur les forums de santé français, Abbe & Falissard [5] ont étudié 2 415 titres de discussions provenant du forum *Doctissimo* afin d'explorer les préoccupations des patients au sujet des antidépresseurs et des anxiolytiques. Ils ont conclu que les préoccupations exprimées étaient cohérentes avec les situations observées en vie réelle et constituaient des demandes et des plaintes conventionnelles à propos de problèmes de santé mentale. Par rapport à la possibilité d'extraire ces informations, Chen et al. [50] ont présenté des résultats préliminaires encourageants de reconnaissance de noms de médicaments (F-score 94%), de concepts médicaux (F-score 81%) dans les messages issus de forums de santé français.

En termes de contextualisation des effets indésirables identifiés sur les médias sociaux, Coloma et al. [45] ont conclu que les données disponibles sur les réseaux sociaux considérés dans leur étude étaient rares et en grande partie non traçables. Ce qui rend difficile de fournir des informations sur des signaux précoces pour la sécurité des médicaments qu'ils ont étudiés.

Pierce et al. [47] corroborent cette conclusion avec une seule paire (*médicament*, *effet indésirable*) apparue sur Twitter avant d'apparaître dans la base FAERS. Dans leur étude, 98 252 couples ont été initialement détectés dans le corpus considéré.

3.3 Approches par filtrage

Les messages de patients sur les médias sociaux ne sont pas seulement factuels [86]. Ces messages peuvent contenir des informations contextuelles (sur l'état de santé du patient par exemple, des avis sur les traitements prescrits, partages d'expériences, discussions autour de nouvelles recherches, etc.). L'un des défis à relever reste la distinction entre les déclarations d'effets indésirables et les autres types de co-occurrences possibles (description de symptômes, partage de notices de médicaments, etc.). Golder et al. [86] ont fait le constat suivant : « *Le vocabulaire utilisé pour déclarer un effet indésirable peut également être utilisé pour la description des indications du traitement, des effets bénéfiques de ce dernier ou encore des effets non expérimentés par le patient.* ». Benton et al. [3] ont fait le même constat. Ils ont analysé des échanges de patients issus de forums de santé évoquant les traitements les plus utilisés contre le cancer du sein. Le but de cette étude était l'identification d'effets indésirables liés à la prise de ces traitements. Le terme « *cancer utérin* » est apparu 374 fois avec le terme « *tamoxifen* ». Il s'avère qu'une large majorité des messages évoquaient l'anxiété face à la prise de *tamoxifen* du fait du possible effet indésirable « *cancer utérin* ». Ces exemples illustrent la nécessité d'éliminer le maximum de faux positifs avant la détection d'effets indésirables.

Sarker et al. [87] ont conclu dans leur revue de la littérature que le filtrage des données après leur extraction constitue un véritable enjeu.

Les méthodes de filtrage pourraient améliorer les performances des approches de détection d'effets indésirables sur les médias sociaux en réduisant la quantité de faux positifs. En nous basant sur une revue des approches de filtrage nous distinguons deux types de filtres.

3.3.1 Les filtres sémantiques

Ces filtres se basent sur la structure et la sémantique de la phrase analysée (règles de négation, analyse de sentiments, etc.).

Liu et al. [88] ont développé des algorithmes basés sur des règles de négation ainsi que des éléments linguistiques et du vocabulaire médical pour filtrer les indications de médicaments et les cas non déclarés dans la base *FAERS*. Ils ont utilisé 1 822 discussions

traitant de bêtabloquants. Cette équipe a identifié 71% des concepts médicaux comme des déclarations d'effets indésirables avérés, 20% comme des indications de traitement et 9% ont été invalidés en tant qu'effets indésirables.

Powell et al. [89] ont présenté « *Social Media Listening* » un outil d'optimisation de détection de cas de pharmacovigilance sur les réseaux sociaux. Cet outil permettait la suppression de publicités qu'ils nommaient « *Junk* » (essentiellement liées aux pharmacies en ligne). Les messages contenant le nom du médicament furent catégorisés « *mention* », les messages contenant au moins un concept médical ont été catégorisés « *Proto-AE* » et les messages qui décrivaient des interactions quelconques ont été catégorisés en « *Health System Interaction* ». Leur étude a montré que seulement 26% des messages considérés contiennent de l'information pertinente. Entre 11 et 51% des messages contenaient des concepts médicaux (*Proto-AE*). Parmi ces messages, entre 3 et 34% évoquaient des médicaments sans ordonnance.

3.3.2 Les filtres statistiques

Le deuxième type de filtres se base sur des méthodes statistique, en particulier sur les *topic models* [90]. Yang et al. [91] ont utilisé l'allocation latente de Dirichlet [92] pour filtrer les thèmes non-pertinents et ainsi réduire l'ensemble des messages à analyser aux posts avec une probabilité élevée de contenir des déclarations d'effets indésirables. Cette méthode a été évaluée par comparaison avec d'autres approches issues de la littérature. Les approches qui ont été utilisées pour la comparaison sont :

- EAT (**E**xample **A**daptation for **T**ext categorization) ;
- PNLH (**P**ositive examples and **N**egative examples **L**abeling **H**euristics) ;
- ACTC (**A**ctive semi-supervised **C**lustering bases **T**wo-stage text **C**lassification) ;
- SVM à noyau Laplacien ;

Cette évaluation a été réalisée à l'aide d'un calcul de *F score*. Ces quatre méthodes ont vu leurs performances améliorées par l'application du filtre proposé. Le gain de *F score* fluctuait entre 1,94% et 6,14%.

La restriction du corpus aux messages avec un contenu pertinent a également été proposée par Korkontzelos et al. [54]. Leur approche consistait en l'application d'une analyse

de sentiments aux tweets analysés afin de filtrer les tweets avec une connotation positive. L'idée était de se restreindre à des tweets contenant des concepts médicaux et un sentiment négatif. Cette approche a permis un gain de F score de 1,08%

Sarker et Gonzalez [93] ont amélioré leur détection d'effets indésirables à l'aide d'approches de filtrage basées sur différentes caractéristiques. Ces caractéristiques ont été déterminées par l'application du score de classification *leave-one-out* et ont été évaluées via *F score* et *Accuracy*. Les caractéristiques sélectionnées pour filtrer les messages sans déclaration d'effet indésirable sont présentées en table 3.6. L'utilisation de l'ensemble de ces caractéristiques de filtrage a pour résultat un *F score* de 67,8%.

Caractéristique de filtrage	Accuracy	F score
N-grams	82,6%	65,4%
Tf-idf	82,6%	65,2%
Polarités des phrases	84%	66,9%
Résultat Positif/Négatif	83,9%	66,5%
Correspondance avec le lexique d'effets indésirables	83,5%	65,9%
Analyse de sentiment des messages	88%	- - - -
Filtrage par thèmes abordés	83,7%	67%

TABLE 3.6: Résultats de l'évaluation de Sarker & Gonzalez.

Bian et al. [67] ont utilisé un SVM pour filtrer les tweets non-pertinents dans leur corpus de deux milliards de tweets. Cette équipe a eu recours à une approche de filtrage, car une large majorité des tweets extraits ne mentionnait pas le nom du médicament. Une approche similaire a été proposée par Adrover et al. [55]. Ils ont utilisé un SVM combiné à une analyse de sentiments pour réduire les tweets analysés et isoler les déclarations d'effets indésirables. Parmi les 39 988 306 tweets analysés, seulement 1 642 étaient pertinents.

3.4 Pharmacovigilance sur les médias sociaux : Synthèse

Cette revue de la littérature nous a permis d'identifier les méthodes employées pour répondre à cette problématique.

À la suite de ce travail, nous identifions les limites suivantes :

Limite 1 : Les études utilisant des réseaux sociaux (et non des forums de santé) comme source de données dans un cadre de détection d'effets indésirables ont rapporté une faible quantité de posts pertinents.

Parmi les deux milliards de tweets extraits par Bian et al. [67], entre 10 et 264 tweets contenaient les noms des médicaments étudiés. Adrover et al. [55] ont extrait 39 988 306 tweets dont 1 642 étaient pertinents pour leur analyse. Comfort et al. [48] ont réalisé une sélection aléatoire de deux jeux de données de 2 500 tweets, les pourcentages de tweets pertinents étaient de 3,7% et de 4,9%. Korkontzelos et al. [54] ont utilisé 8 061 messages de Twitter et du forum *DailyStrength*. Seuls 1 782 posts étaient des tweets. Pierce et al. [47] avaient pour objectif d'identifier les couples (*Médicament*, *Concept médical*) correspondant à des effets indésirables et détectés sur Twitter et Facebook avant leur apparition de la base FAERS. Parmi les 98 252 couples identifiés (dans 935 246 posts), 13 paires (*Médicament*, *Concept médical*) correspondaient à des effets indésirables avérés dont une seule paire était apparue sur ces réseaux sociaux avant la base FAERS.

Nous observons donc que l'utilisation des forums de santé permet l'extraction d'une volumétrie de données pertinentes plus importante que les réseaux sociaux et donc maximise les chances d'identifier des effets indésirables avérés.

Limite 2 : Seules deux publications font état d'un système d'analyse automatique de messages rédigés en français sur les médias sociaux [5,50]. Parmi le reste des publications répertoriées, 84,6% ont analysé des messages rédigés en anglais, une publication décrit l'analyse de blogs tenus en Japonais [51] et une publication décrit l'analyse de messages rédigés en Serbe [60].

Limite 3 : L'ensemble des publications étudiées font état d'un déséquilibre des jeux de données utilisés. Ce déséquilibre se présente comme un taux de faux positifs plus élevé que le nombre d'effets indésirables manuellement identifiés.

3.5 Non-observance au traitement, Web communautaire et Topic Models

Notre recherche d'études ayant traité la question de la non-observance au traitement à partir des données du web communautaire n'a pas permis d'identifier de tels articles. Nous avons effectué notre recherche sur la base de données bibliographiques Medline à l'aide de la requête 3.1 :

$$(\text{« Medication Adherence » [Mesh]}) \text{ AND } \text{« Social Media » [Mesh]} \quad (3.1)$$

Cette recherche a retourné 11 publications. Aucune de ces dernières ne traitait de la problématique de non-observance à partir des données des médias sociaux. Aucune des variantes de cette requête n'a permis d'identifier de telles publications.

Nous avons par ailleurs réalisé revue de la littérature sur l'usage des topic models en nous focalisant sur l'usage des topic models en santé.

Les topic models sont des méthodes d'exploration de textes conçues pour faire émerger les thèmes principaux dans un ensemble de documents. Avec les topic models, les mots qui apparaissent fréquemment ensemble dans le texte sont regroupés en différents thèmes (ou *topics*). Sur la base de ces thèmes, les topic models fournissent un outil pour la classification non supervisée de collections massives de documents. Le modèle Latent Dirichlet Allocation (LDA) a été développé et défini par Blei et al. [94] comme « ... *Un modèle génératif probabiliste pour des collections de données discrètes telles que des corpus de texte ...* ».

Des algorithmes de topic models ont été utilisés pour analyser la composition thématique de corpus extraits depuis des médias sociaux dans divers domaines tels que la politique [95]. Plusieurs équipes ont exploré le contenu des tweets en utilisant le modèle LDA pour identifier des sujets de santé, y compris le tabagisme [96], la grippe saisonnière et les allergies [97], ainsi que l'obésité infantile [98]. Sullivan et al. [99] ont analysé les commentaires d'utilisateurs sur le site amazon pour construire un système de notation des compléments alimentaires.

Des forums de santé ont également été explorés en utilisant le modèle LDA. Yang et al. [91] ont analysé 1 500 messages provenant de forums de patients pour détecter des

déclarations d'effets indésirables de médicaments. Les distributions des thèmes obtenus en appliquant le modèle LDA à ce corpus ont permis d'utiliser des mesures de similarité pour classer des nouveaux messages (qui n'ont pas servi à la modélisation). Les auteurs ont proposé un classifieur de messages basé sur ces mesures. Toutes les études décrites ci-dessus ont utilisé des messages en anglais.

Avec l'objectif d'analyser la qualité de vie de patients atteints de cancer du sein, Tapi Nzali et al. [100] ont étudié des publications de groupes Facebook et d'un forum public français sur le cancer du sein utilisant le modèle LDA. Ils ont analysé des messages en français.

Plusieurs algorithmes peuvent être appliqués pour utiliser les topic models. La version originale de la modélisation LDA proposée par Blei et al [94] a été largement utilisée (Par exemple, [91,95,96,98,100]). Paul et Dredze ont développé une extension du modèle LDA [97,101,102]. Pour établir leur modèle ATAM (Ailment Topic Aspect Model), ils ont ajouté plusieurs composants pour associer un terme à un thème (par exemple, une maladie), ou le considérer comme non-pertinent. Puis, sur la base de 144 millions de tweets, ils ont estimé des thèmes généraux et des thèmes spécifiques à des maladies tels que la grippe, le cancer et les problèmes dentaires. La cohérence sémantique des thèmes obtenus par ATAM était meilleure pour 61,1% (11/18) des thèmes estimés en comparaison avec le modèle LDA [102].

Dans nos travaux, nous nous intéressons à l'utilisation du modèle LDA pour analyser les thèmes dans les messages des patients et identifier les cas de non-observance au traitement. À notre connaissance, il s'agit de la première étude visant à identifier les messages de forums de santé liés aux comportements de non-observance au traitement.

Chapitre 4

Objectifs

4.1 Pharmacovigilance

Les limites identifiées dans le chapitre 3 :

1. Volumétrie de données pertinentes plus importante sur les forums de santé que sur les réseaux sociaux ;
2. Peu de publications ayant analysé des messages rédigés en français ;
3. Taux de faux positifs plus élevé que le taux d'effets indésirables manuellement identifiés dans les jeux de données ;

nous poussent à explorer les approches de filtrage dans le but de réduire la quantité de faux positifs issus des messages rédigés en français sur les forums de santé.

Ce travail vise à proposer une méthode de traitement des données des forums de santé français dans un cadre de pharmacovigilance. Pour ce faire, nous nous sommes intéressés aux positions des mots dans les messages en les modélisant à l'aide de modèles de mélanges de gaussiennes basés sur les distances entre mots.

Après filtrage d'entités médicales à l'aide de ce modèle de mélange, nous avons mis en place un modèle SVM pour la classification des entités conservées après filtrage en deux groupes :

- Effets indésirables ;
- Autre type d'entités ;

4.2 Non-observance du traitement

Les questionnements qu’expriment les patients sur internet et les préoccupations qu’ils induisent (effets secondaires, efficacité des médicaments, etc.) sont des éléments importants de l’adhérence au traitement [103].

L’utilisation des forums de santé pour explorer les témoignages de non-observance au traitement est restée inexploitée jusqu’à présent. Notre objectif était d’évaluer une approche type topic models pour identifier les messages décrivant les comportements de non-observance aux traitements.

Les *topics* correspondent à des groupes de mots qui représentent les thèmes abordés par les patients. Les distributions de ces thèmes dans un corpus de messages devraient permettre l’extraction ciblée de messages correspondant à des comportements de non-observance. Nous nous sommes concentrés sur deux pratiques de non-observance :

- Modification de dose par le patient sans consultation d’un médecin ;
- Arrêt du traitement sans consultation d’un médecin ;

Première partie

Détection d'informations de pharmacovigilance sur les médias sociaux

Chapitre 5

Matériel

Nous présentons, dans les chapitres *matériel* et *méthode* de cette première partie, les éléments (données, dictionnaire médical, approches statistique et informatique) permettant la construction de notre méthode de détection des cas de pharmacovigilance dans des messages postés par des patients sur des forums de santé français.

La base de données Detec't utilisée pour ce travail a été mise à disposition par la société **Kappa Santé**.

5.1 Base de données Detec't

La base de données Detec't [12, 104, 105] a été développée et est maintenue par la société **Kappa Santé** depuis 2012. Cette base de données est constituée de messages postés par des internautes sur 26 forums de santé français. Nous avons utilisé la version datant du mois de janvier 2016 de cette base de données qui été constituée de 26 862 232 messages, postés entre 2002 et 2015, traitant de 389 médicaments.

Les forums surveillés par Detec't n'imposent aucune restriction en nombre de caractères aux utilisateurs.

Ces forums sont sélectionnés en deux temps :

- La première version de cette base de données été constituée de messages provenant des 5 forums français générant le plus de trafic (**Doctissimo**, **atoute.org**, **e-santé**,

Santé Médecine et aufeminin). Ces forums ont accueilli plus de 90 000 visiteurs uniques par jour en 2016 ;

- L'enrichissement de cette base de données a été réalisé en sélectionnant des forums de santé supplémentaires. La sélection de ces nouvelles sources s'est basée sur une grille de décision développée pour évaluer la pertinence des forums de santé selon différents critères [106] ;

La base de données Detec't est alimentée par un Web Crawler [12, 107] développé par Kappa Santé. Ce programme parcourt des pages web et extrait le contenu des balises HTML délimitant les messages postés par les patients. Ce Web Crawler utilise un nom de médicament en entrée et extrait toute discussion, sur les forums surveillés, contenant au moins une citation de ce nom de médicament. La liste de médicaments figurant dans cette base a été définie par la société Kappa Santé en fonction des objectifs du projet Detec't. Les discussions extraites sont traitées pour supprimer les publicités, les citations d'autres messages (afin d'éviter la présence de doublons) ainsi que les signatures des messages.

Les métadonnées liées aux messages extraits comprennent :

- la date du message ;
- le(s) médicament(s) évoqué(s) dans le message ;
- le forum d'où le message est extrait ;
- si le déclarant a pris le médicament cité dans le message ;

Cette dernière information est identifiée à l'aide d'un algorithme de détection de prise de traitement par le déclarant appliqué par le Web Crawler lors de l'extraction des messages. L'algorithme de prise est décrit dans le chapitre *Méthode* en section 6.1. Nous avons développé cet algorithme chez Kappa Santé en collaboration avec Carole Faviez. L'ensemble des messages extraits par le Web Crawler sont stockés dans la base de données Detec't après identification des messages où le patient déclare avoir pris le médicament cité.

5.2 Données

La réalisation de ce travail a nécessité l'utilisation de deux jeux de données (figure 5.1). Le premier jeu de données a été consacré au développement des algorithmes détaillés

dans la suite de ce manuscrit et a permis de présenter les résultats exposés dans le chapitre 7.

Le second jeu de données était dédié à l'évaluation de l'algorithme de détection de mention de prise du traitement par l'auteur d'un message (décrit en section 6.1). Les autres approches proposées dans ce manuscrit ont été développées sur des corpus où la prise du traitement par le déclarant a été vérifiée à l'aide de cette méthode. L'évaluation de cet algorithme a donc nécessité un corpus étendu aux messages sans notion de prise de traitement identifiée.

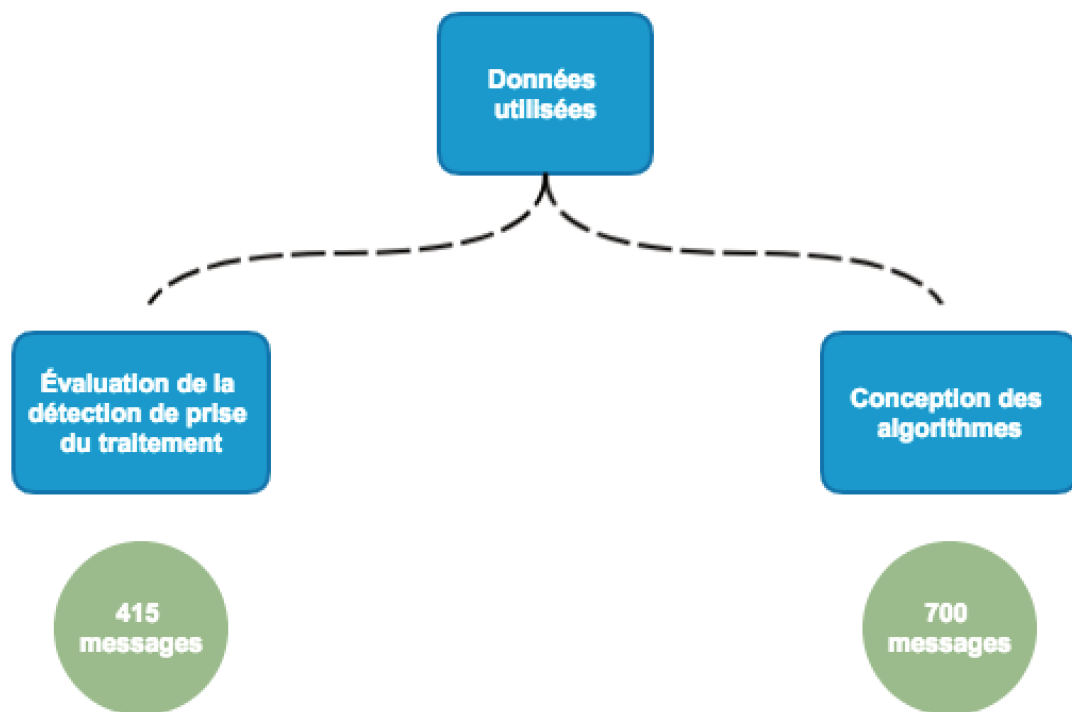


FIGURE 5.1: Données utilisées

5.2.1 Données de conception des algorithmes

Pour la conception de nos algorithmes, nous avons utilisé 700 messages (postés entre 2002 et 2013) sélectionnés aléatoirement dans la base de données Detec't parmi les posts liés aux médicaments *Insuline Glargine*[®], *Teriflunomide*[®] et *Zolpidem*[®] (de classes thérapeutiques différentes) avec prise du traitement par le déclarant.

Les messages utilisés sont extraits depuis les forums de santé via les mots clé correspondant aux principes actifs et noms commerciaux de ces médicaments.

5.2.2 Données d'évaluation de l'algorithme de prise

Cette méthode ayant été appliquée avant les autres approches proposées dans ce travail, l'évaluation de l'algorithme de prise (présenté en 6.1) a été réalisée sur un jeu de données différent incluant des messages avec et sans prise de traitement identifiée.

