



HAL
open science

Analyse des productions dans les médias sociaux : vers une évaluation de la personnalité des patients

Mourad Ellouze

► **To cite this version:**

Mourad Ellouze. Analyse des productions dans les médias sociaux : vers une évaluation de la personnalité des patients. Informatique [cs]. Faculté des Sciences Economiques et de Gestion de Sfax; Université de Sfax (Tunisie), 2023. Français. NNT: . tel-04376647

HAL Id: tel-04376647

<https://theses.hal.science/tel-04376647>

Submitted on 9 Jan 2024

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.

Copyright

THÈSE DE DOCTORAT

POUR L'OBTENTION DU TITRE DE DOCTEUR EN
INFORMATIQUE

*Analyse des productions dans les
médias sociaux : vers une évaluation
de la personnalité des patients*

Présentée et soutenue publiquement par :

Mourad ELLOUZE

Le 28/12/2023

Membres du jury :

M. Ahmed HADJ KACEM	Professeur d'Ens. Sup - FSEG - Sfax	Président
Mme. Lamia HADRICHE BELGUITH	Professeur d'Ens. Sup - FSEG - Sfax	Directeur de thèse
M. Mohamed Adel ALIMI	Professeur d'Ens. Sup - ENI - Sfax	Rapporteur
M. Bilel ELAYEB	Maître de Conférences - ENSI - Tunis	Rapporteur
M. Bilel GARGOURI	Maître de Conférences - FSEG - Sfax	Membre

Année universitaire : 2022-2023

REMERCIEMENTS

D'abord, c'est grâce à Dieu que j'ai pu accomplir ce travail. Je tiens à remercier tous ceux qui m'ont prêté main-forte pour l'élaboration de ce travail. J'adresse mes remerciements, bien sûr, en premier lieu, à ma directrice de thèse, Madame Lamia Hadrich Belguith, Professeur à la Faculté des Sciences Économiques et de Gestion de Sfax (FSEGS) et directrice de l'équipe de recherche ANLPRG du laboratoire MIRACL pour sa rigueur, sa confiance, son soutien et son encouragement tout au long de mes recherches. Je remercie aussi mon co-encadrant Monsieur Seifeddine Mechti, Maître Assistant à l'Institut Supérieur des Sciences Appliquées et de Technologie de Kairouan (ISSAT), pour son aide précieuse et sa contribution permanente dans l'évaluation de ce travail. Je leur suis très reconnaissant pour toutes les discussions fructueuses que j'ai eu avec eux, pour leurs conseils judicieux et aussi pour les corrections qu'ils ont apportées au présent travail. J'ai le plaisir d'adresser les vifs remerciements à l'ensemble des membres de jury qui me font le grand honneur d'avoir accepté de juger mon travail. Je tiens à leur témoigner mon respect et chaleureux remerciements.

Je suis très honoré et reconnaissant à Monsieur Ahmed Hadj Kacem, Professeur à la Faculté des Sciences Économiques et de Gestion de Sfax, pour avoir accepté de présider le jury. Je tiens aussi à remercier les deux rapporteurs de ce travail, Monsieur Mohamed Adel Alimi, Professeur à l'École Nationale d'Ingénieurs de Sfax (ENIS), et Monsieur Bilel Elayeb, Maître de Conférence à l'École nationale des sciences de l'informatique de Tunis (ENSI Manouba), pour le temps et l'attention qu'ils ont consacrés pour rapporter cette thèse. Par la même occasion, je tiens à remercier Monsieur Bilel Gargouri, Maître de Conférence à la Faculté des Sciences Économiques et de Gestion de Sfax, pour avoir fait l'honneur d'examiner ce travail.

Je remercie les membres du laboratoire MIRACL (Multimedia, InfoRmation system and Advanced Computing Laboratory - Sfax) que j'ai pu côtoyer durant la période de ma thèse pour leur soutien et leur encouragement. Enfin, je remercie ma famille qui m'a apporté le soutien dont j'avais besoin pour accomplir ce travail. Je remercie mes parents et ma soeur pour leur confiances en moi et leur encouragements. Papa, Maman, Mariem, merci d'avoir tout fait pour que j'en sois arrivée là aujourd'hui.

Sfax, le 29 novembre 2023.