L'échantillon servant à l'évaluation de l'algorithme de prise était constitué de 415 messages tirés aléatoirement dans la base de données Detec't. Dans le chapitre 7, nous évaluons la quantité de messages mentionnant une prise du médicament vérifiée manuellement par rapport aux messages où cette prise du médicament par le déclarant est identifiée automatiquement.

5.2.3 Annotation manuelle des jeux de données

La conception et l'évaluation des algorithmes que nous proposons dans ce travail ont nécessité différentes revues manuelles des messages. Chaque revue correspond à l'annotation manuelle d'une information particulière dans les messages composant les jeux de données présentés en section 5.2. Les informations manuellement identifiées lors de cette étape d'annotation sont :

1. prise du traitement par le déclarant ;
2. déclaration d'effet indésirable ;

Nous n'avons pas réalisé de calcul de cohérence inter annotateurs. Les différentes annotations réalisées sont décrites dans les paragraphes ci-dessous.

5.2.3.1 Annotation de la prise du traitement

Avec l'objectif d'identifier les messages où le déclarant mentionne avoir pris le médicament cité, trois annotateurs (Redhouane Abdellaoui, Carole Faviez et Paméla Voillot) ont annoté les messages du jeu de données décrit en paragraphe 5.2.2 afin d'étiqueter ces messages.

Les annotateurs ont renseigné chaque message en :

1. « **Prise du médicament** » si le déclarant mentionne une prise du médicament cité ;

2. « **Pas de prise du médicament** » si le déclarant ne mentionne pas de prise du médicament cité ;
3. « **Indéterminé** » si le message ne permettait pas de statuer sur la prise du médicament par le déclarant ;

Les messages annotés en « **Indéterminé** » ont fait l'objet d'un arbitrage entre annotateurs.

5.2.3.2 Annotation des effets indésirables

Dans le chapitre 3, nous avons introduit la notion de « *concept médical* ».

Un concept médical peut être :

- un effet indésirable déclaré, par exemple, « ...*Je prend du Doliprane et ça me donne des **nausées**...* » ;
- Un symptôme ou une maladie qui n'est pas un effet indésirable. Par exemple, un symptôme motivant la prise de médicaments : « ...*J'avais **mal à la tête** donc j'ai pris du Doliprane...* » ;

Dans le but de distinguer les effets indésirables du reste des concepts médicaux cités dans les messages, deux annotateurs (Redhouane Abdellaoui et Carole Faviez) ont revu manuellement les concepts médicaux automatiquement détectés dans les messages du jeu de données de conception (décrit en 5.2.1). Cette étape a permis d'identifier les effets indésirables avérés (voir figure 5.2).

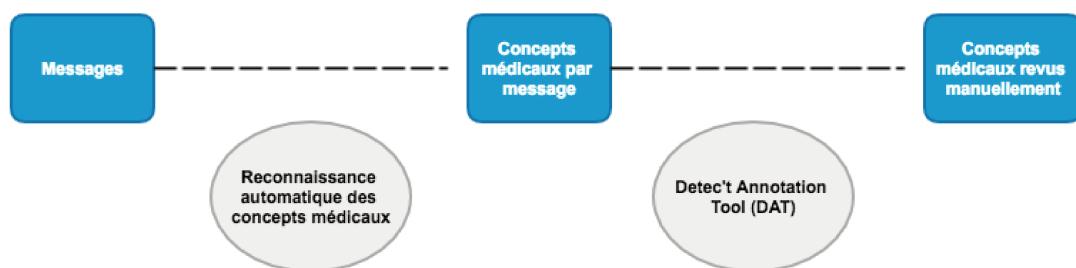


FIGURE 5.2: Processus d'annotation des concepts médicaux

Les annotateurs ont étiqueté chaque concept médical dans les messages en :

1. « **Oui** » si le concept médical a été déclaré par le patient comme un effet indésirable dans son message ;
2. « **Non** » si le concept médical n'a pas été déclaré par le patient comme un effet indésirable dans son message ;
3. « **Indécis** » si le message ne permettait pas de statuer sur la nature du concept médical cité ;

5.3 Dictionnaire Médical

La reconnaissance des concepts médicaux dans les messages a nécessité l'utilisation d'un dictionnaire médical de référence. Nous avons utilisé le *Medical Dictionary for Regulatory Activities* (MedDRA®) [75].

Cette terminologie médicale est organisée en *System Organ Class* (SOC) composés de *High-Level Group Terms* (HLGT) eux-mêmes composés de *High-Level Terms* (HLT) puis de *Preferred Terms* (PT) et de *Lowest Level Terms* (LLT) comme le montre la figure 5.3.

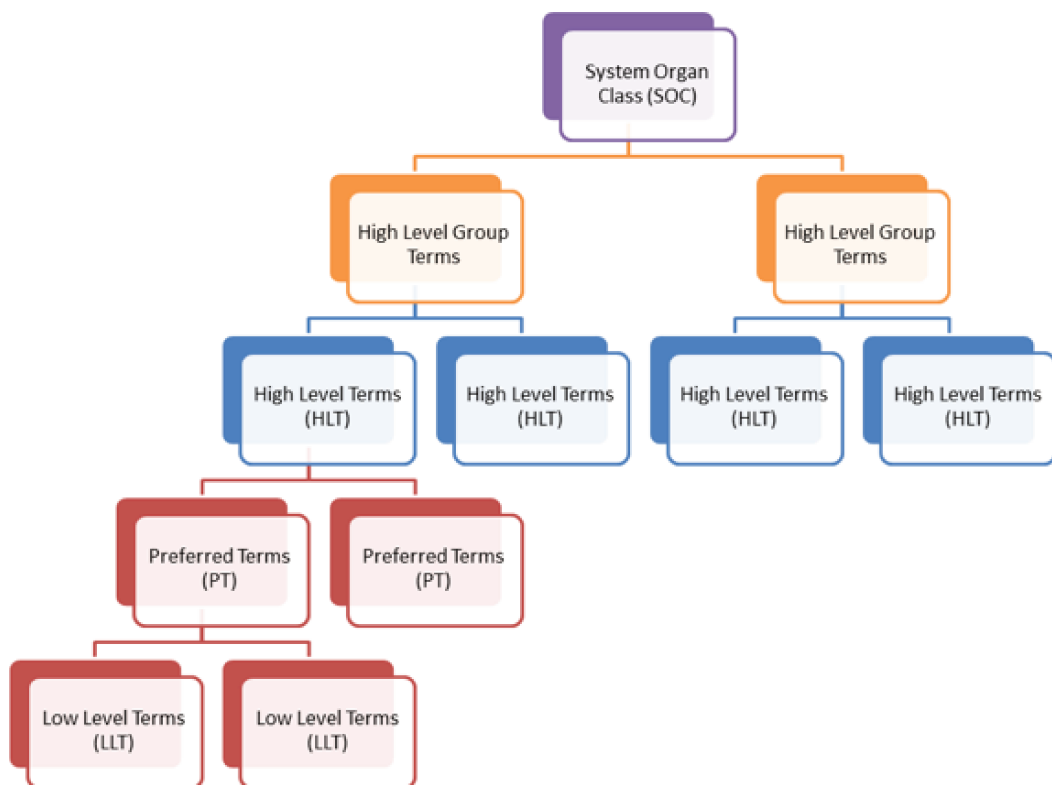


FIGURE 5.3: Medical Dictionary for Regulatory Activities

Nous avons utilisé le dictionnaire médical MedDRA[®] dans sa version 15.0 en français après l'avoir enrichie avec des synonymes issus du langage « profane » utilisé par les patients sur les forums. Notre version étendue de MedDRA[®] incluait également des termes volontairement mal orthographiés. Ces variantes, comportant des fautes d'orthographe, proviennent des messages postés sur les forums.

Ce vocabulaire médical supplémentaire est ajouté via deux approches. La première approche a consisté en l'ajout de synonymes identifiés lors de la lecture des messages par des annotateurs. La seconde approche a consisté en l'utilisation d'un algorithme de *matching approximatif* flou qui permet le regroupement (au niveau LLT) de termes avec différentes orthographes. Cette méthode de regroupement de termes supprime temporairement les voyelles et les consonnes en double ou triple dans un mot et permet de réduire un terme à sa composition primaire de consonnes. C'est ce « squelette » de mots qui permet d'identifier les variantes d'un même terme (masculin/féminin, singulier/pluriel ou encore fautes d'orthographe) [108].

La version originale de MedDRA[®] 15.0 contient 19,550 PT et 70,177 LLT. La version de MedDRA[®] étendue par **Kappa Santé** est formée d'un total de 19,530 PT et 63,392 LLT. La décroissance du nombre de termes dans notre MedDRA[®] enrichi est justifiée par la suppression de certains PT (le PT « Marié » par exemple) considérés comme non-informatifs dans un cadre de détection d'effets indésirables. De plus, certains PT sont regroupés pour cause de synonymie. Par exemple, nous regroupons les PT « alcoolique » et « alcoolisme ».

La version finale enrichie de MedDRA[®] contient 63,392 LLT. Parmi ces derniers, nous retrouvons une majorité de termes issus de la version originelle de MedDRA[®] 15.0 et 259 LLT supplémentaires. Par exemple, « mal de crâne » comme synonyme de « mal de tête ».

Nous recherchons les concepts médicaux dans les messages issus des forums de santé au niveau LLT de MedDRA[®].

5.4 Detec't Annotation Tool

Pour l'annotation manuelle des jeux de données, nous avons utilisé l'interface web Detec't Annotation Tool (DAT). DAT est un outil d'annotation développé par **Kappa Santé** pour les revues de messages. Cette interface permet aux annotateurs de visualiser un message ainsi que les concepts médicaux qui y ont été détectés afin de les annoter.

L'interface DAT (figure 5.4) permet d'annoter chaque concept médical comme décrit en paragraphe 5.2.3.2 :

Oui dans le cas où le concept médical est bien un effet indésirable ;

Non dans le cas où le concept médical n'est pas un effet indésirable ;

Indécis en cas d'incertitude ;

DETEC't

ACCUEIL MESSAGES VALIDATION

Annotation d'un message

Bonjour , j ai 30 ans et une vie tres active et remplie . . . je prends du nom _medic depuis 1 an , depuis 3 mois je ne prend qu un quart de comprime mais je n arrive pas a l arreter ou a diminuer (difficile avec cette forme de comprime) . Il m arrive de m endormir sans mais ne dors pas plus de 2h et me reveille extremement angoissee , en sueur . . . Donc je prend mon quart de comprime et je vais desuite mieux et m endort . je precise que je suis stressee mais heureuse dans ma vie , j ai suivi une dizaine de seances d hypno relaxation chez une psychologue qui m aide a gerer mes emotions mais malheureusement ne resolvent pas mes problemes d insomnies qui me pourrissent la vie depuis une dizaine d annee . , mes questions sont les suivantes : puis je avoir developpe une dependance physique malgres la faible posologie ? y a t il de reels risques sur la sante a prolonger mon traitement plus longtemps ? Merci

Tableau des annotations

Id Message	Utilisateur	Concept	Présence EI			Validation		
			Oui	Non	Indécis	Oui	Non	
1683303	Annotateur 1	relax	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1683303	Annotateur 1	insomni	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1683303	Annotateur 1	depend	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1683303	Annotateur 1	endorm	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1683303	Annotateur 1	arret du trait	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Tableau des concepts

Utilisateur	Concept	Présence EI	Validation
Annotateur 1	risques sur la sante	Indécis	<input type="radio"/> Oui <input type="radio"/> Non

Ajout d'un concept

Enregistrer Annuler

FIGURE 5.4: Detec't Annotation Tool

Cette interface offre également aux annotateurs la possibilité de saisir manuellement les concepts médicaux non détectés automatiquement. L'ajout manuel de ces concepts

médicaux a permis l'enrichissement de notre version du dictionnaire médical MedDRA[®] (voir 5.3).

Chapitre 6

Méthode

Dans le cadre de l'utilisation des forums de santé en pharmacovigilance, la distinction entre les déclarations d'effets indésirables et les autres utilisations possibles de concepts médicaux (description de symptômes, évocation d'indications des médicaments, citations de RCP, discussions autour de scandales médicamenteux, etc.) constitue un enjeu majeur [12]. L'une des limites des approches de détection d'effets indésirables sur les réseaux sociaux observées lors de notre revue de la littérature (voir chapitre 3) est le déséquilibre des jeux de données en faveur des faux positifs (concepts médicaux qui ne sont pas des effets indésirables). Ce déséquilibre augmente les probabilités de commettre des erreurs de classification et affaiblit les performances des systèmes existants [109].

Ce constat nous a poussé à considérer la mise en place d'étapes de filtrage de faux positifs en amont des étapes de détection d'effets indésirables. Le filtre des distances présenté dans les paragraphes suivant a donné lieu à une publication [104].

6.1 Filtre sémantique : Algorithme de détection de prise du médicament par le déclarant

6.1.1 Description de l'algorithme de détection de prise

Nous nous appuyons sur la définition d'un effet indésirable comme « *un signe ou un symptôme dû à une prise médicamenteuse. Un effet indésirable peut apparaître après la* »

prise ponctuelle d'un médicament, après une exposition prolongée ou suite à la combinaison d'au moins deux produits ».

Partant du principe qu'il ne peut pas y avoir d'effets indésirables sans prise du médicament par le patient, nous proposons un algorithme permettant de détecter la notion de prise du traitement dans les messages. Cet algorithme se base sur l'identification d'un groupe de termes caractéristiques d'une prise du médicament. La règle sémantique appliquée est décrite ci-dessous en 6.1.

$$< \text{je} > \& < \text{prendre} > \& < \text{nom du médicament} > \quad (6.1)$$

Nous avons identifié dans les messages les suites de mots faisant référence à l'auteur du message ($< \text{je} >$), une déclinaison du verbe « prendre » ou de l'un de ses synonymes ($< \text{prendre} >$), ainsi que le nom du médicament ($< \text{nom du médicament} >$).

L'identification des composantes sémantiques nécessaires au fonctionnement de cet algorithme a été réalisée par détection d'expressions régulières. Pour chaque composante sémantique ($< \text{je} >$, $< \text{prendre} >$, $< \text{nom du médicament} >$), nous avons utilisé des listes de formes usuelles enrichies avec des versions mal orthographiées.

Pour des raisons d'optimisation d'exécution, cet algorithme est appliqué à chaque message par le Web Crawler développé par **Kappa Santé** au moment de l'extraction de ce dernier depuis le forum.

La détection du nom du médicament dans les messages est réalisée par le Web Crawler au moment de l'extraction. Ce nom de médicament est conservé en tant que métadonnée de chaque message et constitue le mot-clé de la recherche de messages sur les forums. Des versions mal orthographiées du nom du médicament sont identifiées pendant l'extraction des messages à l'aide de l'algorithme de matching décrit en section 5.3 de ce document.

Pour chaque extraction, le nom du médicament est donné avec une liste de versions alternatives :

- noms commerciaux ;
- principe actif ;

6.1.2 Évaluation de l'algorithme de détection de prise

Nous avons évalué notre algorithme de détection de prise du traitement à l'aide d'un jeu de données constitué à cet effet et présenté en 5.2.2. Ces messages sont manuellement annotés en :

- Prise du médicament (Prise) ;
- Pas de Prise du médicament ($\overline{\text{Prise}}$) ;

		Gold Standard		Les indicateurs utilisés pour cette évaluation sont :
		Prise	$\overline{\text{Prise}}$	
Algorithme	Prise	VP	FP	$Precision = \frac{VP}{VP+FP}$
	$\overline{\text{Prise}}$	FN	VN	$Rappel = \frac{VP}{VP+FN}$

Le tirage de ces messages est effectué parmi des posts d'utilisateurs, évoquant explicitement le nom du médicament.

6.2 Reconnaissance d'entités médicales

Nous définissons une *entité médicale* dans ce travail (voir chapitre 3) comme :

1. Un concept médical (terme évoqué dans un message et constituant un potentiel effet indésirable). Par exemple, « nausées », « mal de tête », « acouphène » etc ;
2. Un nom de médicament ;

6.2.1 Prétraitement des données

Notre approche a nécessité de prétraiter les données [104]. Les prétraitements que nous avons appliqués sont :

1. Espacement des caractères de ponctuation ;
2. Mise en minuscules des caractères ;
3. Suppression des espaces abusifs ;
4. Segmentation des messages ;
5. Racinisation des termes ;

Espacement des caractères de ponctuation

Cette étape a consisté en l'introduction d'un caractère « espace » avant et après chaque élément de ponctuation. Cette étape a été réalisée afin d'optimiser la détection des concepts médicaux (voir section 6.2.3). Des concepts médicaux apparaissaient dans certains messages avec un caractère de ponctuation collé. Par exemple, «... *J'ai eu une migraine, c'était horrible ...* ».

Ce prétraitement a été appliqué aux messages de patients ainsi qu'au dictionnaire MedDRA[®] utilisé (présenté en 5.3). Ce prétraitement a permis de maximiser l'identification de concepts médicaux. Les espaces supplémentaires ajoutés ont été supprimés lors de l'étape de *suppression des espaces abusifs* présentée ci-après.

Mise en minuscules des caractères

Le traitement des données textuelles à l'aide du logiciel R fait une distinction entre caractères minuscules et majuscules. Cette distinction nous a obligés à traiter les messages et les mots du dictionnaire MedDRA[®] (voir 5.3) afin d'uniformiser les chaînes de caractères en mettant l'ensemble des textes en minuscules. Ce traitement est réalisé à l'aide du package « `tm` » [110].

Suppression des espaces abusifs

L'algorithme des distances proposé dans ce travail (voir paragraphe 6.3) s'appuie sur le nombre de mots séparant le nom du médicament des concepts médicaux détectés dans un message. La mise au point de ce filtre a nécessité la normalisation des espaces dans les textes. Par « normalisation des espaces » nous entendons un seul caractère d'espacement entre deux mots d'un message. Cette normalisation nous a permis d'utiliser les caractères d'espacements comme séparateurs de mots.

Nous avons supprimé les espaces abusifs dans les textes analysés. Ce traitement a été réalisé à l'aide du package « `tm` » [110].

Segmentation des messages

Cette segmentation (ou « tokenisation ») a permis d’obtenir des messages sous forme de listes de mots. Ce prétraitement a été appliqué par séparation des mots d’un message en se basant sur les caractères d’espacements. Nous avons obtenu la position de chaque mot du message via sa position dans la liste de mots générée par tokenisation. Cette information a été nécessaire pour le calcul du nombre de mots séparant les entités médicales afin d’appliquer notre algorithme des distances présenté en 6.3.

Racinisation des termes

Nous avons appliqué une méthode de racinisation (algorithme de Porter [3, 93, 111]) du vocabulaire. La racinisation des termes a permis l’élimination des préfixes et suffixes en réduisant un mot à sa racine. Savoy [112] a montré que l’utilisation de l’algorithme de racinisation de Porter [111] améliore de 30.5% l’extraction d’informations depuis des textes rédigés en français.

6.2.2 Détection des noms de médicaments

Lors de l’extraction des messages par le Web Crawler, les noms de médicaments cités sont conservés en métadonnées (voir section 5.1). Nous avons identifié ces derniers et les avons remplacés dans les messages analysés par la chaîne de caractères invariante « `nom_medicament` » pour standardiser l’identification de la position du nom de médicament pour le calcul du nombre de mots le séparant des concepts médicaux détectés. Cette information est nécessaire pour l’application de la modélisation présentée dans le paragraphe suivant (6.3).

6.2.3 Détection de concepts médicaux

Nous définissons un concept médical dans ce paragraphe comme une entité médicale (voir section 6.2) qui n’est pas un nom de médicament.

L’objectif de cette étape était l’identification des concepts médicaux dans les messages analysés. Le dictionnaire MedDRA[®] utilisé pour l’identification de ces concepts médicaux (présenté en 5.3) a été enrichi avec diverses variantes des mots et expressions le

composant. La détection de ces derniers a été réalisée par identification d'expressions régulières dans les messages ayant subi les prétraitements décrits en 6.2.1. La détection des concepts médicaux a été appliquée uniquement sur les messages où une prise du médicament par le déclarant a été identifiée par l'algorithme décrit en section 6.1.

6.2.4 Issue de la détection d'entités médicales

L'étape de détection d'entités médicales après les prétraitements a permis d'identifier les positions de ces entités dans les messages. Ces données de positions étaient nécessaires pour la conception de l'algorithme des distances présenté ci-après (en section 6.3). Ce paragraphe présente les données recueillies pendant cette étape de détection d'entités médicales.

Les informations récoltées pour chaque concept médical pendant cette étape sont :

- Le concept médical identifié dans le message (codé au niveau LLT de MedDRA[®]) ;
- Le PT correspondant ;
- La position du concept médical dans le message (par exemple, le concept identifié est le *10ème* mot du message) ;
- La position relative du concept médical par rapport au nom de médicament (situé avant ou après ce dernier) ;
- La distance (en nombre de mots) séparant le concept médical et le nom du médicament cités dans le message ;
- La taille du message (en nombre de mots) ;

Ces données nous ont permis de concevoir le filtre des distances présenté dans la section suivante (6.3) en testant l'hypothèse selon laquelle plus un concept médical est éloigné du nom du médicament dans un message, plus la probabilité que ce dernier soit un effet indésirable est faible.

6.3 Filtre statistique : Algorithme des distances entre entités médicales

Nous avons développé un filtre basé sur un modèle de mélange gaussien afin d'identifier les concepts médicaux ayant une faible probabilité d'être des effets secondaires.

L'identification de ces derniers a pour but leur filtrage et ainsi la réduction du taux de faux positifs. Par « faux positif », nous entendons un concept médical considéré à tort comme un effet indésirable.

6.3.1 Hypothèse

Nous avons défini la « *distance* » entre deux entités médicales dans un message par le nombre de mots séparant ces entités.

Notre hypothèse est la suivante :

La distance séparant un concept médical et un nom de médicament dans un message est inversement proportionnelle à la probabilité que ce concept médical soit un effet indésirable de ce traitement.

6.3.2 Calcul des distances

La distance entre un nom de médicament « *m* » et un concept médical « *c* » au sein d'un message est définie dans ce travail comme le nombre de mots séparant ces deux entités médicales. Ces distances ont été calculées pendant l'étape d'identification d'entités médicales. Ce calcul a été réalisé comme défini dans l'équation 6.2 :

$$Distance(m, c) = Position(c) - Position(m) \quad (6.2)$$

Le nom d'un médicament pouvant être cité plus d'une fois dans un message, nous avons calculé ces distances pour l'ensemble des occurrences des couples (nom du médicament, concept médical) identifiées. L'unique couple considéré pour chaque concept et chaque médicament est le couple pour lequel la distance est minimale (en valeur absolue).

6.3.3 Principe de la classification par mélange gaussien

La densité d'une famille d'observations $X \in R^d$ selon la loi normale (de moyenne μ et de matrice de covariance Σ) est définie en équation 6.3.

$$\mathcal{N}(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu)} \quad (6.3)$$

Un mélange de I gaussiennes (défini en équation 6.4) est une combinaison linéaire de plusieurs gaussiennes. Chaque distribution prise en compte pour ce modèle est pondérée par un coefficient α_i et caractérisée sa moyenne μ_i et sa matrice de covariance Σ_i ($i \in I$).

$$\sum_{i=1}^I \alpha_i \mathcal{N}(\mu_i, \Sigma_i). \quad (6.4)$$

L'ajustement du modèle a été réalisé à l'aide de l'algorithme EM [113] et a été utilisé pour l'évaluation de la part de chaque sous-modèle gaussien dans la distribution générale des distances.

L'application du Gaussian Mixture Model (GMM) pour la classification des concepts médicaux a été réalisée à l'aide du package R « `mclust` » [114]. La distribution des distances calculées entre les entités identifiées a été modélisée par le mélange de deux distributions gaussiennes ($I = 2$). L'association d'un concept à l'une des I composantes du mélange a été faite par maximisation de fonction de vraisemblance.

6.4 Caractérisation des effets indésirables : L'algorithme de pertinence

Nous définissons dans ce chapitre la notion de « *pertinence* » d'un concept médical. Cette notion décrit la caractérisation d'un concept médical en « effet indésirable » ou non. Cette caractérisation est opérée automatiquement via la méthode de classification que nous avons conçue et que nous décrivons dans cette section.

Nous formalisons cette pertinence par une variable binaire définie comme :

$$\begin{cases} pertinence = 1 & \text{si le concept médical est un effet indésirable} \\ pertinence = 0 & \text{sinon} \end{cases}$$

6.4.1 Support Vector Machine (SVM)

Les SVM sont des méthodes de classification binaire par apprentissage statistique supervisé [115]. Ces méthodes reposent sur l'hypothèse selon laquelle il existe d'un classifieur linéaire dans un espace approprié. Elles se basent sur l'utilisation d'une fonction noyau permettant une séparation optimale des données via un hyperplan [116]. Les points les plus proches de l'hyperplan (vecteurs de support) sont utilisés pour la détermination de ce dernier. Il existe une multitude d'hyperplans valides, mais l'avantage des SVM est que cet hyperplan doit séparer les clusters de données de manière optimale. L'hyperplan séparateur est celui qui maximise la marge, à savoir, la distance minimale aux vecteurs de support de l'ensemble d'apprentissage.

La marge entre les points les plus proches de l'hyperplan est maximisée et la fonction de décision est spécifiée par l'utilisation d'un sous-ensemble de l'espace d'apprentissage.

Soient y_i les marqueurs de classe et $x_i \in \mathbb{R}^d$ le vecteur contenant les données à classifier. Nous utilisons une classification *C-SVM* dans laquelle la fonction 6.5 est minimisée (sur les données d'apprentissage)

$$\frac{1}{2}w^t w + C \sum_{i=0}^N \xi_i \quad (6.5)$$

avec les contraintes : $y_i(w^t \phi(x_i, x'_i) + b) \geq 1 - \xi_i$ et $\xi_i \geq 0$, $i \in \{1, \dots, N\}$, où C est la constante de capacité, w le vecteur de coefficients inconnu, b une constante permettant le contrôle des données non séparables, N est le nombre d'individus considérés dans l'ensemble d'apprentissage du modèle. Les ξ_i représentent les variables indépendantes. Le noyau utilisé $\phi(.) : \mathbb{R}^d \mapsto \mathbb{R}^{d_h}$ projette les données à classifier dans un espace de représentation de plus grande dimension ($d_h > d$). La constante C permet de contrôler les erreurs de classification en régulant la largeur de la marge. Nous utilisons une fonction de projection basée sur un noyau gaussien. La forme d'un noyau gaussien est présentée en (6.6).

$$\phi(x_i, x'_i) = e^{-\gamma \|x - x'\|^2} \quad (6.6)$$

Les noyaux gaussiens sont un choix standard de fonction de projection d'usage général qui peut traiter les relations non linéaires entre les marqueurs de classes y_i et les données à classifier x_i .

Dans la majorité des problèmes réels, il n'y a pas de séparation linéaire possible des données. Pour résoudre ce problème, un changement d'espace de représentation de données est réalisé via la fonction noyau. Plus la dimension de l'espace de description est grande, plus la probabilité de trouver un hyperplan séparateur est élevée [117].

6.4.2 Stratégie de classification

Le modèle SVM a été utilisé dans ce travail après filtrage des concepts médicaux par l'algorithme des distances présenté en 6.3. Nous avons appliqué cette approche de classification aux couples (médicament, concept médical) qui ont été conservés par ce filtre.

Nous avons divisé l'ensemble de données en 2 sous-ensembles afin de constituer un ensemble d'apprentissage et un ensemble de test. Chacun des 2 sous-ensembles est formé à partir de 50% des observations.

Nous appliquons une pondération des données pour faire face au déséquilibre de l'échantillon décrit en 3.4. Ce déséquilibre concerne les quantités d'effets indésirables vérifiés manuellement par rapport aux autres concepts médicaux. L'optimisation des paramètres du modèle est réalisée par *grid search method* [118] sur les paramètres C , γ et la fonction noyau.

Nous appliquons cette classification sur les variables suivantes mesurées pendant l'étape de détection de concepts médicaux (voir 6.2.3) :

1. La position du concept médical détecté dans le message ;
2. Le nombre de concepts médicaux différents détectés dans le message ;
3. La position du nom du médicament dans le message ;
4. La valeur absolue de la distance entre le nom du médicament et le concept médical ;

6.4.3 Post-traitement des données

6.4.3.1 Rétrogradation des SOC's non informatifs

La pertinence établie est rétrogradée (i.e. passée de 1 à 0) pour les SOC's « Actes médicaux et chirurgicaux » et « Caractéristiques Socio-Environnementales ». Nous avons

considéré ces SOCs comme non-informatifs car les PTs affiliés ne correspondent pas à des effets indésirables potentiels. Quelques exemples de PTs affiliés à ces SOCs : « *Crise de la quarantaine* », « *foyer d'accueil* », « *logement humide* », « *inondation* », etc.

6.4.3.2 Dédoublonnage des concepts

Nous avons choisi de ne conserver qu'une seule occurrence de chaque concept médical détecté dans un même message. Par exemple, si le concept médical « *nausées* » était détecté plusieurs fois dans un message, nous ne le considérons qu'une seule fois pour éviter sa surreprésentation quantitative dans la liste des effets indésirables identifiés.

Nous avons favorisé le concept médical avec une pertinence égale à 1 (i.e. celui qui avait la probabilité la plus élevée d'être un effet indésirable).

6.4.3.3 Contextualisation des concepts médicaux polysémiques

La polysémie des termes constitue une source de faux positifs dans les messages postés par les patients sur les forums de santé. Par exemple, le concept médical « **mal** » peut être utilisé pour décrire un potentiel effet indésirable « ... **mal** de tête... » ou comme autre type de verbatim « ...j'ai **du mal** à... » ou encore « ...il y a pas **mal** de temps... ». Pour cette raison, nous avons conçu un programme de contextualisation des concepts médicaux considérés comme « *polysémiques* ». Ce programme consiste en la détection des mots voisins des concepts médicaux polysémiques pour leur contextualisation.

Les concepts médicaux polysémiques ont été éliminés de la liste de concepts médicaux détectés dans chaque message, car ces derniers ne sont pas porteurs d'information médicale.

6.5 Synthèse de notre approche de détection de cas de pharmacovigilance

La figure 6.1 présente une synthèse des étapes de traitement des messages de patients proposé dans ce travail.

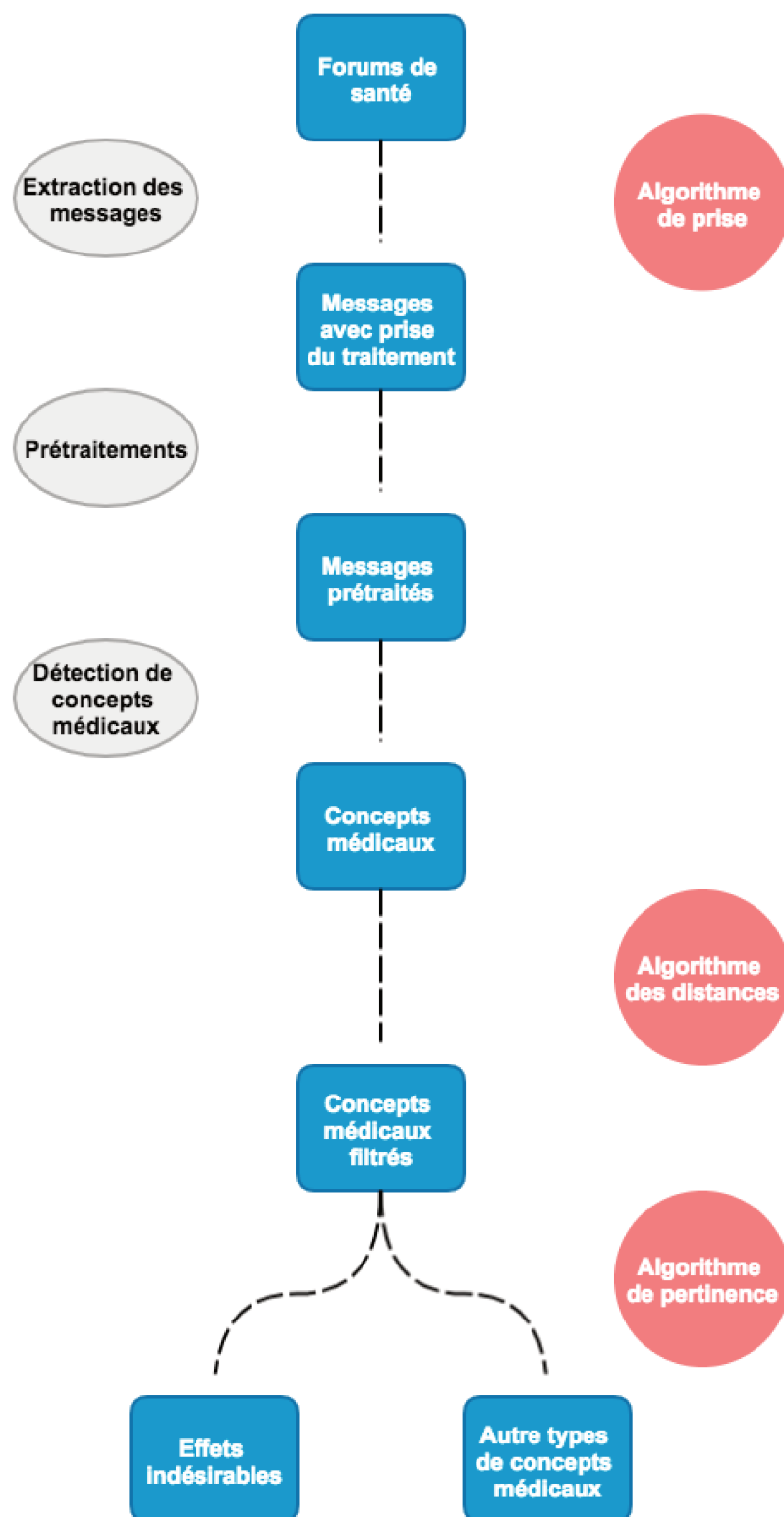


FIGURE 6.1: Synthèse de l'approche proposée

Chapitre 7

Résultats

Ce chapitre présente les résultats des algorithmes de détection de cas de pharmacovigilance sur les forums de santé que nous proposons. Nous présentons les résultats obtenus à l'aide du jeu de donnée décrit en 5.2.1.

Parmi les résultats de ces travaux, figure le package R « *DetectAnalyzer* » présenté ci-après. Ce package permet d'appliquer l'approche présentée dans ce manuscrit à des messages rédigés en français issus de forums de santé. Le package *DetectAnalyzer* est aujourd'hui utilisé par la société **Kappa Santé** dans la chaîne d'exécution du projet Detec't.

7.1 Package DetectAnalyzer

Le package *DetectAnalyzer* est un programme que nous avons conçu (Sous la forme d'un package R [119]) pour la détection d'effets indésirables de médicaments dans les posts issus des forums de santé. Ce package prend en charge les tâches suivantes détaillées dans le chapitre précédent :

1. Chargement des messages extraits par le Web Crawler de **Kappa Santé** et stockés dans la base de données Detec't ;
2. Détection d'entités médicales (voir 6.2) ;
3. Application de l'algorithme des distances (voir 6.3) ;
4. Application de l'algorithme de pertinence (voir 6.4) ;

5. Application des post-traitements (voir 6.4.3) ;

Dans ce chapitre, nous détaillons les résultats obtenus à l'aide de ce package.

7.2 Algorithme de détection de prise du traitement dans les messages

7.2.1 Description des données d'évaluation de l'algorithme de prise

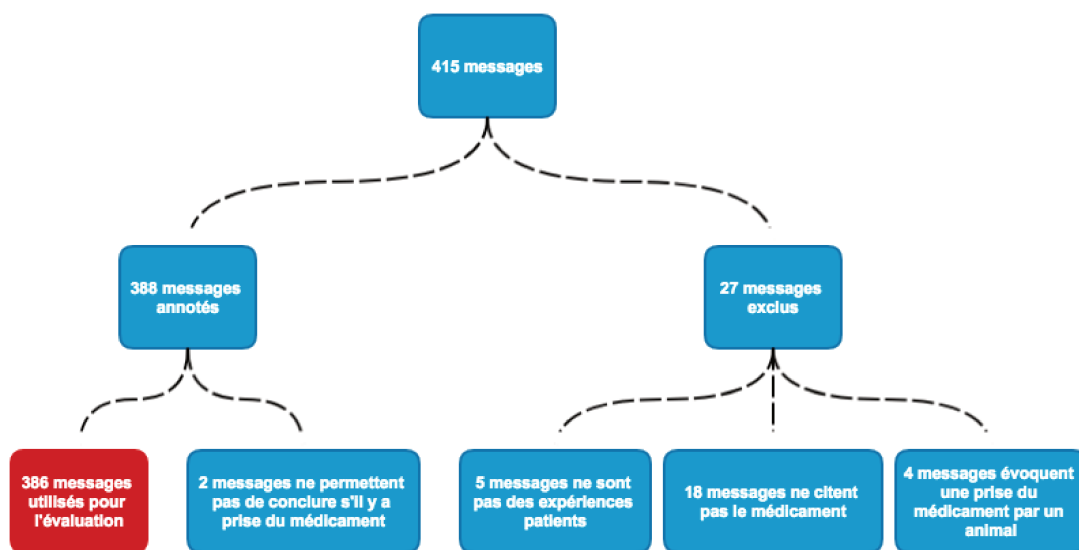


FIGURE 7.1: Description des données d'évaluation de la détection de prise du traitement.

Parmi les 415 messages utilisés pour l'évaluation de l'algorithme de détection de prise du traitement par le déclarant (voir 5.2.2), 27 ont été supprimés pour trois raisons :

- Le message ne correspondait pas à une expérience patient, mais à un RCP partagé par un utilisateur du forum (5 messages) ;
- Pas de référence au nom du médicament (18 messages). Ces messages ont été reconnus, à tort, comme citant le médicament ;
- Le message décrit l'administration d'un médicament à un animal (4 messages) ;

Un total de 388 messages a été utilisé pour cette évaluation de l'algorithme de prise. Parmi ces 388 messages, 244 ont été annotés manuellement en « **Prise du médicament**

», 133 ont été annotés en « **Pas de prise du médicament** » et 11 sont annotés en « **Indéterminé** ». Ces 11 messages ont été catégorisés en « **Indéterminé** », car la qualité de rédaction du message ne permettait pas de conclure à une prise du médicament par le déclarant. Le message suivant constitue un exemple de message qualifié en « **Indéterminé** » :

« le cicalfate assèche, tout comme biseptine.... »

Après discussion et arbitrage entre les annotateurs sur ces 11 messages, deux messages ont été conservés dans la catégorie « **Indéterminé** », 6 ont été annotés « **Pas de prise du médicament** » et 3 ont été annotés en « **Prise du médicament** ». Les 2 messages annotés en « **Indéterminé** » ont été exclus de l'évaluation (voir figure 7.1).

7.2.2 Évaluation de l'algorithme de Prise

La conception méthodologique de l'algorithme de prise, présenté en 6.1, ainsi que les évaluations ci-dessous sont réalisées au sein de la société **Kappa Santé** en collaboration avec **Carole Faviez**.

La table 7.1 présente le résultat de l'annotation manuelle du jeu de données (voir paragraphe 5.2.2) utilisé pour l'évaluation de l'algorithme de prise :

Gold Standard	Fréquence
Prise du produit	247
Pas de prise du produit	139
Indéterminé	2
Total	388

TABLE 7.1: Résultats de l'annotation manuelle des messages de prise.

Le résultat ci-après est obtenu en excluant les messages annotés en "Indéterminé".

		Gold Standard		Nous obtenons les indicateurs : <i>Precision</i> = 72% <i>Rappel</i> = 96%
		Prise	$\overline{\text{Prise}}$	
Algorithme	Prise	238	94	
	$\overline{\text{Prise}}$	9	45	

7.3 Algorithme des distances

7.3.1 Description des données

Parmi les 700 messages utilisés pour la conception des algorithmes (présentés en 5.2.1), 52 ne citaient pas les produits étudiés et ne pouvaient donc pas décrire d'effets indésirables. Les 648 posts restants ont été analysés automatiquement à l'aide du package *DetectAnalyzer* que nous avons développé. Les concepts médicaux automatiquement identifiés ont ensuite été manuellement annotés à l'aide de l'interface DAT décrite en 5.4. L'analyse des concepts médicaux identifiés dans ces 648 messages est présentée ci-après.

Ces messages contenaient 320 concepts médicaux uniques correspondant à 268 PT du dictionnaire médical MedDRA® (voir 5.3). Parmi ces 648 messages, 265 (soit 41%) ne contenaient pas de concepts médicaux. Nous avons identifié 1 654 couples (*médicament*, *concept médical*) dans les 383 messages restants.

La figure 7.2 décrit les concepts médicaux automatiquement identifiés dans ces 383 messages par notre package *DetectAnalyzer*. Ces derniers sont présentés au niveau SOC du dictionnaire MedDRA® (voir figure 5.3).

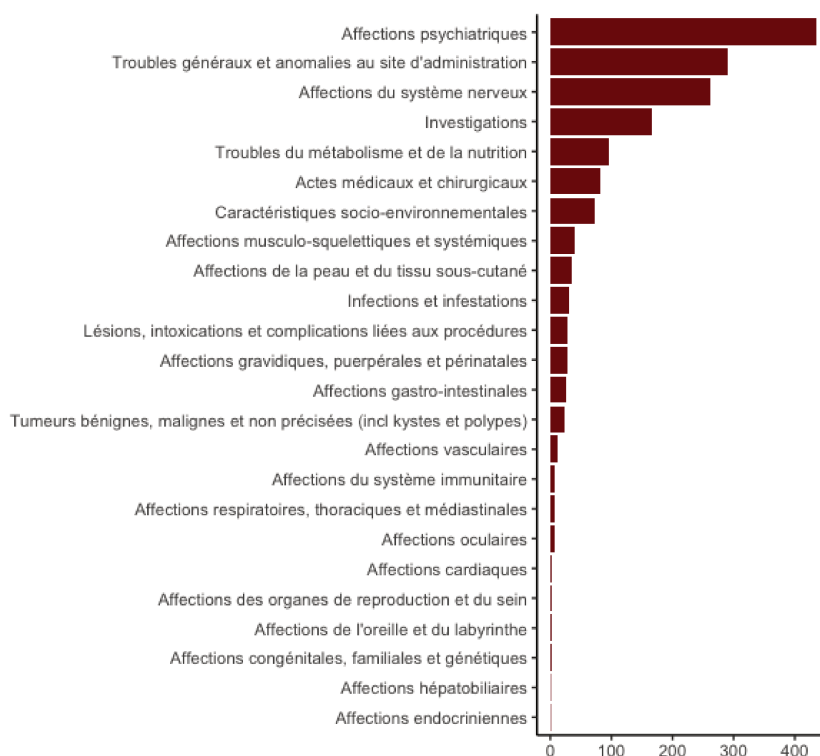


FIGURE 7.2: Concepts médicaux automatiquement identifiés (au niveau SOC).

Après la vérification manuelle des 1 654 concepts médicaux détectés, nous avons identifiés 189 concepts (11% des concepts médicaux) comme des effets indésirables et 1 465 concepts (89% des concepts détectés) comme autre type de concepts médicaux.

La figure 7.3 décrit la répartition des messages dans le jeu de données utilisé.