TABLE DES MATIÈRES

TABLE DES MATIÈRES	iv
LISTE DES FIGURES	viii
LISTE DES TABLEAUX	x
INTRODUCTION GÉNÉRALE	1
1 LES CONCEPTS DE BASE DE LA DÉTECTION DE LA PERSONNALITÉ	6
1.1 INTRODUCTION	7
1.2 QU'EST CE QU'UNE PERSONNALITÉ?	7
1.3 QU'EST CE QU'UN TROUBLE DE PERSONNALITÉ?	8
1.4 LES CARACTÉRISTIQUES DE LA PERSONNALITÉ	9
1.4.1 Tempérament	9
1.4.2 Caractère	9
1.4.3 Cognitive	10
1.4.4 Humaniste	12
1.4.5 Psychobiologique	13
1.5 LES TROUBLES DE PERSONNALITÉS EN TUNISIE	13
1.6 MOTIVATIONS POUR L'AUTOMATISATION DE LA DÉTECTION ET DU SUIVI DES PERSONNES AVEC DES TP	14
1.7 LES PRINCIPAUX TRAITS STYLISTIQUES EMPLOYÉS POUR LA DÉTECTION DE LA PERSONNALITÉ	14
1.7.1 Traits lexicaux	15
1.7.2 Traits Morpho-Syntaxiques	15
1.7.3 Traits Sémantiques	16
1.7.4 Traits Numeriques	16
1.8 LE RÔLE DE L'INTELLIGENCE ARTIFICIELLE DANS LA DÉTECTION DE LA PERSONNALITÉ	17
1.8.1 Les outils de TALN	19
1.8.2 Les ontologies	19
1.8.3 Les techniques de fouille de données et de fouille de textes	20
1.8.4 Apprentissage profond	22
1.8.5 Modèle de langage	23
1.8.6 Les techniques de traitement de données massives	24
1.8.7 Vers le Web sémantique 3.0	25
1.9 CONCLUSION	26

2	ÉTAT DE L'ART DES MÉTHODES D'ANALYSE DE LA PERSONNALITÉ SUR LES RÉSEAUX SOCIAUX	27
2.1	INTRODUCTION	28
2.2	DÉTECTION DES TRAITS DE PERSONNALITÉ	30
2.2.1	Méthodes basées sur les techniques classiques d'apprentissage automatique	30
2.2.2	Méthodes basées sur l'apprentissage profond	31
2.2.3	Méthodes basées sur les techniques de traitement linguistiques	33
2.2.4	Méthodes basées sur des techniques statistiques	34
2.2.5	Méthodes hybrides	34
2.3	DÉTECTION DES TROUBLES DE PERSONNALITÉ	35
2.3.1	Manière explicite	37
2.3.2	Manière implicite	42
2.4	DÉTECTION DES LIENS ENTRE LES ÉLÉMENTS DE LA PERSONNALITÉ	43
2.5	SURVEILLANCE DE L'ÉTAT DU PATIENT	46
2.6	DISCUSSION	47
2.6.1	Détection des traits de personnalité	47
2.6.2	Détection des troubles de personnalité	48
2.7	CONCLUSION	49
3	DÉTECTION DES PERSONNES SOUFFRANT DE TROUBLES DE PERSONNALITÉ	51
3.1	INTRODUCTION	52
3.2	COLLECTE DU CORPUS POUR LA DÉTECTION DES UTILISATEURS AYANT DES TP	53
3.3	MÉTHODE NEURONALE PROPOSÉE POUR LA DÉTECTION DES PERSONNES AYANT DES TP	55
3.3.1	Prétraitement du corpus collecté	56
3.3.2	Génération des caractéristiques	57
3.3.3	Réseau de neurones convolutifs pour l'extraction des caractéristiques	60
3.3.4	Génération synthétique des données observées	60
3.3.5	Capture des dépendances à long terme entre les unités lexicales pour la tâche de classification des symptômes de TP	62
3.3.6	Classification binaire pour la tâche de détection de trouble paranoïde	63
3.4	MÉTHODE PROPOSÉE BASÉE SUR LES TECHNIQUES D'APPRENTISSAGE CLASSIQUES	64
3.4.1	Prétraitement	65
3.4.2	Extraction des caractéristiques	66
3.4.3	Réduction de la dimensionnalité	69
3.4.4	Classification basée sur la notion de logique floue pour la détection des personnes ayant des TP	70
3.5	CONCLUSION	73
4	IDENTIFICATION DES CAUSES AFFECTANT LES TROUBLES DE LA PERSONNALITÉ	75
4.1	INTRODUCTION	76
4.2	CORPUS UTILISÉ	77
4.3	GESTION DES SPÉCIFICITÉS DES MULTI-VARIATIONS DES DOMAINES	78