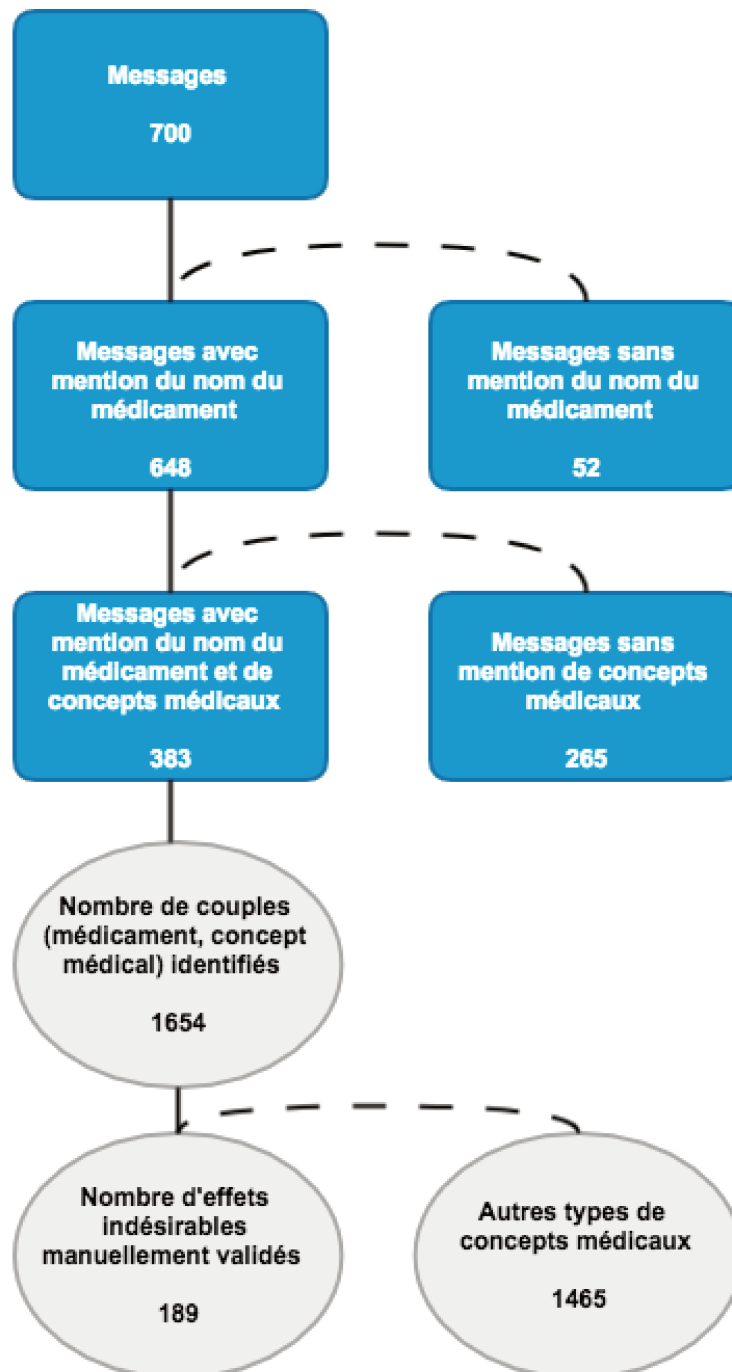


FIGURE 7.3: Description du jeu de données utilisé

7.3.2 Distribution des concepts médicaux dans le corpus

La figure 7.4 présente la distribution des distances calculées entre concepts médicaux et noms de médicaments.

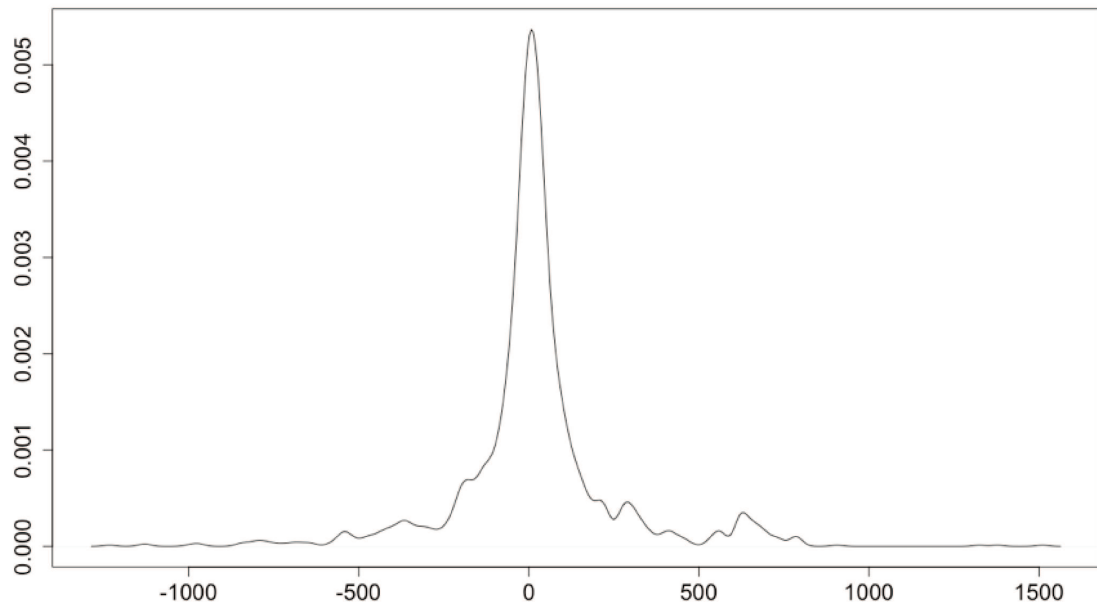


FIGURE 7.4: Densité des distances observée entre concepts médicaux et noms de médicaments.

La densité des distances mesurées semble suivre une distribution gaussienne. Un test de Shapiro-Wilk a permis de rejeter l'hypothèse d'une distribution normale $p < 2,2 \times 10^{-16}$.

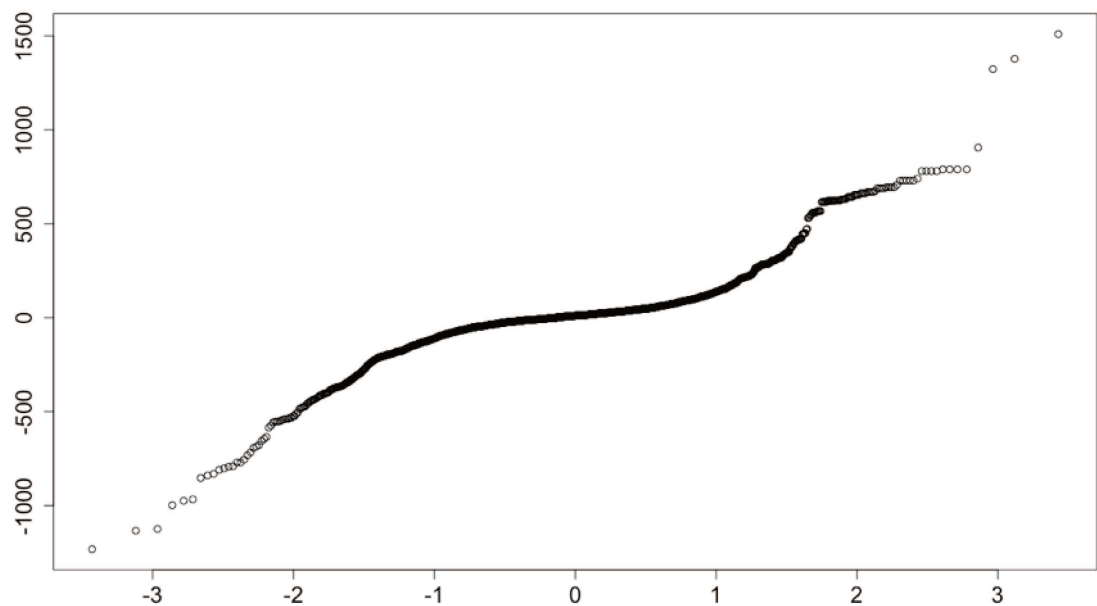


FIGURE 7.5: Q-Q plot normal.

Le Q-Q plot en figure 7.5 montre l'hétérogénéité de la répartition des distances calculées. Cette hétérogénéité renforce l'hypothèse d'un mélange de lois gaussiennes [120, 121].

Les valeurs prises par les distances calculées étaient variables avec une moyenne de 20,3 et une médiane de 11. Ces distances variaient entre 1 233 mots avant les mentions au nom du médicament et 1 510 mots après.

La figure 7.6 montre la concentration des effets indésirables dans un intervalle restreint autour du nom du médicament.

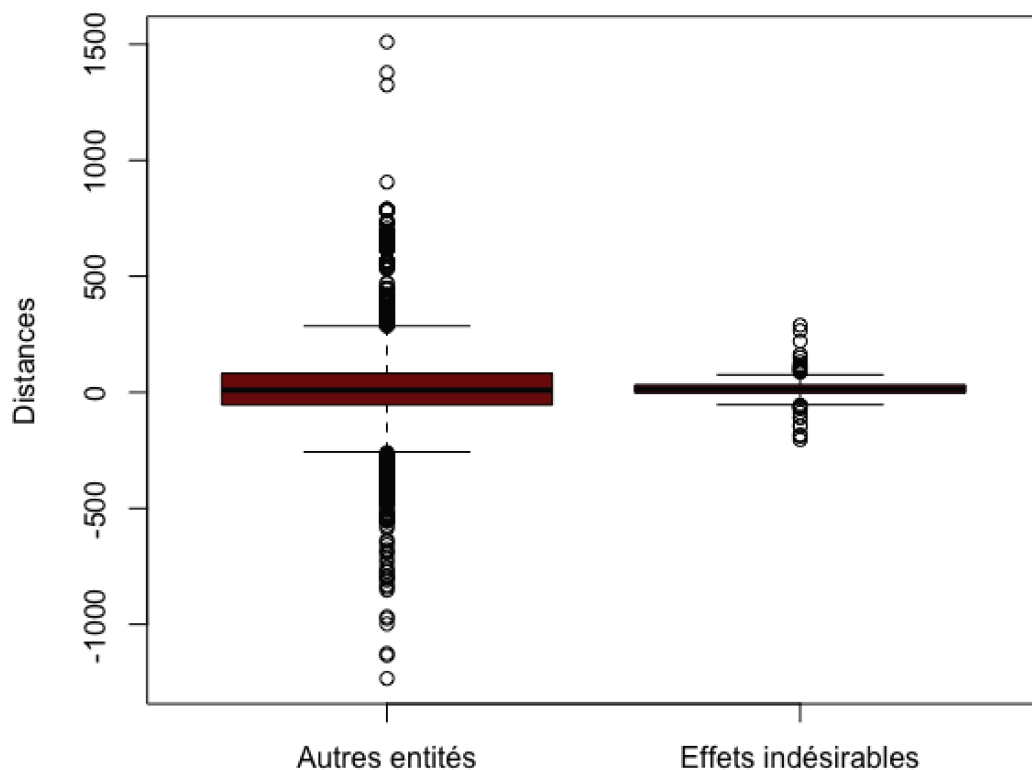


FIGURE 7.6: Répartition des effets indésirables et des autres types de concepts médicaux.

Les effets indésirables étaient compris dans un intervalle de 204 mots avant le nom du médicament et 289 mots après. La totalité des concepts en dehors de cet intervalle n'étaient pas identifiés comme effets indésirables lors de l'annotation manuelle.

7.3.3 Classification des concepts médicaux

La méthode de classification utilisée a retourné des classes (clusters) de concepts médicaux regroupés sur la base des distances les séparant des noms de médicaments. Nous nous sommes basé sur l'utilisation d'un modèle de mélange gaussien et d'un algorithme EM [122–124].

La figure 7.7 présente le résultat de cette classification.

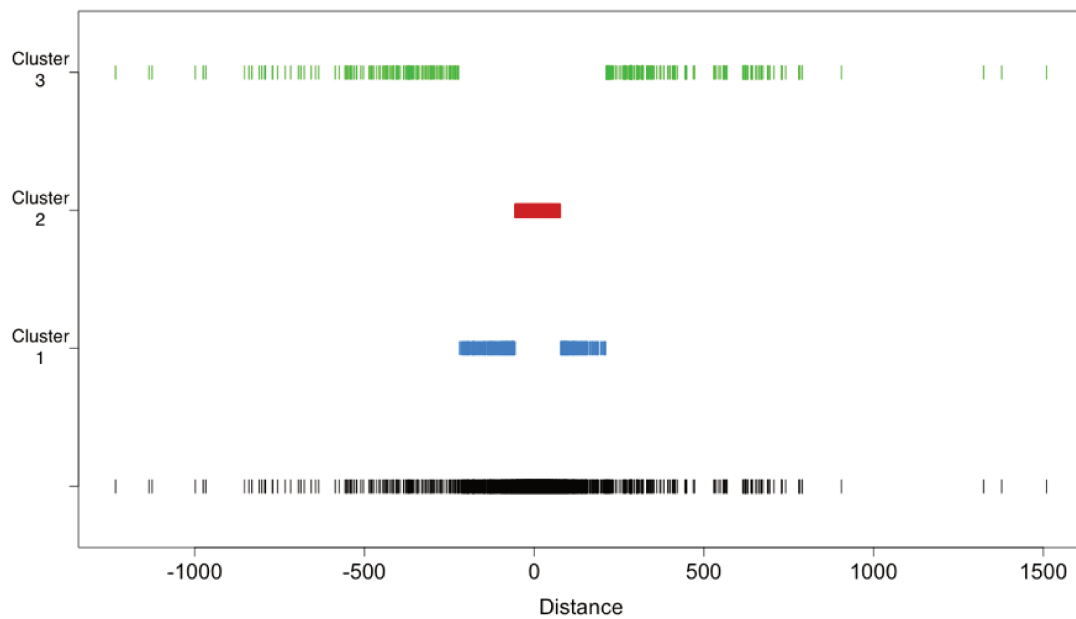


FIGURE 7.7: Les classes de concepts médicaux obtenues.

7.3.3.1 Cluster 1

Ce groupe contenait les concepts médicaux dans l'intervalle de distances suivant :

$$[-220, -57] \cup [+78, +211]$$

Autrement dit, les concepts entre 220 et 57 mots avant le nom du médicament ou entre 78 et 221 mots après le nom du médicament dans les messages. Cette classe contenait 441 concepts médicaux parmi lesquels nous avons identifié 29 effets indésirables (6,6% des concepts de cette classe).

7.3.3.2 Cluster 2

Ce groupe contenait les concepts médicaux dans l'intervalle de distances suivant :

$$[-56, -1] \cup [+2, +77]$$

À savoir les concepts médicaux situés à moins de 56 mots avant le nom du médicament et à moins de 77 mots après le nom du médicament dans les messages. Cette classe contenait 889 concepts médicaux dont 17,7% étaient des effets indésirables identifiés lors de l'annotation manuelle.

7.3.3.3 Cluster 3

Ce groupe contenait les concepts médicaux dans l'intervalle de distances suivant :

$$[-1233, -222] \cup [+212, +1510]$$

Ce qui correspond aux concepts situés entre 1 233 et 222 mots avant le nom du médicament dans le message et entre 212 et 1 510 mots après. La classe 3 contenait 324 concepts médicaux. Parmi ceux-ci, 3 étaient des effets indésirables (moins de 1%).

7.3.4 Stratégies de filtrage

Nous avons testé deux stratégies de filtrage. L'objectif était de filtrer les concepts médicaux qui n'étaient pas des effets indésirables afin de réduire la quantité de faux positifs. La table 7.2 présente la classification des concepts médicaux par notre modèle.

	Effets Indésirables (%)	Effets Indésirables (%)	Total
Classe 1	412 (93,4)	29 (6,6)	441
Classe 2	732 (82,3)	157 (17,7)	889
Classe 3	321 (99,1)	3 (0,9)	324
Total	1465	189	1654

TABLE 7.2: Répartition des concepts médicaux après classification.

7.3.4.1 Stratégie 1

La première stratégie de filtrage a consisté en la fusion des classes 1 et 3 présentées en table 7.2. Cette fusion de classes a permis de regrouper les concepts situés à une distance au nom du médicament dans l'intervalle :

$$[-1233, -57] \cup [+78, +1510]$$

L'objectif de cette stratégie de classification était de regrouper les concepts médicaux qui n'étaient pas des effets indésirables (412 en classe 1 et 321 en classe 3) dans une seule classe afin de les filtrer. L'union de ces deux classes ne contenait que 4,2% des effets indésirables (32 effets indésirables) validés par annotation manuelle (voir table 7.3).

Comme le montre la table 7.3, 95,8% des concepts médicaux présents dans la fusion des classes 1 et 3 n'étaient pas des effets indésirables (733 concepts médicaux).

	Effets Indésirables (%)	Effets Indésirables (%)	Total
Classe 1 et 3	733 (95,8)	32 (4,2)	765
Classe 2	732 (82,3)	157 (17,7)	889
Total	1465	189	1654

TABLE 7.3: Filtrage par union des classes 1 et 3.

Dans un contexte de détection d'effets indésirables, l'utilisation de cette approche pour la suppression des concepts médicaux des classes 1 et 3 a permis une réduction de 50,03% de faux positifs. Avec cette stratégie de filtrage, nous avons obtenu une précision de 95,8% et un rappel de 50%.

Plus de 95% des concepts médicaux filtrés via cette stratégie n'étaient pas des effets indésirables, mais l'approche n'a détecté que 50% des faux positifs.

7.3.4.2 Stratégie 2

La seconde stratégie de filtrage a consisté en la fusion des classes 1 et 2 présentées en table 7.2. L'objectif de cette stratégie était le regroupement des effets indésirables dans

une seule classe afin de se restreindre à cette dernière. Nous avons fusionné les classes 1 et 2 (412 en classe 1 et 732 en classe 2) puis nous avons filtré la classe 3.

	$\overline{\text{Effets Indésirables}}$ (%)	Effets Indésirables (%)	Total
Classe 3	321 (99,1)	3 (0,9)	324
Classe 1 et 2	1144 (86)	186 (14)	1330
Total	1465	189	1654

TABLE 7.4: Filtrage par union des classes 1 et 2.

L'union des classes 1 et 2 contenait 98,4% des effets indésirables présents dans le jeu de données. La classe 3 contenait 1,6% des effets indésirables. L'exclusion de la classe 3 a entraîné une erreur de classification de 3 effets indésirables validés lors de l'annotation manuelle comme le montre la table 7.4.

Cette stratégie a permis d'obtenir une précision de 99,07% et un rappel de 21%.

7.4 Algorithme de pertinence

Avec l'objectif de classifier les concepts médicaux conservés par notre algorithme des distance (section 6.3), nous avons utilisé les données présentées en table 7.5 obtenues à partir de la classification présentée en table 7.2 après filtrage la classe 3 :

	Effets Indésirables	$\overline{\text{Effets Indésirables}}$	Total
Classe 1	29	412	441
Classe 2	157	732	889
Total	186	1144	1330

TABLE 7.5: Données de conception du modèle SVM

Nous avons testé différents ajustements des paramètres *Noyau*, *C*, γ du modèle SVM. Une présélection des valeurs de ces paramètres a été obtenue par la méthode *grid search* décrite en [118]. Nous avons également centré et réduit notre échantillon d'observation. La table 7.6 expose les résultats des meilleurs ajustements obtenus.

Paramètres du modèle SVM				Performances			
Noyau	Normalisation	γ	C	Precision	Recall	Spécificité	AUC
Gaussien		1.6	6	19.2%	87.2%	57.5%	0.606
Gaussien		78.1	43	19.3%	81.4%	60.6%	0.628
Gaussien	✓	1	10	16.4%	81.4%	52.0%	0.551
Gaussien	✓	1	1.0×10^5	17.1%	82.6%	53.6%	0.566
Laplacien		0.006	270	18.3%	84.9%	56.1%	0.591
Laplacien		0.1	10	16.9%	82.6%	53.1%	0.562
Laplacien	✓	0.00056	14 222	16.8%	66.3%	62.0%	0.624
Laplacien	✓	1.6	0.5	16.8%	82.6%	52.6%	0.557

TABLE 7.6: Calibrage du modèle SVM

Nous notons une précision faible pour l'ensemble des configurations observées (entre 16,4% et 19,3%) avec un rappel entre 66,3% et 87,2%.

Cette faible précision nous a poussés à considérer plus d'indicateurs métrologiques (à défaut d'informations supplémentaires à fournir au modèle) pour statuer sur l'ajustement du modèle à adopter. Nous avons donc étudié les spécificités et AUC des différentes configurations sélectionnées.

Nous avons basé notre choix sur le rappel (sensibilité) en privilégiant la configuration la plus sensible afin de capter un maximum d'effets indésirables au risque de capter également des faux positifs. Cette approche nous garantit l'exhaustivité de nos résultats d'effets indésirables et nous semble être l'approche la plus sécurisée, au vu des résultats de ce modèle, dans un cadre de surveillance sanitaire.

Deuxième partie

Détection de pratiques de non-observance au traitement des patients sur les réseaux sociaux

Chapitre 8

Matériel et Méthode

L’approche présentée ci-après a fait l’objet d’une publication [125].

8.1 Matériel

Nous avons extrait deux corpus de la base de données Detec’t (décrite en 5.1) : le premier correspondait aux messages relatifs à l’escitalopram, un antidépresseur, le second à l’aripiprazole, un antipsychotique. La raison du choix de ces médicaments est que les cas de non-observance sont plus susceptibles d’apparaître dans un contexte de maladies chroniques et constituent une préoccupation majeure dans la prise en charge des maladies psychiatriques [126]. De plus, ces médicaments appartiennent à deux classes thérapeutiques différentes : l’escitalopram fait partie d’une classe d’antidépresseurs appelés *Inhibiteurs sélectifs de la recaptures de la sérotonines* ; L’aripiprazole appartient à la classe des *antipsychotiques atypiques de deuxième génération* et agit comme un agoniste partiel de la dopamine.

Les messages utilisés pour cette étude ont été publiés entre 2004 et 2013 sur trois des forums français les plus populaires (Doctissimo, atoute.org et Santé Médecine). Les métadonnées qui accompagnent les messages du corpus sont décrites dans la section 5.1 de ce manuscrit.

Les messages ont été extraits en se basant sur les noms commerciaux des médicaments étudiés : Seroplex[®] et Abilify[®]. Les posts ont été sélectionnés sur le critère de présence du nom du médicament dans le message.

8.2 Méthode

8.2.1 Traitement préliminaire des données

Notre objectif dans ce travail était l'identification des messages décrivant un comportement de non-observance au traitement. La différence d'objectif avec la partie I de ce manuscrit (détection de cas de pharmacovigilance) nous a poussés à considérer certains prétraitements différents.

8.2.1.1 Prétraitements standards

Le but de l'étape de prétraitement est le nettoyage des données pour réduire le bruit et l'incohérence [127]. Le prétraitement a été effectué en six étapes :

1. Étant donné que le logiciel R fait la distinction entre les caractères minuscules et majuscules, tous les messages ont été convertis en minuscules ;
2. La ponctuation et les mots vides ont été supprimés ;
3. Nous avons supprimé toutes les occurrences des noms de médicaments utilisés pour construire le corpus (par exemple, « seroplex »). Comme ils étaient présents dans chaque message, ces derniers étaient surreprésentés et non-informatifs ;
4. Les espaces abusifs ont été supprimés pour créer nos segments lexicaux (*tokens*) (voir 6.2.1) ;
5. La racination des mots a été effectuée en utilisant l'algorithme de Porter [111, 112] ;
6. Nous avons décidé de conserver les unigrams et les bigrams. Les bigrams ont permis de traiter des séquences contiguës fréquentes de deux éléments, tels que « effets secondaires » ;

8.2.1.2 Standardisation des mentions de dosage

Comme des variations dans les descriptions de dosages dans les messages sont possibles (par exemple, utilisation de *milligramme* ou *mg*), nous les avons remplacés par une expression standard dans les posts. Nous avons identifié des mentions de dosages (par exemple *10 mg*) en recherchant chaque séquence de nombres suivie d'une unité de dosage.

Ensuite, nous avons remplacé la mention de dosage par une chaîne de caractères neutre « *dosemilligrammes* ».

8.2.1.3 Estimation du modèle

Pondération de la Matrice Termes-Documents :

La Document-Term Matrix (DTM) décrit la fréquence des termes qui apparaissent dans une collection de messages : les lignes correspondent aux messages (documents) et les colonnes correspondent aux termes contenus dans ces derniers (tokens). Si un terme apparaît dans un message particulier, alors l'entrée de la matrice correspondant à ces lignes et colonne vaut 1, sinon elle vaut 0. La « sparsité » (caractère creux) de la matrice correspond à la fréquence des éléments de valeur nulle dans cette dernière.

Un seuil maximal de sparsité, au-dessus duquel les tokens ont été supprimés, a été déterminé empiriquement. La sparsité totale de la matrice a été calculée pour un intervalle de seuils de sparsité appliqué aux colonnes (au termes apparaissant dans les messages). Ces valeurs allaient de 99,95% à 80% et ont été décrémentées de 0,025%. Nous avons conservé dans la DTM des tokens correspondant à une sparsité d'au moins 97%. Ensuite, pour éviter la surreprésentation des tokens fréquents, nous avons appliqué une pondération à notre DTM basée sur l'approche Term-frequency Inverse-document-frequency (Tf-Idf) [128].

Après traitement de la sparsity, une DTM a été générée pour chaque corpus (escitalopram et aripiprazole, respectivement) et utilisée comme entrée de la modélisation LDA.

Afin de supprimer les tokens correspondant à des fautes d'orthographe ou des abréviations non-pertinentes et ne prendre en compte que les mots fréquemment utilisés par les patients, nous avons supprimé les tokens peu fréquents en nous basant sur la sparsité de la DTM obtenue pour chaque corpus.

La figure 8.1 résume la méthode utilisée dans cette étude.

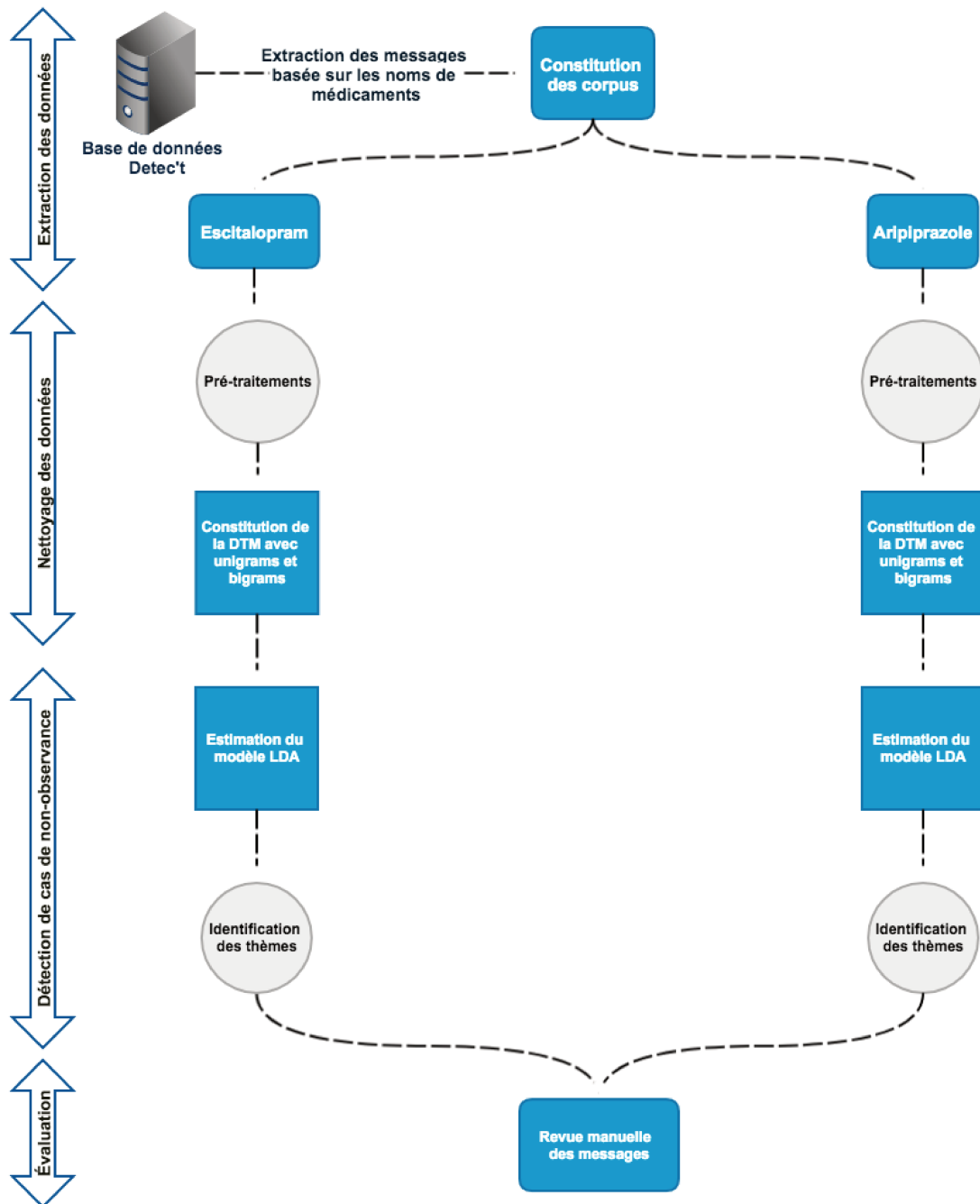


FIGURE 8.1: Récapitulatif de l'approche proposée

Modèle LDA :

Notre méthode s'appuie sur l'algorithme LDA. Ce modèle a été développé par David Blei, Andrew Ng et Michael Jordan [94]. Le LDA est un modèle génératif, mais dans l'exploration de texte, il introduit un moyen d'attacher un contenu thématique aux documents textuels analysés. Chaque document est considéré comme un mélange de

plusieurs thèmes distincts. Un avantage de la méthode LDA est que l'on n'a pas à connaître à l'avance quels sont les thèmes ni comment ils sont décrits dans le corpus. En ajustant les paramètres du modèle LDA, on peut explorer la formation des thèmes et les groupes de documents qui y sont rattachés.

Le nombre de thèmes (*topics*) K doit être spécifié par avance. Le corpus contient les documents $d_i = (w_{i,1}, \dots, w_{i,n_i})$ de longueur n_i . Chaque mot $w_{i,j}$ provient d'un vocabulaire consistant en V termes différents.

La distribution des termes pour chaque topic est modélisée par la formule 8.1 :

$$\beta_i \sim \text{Dirichlet}(\eta) \quad (8.1)$$

où : $\text{Dirichlet}(\eta)$ correspond à une distribution de Dirichlet de paramètre η .

La proportion de la distribution des thèmes pour chaque document est distribuée suivant le formule 8.2 :

$$\omega_i \sim \text{Dirichlet}(\alpha) \quad (8.2)$$

Chaque mot $w_{i,j}$ est associé à un topic $z_{i,j}$ comme décrit dans le formule 8.3 :

$$z_{i,j} \sim \text{Multinomial}(\omega_i) \quad (8.3)$$

où : $\text{Multinomial}(\omega_i)$ décrit une distribution multinomiale.

Un document est un mélange de thèmes. C'est-à-dire qu'il correspond à une distribution de probabilité sur l'ensemble des thèmes identifiés dans le corpus. Autrement dit, lorsqu'un patient écrit un message, il choisit de parler d'un certain nombre de thèmes. Quand il parle d'un thème dans un message, il utilise des mots avec une certaine probabilité (sur l'ensemble du vocabulaire) de correspondre à ce thème. L'hypothèse de base du modèle est la suivante : chaque message contient plusieurs thèmes parmi l'ensemble des thèmes identifiés, et la distribution de probabilité indique la prédominance des thèmes identifiés dans les documents (les messages).

D'un point de vue technique, les raisons du choix du modèle LDA étaient les suivantes :

1. Comparée à d'autres algorithmes de topic models (analyse sémantique latente LSA, factorisation matricielle non-négative NMF ou décomposition de valeurs singulières dans le contexte de LSA), la méthode LDA est plus adaptée aux domaines où les données sont en unités sémantiques ;
2. Le modèle LDA fournit une meilleure interprétabilité des thèmes que d'autres algorithmes de topic models (tels que NMF) ;
3. Le modèle LDA fournit également une meilleure cohérence sémantique des thèmes estimés que la LSA [129] ;

Nous avons utilisé l'algorithme LDA développé par Blei et al. [94, 130]. Le modèle LDA a été estimé en utilisant l'algorithme Maximum A Posteriori (MAP) décrit par Taddy [131, 132]. L'algorithme MAP est une variante de l'algorithme d'espérance-maximisation (EM) avec un coût de calcul plus faible et des résultats plus stables que les algorithmes couramment utilisés pour les estimations (échantillonnage de Gibbs, EM variationnelle). À chaque itération, au lieu d'approcher la maximisation de la vraisemblance marginale, une estimation combinée des paramètres est calculée par diagonalisation en bloc de la matrice de Hessienne. Cela conduit à une estimation exacte de la distribution des thèmes, plutôt qu'une approximation. Le nombre de thèmes a été sélectionné en utilisant le facteur log Bayes [131]. Le log-Bayes factor est un rapport de vraisemblance utilisé pour les comparaisons de modèles. En le calculant sur un topic models pour des cas de nombres de thèmes différents, il permet de sélectionner un nombre de thèmes approprié pour décrire l'ensemble du corpus.

Dans le but d'optimiser l'interprétabilité et la cohérence sémantique des thèmes, nous avons considéré un message significativement associé à un thème quant au moins 25% du vocabulaire (des *tokens*) qu'il contenait étaient associés à ce thème. Le seuil de 25% a été fixé empiriquement.

8.2.1.4 Évaluation

L'objectif de cette étape d'évaluation était d'estimer le nombre de messages correctement identifiés par notre approche. L'évaluation manuelle a été effectuée en deux étapes :

1. Nous avons revu manuellement tous les messages relatifs aux thèmes d'intérêt (variation de dosage et interruption du traitement) dans les deux corpus (Escitalopram et Aripiprazole). Un message est considéré comme correctement classé s'il décrit un comportement non-observant correspondant au thème reconnu par la méthode des topic models. L'évaluation de notre classification a été mesurée par le ratio de messages correctement classés pour chaque thème d'intérêt. Deux annotateurs (Redhouane Abdellaoui et Pierre Foulquié) ont participé à l'examen des messages. Pour mesurer l'accord inter annotateurs, les deux annotateurs ont annoté une sélection aléatoire commune de 20% des messages de chaque ensemble identifié par les thèmes de non-observance. L'accord inter annotateurs a été calculé en utilisant le coefficient Kappa de Cohen [133];
2. Pour estimer le rappel de notre méthode, nous avons extrait au hasard 20% des messages liés à des thèmes autres que la non-observance (345/1723 messages pour l'aripiprazole et 650/3246 pour l'escitalopram). Nous les avons classés manuellement dans deux catégories : les messages avec des comportements de non-observance et sans comportements de non-observance;

Chapitre 9

Résultats

Nous présentons dans ce chapitre les éléments de modélisation de discussions entre patients sur les forums de santé. Ce travail est réalisé en collaboration avec **Pierre Foulquié**. Notre apport à ce travail a consisté en la conception méthodologique de cette approche et l’encadrement de **Pierre Foulquié** pendant son stage de fin d’études au sein de la société **Kappa Santé**.

9.1 Caractéristiques du jeu de données

Le table **9.1** expose le nombre de messages dans chaque corpus extrait de la base de données Detec’t.

Médicament	Classe thérapeutique	Nombre de messages avec le nom du médicament	Date de publication
Escitalopram	Antidépresseur	3 650	Entre 2004 et 2013
Aripiprazole	Antipsychotique	2 164	Entre 2005 et 2013

TABLE 9.1: Description des corpus.

Le prétraitement préliminaire du corpus escitalopram a donné une DTM de 3 650 messages et 155 883 tokens (unigrammes et bigrammes). En fixant le seuil de sparsity à 99,35%, nous avons obtenu une DTM de 3 649 messages et 1 497 tokens. Un message a été automatiquement supprimé, car les termes qu’il contenait étaient particulièrement mal orthographiés.

Le traitement du corpus aripiprazole a donné une DTM de 2 164 messages et 81 371 tokens. Sur la base d'un seuil de sparsity fixé à 99,25%, nous avons obtenu une DTM de 2 164 messages et 1 062 termes.

9.2 Variations de doses et arrêts de traitements

9.2.1 Estimation du modèle

La méthode de sélection du nombre de thèmes par le log-Bayes Factor a retourné un total de 13 thèmes pour le corpus d'escitalopram, comme le montre la figure 9.1. La même approche nous a conduit à identifier 11 thèmes pour le corpus de l'aripiprazole.

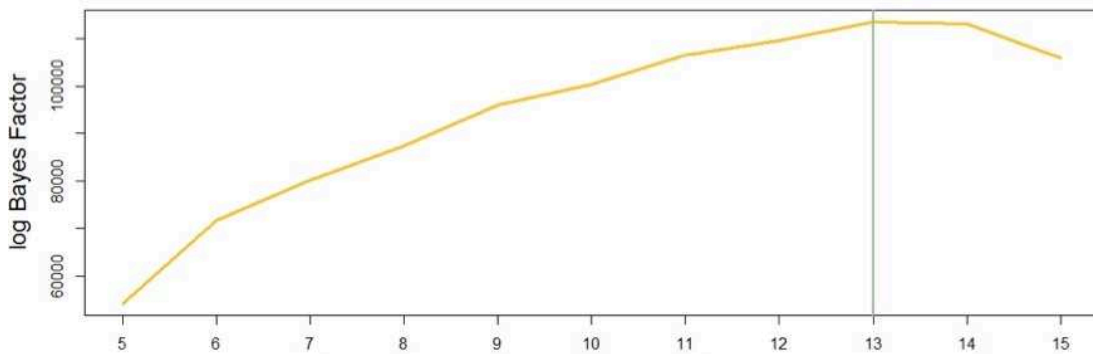


FIGURE 9.1: Nombre de thèmes sélectionnés pour Escitalopram en utilisant le log-Bayes Factor.

Nous avons obtenu un total de 2 691 messages évoquant l'escitalopram répartis entre 13 thèmes. Les 958 messages restants étaient en dessous du seuil concernant l'association entre termes et thèmes, qui était fixé à 25%. Le nombre moyen de thèmes par message était de 1,22 et le nombre médian de thèmes est à 1.

Pour les données sur l'aripiprazole, nous avons obtenu un total de 1 778 messages mentionnant le médicament et distribués parmi 11 thèmes. Les 396 messages restants étaient inférieurs au seuil d'association entre les termes et les thèmes du message. Le nombre moyen de thèmes par message était de 1,31 et le nombre médian de thèmes est à 1.

9.2.2 Interprétation des thèmes identifiés

Comme un thème doit être interprétable avec les premiers termes obtenus (classés par leur probabilité d'apparition) [130], les thèmes identifiés ont été étiquetés manuellement à partir des 15 premiers mots les décrivant.

9.2.2.1 Thèmes liés à l'Escitalopram

La liste des thèmes et la distribution des messages dans ces derniers sont présentées en table 9.2 pour le corpus Escitalopram. Ce résultat est exprimé en fréquence et en proportions de messages associés à chaque thème (par rapport aux 3 649 messages d'escitalopram). Un message est associé à un thème s'il contient au moins 25% du vocabulaire pour lesquels la variabilité latente correspondante décrit une association avec le thème en question.

TABLE 9.2: Thèmes liés au corpus Escitalopram.

Thème	Label	Mots caractéristiques	Messages significativement associés	Proportion (%)
1	Thèmes Généraux	rien, peur, fois, arriver, sait, vie, faire, plus, peux, chose, fait, dire, vraiment, continuer, arriver a	451	12,36
2	Prescription	médecin, bon courage, courage, bon, prescrit, médecin a, dit, psi, médicament, a prescrit, vie, a dit, tout, prendre, dépendre	297	8,14
5	Angoisse, Attaque de panique	angoisse, crise, matin, soir, crise angoisse, xanax, grosse, prendre, jour, dormir, hier, panique, journée, depuis, nausées	285	7,81
3	État quotidien	mal, journée, bonne, nuit, mal a, bisous, super, gros, tout, fille, message, a cause, enfin, noire, sommeil	266	7,29
7	Expériences patients	depuis, merci, sous, prend, savoir, prend depuis, mois, depuis mois, réponse, depuis an, aimerais, sous depuis, seropram, savoir si, avancer	225	6,17
9	Prise de poids	deux, poids, prise, autre, effexor, marche, antidépresseur, dépression, différent, médicament, prise, prise poids, prozac, chose, kilo	218	5,97
4	Interruption de traitement	arrêt, mois, pendant, sevrage, pris, traitement, suit, année, debut, demi, depuis, suit à, nouveau, senti, plusieurs	216	5,92
6	Amélioration de la condition	mieux, ça va, bien, vais, beaucoup, espère, voir, aide, dit, aller, va mieux, compte, va bien, donne, psi	206	5,65
10	Effets indésirables	effet, secondaire, effet secondaire, aucun, médicament, aucun effet, phobie, beaucoup, part, libido, social, moins, négatif, plus effet, a part	199	5,45
11	Modifications de dosage	dosemilligrammes, a dosemilligrammes, dose, trouble, augmenter, anxiété, prend dosemilligrammes, dosemilligrammes jour, pendant, diminuer, Temesta, sevrage, dosemilligrammes a, dosemilligrammes depuis, jour	187	5,12
8	Activités	commencer, quelques, commencer a, encore, avant, temps, aider, besoin, penser, grand, semaine, question, plusieurs jour, sortir, physique	167	4,58
12	Effets du traitement	petit, semaines, temps, positif, début, combien, bonsoir, côté, plutôt, fin, maux, combien temps, sentir, tête, sommeil	100	2,74
13	Durée du traitement	jours, aujourd'hui, ça fait, gens, depuis jours, toujours, anti, passer, encore, gérer, dépresseur, demande si, anti dépresseur, a jour, jour a	93	2,55

Nous avons remarqué l'émergence d'une classe contenant des messages décrivant les expériences de l'utilisateur avec le médicament d'une manière générale (thème 7) et décrivant comment cela affecte leur état de santé (thème 6). Les thèmes 3 et 8 ont trait aux sentiments quotidiens des patients et à leurs activités. Le thème 2 était lié à la prescription de médicaments par les professionnels de santé et le thème 5 aux attaques de panique et à l'anxiété. Les thèmes 9, 10 et 12 correspondent à des messages concernant les effets indésirables vécus ou redoutés par les utilisateurs, ainsi que les effets globaux du médicament. Le thème 13 était lié à la durée du traitement.

Le thème 1 a été étiqueté comme « thème général ». Il décrit des thèmes associés à des discussions entre individus et correspond à un vocabulaire très « généraliste » (formules de politesse par exemple) et donc non-informatif. Néanmoins, ce vocabulaire était largement utilisé dans les messages, ce qui explique la proportion relativement élevée de messages associés à ce thème.

Les messages sur les problèmes d'arrêt du traitement et les variations de doses ont été regroupés respectivement dans les thèmes 4 et 11. L'intersection des deux thèmes de non-observance correspondait à 7 messages.

9.2.2.2 Thèmes liés à l'Aripiprazole

Les thèmes obtenus en reproduisant les étapes de modélisation avec le corpus de l'aripiprazole sont décrits en table 9.3.

TABLE 9.3: Thèmes liés au corpus Aripiprazole.