4.4	IDENTIFICATION DES SUJETS DES TWEETS	80
4.4.1	Prétraitement des données et traitement des caractéristiques	80
4.4.2	Regroupement des sujets d'étude	81
4.5	CONSTRUCTION AUTOMATIQUE D'ONTOLOGIES DE DOMAINES	84
4.5.1	Sélection, normalisation et enrichissement des concepts	85
4.5.2	Extraction des relations sémantiques	86
4.5.3	Attribution des propriétés aux données extraites	88
4.5.4	Affectation des individus aux données extraites	88
4.6	CONSTRUCTION AUTOMATIQUE DES RÈGLES D'INTERROGATION L'ONTOLOGIE	90
4.6.1	Construction des règles	90
4.6.2	Interrogation de l'ontologie	92
4.7	CONCLUSION	94
5	SURVEILLANCE DES PERSONNES ATTEINTES DU TROUBLE DE LA PERSONNALITÉ	95
5.1	INTRODUCTION	96
5.2	CORPUS UTILISÉ	97
5.3	MÉTHODE PROPOSÉE POUR SURVEILLER L'ÉTAT DES PERSONNES AYANT UN TP	99
5.3.1	Méthode basée sur des techniques d'apprentissage supervisé visant à détecter les personnes atteintes de troubles de la personnalité	100
5.3.2	Méthode fondée sur des techniques d'apprentissage supervisé	101
5.3.3	Méthode basée sur les ontologies	103
5.3.4	Méthode basée sur l'apprentissage profond	107
5.4	MÉTHODE POUR SURVEILLER LES UTILISATEURS	112
5.5	CONCLUSION	114
6	EXPÉRIMENTATIONS ET DISCUSSION DES RÉSULTATS OBTENUS	116
6.1	INTRODUCTION	118
6.2	ENVIRONNEMENT DE PROGRAMMATION	118
6.3	CORPUS DE RÉFÉRENCE	119
6.3.1	Dataset pour l'analyse de la dépression sur Twitter Infamouscoder [2022b]	119
6.3.2	Dataset pour l'analyse du suicide et de la dépression Komati [2021]	120
6.3.3	Dataset pour l'analyse de la dépression sur Reddit Infamouscoder [2022a]	120
6.3.4	Dataset pour l'analyse de la santé mentale Namdari [2023]	120
6.3.5	Dataset pour l'analyse du comportement anti-social sur Twitter pour le public indonésien Astuti [2021]	121
6.3.6	Dataset provenant d'un groupe de témoignage sur Facebook : Trouble Personnalité Borderline	121
6.4	VALIDATION DU CORPUS COLLECTÉ POUR LA DÉTECTION DES PERSONNES AYANT DES TP	122
6.5	DÉTECTION DES PERSONNES SOUFFRANT DES TROUBLES DE LA PERSONNALITÉ SUR LES RÉSEAUX SOCIAUX	125
6.5.1	Résultats des méthodes de détection des troubles	125

6.5.2	Évaluation des méthodes de détection des troubles	127
6.5.3	Discussion des méthodes de détection des troubles	130
6.6	IDENTIFICATION DES CAUSES AFFECTANT LA MALADIE DE TP	132
6.6.1	Résultats de la méthode d'identification des causes du TP	133
6.6.2	Évaluation de la méthode d'identification des causes du TP	136
6.6.3	Discussion de la méthode d'identification des causes du TP	137
6.7	SURVEILLANCE DES PERSONNES ATTEINTES DU TP SUR LES RÉSEAUX SOCIAUX : ÉTUDE DE CAS COVID-19	138
6.7.1	Résultats de la méthode de surveillance des cas des personnes ayant un trouble	138
6.7.2	Évaluation de la méthode de surveillance des cas des personnes ayant un trouble	139
6.7.3	Discussion de la méthode de surveillance des cas des personnes ayant un TP	142
6.8	CONCLUSION	147
	MES CONTRIBUTIONS SCIENTIFIQUES	149
	CONCLUSION GÉNÉRALE	150
	BIBLIOGRAPHIE	153

LISTE DES FIGURES

1.1	Les différents domaines de l'intelligence artificielle De Ganay et Gilot [2017]	18
1.2	Processus KDD dans l'exploration de données Oprean [2010]	21
1.3	Chaîne de valeur du Big Data : de la collecte à l'exploitation des données Bahri et al. [2018]	24
2.1	Synthèse des méthodes d'analyse des traits et des troubles de la personnalité.	29
2.2	Diagramme du modèle MLP pour découvrir les traits de la personnalité An et Levitan [2018]	32
2.3	Présentation du diagramme du modèle WE-LSTM An et Levitan [2018]	32
2.4	Extrait de l'ontologie proposé en vue de la détection des traits de personnalité Alamsyah et al. [2020]	33
2.5	Architecture générale pour la détection du stress Lin et al. [2017a]	38
2.6	Diagramme du modèle MLP pour l'apprentissage multitâches : prédire conjointement la personnalité et la déception An et al. [2018]	39
2.7	Diagramme du modèle MLP pour l'apprentissage multitâches : alimentation des étiquettes de personnalité dans la couche de sortie An et al. [2018]	40
2.8	Architecture basée sur <i>Tree-LSTM network</i> pour la détection des émotions Wang et al. [2019a]	43
3.1	Méthode proposée pour la détection de la maladie paranoïaque au moyen de la détection de ses symptômes.	56
3.2	Proposition de modèle d'apprentissage profond CNN-BiLSTM-SVM pour la détection de la maladie paranoïde ainsi que ses symptômes.	56
3.3	Architecture générale du réseau de neurones à convolution appliqué à une image Haque et al. [2020]	61
3.4	Illustration du processus de génération d'instances synthétiques à l'aide de la méthode SMOTE Soufan et al. [2015]	61
3.5	Extrait d'une méthode basée sur le réseau de neurones BiLSTM Gao et al. [2018]	63
3.6	Proposition de modèle CNN-LSTM profond pour la vérification de l'existence de la maladie, de sa catégorie et de son type.	65
3.7	Méthode fondée sur la notion de logique floue pour détecter les personnes souffrant de TP sur Twitter.	66

4.1	Méthode hybride basée sur des techniques linguistiques et statistiques pour la détection des causes provoquant une situation instable.	79
4.2	Différents types de méthodes d'apprentissage non supervisée Tyagi [2018] .	82
4.3	Exemple de résultat d'une méthode d'apprentissage non supervisée à base de partitionnement Gitansh Chadha et Karnin [2018] .	83
4.4	Exemple de résultat d'une méthode d'apprentissage hiérarchique non supervisée Randriamihamison [2021] .	84
4.5	Exemple de résultat d'une méthode d'apprentissage non supervisée fondée sur la densité Campello et al. [2020] .	84
4.6	Aperçu des exemples liés à la typologie des problèmes humains dans le monde.	85
4.7	les trois principaux composants d'un concept Ogden et Richards [1923] .	85
5.1	Méthode hybride pour assurer la surveillance des personnes souffrant de TP en filtrant et en recommandant des tweets en fonction de leurs états envers COVID-19.	100
5.2	Exemple d'entraînement d'un modèle d'intégration de phrases Reimers et Gurevych [2019] .	108
5.3	Réseau neuronal ConvAE-LSTM-SVM pour la tâche de filtrage ou de recommandation des tweets.	109
5.4	Architecture de base d'un auto-encodeur monocouche Sublime et Kalinicheva [2019] .	110
5.5	Méthode proposée pour la détection et le suivi des personnes souffrant de TP.	114
6.1	Architecture du modèle proposé.	127
6.2	Extrait du résultat de l'application récursive de notre méthode d'apprentissage automatique non supervisée.	133
6.3	Notre ontologie élaborée pour le sujet Covid-19.	136
6.4	Comparaison des différentes méthodes utilisées pour intégrer les résultats des différents modèles de langage.	143
6.5	Évaluation des performances des différentes méthodes pour le filtrage des tweets inappropriés.	144