Thème	Label	Mots caractéristiques	Messages significativement associés	Proportion (%)
4	Interruption de traitement	effet, secondaire, effet secondaire, poids, arrêt, prise, kilo, prise poids, baisse, perdre, a arrête, médicament, perdre poids, moins, reprendre	265	12,25
2	Expériences patients (témoignages)	petit, vie, envie, zyprexa, vraiment, temps, plus, sais, assez, impression, nuit, sais si, temps, vivre, dur	253	11,7
1	Thèmes Généraux - 1	psi, angoisses, chose, peut, cause, voir, autre, médecin, penser, a cause, faut, quand, crise, psi a, problème	246	11,4
6	Durée du traitement	jours, semaine, mois, deux, pendant, aujourd’hui, premier, arrêt, quelqu’un, xanax, trois, nouveau, psychiatre, dernier, fois	218	10,1
5	Interactions sociales divers	dit, bon, forum, courage, personne, maladie, bon courage, a dit, fille, fils, peut, donner, ami, souhait, nouvelles	214	9,9
7	Prescription	depuis, prends, prescrit, prends depuis, merci, depuis ans, bipolaire, a prescrit, troubles, traitement, dépression, psi, sous, semaines, depuis mois	213	9,9
8	Insomnie	solian, dormir, bonne, prend, sommeil, journée, problème, tercian, tout, nuit, somnifère, convient, insomnie, heure, arrive a	198	9,1
3	Échanges entre patients	bien, sous, très, rien, sens, ça va, neuro, aucun, support, très bien, mal, a fait, bien a, contre, a rien	190	8,8
9	Effets du traitement	neuroleptique, effet, patient, dose, étude, risque, autre, indésirable, effet indésirable, clinique, dopamine, traitement, agit, médicament, différent	190	8,8
11	Modifications de dosage	<i>dosemilligrammes, a dosemilligrammes, jour, dosemilligrammes, matin, prendre, prandre dosemilligrammes, soir, dosemilligrammes jour, comprimé, augmenter, dosemilligrammes a, dosemilligrammes dosemilligrammes, Zyprexa dosemilligrammes, dosemilligrammes matin</i>	176	8,1
10	Thèmes Généraux - 2	dire, risperdal, merci, savoir, temps, médicament, aimerais, savoir si, combien, réponse, voudrais, peux, avancer, libido, combien temps	158	7,3

Parmi les thèmes estimés pour le corpus aripiprazole, nous avons identifié le thème correspondant aux expériences des patients (thème 2). Deux thèmes ont décrit les effets du traitement (thèmes 8 et 9), et un thème est lié à la durée du traitement (thème 6). Deux thèmes ont porté sur les relations des patients avec d'autres patients (thèmes 3 et 5) et un thème traite des prescriptions faites aux patients par les professionnels de la santé (thème 7).

Le thème 4 décrivait l'interruption du traitement. Les variations de dosage ont été décrites dans le thème 11. L'intersection des deux thèmes de non-observance (4 et 11) correspondait à 6 messages. Comme pour l'escitalopram, deux thèmes étaient composés de mots non-informatifs (thèmes généraux 1 et 10).

La table 9.4 montre les thèmes identifiés dans les deux corpus présentés par ordres décroissants (en nombre de messages).

Escitalopram	Aripiprazole
Thèmes Généraux	Interruption de traitement
Prescription	Expériences patients (témoignages)
Angoisse, Attaque de panique	Thèmes Généraux - 1
État quotidien	Durée du traitement
Expériences patients	Interactions sociales divers
Prise de poids	Prescription
Interruption de traitement	Insomnie
Amélioration de la condition	Échanges entre patients
Effets indésirables	Effets du traitement
Modifications de dosage	Modifications de dosage
Activités	Thèmes Généraux - 2
Effets du traitement	-
Durée du traitement	-

TABLE 9.4: Thèmes associés aux deux corpus.

9.2.3 Évaluation

La cohérence inter annotateurs a été mesurée sur 20% (169/844) des messages identifiés par les thèmes de non-observance en utilisant le coefficient Cohen kappa, nous avons obtenu une cohérence inter annotateurs de 0,90 (152/169).

Nous avons calculé le ratio de messages correspondant à des cas de non-observance associé à chaque thème d'intérêt. Les résultats sont présentés dans les tables 9.5 et 9.6. La précision obtenue pour l'identification de cas de non-observance était de 32,6% (272/844). Nous avons obtenu la précision la plus basse (23,4%, 62/265) pour le thème d'arrêt d'aripiprazole et la précision la plus élevée (46,3%, 100/216) pour l'arrêt d'escitalopram.

Comportement identifié	Nombre de messages (n)	Nombre de messages correctement classés (n)	Precision (%)
Variation de dosage	187	54	28,9
Arrêt de traitement	216	100	46,3

TABLE 9.5: Annotations du corpus Escitalopram.

Comportement identifié	Nombre de messages (n)	Nombre de messages correctement classés (n)	Precision (%)
Variation de dosage	176	56	31,8
Arrêt de traitement	265	62	23,4

TABLE 9.6: Thèmes associés aux deux corpus.

L'analyse de 20% des messages (345/1723 pour l'aripiprazole, 650/3246 pour l'escitalopram) liés à d'autres sujets que la non-observance n'a révélé que quatre messages décrivant un comportement de non-observance et non détectés par notre approche. Les quatre faux négatifs concernaient l'arrêt d'aripiprazole. Nous avons donc obtenu un score de rappel de 94% (62/66) pour le thème d'arrêt d'aripiprazole et 100% pour l'arrêt d'escitalopram et les changements de doses pour les deux médicaments. Le rappel moyen estimé était de 98,5% (272/276).

Nous présentons ci-après une analyse détaillée des résultats pour le corpus Aripiprazole.

9.2.3.1 Variations de doses

Le modèle LDA utilisé a permis d'identifier 176 messages comme traitant de variations de dosage. La revue manuelle de ces messages a révélé que seulement 56 messages (31,8%, 56/176) contenaient une véritable déclaration de non-observance (soit 2,6% des 2 164 messages initialement inclus dans le corpus).

Parmi les 120 messages restants, 68 (56,7%, 68/120) étaient des discussions entre patients comparant leurs doses d'aripiprazole. Au total, 13 messages (10,8%, 13/120) contenaient des renseignements sur les doses d'autres médicaments prescrits en plus de l'aripiprazole. Les médicaments les plus cités étaient l'amisulpride et l'olanzapine. Huit messages (6,7%, 8/120) étaient des questions sur les posologies de l'aripiprazole, sept messages (5,8%, 7/120) évoquaient un changement de doses, six messages (5%, 6/120) correspondant à des modifications de dosages en accord avec le médecin, et quatre messages (3,3%, 4/120) étaient des conseils.

Huit messages (6,7%, 8/120) ne contenaient pas de mentions de dosage, mais seulement des mots évoquant une variation tels que « *augmentation* » ou encore « *diminution* », accompagnant généralement des références au dosage.

Les six posts restants (5%, 6/120) mentionnaient des changements de posologies prévues ou susceptibles de se produire à l'avenir. Par exemple :

« ...je prends 5 mg pour le moment le psy veut augmenter la dose à 10 mg au prochain rendez-vous... »

Dans tous les cas de non-observance (56 cas), le patient a diminué la dose en raison d'effets indésirables. Les effets indésirables les plus fréquemment mentionnés dans ces messages étaient l'insomnie, l'asthénie et les problèmes de libido.

9.2.3.2 Arrêts de traitements

Nous avons identifié 62 messages correspondant à des comportements de non-observance sur les 265 messages liés au thème d'arrêt d'aripiprazole (23,4%, 62/265). En d'autres termes, 2,86% des 2 164 messages du corpus sont des messages de patients prenant de l'aripiprazole qui ont décidé de mettre fin à leur traitement.

Parmi les 203 postes restants, treize (6,4%, 13/203) correspondent à l'interruption d'aripiprazole, mais ne sont pas des cas de non-observance : soit l'aripiprazole est arrêté pour débiter un autre traitement, soit l'arrêt définitif du traitement est décidé par le médecin.

Un total de 89 postes (43,9%, 89/203) ont été rédigés par des patients qui ont reçu ce traitement par le passé.

Dans 55 messages (27,1%, 55/203), le patient a mentionné qu'il ou elle était réticent à poursuivre le traitement, principalement à cause des effets indésirables (74,6%, 151/203). Les symptômes les plus cités étaient l'insomnie, la fatigue, les problèmes de libido et les nausées.

Dans 23 posts (11,3%, 23/203), le patient a reçu plus d'un traitement, et le post décrit l'interruption de l'un des autres médicaments (par exemple, dans quatorze cas, l'olanzapine a été arrêté).

Dix-huit messages ont été attribués de manière erronée au thème d'arrêt de traitement parce qu'ils contenaient des termes comme « *arrêt* », bien qu'ils ne mentionnent pas l'arrêt d'aripiprazole. Parmi ceux-ci, il y avait 11 messages (5,4%, 11/203) où l'interruption n'était pas liée à un sujet de santé « ... *Je me suis arrêté, comme gelé...* », cinq messages correspondant à un arrêt d'alcool, de narcotiques ou de tabac et non d'aripiprazole (2,5%, 5/203). Deux posts (<1%, 1/203) où le patient a arrêté un type d'alimentation ou d'autres activités.

Les cinq messages restants (2,5%, 5/203) étaient des demandes de conseils à propos d'un éventuel arrêt de traitement.

Chapitre 10

Discussion

10.1 Principaux résultats

Pharmacovigilance

L'algorithme de détection de mentions de prise du traitement par le déclarant (section 6.1) réalise une précision de 72% et un rappel de 96%. Cette méthode réduit la quantité de messages à considérer en tenant compte de la pertinence de l'information exclue de l'analyse. Ce qui réduit la probabilité de commettre une erreur lors de la classification des entités en effets indésirables ou en autre type d'entités. Un autre avantage de cet algorithme est la rapidité d'exécution du package « *DetectAnalyzer* ». Cet algorithme est appliqué au moment de l'extraction des messages depuis les forums de santé (par le Web Crawler), et notre package prend en entrée les messages mentionnant une prise de traitement. Ce filtrage (dès l'étape d'extraction) entraîne une réduction de la volumétrie de messages à analyser ce qui permet un gain de temps d'exécution pour les modules suivants de notre package.

L'algorithme des distances entre entités que nous proposons (section 6.3) atteint une précision de 95,8% et un rappel de 50%. Il existe donc une limite d'éloignement (en nombre de mots) à partir du nom de médicament au-delà de laquelle il est possible d'affirmer qu'un concept médical n'est pas un effet indésirable. Cette approche s'adapte automatiquement au corpus en identifiant les queues de la distribution correspondant aux concepts les plus éloignés du nom du médicament. L'utilisation de cette approche

en routine consiste à appliquer la classification basée sur les distances pour un nouveau corpus de messages puis à filtrer les groupes de concepts médicaux en identifiant ceux avec les moyennes de distances les plus élevées.

L'algorithme de pertinence (section 6.4) proposé pour la caractérisation des effets indésirables déclarés réalise une précision de 19,3% et un rappel de 87,2%. Cette approche nous a permis d'identifier, avec une sensibilité acceptable, les effets indésirables avérés parmi les déclarations présentes dans le corpus et validées manuellement. Néanmoins, en nous basant sur cette approche, nous n'avons pas été en mesure d'obtenir des prédictions satisfaisantes du fait de la faible précision de cette dernière.

Non Observance

À notre connaissance, l'approche que nous proposons pour étudier la non-observance au traitement à partir des forums de santé constitue la première initiative pour traiter ce sujet. Notre méthode permet d'isoler les messages décrivant les arrêts de traitements et les modifications de doses avec un rappel moyen de 98,5% et une précision moyenne de 32,6%. Notre étude montre que les topic models sont donc utiles pour identifier avec une bonne sensibilité les messages signalant des comportements de non-observance au traitement.

La revue manuelle des messages affectés aux différents topics a permis d'identifier les situations qui sont à l'origine des erreurs de classification (faux positifs) [125].

1. Deux expériences avec des médicaments différents ou des personnes différentes rapportées dans un même message (Par exemple, « ... *Jean prend 100 mg, alors que Marie prend 200 mg ...* »);
2. Description d'une tentative future à l'initiative du médecin ou du patient (par exemple, le patient signale que le médecin veut augmenter ou diminuer la dose au prochain rendez-vous.);
3. L'action concerne un autre produit utilisé par le patient, et dans certains cas ce produit n'est pas un médicament. Par exemple, l'arrêt du tabac, l'arrêt d'un autre médicament pris par le patient, etc.;

La cause d’erreurs de classification la plus fréquemment observée dans les messages correspond à des cas d’interruption du traitement en accord avec le médecin (pour l’arrêt de traitement) et des discussions autour des posologies prescrites (pour la variation de dosage).

10.2 Comparaison avec la littérature

Il est difficile de construire une comparaison de nos travaux avec le reste des approches de la littérature, car de nombreuses différences sont à noter.

La première est le développement des approches sur différents corpus provenant de différentes sources (Twitter, Facebook, Forums, Blogs). Les caractéristiques propres à chaque type de source ont un impact sur les messages postés (nombre de caractères limités, partage via un compte identifié, anonymat et liberté d’écriture).

Une autre différence est le développement des approches pour des langues différentes. La plus part des méthodes identifiées dans la littérature sont conçues pour traiter des messages en anglais. Trois publications font état de messages analysés en français [5, 50, 100].

La dernière différence observée, propre à notre objectif de pharmacovigilance, est l’utilisation de terminologies médicales différentes (MedDRA[®], UMLS[®], SIDER[®], COSTART[®]) pour la reconnaissance d’entités médicales.

Filtrage de données en pharmacovigilance

Les approches de filtrage des données non-pertinentes ont déjà été explorées dans la littérature. Powell et al. [89] montrent que seulement 26% des posts contiennent de l’information pertinente. Cette observation du faible taux d’information utilisable est appuyée par les résultats que nous obtenons avec notre algorithme des distances [104] où seulement 11,4% des concepts médicaux identifiés étaient des effets indésirables (189 concepts parmi 1 654).

Le filtrage de ce bruit améliore la détection de cas avérés de déclaration d’effets indésirables. Yang et al. [91] utilisent une approche type topic models pour filtrer les messages

et réduire leur ensemble de données à un sous-ensemble de posts évoquant de potentiels effets indésirables. Cette approche de filtrage a été testée sur différents modèles provenant de la littérature. L'utilisation du filtre proposé par Yang et al. [91] a permis un gain de F-score qui fluctuait entre 1,94% et 6,14%. Sarker & Gonzalez [93] évaluent différentes caractéristiques de filtrage dont une approche type n-grams qui réalise une accuracy 80,7% et une approche d'évaluation de topic models avec une accuracy de 86,1%.

L'algorithme des distances proposé dans ce travail constitue à notre connaissance la première étude à s'appuyer sur la proximité entre noms de médicaments et potentiels effets indésirables au sein d'un message.

Non Observance

Notre travail de revue de la littérature (Chapitre 3) n'a pas permis d'identifier des recherches antérieures traitant de la problématique de la non-observance au traitement à partir des témoignages de patients sur les forums de santé. Ne pouvant donc étudier des articles traitant de ce sujet, nous basons notre comparaison sur l'usage des topic models (et notamment du modèle LDA) avec les autres études ayant utilisé ce modèle.

De multiples études se sont appuyées sur les topic models afin d'étiqueter automatiquement des ensembles de tweets [95–98, 102]. Seules deux études utilisent des données issues de forums de santé [91, 100]. Ces deux études utilisent le modèle LDA.

Tapi Nzali et al. [100] ont utilisé un LDA sur des messages rédigés en français issues de groupes publics sur Facebook et du forum de santé de cancerdusein.org. Nos deux études ne sont pas comparables, car leur objectif était la mesure de la correspondance entre thèmes identifiés dans les messages et les réponses à des questionnaires de qualité de vie. Notre objectif était la qualification de groupes de messages décrivant des comportements de non-observance au traitement.

Yang et ses collaborateurs ont rapporté un F score de 89% dans leurs travaux [91]. Toutefois, leur approche se concentrait sur la détection de déclarations d'effets indésirables (sur des messages en anglais) et non la non-observance au traitement (sur des messages rédigés en français).

10.3 Limites et perspectives

Des limites des approches proposées dans ce manuscrit sont à noter.

Pharmacovigilance

L'algorithme des distances proposé dans ce manuscrit n'est pas efficace pour classer les concepts médicaux « proches » du nom du médicament dans les messages analysés. Les performances de notre algorithme sont correctes en termes de spécificité puisque, parmi les concepts médicaux éliminés par cet algorithme, 95,8% n'étaient pas des effets indésirables. Cependant, notons que 82,3% des concepts médicaux restant après filtrage n'étaient pas non plus des effets indésirables. Ces concepts médicaux conservés, considérés comme « proches », se situaient entre 56 mots avant le nom du médicament et 77 mots après ce dernier. Nous avons privilégié une approche qui permettait de limiter l'exclusion des effets indésirables potentiels (« proches » du nom du médicament) tout en maximisant le filtrage des faux positifs (« éloignés » du nom du médicament).

Cette limite est également visible lorsque l'approche des distances entre mots est utilisée sur des posts provenant de Twitter. La limite, de 140 caractères par tweet, imposée par le réseau social rend le nombre de mots possibles dans un tweet trop faible pour le bon fonctionnement de notre méthode.

La détection des effets indésirables non filtrés par l'algorithme des distances est prise en charge par le modèle SVM proposé en section 6.4. La faible précision de ce modèle (19,3%) constitue une limite supplémentaire des travaux que nous proposons. Nous identifions deux raisons possibles à ces faibles performances. La première est l'ensemble d'entraînement utilisé qui est trop petit (1 561 concepts médicaux annotés manuellement par nos experts). L'annotation des jeux de données est un travail lourd et chronophage. La disponibilité des collaborateurs qualifiés est une limite supplémentaire pour la constitution d'ensembles d'apprentissage et de tests afin de renforcer ce modèle. La seconde faiblesse de ce modèle est due aux variables utilisées (position du concept dans le message, nombre de concepts dans le message, position du nom du médicament, valeur absolue de la distance entre le nom du médicament et le concept médical). Cette sélection de variables avait pour but de distinguer les effets indésirables à partir des positions

des entités dans le message. L'étude des faux positifs générés par ce modèle n'a pas permis de distinguer une cause précise à ces performances faibles.

Nous avons testé de notre modèle SVM sur un corpus anglais. Ce corpus était constitué de 730 messages provenant de trois forums de santé anglophones ([CancerCompass](#), [breastcancer.org](#) et [HealingWell.com](#)). Ces messages ont été postés entre 2005 et 2016 et concernaient 44 médicaments. Ce test a permis de réaliser de meilleures performances (63,4% de précision et 57,2% de rappel). Et ce, malgré un ensemble d'apprentissage plus petit (1 153 concepts médicaux).

Non Observance

Une première limite à ce travail de détection de cas de non-observance au traitement est la restriction aux données issues des forums de santé français. Lors de notre travail de revue de la littérature pour notre objectif de pharmacovigilance, nous avons conclu que l'usage des forums de santé donnait accès à une volumétrie de messages plus importante. Ce constat reste à vérifier pour une problématique différente comme l'étude de la non-observance au traitement. Un travail similaire réalisé à partir des données provenant d'autres sources comme les réseaux sociaux (Twitter, Facebook, etc.) ou les communautés de patients ([PatientsLikeMe](#), [ComPaRe](#), etc.) pourrait faire émerger des informations supplémentaires sur les problèmes rencontrés par les patients avec leurs traitements et les pratiques de non-observance qui en découlent.

Notre méthode est limitée par l'usage du même vocabulaire pour décrire des cessations de traitement ou des variations de doses et pour décrire des comportements d'arrêts ou de variations plus généraux non liés au médicament (arrêt de tabac ou arrêt de régime par exemple).

Une autre limite de notre étude est la définition empirique de certains seuils. Notamment la borne supérieure de sparsité de la DTM à 97% et le seuil d'association minimum entre mots et topics fixé à 25% et qui nous pousse à exclure 23,2% des messages qui ne sont reliés à aucun topic. La fixation empirique de ces seuils est courante dans la littérature. Prier et ses collaborateurs [96] sélectionnent le nombre de topics dans leur modèle en testant différents seuils par séries de 50 topics.

Notre approche bénéficierait d'un modèle plus développé comme le Structural Topic models (STM) [134]. Le modèle STM prend en compte les corrélations entre topics et fait ainsi émerger des relations entre les thèmes. Ces corrélations permettraient l'identification de potentielles causes de non-observance. L'utilisation du modèle STM permettrait une étude approfondie du lien entre non-observance au traitement et effets indésirables de médicaments à travers la détection de messages, comme l'exemple ci-après, traitant de ces deux thèmes :

« Bonsoir, Je souhaitais également vous livrer mon avis sur cet antidépresseur. J'ai pris 20 mg durant deux mois, et je ne les ai pas supportées. J'étais en proie à des hypersomnies qui ne me permettaient plus de vivre normalement (je me couchais, et me dormais 18 heures par jour, je ne pouvais plus avoir d'activité physique). En deux mois, j'ai également pris 6 kg, et j'ai observé une augmentation de la peau. Ces raisons m'ont fait arrêter immédiatement le traitement... »

De plus, ce modèle s'accompagne d'une série de mesures telles que l'*exclusivité* entre mots et topics ou encore la *cohésion sémantique* des topics. Ces indicateurs permettent un meilleur contrôle de la modélisation à travers la sélection automatique du nombre de topics approprié. Dans le cas où la mesure d'intérêt est la *cohérence sémantique*, les topics obtenus sont moins nombreux et agrègent les messages en se basant sur leur vocabulaire commun. Il en résulte un faible nombre de topics, décrits par ce vocabulaire commun et ces derniers sont représentatifs des grands thèmes abordés dans un corpus. À l'inverse, l'usage de la mesure d'*exclusivité* entre mots et topics renvoie un grand nombre de topics constitués à partir du vocabulaire propre à chaque groupe de messages. Ce qui donne une vision plus segmentée du corpus analysé.

Avec un taux observé de 6% des messages du corpus qui décrivent des comportements de non-observance au traitement, le fait de privilégier le critère d'*exclusivité* pourrait favoriser l'émergence de ces thèmes. Maximiser le nombre de topics obtenus augmenterait les chances de voir des thèmes si faiblement représentés apparaître. De plus, utiliser un STM avec un critère d'*exclusivité* permettrait de constituer des groupes de messages basés sur le vocabulaire propre aux messages de non-observance ce qui pourrait mieux isoler ces derniers.