LISTE DES TABLEAUX

1.1	Comparaison des signes de différents types de troubles de la personnalité pour le groupe A selon DSM-5.	10
1.2	Comparaison des signes de différents types de troubles de la personnalité pour le groupe B selon DSM-5.	11
1.3	Comparaison des signes de différents types de troubles de la personnalité pour le groupe C selon DSM-5.	12
1.4	Exemple des traits numériques Saha et al. [2021]	17
1.5	Les points de différence entre l'apprentissage classique et l'apprentissage profond selon Data Flair.	23
2.1	La comparaison de différents travaux de détection de la personnalité.	36
2.2	Comparaison de divers travaux sur la détection de troubles psychologiques.	44
2.3	Les critères de comparaison entre les différentes travaux de détection des troubles psychologiques.	50
3.1	La répartition des instances par classe.	54
3.2	Distribution des instances pour la classification des symptômes.	55
3.3	Comparaison des différents modèles de <i>sentence embedding</i> Das et al. [2021]	59
3.4	Comparaison des types de réseaux de neurones (MLP (ANN) vs. RNN vs. CNN) thetechplatform [2021]	62
3.5	Les caractéristiques des métadonnées	68
3.6	Description des caractéristiques du plan temporel	68
3.7	Extraction des caractéristiques liées au style d'écriture	69
3.8	Avantages et inconvénients des différents modèles d'algorithmes de machine learning Liu et Lang [2019]	71
4.1	La distribution de notre corpus au cours de chaque étape.	78
4.2	Le principe de fonctionnement de chaque méthode d'apprentissage non supervisé Tyagi [2018]	83
4.3	Exemples d'instances extraites pour chacun des concepts en fonction de la ressource utilisée	89
4.4	Exemples d'adjectifs extraits associés à chaque concept indiquant le type d'humeur exprimée	90
4.5	Comparaison des terminologies relatives à RST et ARM Gardiner et Gillet [2015]	91

5.1	Exemple de tweets indiquant un éventuel comportement envers COVID-19.	98
5.2	La répartition de notre corpus par étape.	99
5.3	Les critères de comparaison entre les méthodes numériques et linguistiques Maâloul [2012]	107
6.1	Liste des principales bibliothèques utilisées de python.	119
6.2	Aperçu général sur la taille du corpus collecté à partir Facebook. . .	121
6.3	Aperçu général sur les corpus de références sélectionnés.	122
6.4	Analyse du style d'écriture pour les différents corpus.	123
6.5	Comparaison des résultats de notre analyse du style d'écriture avec les résultats d'ASHA.	124
6.6	Exemple de détection de degrés de symptômes liés à la maladie paranoïaque.	126
6.7	Comparaison des méthodes d'apprentissage profond Mathew et al. [2020]	128
6.8	Résultats de la détection du trouble, des catégories et des types de troubles.	129
6.9	Résultats de classification du trouble paranoïaque en se basant sur la notion de logique floue.	130
6.10	Comparaison des résultats obtenus pour la classification des symptômes liés à la maladie paranoïaque à l'aide du modèle CNN+BiLSTM.	131
6.11	Comparaison des résultats obtenus pour la détection de la maladie paranoïaque sur la base des symptômes.	131
6.12	Comparaison des résultats obtenus pour la classification des TP. . .	132
6.13	Comparaison des résultats obtenus pour la classification des TP selon la catégorie suspecte.	132
6.14	Comparaison des résultats obtenus pour la classification des TP selon la catégorie émotionnelle.	133
6.15	Comparaison des résultats obtenus pour la classification des TP selon la catégorie anxieuse.	133
6.16	Comparaison des résultats obtenus pour la détection des types de TP à l'aide du modèle LSTM+SVM.	134
6.17	Comparaison des résultats obtenus pour la détection des types de TP à l'aide du modèle BiLSTM+SVM.	134
6.18	Comparaison des résultats obtenus pour la détection du trouble paranoïaque (sans ou avec le passage par la détection des symptômes).	134
6.19	Comparaison des résultats obtenus pour la détection du trouble paranoïaque en se basant sur la combinaison des type d'attributs.	135
6.20	Comparaison des résultats obtenus par notre méthode et la méthode <i>baseline</i> pour la classification des personnes paranoïaques.	135
6.21	Évaluation de notre méthode par rapport à d'autres méthodes proposées pour la détection des personnes ayant des TP.	135

6.22	Évaluation des diverses méthodes proposées pour la détection du trouble paranoïaque.	135
6.23	Évaluation de la tâche d'apprentissage récurrent non supervisée.	137
6.24	Évaluation de la tâche d'identification du sujet et d'humeur.	137
6.25	Résultats de la partie de détection du comportement des personnes envers Covid-19.	138
6.26	Résultats de la tâche de filtrage des tweets inappropriés en fonction des préoccupations du malade.	139
6.27	Résultats de la tâche de recommandation des tweets appropriés en fonction des préoccupations du malade.	140
6.28	Comparaison des résultats obtenus pour la classification du comportement lié à la crise COVID-19.	141
6.29	Comparaison des résultats obtenus par notre méthode et la méthode baseline pour la classification du comportement.	141
6.30	Comparaison des résultats obtenus pour le filtrage des tweets inappropriés sur la base d'un auto-encodeur convolutif profond.	142
6.31	Comparaison des résultats obtenus pour le filtrage des tweets inappropriés sur la base d'un auto-encodeur profond.	142
6.32	Comparaison des résultats obtenus pour la recommandation des tweets appropriés sur la base d'un auto-encodeur convolutif profond.	144
6.33	Comparaison des résultats obtenus pour la recommandation des tweets appropriés sur la base d'un auto-encodeur profond.	145
6.34	Analyse comparative de la qualité des réponses fournies par notre chatbot et chatGPT.	146

INTRODUCTION GÉNÉRALE

L'analyse des données des réseaux sociaux est devenue de plus en plus cruciale pour la compréhension des besoins et des préférences des individus à l'échelle mondiale. En effet, ces plateformes servent à exprimer des opinions, des sentiments et des expériences personnelles relatives aux utilisateurs. Ces données peuvent être analysées pour identifier les tendances émergentes et les changements dans les attitudes et les comportements des gens.

En outre, l'analyse des données des réseaux sociaux peut également être employée pour surveiller l'état psychologique des individus, notamment leur niveau de stress, d'anxiété ou de dépression. Cette surveillance peut être particulièrement utile pour les entreprises qui offrent des services liés à la santé mentale de ses employés ou pour les organisations qui travaillent dans des secteurs sensibles, tels que la prévention du suicide ou la lutte contre les abus. En résumé, travailler sur les données des réseaux sociaux est important pour aider les spécialistes (entreprises et organisations) à développer des services plus efficaces et à améliorer la qualité de vie des gens.

Les problèmes psychologiques peuvent clairement affecter à la fois les individus et l'état du pays. Selon l'Organisation Mondiale de la Santé (OMS), la prévalence européenne des troubles mentaux est considérablement élevée OMS [2016]. Dans ce contexte, cette organisation déclare que sur les 880 millions d'habitants de la région européenne, il est estimé qu'environ 100 millions de personnes souffrent d'anxiété et de dépression ; plus de 21 millions de personnes atteintes de troubles liés à l'alcool ; plus de 7 millions de personnes souffrant de la maladie d'alzheimer et d'autres démences ; à peu près 4 millions de personnes schizophréniques ; 4 millions de personnes ont des troubles affectifs et bipolaires ; et 4 millions de personnes en état de panique¹.

Selon les estimations de l'OMS, le taux de dépression et de troubles anxieux coûtent annuellement 1000 milliards de dollars (USD) à l'économie mondiale². Cependant, le retour sur cet investissement est nettement supérieur à ces dépenses. En effet, les problèmes que ces patients peuvent provoquer dans le monde entier, en termes de perte de productivité, de coûts sociaux et de pertes de revenus fiscaux, peuvent freiner la croissance économique et restreindre le potentiel de développement. Malgré cela, les concepts liés à la santé mentale sont souvent négligés dans les politiques de santé et les budgets alloués à la santé. Dans ce contexte, le rapport de cette estimation met en évidence que les gouvernements consacrent en moyenne 3% de leur budget au secteur de la santé lié à ce domaine. Pour cette

1. https://www.euro.who.int/_data/assets/pdf_file/0008/98918/E88538.pdf

2. <https://www.who.int/fr/news/item/13-04-2016-investing-in-treatment-for-depression-and-anxiety-leads-to-fourfold-return>

raison, le rapport souligne la nécessité d'investir plus dans la santé mentale, tant au niveau national qu'international.

Les réseaux sociaux peuvent offrir un moyen de connecter les gens, mais ils peuvent également contribuer à l'isolement social et à la dépression. Les utilisateurs de réseaux sociaux qui sont confrontés à des problèmes de santé mentale peuvent être plus vulnérables aux effets négatifs de ces plateformes, tels que le cyberharcèlement et la propagation de fausses informations. Par exemple, une étude publiée dans la revue internationale de recherche environnementale et de santé publique (*International journal of environmental research and public health*) a révélé que les personnes atteintes de troubles de la personnalité narcissique ont tendance à passer plus de temps sur Facebook que les personnes n'en souffrant pas [Kuss et Griffiths \[2017\]](#). De même, une étude publiée dans le journal rapports scientifiques (*Scientific reports*) a affirmé que les personnes atteintes de troubles de la personnalité borderline ont généralement plus d'amis sur Facebook que les personnes n'en souffrant pas [Ooi et al. \[2020\]](#).