Une caractéristique supplémentaire du modèle STM est qu'avec ce dernier, il est possible de générer des thèmes en tenant compte des métadonnées accompagnant les documents

textuels à classer. Cet avantage fait du STM un modèle très utilisé en sciences sociales [135] et en analyse de discours politiques [136]. Notamment pour des sujets d'analyse de discours avec de multiples intervenants. Ce qui permet d'identifier et d'imputer des thèmes à des personnes ainsi que de quantifier la part des thèmes abordés et du vocabulaire utilisé par chaque intervenant.

L'utilisation de ce modèle dans le contexte de notre problématique de détection de cas de non-observance au traitement permettrait d'une part de constituer des topics mieux segmentés en privilégiant le critère d'*exclusivité* entre mots et topics lors de la classification générant les thèmes. D'autre part, il serait possible d'intégrer le médicament surveillé, le pseudonyme du patient sur le forum, la date et d'autres informations afin de constituer un « historique médical » par patient à partir des informations rapportées par ce dernier sur les forums de santé. Pour ce faire, il suffirait de considérer les pseudonymes des patients ainsi que les dates des messages postés par ces derniers comme métadonnées d'un tel modèle. Ce qui permettrait d'analyser les témoignages du patient au fur et à mesure de l'évolution de son état de santé et des traitements qui lui sont prescrits. Néanmoins, une telle approche souffrirait d'un manque d'exhaustivité lié au caractère laconique de certaines informations sur les forums de santé.

10.4 Discussion générale

Les échanges entre patients sur les médias sociaux sont moins formels et moins structurés que dans le reste des textes biomédicaux (par exemple, les comptes rendus des dossiers patients, ou la littérature biomédicale). Il s'agit du principal facteur d'erreur de classification lorsque ces textes sont utilisés dans un contexte de détection d'informations de santé sur les médias sociaux. Ce manque de structure est dû aux choix de communication des déclarants qui se passent des règles de syntaxe usuelles au profit de syntaxes non standardisées. La principale raison de ce choix fait par les patients est le gain de temps dans leurs échanges.

L'extraction des messages depuis les forums de santé est une limite commune à nos deux objectifs. Dans le cas où la source de données était un forum de santé, des études (parmi celles considérées lors de notre revue de la littérature) ont utilisé un Web Crawler. Nous avons distingué deux types de Web Crawler. Le premier type se basait sur l'extraction

de l'ensemble de la page Web identifiée puis sur le nettoyage du contenu extrait afin de supprimer les éléments indésirables (code HTML, publicités, etc.). Le second type consistait à cibler le contenu des balises HTML délimitant les messages de patients. Ce second type de Web Crawler (utilisé pour nos travaux) exige un développement supplémentaire pour l'intégration d'une nouvelle source, car la structure propre à chaque site Web nécessite de cibler des emplacements différents pour extraire les messages. Dans un contexte de veille sanitaire sur les forums de santé, cette contrainte dans l'étape d'intégration de sources de données limiterait le périmètre de surveillance sanitaire.

Le travail d'Audeh et al. [107] constitue une solution possible à ce problème. Cette équipe propose un système permettant l'extraction structurée des messages issus des forums de santé français. L'utilisateur renseigne les caractéristiques du forum visé dans un fichier de configuration et l'extraction est réalisée sans développement supplémentaire. Ce qui constitue un accès rapide et structuré aux données des forums de santé français.

La qualité de rédaction des messages constitue une autre limite commune à nos deux objectifs. Comme nous avons pu l'observer lors de la réalisation de ce travail, des messages peuvent être trop mal rédigés pour qu'une personne statue sur la nature de l'information que ces derniers contiennent. Lors de notre étape d'annotation manuelle, nous avons été confrontés à des cas où le message ne permettait pas de statuer sur l'information de prise du traitement par le déclarant. L'un de ces messages est présenté en exemple : « *le cicalfate assèche, tout comme biseptine...* ». Lors de l'étape de prétraitement automatique des corpus d'étude de la non-observance au traitement, un message a été vidé des termes qui le composait, car ces derniers étaient particulièrement mal orthographiés.

L'utilisation des données des réseaux sociaux et leur intégration dans un système de remonter d'incidents médicamenteux constituent un enjeu réel pour les différents systèmes de santé à travers le monde. L'une des limites observées est le manque d'informations de contextualisation des cas extraits depuis les forums de santé et les réseaux sociaux. Par exemple, Coloma et al. [45] ont essayé de caractériser les événements indésirables identifiés sur Twitter, Facebook et Google+ à l'aide des données de géolocalisation. Ils ont conclu que ces informations de géolocalisation n'étaient pas utilisables dans le cadre d'une surveillance sanitaire car ces données sont « *...sparse and largely untrackable...* ».

Les conclusions de Coloma et al. [45] constituent un exemple du manque de métadonnées pour caractériser les informations issues des médias sociaux. Ce constat, appuyé par les

travaux de Lardon et al. [137] et d'Izem et al. [138], symbolise les problèmes d'intégration des cas de pharmacovigilance issues des réseaux sociaux dans une chaîne de remontée d'effets indésirables institutionnelle où les effets indésirables déclarés sont parfaitement caractérisés (âge et sexe du déclarant, date de l'évènement etc.). Ces informations sont soit indisponibles soit décentralisées sur les médias sociaux. De même, l'utilisation de ces informations dans un cadre de détection automatique de signaux de pharmacovigilance pousse à s'interroger sur la prise en compte de certains facteurs pouvant constituer un biais. L'étude d'Abou Taam et al. [139] montre également l'augmentation du nombre de messages liés au Mediator® pendant la période de scandale autour de ce médicament. Ce qui montre l'existence d'un biais de notoriété [140]. La prise en compte de ces biais depuis les données issues des médias sociaux constitue l'un des prochains défis de ce champ de recherche.

La prise en compte des indications des médicaments dans les messages constitue un nouvel enjeu. Pendant la réalisation de ce travail, nous avons exploré une possible prise en compte des indications de médicaments dans l'étape de caractérisation des concepts médicaux identifiés effets indésirables. Ces indications devaient être identifiées pour améliorer notre algorithme de pertinence et ainsi obtenir de meilleurs résultats de détection d'effets indésirables. La prise en compte des indications des traitements s'est avérée problématique. Pour chaque médicament, il fallait identifier manuellement dans MedDRA® tout ce qui pouvait constituer une indication. Et ce, dans les différents niveaux du dictionnaire (*SOC*, *HLGT*, *HLT*, *PT*, *LLT*). Une autre limite de la reconnaissance des indications dans les messages de patients est la prescription hors AMM de certains traitements. Par exemple, la prescription d'une pilule contraceptive pour traiter un syndrome d'ovaires micropolykystiques. Nous avons conclu que la prise en charge des indications des traitements nécessitait l'utilisation d'une ressource médicale où ces indications seraient renseignées. Ce travail est en cours de réalisation chez **Kappa Santé**.

Notre package *DetectAnalyzer* est aujourd'hui utilisé en routine par la société **Kappa Santé** dans la chaîne de traitement Detec't. Ce dernier a été utilisé pour réaliser une analyse de signaux de pharmacovigilance issus des forums de santé dans le cadre d'une étude pharmacoépidémiologique rétrospective menée par **Kappa Santé** pour les laboratoires Sanofi [105]. Trois produits ont été ciblés par le laboratoire pour cette étude : Lantus®, Aubagio® et Stilnox®.

Les données considérées pour cette étude étaient des messages provenant de la base de données Detec't qui ont été postés sur des forums de santé sur une période de 10 ans. Ces messages ont été postés entre le 1 janvier 2005 et le 31 décembre 2015 pour Lantus[®] et le Stilnox[®], entre le 1 avril 2014 et le 31 décembre 2015 pour Aubagio[®]. L'absence de messages avant 2014 pour Aubagio[®] est due à son AMM datée de septembre 2013 (mise sur le marché en mars 2014).

La décision d'exclure Twitter des sources de données a été prise pour deux raisons :

1. le faible nombre de tweets évoquant les médicaments d'intérêt ;
2. notre méthode n'est pas adaptée pour traiter les données issues de Twitter. Particulièrement l'algorithme des distances présenté en 6.3 ;

Les données utilisées comme « base de comparaison » pour l'application des méthodes de détection de signaux étaient constituées de messages liés à 327 médicaments sélectionnés aléatoirement dans la base de données Detec't. Les messages de cette base de comparaison ont été postés sur les mêmes forums de santé que les trois médicaments d'intérêt et dans le même intervalle de dates.

Les données de Vigibase[®] qui ont été considérées pour cette étude étaient des effets indésirables rapportés entre le premier janvier 2005 et le 31 décembre 2015. Ce qui correspondait à 7 618 cas pour les trois médicaments d'intérêt.

Parmi les mesures d'associations qui ont été utilisées figurent les :

- | | |
|--|---|
| 1. <i>Proportional Reporting Ratio</i> (PRR) | 4. <i>Bayesian Confidence Propagation</i> |
| 2. <i>Reporting Odds Ratio</i> (ROR) | <i>Neural Network</i> (BCPNN) |
| 3. <i>mid-P-values calculée à partir du test exact de Fisher</i> (midRFET) | 5. <i>Empirical Bayes Geometric Mean</i> (EBGM) |

Les seuils qui ont été appliqués dans ce travail sont décrits en [105].

Une base de 545 paires (médicament, effet indésirable) constituant les signaux issus des forums de santé a été comparée aux 7 618 paires extraites depuis Vigibase[®]. Parmi celles-ci, 422 paires étaient présentes dans les deux bases.

La comparaison entre ces deux bases de signaux a été réalisée à l'aide des indicateurs de précision, rappel, valeur prédictive négative, accuracy et spécificité pour chacune des mesures d'associations utilisées.

La spécificité élevée observée (entre 86,09% et 95,53%) montre qu’une majorité de paires (médicament, effet indésirables) non considérées comme signaux dans *VigiBase*[®] ne le sont pas non plus sur les forums de santé. Le faible rappel (entre 29,1% et 50,6%) indique qu’une part importante des signaux identifiés dans *VigiBase*[®] ne sont pas identifiés comme des signaux sur les forums de patients. La précision mesurée entre 51,2% et 75,4% montre que la plupart des signaux identifiés sur les forums sont bien des paires considérées comme signaux sur *VigiBase*[®]. La méthode *Empirical Bayes Geometric Mean* (EBGM) (utilisée avec le seuil $EBGM \geq 4$) produit la détection la plus concordante avec *VigiBase*[®] et ce parmi l’ensemble des métriques utilisées.

Les périodes d’apparitions des signaux dans chacune des deux bases considérées (*Detect* et *VigiBase*[®]) ont été comparées. Plus particulièrement, les dates d’apparitions des signaux dans chaque base et la différence des dates d’apparitions (en mois).

Cette étude a montré que 29,5% des signaux considérés sont identifiés jusqu’à 128 mois avant leur recensement dans *VigiBase*[®].

Ce travail a été réalisé au sein de *Kappa Santé* à l’aide de notre package *DetectAnalyzer* malgré les limites de ce dernier que nous avons énumérées. Des études complémentaires de détection de signaux de pharmacovigilance depuis les forums de santé et les réseaux sociaux doivent être réalisées afin d’affiner ce résultat.

Un autre défi reste à relever, la constitution d’un corpus en français permettant le développement et la comparaison des méthodes. Sarker, Nikfarjam & Gonzalez [46] ainsi que Leaman et al. [1] sont à l’origine de telles initiatives. Ce qui a permis à différentes équipes de comparer leurs approches d’analyse de médias sociaux en anglais. Un corpus de travail commun permettrait également la compilation des différentes approches de filtrage répertoriées dans un seul module permettant l’épuration des données issues des réseaux sociaux. La notion de corpus partagé s’inscrit dans le mouvement Findable Accessible Interoperable Reusable (FAIR) data [141] et la volonté de la communauté scientifique de rendre reproductible et transparente l’évaluation des algorithmes développés par les différentes équipes.

Chapitre 11

Conclusion

Les travaux d’Abou Taam et al. [139] ont permis, dès 2013, de montrer l’existence d’information exploitable dans les échanges entre patients sur les forums de santé français. Cependant, il restait des défis méthodologique à relever pour automatiser l’extraction de cette information.

La pertinence de l’information présente sur les réseaux sociaux est également évaluée. Abbe & Falissard [5] ont exploré des titres de discussions sur le forum *Doctissimo* autour des antidépresseurs et des anxiolytiques. Leur conclusion était que les préoccupations exprimées sur ce forum étaient cohérentes avec les situations observées en vie réelle en santé mentale. Chary et al. [19] ont conclu que les mentions d’abus de drogues sur Twitter étaient fortement corrélées avec les estimations du National Survey on Drug Usage and Health (NSDUH) sur l’abus d’opioïdes. Cette conclusion est faite après étude des corrélations entre la distribution géographique de fortes consommations d’opioïdes sur ordonnance estimé à partir des médias sociaux et les estimations du NSDUH.

De nombreuses méthodes ont été proposés pour extraire ces informations (voir 3) dans différentes langues. Parmi ces méthodes, Chen et al. [50] ont présenté des résultats préliminaires encourageants de reconnaissance de noms de médicaments (F-score 94%), de concepts médicaux (F-score 81%) dans les messages rédigés en français sur les forums de santé.

La présence d’un contenu pertinent sur les médias sociaux ainsi que l’exploitation de ces données a donc été démontrée dans la littérature.

En termes de filtrage de données issues des réseaux sociaux, différentes études ont conclu que ces approches étaient prometteuses. Sarker et al. [87], après une revue de la littérature, sont arrivés à la conclusion que le filtrage des données après leur extraction constitue un enjeu pour l'exploitation des données issues des médias sociaux. Ce constat a été vérifié par plusieurs autres études [54, 55, 88, 89, 91, 93].

Nous avons démontré que la distance entre concepts médicaux et noms de médicaments dans un message influence la probabilité qu'un concept médical soit un effet indésirable avéré. L'utilisation de la distance entre les entités médicales pour l'analyse des messages de patients à partir des médias sociaux nous a permis de filtrer une partie des faux positifs des concepts médicaux détectés, et ainsi, d'optimiser le dépistage d'effets indésirables dans les messages issus des forums de santé. Cette méthode fait partie du package *DetectAnalyzer* qui est aujourd'hui utilisé en routine par la société **Kappa Santé** dans le cadre du projet Detec't. Ce qui constitue l'un des livrables industriels de nos travaux.

De plus, notre étude de la non-observance au traitement à partir des messages issus des forums de santé a permis de démontrer l'existence d'un contenu pertinent pour traiter cette problématique. L'usage des topic models pour isoler des sous-groupes de messages contenant des cas de non-observance au traitement est efficace (rappel moyen observé de 98.5%). Cette méthode est également utilisée en routine par la société **Kappa Santé** et constitue un second livrable industriel de ce travail de recherche. Cette famille de méthodes permet également de traiter d'autres problématiques que la non-observance au traitement. Tapi Nzali et al. [100] ont étudié des messages sur le cancer du sein postés par des patients sur des médias sociaux français à l'aide du modèle LDA. Leur étude a confirmé que les médias sociaux pourraient être une source d'information importante pour l'étude de la qualité de vie dans le domaine du cancer.

Les topic models peuvent être optimisés avec des modèles plus évolués ou encore via des approches sémantiques pour opérer un filtrage plus fin et isoler les messages plus efficacement.

Des questions restent en suspens concernant les données collectées et les informations extraites depuis les médias sociaux. Le rôle des médias sociaux dans le système de pharmacovigilance institutionnelle reste à définir. L'étude menée par Coloma et al. [45] montre le manque de métadonnées permettant de contextualiser les effets indésirables détectés sur les médias sociaux.

Pour conclure, les résultats des approches observées dans la littérature et proposées dans cette thèse viennent confirmer la faisabilité d'utiliser les médias sociaux comme source de données à des fins de surveillance de l'usage de médicaments. De ce fait, en parallèle des perspectives de recherche méthodologiques citées plus haut, il paraît essentiel de réunir l'ensemble des disciplines concernées et de travailler à la mise en place d'un système de détection automatique d'information de santé depuis les médias sociaux en France.

Bibliographie

- [1] Leaman R, Wojtulewicz L, Sullivan R, Skariah A, Yang J, Gonzalez G. Towards internet-age pharmacovigilance : extracting adverse drug reactions from user posts to health-related social networks. In : Proceedings of the 2010 workshop on biomedical natural language processing. Association for Computational Linguistics ; 2010. p. 117–125.
- [2] Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. *Clinical Pharmacology & Therapeutics*. 2012 ;91(6) :1010–1021.
- [3] Benton A, Ungar L, Hill S, Hennessy S, Mao J, Chung A, et al. Identifying potential adverse effects using the web : A new approach to medical hypothesis generation. *Journal of biomedical informatics*. 2011 ;44(6) :989–996.
- [4] Chee BW, Berlin R, Schatz B, et al. Predicting adverse drug events from personal health messages. In : *AMIA Annu Symp Proc*. vol. 2011 ; 2011. p. 217–26.
- [5] Abbe A, Falissard B. Stopping Antidepressants and Anxiolytics as Major Concerns Reported in Online Health Communities : A Text Mining Approach. *JMIR mental health*. 2017 ;4(4).
- [6] Hauben M, Hung EY. Pneumothorax as an adverse drug event : an exploratory aggregate analysis of the US FDA AERS database including a confounding by indication analysis inspired by Cornfield’s Condition. *Int J Med Sci*. 2013 ;10(8) :965–73.
- [7] Hoffman KB, Kraus C, Dimbil M, Golomb BA. A survey of the FDA’s AERS database regarding muscle and tendon adverse events linked to the statin drug class. *PloS one*. 2012 ;7(8) :e42866.

- [8] Sabaté E. Adherence to long-term therapies : evidence for action. World Health Organization ; 2003.
- [9] Haynes RB, McDonald H, Garg AX, Montague P. Interventions for helping patients to follow prescriptions for medications. *Cochrane Database Syst Rev.* 2002;2(11) :7–9.
- [10] Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keenanasseril A, et al. Interventions for enhancing medication adherence. *The Cochrane Library.* 2014;.
- [11] Fox S, Duggan M. Health online 2013. Washington, DC : Pew Internet & American Life Project. 2013;.
- [12] Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignet J, Texier N, et al. Adverse drug reaction identification and extraction in social media : a scoping review. *Journal of medical Internet research.* 2015;17(7) :e171.
- [13] Househ M, Borycki E, Kushniruk A. Empowering patients through social media : the benefits and challenges. *Health informatics journal.* 2014;20(1) :50–58.
- [14] Horvath KJ, Danilenko GP, Williams ML, Simoni J, Amico KR, Oakes JM, et al. Technology use and reasons to participate in social networking health websites among people living with HIV in the US. *AIDS and Behavior.* 2012;16(4) :900–910.
- [15] Taggart T, Grewe ME, Conserve DF, Gliwa C, Isler MR. Social media and HIV : a systematic review of uses of social media in HIV communication. *Journal of medical Internet research.* 2015;17(11) :e248.
- [16] Wang W. Mining adverse drug reaction mentions in twitter with word embeddings. In : *Proceedings of the Social Media Mining Shared Task Workshop at the Pacific Symposium on Biocomputing* ; 2016. .
- [17] Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults : a systematic review. *Journal of medical Internet research.* 2013;15(2) :e35.
- [18] Mao JJ, Chung A, Benton A, Hill S, Ungar L, Leonard CE, et al. Online discussion of drug side effects and discontinuation among breast cancer survivors. *Pharmacoepidemiology and drug safety.* 2013;22(3) :256–262.

- [19] Chary M, Genes N, Giraud-Carrier C, Hanson C, Nelson LS, Manini AF. Epidemiology from Tweets : Estimating Misuse of Prescription Opioids in the USA from Social Media. *Journal of Medical Toxicology*. 2017 ;13(4) :278–286.
- [20] Topaz M, Radhakrishnan K, Lei V, Zhou L. Mining Clinicians’ Electronic Documentation to Identify Heart Failure Patients with Ineffective Self-Management : A Pilot Text-Mining Study. *Studies in health technology and informatics*. 2016 ;225 :856.
- [21] mondiale de la Santé O. Pharmacovigilance : assurer la sécurité d’emploi des médicaments. 2004 ;.
- [22] ; 2015 [cited World Health Organization. Geneva, Switzerland : World Health Organization English Glossary]. Available from : http://www.who.int/medicines/areas/coordination/English_Glossary.pdf.
- [23] Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al. ‘Global trigger tool’ shows that adverse events in hospitals may be ten times greater than previously measured. *Health affairs*. 2011 ;30(4) :581–589.
- [24] Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, et al. Adverse drug reactions as cause of admission to hospital : prospective analysis of 18 820 patients. *Bmj*. 2004 ;329(7456) :15–19.
- [25] Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients : a meta-analysis of prospective studies. *Jama*. 1998 ;279(15) :1200–1205.
- [26] Mitchell AS, Henry DA, Sanson-Fisher R, O’Connell DL. Patients as a direct source of information on adverse drug reactions. *Bmj*. 1988 ;297(6653) :891–893.
- [27] Medawar C, Herxheimer A. A comparison of adverse drug reaction reports from professionals and users, relating to risk of dependence and suicidal behaviour with paroxetine. *International Journal of Risk & Safety in Medicine*. 2004 ;16(1) :5–19.
- [28] Van Grootheest K, de Jong-van den Berg L. Patients’ role in reporting adverse drug reactions. *Expert opinion on drug safety*. 2004 ;3(4) :363–368.
- [29] Blenkinsopp A, Wilkie P, Wang M, Routledge PA. Patient reporting of suspected adverse drug reactions : a review of published literature and international experience. *British journal of clinical pharmacology*. 2007 ;63(2) :148–156.