Ce travail vise à proposer une méthode d'analyse des données des utilisateurs susceptibles d'avoir des troubles de la personnalité sur les réseaux sociaux afin de réduire leurs répercussions sociales négatives sur d'autres personnes. Pour ce faire, nous proposons dans le cadre de cette thèse les objectifs à atteindre suivants :

- Proposer différentes méthodes permettant de détecter les utilisateurs susceptibles de souffrir de troubles de la personnalité sur les réseaux sociaux en se basant sur l'analyse de ses données.
- Étudier les sujets provocateurs qui ont un impact négatif sur la stabilité psychologique des utilisateurs des médias sociaux.
- Assurer la surveillance des personnes atteintes de troubles psychologiques sur les réseaux sociaux pendant la période de COVID-19.
- Évaluer les méthodes proposées et discuter des résultats obtenus et aborder les difficultés rencontrées.

Comme il est démontré, le traitement des personnes souffrant de troubles de la personnalité revêt une importance considérable. Cependant, la mise en place d'outils efficaces pour les détecter et les aider est une tâche ardue qui nécessite des ressources matérielles, humaines et logistiques considérables. Le défi dans la lutte contre cette maladie, c'est qu'il s'agit d'informations implicites. En effet, la présence des symptômes chez un individu n'est pas clairement définie. Cette nuance peut permettre à chacun d'interpréter la situation de manière subjective, ce qui signifie que dans certains cas, une personne peut ne pas se rendre compte qu'elle est malade. Dans ce contexte, l'analyse des données des réseaux sociaux peut offrir une solution en prenant en compte simultanément plusieurs critères à différents moments, sans causer de gêne au patient. Dans la littérature, plusieurs chercheurs ont cherché à détecter certains symptômes liés au trouble de la personnalité tels que la dépression au niveau d'un tweet [Kayalvizhi et Thenmozhi \[2022\]](#), [Senn et al. \[2022\]](#). Mais réellement, détecter un tweet qui reflète la dépression suffit pour déclencher un avertissement? D'autres chercheurs ont également travaillé sur la détection de personnes terroristes ou violentes à l'aide d'un ensemble de mots-clés indiquant ces situations [Rekik et al. \[2019\]](#), [Mbarek et al. \[2022\]](#). Cependant, comment ces méthodes peuvent-elles détecter la présence d'une maladie comme la schizophrénie, qui se caractérise par des symptômes tels que le chan-

gement fréquent d'avis relatif au même sujet ? Cela signifie qu'une personne peut exprimer ce comportement inhabituel sans avoir recours à un langage vulgaire. En général, la plupart des travaux dans la littérature se concentrent uniquement sur la détection de problèmes psychologiques spécifiques tels que des symptômes ou des troubles particuliers [Rashida et Suresh Kumar \[2023\]](#), [Reddy et al. \[2023\]](#). Cependant, est-ce que cette tâche permet de tirer des connaissances utiles ou d'aider les patients à surmonter leurs problèmes ?

Nous visons au niveau de cette thèse à travailler sur la fréquence d'apparition de symptômes d'une manière individuelle ainsi que sur leur combinaison pour établir un diagnostic précis et déduire un véritable état de problème. De même, notre objectif consiste à considérer les aspects linguistiques comme le contexte afin de mettre en évidence les spécificités propres à chaque cas d'étude. Nous visons également à extraire d'autres informations pertinentes telles que la cause du trouble, afin de mieux comprendre les principales préoccupations de ces personnes. Par la suite, nous pourrions répondre de manière personnalisée à leurs besoins. En général, notre contribution principale ne consiste pas à détecter les personnes malades, car cela reste un sujet sensible. Mais elle consiste à prévenir l'état de santé des personnes qui pourraient être instables en assurant la surveillance de leurs états psychologiques dans les réseaux sociaux.

Le présent rapport s'articule autour de six chapitres.