- [30] van Hunsel F, Talsma A, van Puijenbroek E, de Jong-van den Berg L, van Grootheest K. The proportion of patient reports of suspected ADRs to signal detection in the Netherlands : case-control study. *pharmacoepidemiology and drug safety*. 2011 ;20(3) :286–291.
- [31] Hughes S, Cohen D. Can online consumers contribute to drug knowledge? A mixed-methods comparison of consumer-generated and professionally controlled psychotropic medication information on the internet. *Journal of medical Internet research*. 2011 ;13(3) :e53.
- [32] Margraff F, Bertram D. Adverse drug reaction reporting by patients : an overview of fifty countries. *Drug Safety*. 2014 ;37(6) :409–419.
- [33] Bates DW, Evans RS, Murff H, Stetson PD, Pizziferri L, Hripcsak G. Detecting adverse events using information technology. *Journal of the American Medical Informatics Association*. 2003 ;10(2) :115–128.
- [34] Tubert P, Bégaud B, Péré JC, Haramburu F, Lellouch J. Power and weakness of spontaneous reporting : a probabilistic approach. *Journal of clinical epidemiology*. 1992 ;45(3) :283–286.
- [35] Tubert P, Bégaud B. Random models for margins of a 2×2 contingency table and application to pharmacovigilance. *Statistics in medicine*. 1991 ;10(6) :991–999.
- [36] Chen ES, Hripcsak G, Xu H, Markatou M, Friedman C. Automated acquisition of disease-drug knowledge from biomedical and clinical documents : an initial study. *Journal of the American Medical Informatics Association*. 2008 ;15(1) :87–98.
- [37] Trifirò G, Pariente A, Coloma PM, Kors JA, Polimeni G, Miremont-Salamé G, et al. Data mining on electronic health record databases for signal detection in pharmacovigilance : which events to monitor? *Pharmacoepidemiology and drug safety*. 2009 ;18(12) :1176–1184.
- [38] Aramaki E, Miura Y, Tonoike M, Ohkuma T, Masuichi H, Waki K, et al. Extraction of adverse drug effects from clinical records. *Stud Health Technol Inform*. 2010 ;160(Pt 1) :739–43.
- [39] Doğan RI, Névél A, Lu Z. A context-blocks model for identifying clinical relationships in patient records. *BMC bioinformatics*. 2011 ;12(3) :S3.
- [40] Gurulingappa H, Mateen-Rajpu A, Toldo L. Extraction of potential adverse drug events from medical case reports. *Journal of biomedical semantics*. 2012 ;3(1) :15.

- [41] Warrer P, Hansen EH, Juhl-Jensen L, Aagaard L. Using text-mining techniques in electronic patient records to identify ADRs from medicine use. *British journal of clinical pharmacology*. 2012;73(5) :674–684.
- [42] Eriksson R, Jensen PB, Frankild S, Jensen LJ, Brunak S. Dictionary construction and identification of possible adverse drug events in Danish clinical narrative text. *Journal of the American Medical Informatics Association*. 2013;20(5) :947–953.
- [43] Li Y, Salmasian H, Vilar S, Chase H, Friedman C, Wei Y. A method for controlling complex confounding effects in the detection of adverse drug reactions using electronic health records. *Journal of the American Medical Informatics Association*. 2014;21(2) :308–314.
- [44] Micoulaud-Franchi JA. One step more toward pharmacovigilance 2.0. Integration of web data community for a pharmacovigilance more alert ; 2011.
- [45] Coloma PM, Becker B, Sturkenboom MC, Van Mulligen EM, Kors JA. Evaluating social media networks in medicines safety surveillance : two case studies. *Drug safety*. 2015;38(10) :921–930.
- [46] Sarker A, Nikfarjam A, Gonzalez G. Social media mining shared task workshop. In : *Proceedings of the Pacific Symposium on Biocomputing* ; 2016. p. 581–592.
- [47] Pierce CE, Bouri K, Pamer C, Proestel S, Rodriguez HW, Van Le H, et al. Evaluation of Facebook and Twitter Monitoring to Detect Safety Signals for Medical Products : An Analysis of Recent FDA Safety Alerts. *Drug Safety*. 2017;p. 1–15.
- [48] Comfort S, Perera S, Hudson Z, Dorrell D, Meireis S, Nagarajan M, et al. Sorting Through the Safety Data Haystack : Using Machine Learning to Identify Individual Case Safety Reports in Social-Digital Media. *Drug safety*. 2018;p. 1–12.
- [49] Liu J, Zhao S, Wang G. SSEL-ADE : A semi-supervised ensemble learning framework for extracting adverse drug events from social media. *Artificial intelligence in medicine*. 2017;.
- [50] Chen X, Deldossi M, Aboukhamis R, Faviez C, Dahamna B, Karapetiantz P, et al. Mining Adverse Drug Reactions in Social Media with Named Entity Recognition and Semantic Methods. *Studies in health technology and informatics*. 2017;245 :322–326.
- [51] Matsuda S, Aoki K, Tomizawa S, Sone M, Tanaka R, Kuriki H, et al. Analysis of patient narratives in disease blogs on the internet : an exploratory study of social pharmacovigilance. *JMIR public health and surveillance*. 2017;3(1).

- [52] Duh MS, Cremieux P, Audenrode MV, Vekeman F, Karner P, Zhang H, et al. Can social media data lead to earlier detection of drug-related adverse events? *Pharmacoepidemiology and drug safety*. 2016 ;25(12) :1425–1433.
- [53] Eshleman R, Singh R. Leveraging graph topology and semantic context for pharmacovigilance through twitter-streams. In : *BMC bioinformatics*. vol. 17. BioMed Central ; 2016. p. 335.
- [54] Korkontzelos I, Nikfarjam A, Shardlow M, Sarker A, Ananiadou S, Gonzalez GH. Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. *Journal of biomedical informatics*. 2016 ;62 :148–158.
- [55] Adrover C, Bodnar T, Huang Z, Telenti A, Salathé M. Identifying adverse effects of HIV drug treatment and associated sentiments using twitter. *JMIR public health and surveillance*. 2015 ;1(2).
- [56] Cocos A, Fiks AG, Masino AJ. Deep learning for pharmacovigilance : recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. *Journal of the American Medical Informatics Association*. 2017 ;24(4) :813–821.
- [57] Moncrieff J, Cohen D, Mason J. The subjective experience of taking antipsychotic medication : a content analysis of Internet data. *Acta Psychiatrica Scandinavica*. 2009 ;120(2) :102–111.
- [58] Pages A, Bondon-Guitton E, Montastruc JL, Bagheri H. Undesirable effects related to oral antineoplastic drugs : comparison between patients' internet narratives and a national pharmacovigilance database. *Drug safety*. 2014 ;37(8) :629–637.
- [59] Sarrazin MSV, Cram P, Mazur A, Ward M, Reisinger HS. Patient perspectives of dabigatran : analysis of online discussion forums. *The Patient-Patient-Centered Outcomes Research*. 2014 ;7(1) :47–54.
- [60] Hadzi-Puric J, Grmusa J. Automatic drug adverse reaction discovery from parenting websites using disproportionality methods. In : *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*. IEEE Computer Society ; 2012. p. 792–797.
- [61] Jiang Y, Liao QV, Cheng Q, Berlin RB, Schatz BR. Designing and evaluating a clustering system for organizing and integrating patient drug outcomes in personal health messages. In : *AMIA Annual Symposium Proceedings*. vol. 2012. American Medical Informatics Association ; 2012. p. 417.

- [62] Li YA. Medical data mining : Improving information accessibility using online patient drug reviews. Massachusetts Institute of Technology ; 2011.
- [63] Liu X, Chen H. AZDrugMiner : an information extraction system for mining patient-reported adverse drug events in online patient forums. In : International Conference on Smart Health. Springer ; 2013. p. 134–150.
- [64] Wu H, Fang H, Stanhope S, et al. Exploiting online discussions to discover unrecognized drug side effects. *Methods Inf Med.* 2013 ;52(2) :152–9.
- [65] Yang CC, Jiang L, Yang H, Tang X. Detecting signals of adverse drug reactions from health consumer contributed content in social media. In : Proceedings of ACM SIGKDD Workshop on Health Informatics; 2012. .
- [66] Yates A, Goharian N, Frieder O. Extracting adverse drug reactions from forum posts and linking them to drugs. In : Proceedings of the 2013 ACM SIGIR Workshop on Health Search and Discovery ; 2013. .
- [67] Bian J, Topaloglu U, Yu F. Towards large-scale twitter mining for drug-related adverse events. In : Proceedings of the 2012 international workshop on Smart health and wellbeing. ACM; 2012. p. 25–32.
- [68] Bodenreider O. The unified medical language system (UMLS) : integrating biomedical terminology. *Nucleic acids research.* 2004;32(suppl_1) :D267–D270.
- [69] Sampathkumar H, Chen Xw, Luo B. Mining adverse drug reactions from online healthcare forums using hidden Markov model. *BMC medical informatics and decision making.* 2014;14(1) :91.
- [70] Porter MF. Snowball : A language for stemming algorithms ; 2001.
- [71] Nikfarjam A, Gonzalez GH. Pattern mining for extraction of mentions of adverse drug reactions from user comments. In : AMIA Annu Symp Proc. vol. 2011 ; 2011. p. 1019–1026.
- [72] Yeleswarapu S, Rao A, Joseph T, Saipradeep VG, Srinivasan R. A pipeline to extract drug-adverse event pairs from multiple data sources. *BMC medical informatics and decision making.* 2014;14(1) :13.
- [73] Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic acids research.* 2015 ;44(D1) :D1075–D1079.
- [74] Food U, Administration D, et al. COSTART : coding symbols for thesaurus of adverse reaction terms. Publication NTIS PD. 1995 ;p. 90–114026.

- [75] Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). *Drug safety*. 1999 ;20(2) :109–117.
- [76] Yates A, Goharian N. ADRTrace : detecting expected and unexpected adverse drug reactions from user reviews on social media sites. In : *European Conference on Information Retrieval*. Springer ; 2013. p. 816–819.
- [77] MetaMap ; 2015. Available from : <https://metamap.nlm.nih.gov>.
- [78] Food U. Drug Administration Silver Spring. MD US FDA FDA Adverse Event Reporting System (FAERS) [http ://owly/4wkv100Efhc htm](http://owly/4wkv100Efhc.htm) 6Sfxrss3N. 2000 ;.
- [79] Zhou ZH. Ensemble Learning. *Encyclopedia of biometrics*. 2009 ;1 :270–273.
- [80] Rastegar-Mojarad M, Elayavilli RK, Yu Y, Liu H. Detecting signals in noisy data-can ensemble classifiers help identify adverse drug reaction in tweets. In : *Proceedings of the Social Media Mining Shared Task Workshop at the Pacific Symposium on Biocomputing* ; 2016. .
- [81] Zhang Z, Nie J, Zhang X. An ensemble method for binary classification of adverse drug reactions from social media. In : *Proceedings of the Social Media Mining Shared Task Workshop at the Pacific Symposium on Biocomputing* ; 2016. .
- [82] Ofoghi B, Siddiqui S, Verspoor K. Read-biomed-ss : Adverse drug reaction classification of microblogs using emotional and conceptual enrichment. In : *Proceedings of the Social Media Mining Shared Task Workshop at the Pacific Symposium on Biocomputing* ; 2016. .
- [83] Jonnagaddala J, Jue TR, Dai H. Binary classification of Twitter posts for adverse drug reactions. In : *Proceedings of the Social Media Mining Shared Task Workshop at the Pacific Symposium on Biocomputing, Big Island, HI, USA* ; 2016. p. 4–8.
- [84] Egger D, Uzdilli F, Cieliebak M, Derczynski L. Adverse drug reaction detection using an adapted sentiment classifier. In : *Proceedings of the Social Media Mining Shared Task Workshop at the Pacific Symposium on Biocomputing* ; 2016. .
- [85] Wang C, Singh O, Dai H, Jonnagaddala J, Jue TR, Iqbal U, et al. NTTMUNSW system for adverse drug reactions extraction in Twitter data. In : *Proceedings of the Social Media Mining Shared Task Workshop at the Pacific Symposium on Biocomputing, Big Island, HI, USA* ; 2016. p. 4–8.
- [86] Golder S, Norman G, Loke YK. Systematic review on the prevalence, frequency and comparative value of adverse events data in social media. *British journal of clinical pharmacology*. 2015 ;80(4) :878–888.

- [87] Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance : A review. *Journal of biomedical informatics*. 2015 ;54 :202–212.
- [88] Liu X, Chen H. A research framework for pharmacovigilance in health social media : Identification and evaluation of patient adverse drug event reports. *Journal of biomedical informatics*. 2015 ;58 :268–279.
- [89] Powell GE, Seifert HA, Reblin T, Burstein PJ, Blowers J, Menius JA, et al. Social media listening for routine post-marketing safety surveillance. *Drug safety*. 2016 ;39(5) :443–454.
- [90] Blei DM. Probabilistic topic models. *Communications of the ACM*. 2012 ;55(4) :77–84.
- [91] Yang M, Kiang M, Shang W. Filtering big data from social media—Building an early warning system for adverse drug reactions. *Journal of biomedical informatics*. 2015 ;54 :230–240.
- [92] Canini KR, Shi L, Griffiths TL. Online Inference of Topics with Latent Dirichlet Allocation. In : *AISTATS*. vol. 9 ; 2009. p. 65–72.
- [93] Sarker A, Gonzalez G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *Journal of biomedical informatics*. 2015 ;53 :196–207.
- [94] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *Journal of machine Learning research*. 2003 ;3(Jan) :993–1022.
- [95] Yıldırım A, Üsküdarlı S, Özgür A. Identifying topics in microblogs using Wikipedia. *PloS one*. 2016 ;11(3) :e0151885.
- [96] Prier KW, Smith MS, Giraud-Carrier C, Hanson CL. Identifying health-related topics on twitter. In : *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer ; 2011. p. 18–25.
- [97] Paul MJ, Dredze M. A model for mining public health topics from Twitter. *Health*. 2012 ;11 :16–6.
- [98] Ghosh D, Guha R. What are we "tweeting" about obesity ? Mapping tweets with topic modeling and Geographic Information System. *Cartography and geographic information science*. 2013 ;40(2) :90–102.

- [99] Sullivan R, Sarker A, O'CONNOR K, Goodin A, Karlsrud M, Gonzalez G. Finding potentially unsafe nutritional supplements from user reviews with topic modeling. In : Biocomputing 2016 : Proceedings of the Pacific Symposium. World Scientific ; 2016. p. 528–539.
- [100] Tapi MN, Bringay S, Lavergne C, Mollevi C, Opitz T. What Patients Can Tell Us : Topic Analysis for Social Media on Breast Cancer. JMIR medical informatics. 2017;5(3) :e23–e23.
- [101] Paul MJ, Dredze M. You are what you Tweet : Analyzing Twitter for public health. Icwsm. 2011;20 :265–272.
- [102] Paul MJ, Dredze M. Discovering health topics in social media using topic models. PloS one. 2014;9(8) :e103408.
- [103] Abbé A. Analyse de données textuelles d'un forum médical pour évaluer le ressenti exprimé par les internautes au sujet des antidépresseurs et des anxyolitiques. Paris Saclay ; 2016.
- [104] Abdellaoui R, Schück S, Texier N, Burgun A. Filtering entities to optimize ADR identification from social media : how can the number of words between entities in the messages help ? JMIR Public Health Surveill. 2017;.
- [105] Schück S, Kürzinger M, Abdellaoui R, Texier N, Pouget J, Faviez C, et al. Comparison Of Web-Based Signal Detection Using Medical Forums Data In France From 2005-2015 With Signals From Vigibase®. Value in Health. 2016;19(7) :A474.
- [106] Katsahian S, Simond ME, Leprovost D, Lardon J, Bousquet C, Kerdelhué G, et al. Evaluation of Internet Social Networks using Net scoring Tool : A Case Study in Adverse Drug Reaction Mining. Studies in health technology and informatics. 2014;210 :526–530.
- [107] Audeh B, Beigbeder M, Zimmermann A, Jaillon P, Bousquet C. Vigi4Med Scraper : A Framework for Web Forum Structured Data Extraction and Semantic Representation. PloS one. 2017;12(1) :e0169658.
- [108] Gupta V, Lehal GS, et al. A survey of text mining techniques and applications. Journal of emerging technologies in web intelligence. 2009;1(1) :60–76.
- [109] Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance : A review. Journal of biomedical informatics. 2015;54 :202–212.

- [110] Feinerer I. Introduction to the tm Package Text Mining in R ; 2015.
- [111] Porter MF. An algorithm for suffix stripping. *Program*. 1980 ;14(3) :130–137.
- [112] Savoy J. Light stemming approaches for the French, Portuguese, German and Hungarian languages. In : *Proceedings of the 2006 ACM symposium on Applied computing*. ACM ; 2006. p. 1031–1035.
- [113] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society Series B (methodological)*. 1977 ;p. 1–38.
- [114] Fraley C, Raftery AE, et al. Model-based methods of classification : using the mclust software in chemometrics. *Journal of Statistical Software*. 2007 ;18(6) :1–13.
- [115] Cortes C, Vapnik V. Support-vector networks. *Machine learning*. 1995 ;20(3) :273–297.
- [116] Hasan M, Boris F. Svm : Machines à vecteurs de support ou séparateurs à vastes marges. *Rapport technique*, Versailles St Quentin, France Cité. 2006 ;p. 64.
- [117] Hasan M, Boris F. Svm : Machines à vecteurs de support ou séparateurs à vastes marges. *Rapport technique*, Versailles St Quentin, France Cité. 2006 ;p. 64.
- [118] Hornik K, Meyer D, Karatzoglou A. Support vector machines in R. *Journal of statistical software*. 2006 ;15(9) :1–28.
- [119] Team RC, et al. R : A language and environment for statistical computing. 2013 ;.
- [120] Choi B. A Graphical Method to Assess Goodness-of-Fit for Inverse Gaussian Distribution. *Korean Journal of Applied Statistics*. 2013 ;26(1) :37–47.
- [121] Grego JM, Yates PA. Point and standard error estimation for quantiles of mixed flood distributions. *Journal of hydrology*. 2010 ;391(3) :289–301.
- [122] Torres-Carrasquillo PA, Reynolds DA, Deller JR. Language identification using Gaussian mixture model tokenization. In : *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*. vol. 1. IEEE ; 2002. p. I–757.
- [123] Torres-Carrasquillo PA, Singer E, Kohler MA, Greene RJ, Reynolds DA, Deller Jr JR. Approaches to language identification using Gaussian mixture models and shifted delta cepstral features. In : *Interspeech* ; 2002. .

- [124] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society Series B (methodological)*. 1977;p. 1–38.
- [125] Abdellaoui R, Foulquié P, Texier N, Faviez C, Burgun A, Schück S. Detection of Cases of Noncompliance to Drug Treatment in Patient Forum Posts : Topic Model Approach. *Journal of medical Internet research*. 2018;20(3).
- [126] McHorney CA, Spain CV. Frequency of and reasons for medication non-fulfillment and non-persistence among American adults with chronic disease in 2008. *Health Expectations*. 2011;14(3) :307–320.
- [127] Iavindrasana J, Cohen G, Depeursinge A, Müller H, Meyer R, Geissbuhler A. Clinical data mining : a review. *Yearbook of medical informatics*. 2009;18(01) :121–133.
- [128] Drazic M, Kukolj D, Vitas M, Pokric M, Manojlovic S, Tekic Z. Technology matching of the patent documents using clustering algorithms. In : *Computational Intelligence and Informatics (CINTI), 2013 IEEE 14th International Symposium on*. Ieee; 2013. p. 405–409.
- [129] Stevens K, Kegelmeyer P, Andrzejewski D, Buttler D. Exploring topic coherence over many models and many topics. In : *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Association for Computational Linguistics; 2012. p. 952–961.
- [130] Blei DM, Lafferty JD. Topic models. *Text mining : classification, clustering, and applications*. 2009;10(71) :34.
- [131] Taddy M. On Estimation and Selection for Topic Models. In : *AISTATS*; 2012. p. 1184–1193.
- [132] Sievert C, Shirley K. LDAvis : A method for visualizing and interpreting topics. In : *Proceedings of the workshop on interactive language learning, visualization, and interfaces*; 2014. p. 63–70.
- [133] Cohen J. A coefficient of agreement for nominal scales. *Educational and psychological measurement*. 1960;20(1) :37–46.
- [134] Roberts ME, Stewart BM, Tingley D. STM : R package for structural topic models, 2014. URL <http://www.structuraltopicmodel.com> R package version. 2016;1(8).

- [135] Roberts ME, Stewart BM, Tingley D, Airolidi EM, et al. The structural topic model and applied social science. In : *Advances in Neural Information Processing Systems Workshop on Topic Models : Computation, Application, and Evaluation* ; 2013. .
- [136] Roberts ME, Stewart BM, Tingley D, Lucas C, Leder-Luis J, Gadarian SK, et al. Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*. 2014;58(4) :1064–1082.
- [137] Lardon J, Bellet F, Aboukhamis R, Asfari H, Souvignet J, Jaulent MC, et al. Evaluating Twitter as a complementary data source for pharmacovigilance. *Expert opinion on drug safety*. 2018;17(8) :763–774.
- [138] Izem R, Sanchez-Kam M, Ma H, Zink R, Zhao Y. Sources of safety data and statistical strategies for design and analysis : postmarket surveillance. *Therapeutic innovation & regulatory science*. 2018;52(2) :159–169.
- [139] Abou Taam M, Rossard C, Cantaloube L, Bouscaren N, Roche G, Pochard L, et al. Analysis of patients' narratives posted on social media websites on benfluorex's (Mediator®) withdrawal in France. *Journal of clinical pharmacy and therapeutics*. 2014;39(1) :53–55.
- [140] Pariente A, Gregoire F, Fourrier-Reglat A, Haramburu F, Moore N. Impact of safety alerts on measures of disproportionality in spontaneous reporting databases the notoriety bias. *Drug safety*. 2007;30(10) :891–898.
- [141] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*. 2016;3.