Le premier chapitre présente une vue d'ensemble des différents concepts de base des traits et des troubles de la personnalité. Nous énumérons les différents types des traits à savoir l'extraversion, la stabilité, l'agréabilité, la conscience, l'ouverture, ainsi que les différents types de catégories des troubles tels que le paranoïde, le schizoïde, le schizotypique, l'antisociale, le borderline, l'histrionique, le narcissique, l'évitante, la dépendante et l'obsessionnelle compulsive. Ensuite, nous distinguons les caractéristiques de la personnalité, notamment le tempérament, le caractère, la cognition, l'humanisme, la psychobiologie. Ensuite, nous distinguons les principales techniques employées pour la détection de concepts liés à la personnalité, y compris : (i) les traits lexicaux qui consistent à catégoriser les mots simples et composés liés à chaque groupe, (ii) les traits morpho-syntaxiques qui se focalisent sur l'étude des catégories grammaticales les plus fréquemment utilisées au niveau de chaque groupe représentant un type de personnalité spécifique, (iii) les traits sémantiques qui s'intéressent à étudier le contexte de la phrase, (iv) les traits numériques qui mettent en valeur les critères stylistiques figurant au niveau du corpus.

Après, nous montrons l'intérêt de l'usage des techniques d'intelligence artificielle dans le traitement des données liées à la personnalité. Nous pouvons citer dans ce contexte, les outils de TALN, les ontologies, l'apprentissage automatique, etc. Finalement, nous clôturons ce chapitre par une discussion sur le besoin de basculer vers le Web sémantique.

Nous présentons, dans le deuxième chapitre, les différents travaux proposés dans la littérature qui portent sur l'analyse des concepts liés à la personnalité à partir des données des médias sociaux. Nous catégorisons les différents travaux analysés en quatre catégories.

- Nous entamons par les méthodes proposées pour la détection des traits de personnalité. Parmi ces méthodes nous pouvons énumérer celles fondées sur des modèles de classification, que nous pouvons les répartir en deux

classes : des méthodes basées sur l'apprentissage classique et des méthodes basées sur la notion d'apprentissage profond. En outre, nous distinguons d'autres méthodes qui tirent profit des techniques de TALN basées sur l'utilisation des règles et des ressources linguistiques, ainsi que des techniques statistiques fondées sur une analyse quantitative des données. Enfin, il convient de noter qu'il existe des méthodes fondées sur l'hybridation qui impliquent la combinaison de plusieurs méthodes.

- Nous poursuivons en présentant les méthodes de détection des troubles de la personnalité d'une manière explicite, c'est-à-dire ceux qui sont directement liées à la maladie. Nous présentons ensuite les travaux qui portent sur la détection des troubles de façon implicite, c'est-à-dire ceux qui ont un rapport indirect avec le trouble de la personnalité.
- Nous présentons ensuite les méthodes permettant d'identifier les dépendances entre certains concepts reflétant la personnalité. En effet, cela vise à mieux comprendre la situation de manière à extraire de nouvelles connaissances.
- Nous décrivons par la suite les méthodes utilisées pour surveiller les personnes ayant des troubles afin de mieux gérer certaines situations et d'éviter les cas dangereux. Finalement, nous clôturons ce deuxième chapitre par une discussion des méthodes présentées afin de mettre en évidence à la fois leurs avantages et leurs inconvénients.

Le troisième chapitre aborde les méthodes proposées pour la détection des troubles de la personnalité. Nous détaillons dans ce contexte deux méthodes : (i) la première repose sur une architecture neuronale profonde pour la détection générale de la présence du trouble, ainsi que les catégories, les types spécifiques et les symptômes associés à ce dernier, (ii) la deuxième consiste à détecter un type spécifique de trouble de personnalité en se basant sur des caractéristiques personnalisées.

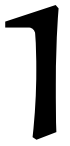
Dans le quatrième chapitre, nous nous concentrons sur l'explicabilité et l'enrichissement de la tâche de diagnostic précédemment abordée en identifiant les causes sous-jacentes à l'apparition de troubles. Pour ce faire, nous proposons dans ce chapitre trois méthodes complémentaires : (i) la première consiste à regrouper récursivement les tweets ayant des thèmes similaires en utilisant la technique d'apprentissage non supervisée, (ii) la deuxième repose sur la création de façon automatique une ontologie qui permet de modéliser le lexique et l'humeur liés à chaque sujet détecté, (iii) la troisième consiste à élaborer automatiquement un ensemble de règles permettant d'interroger l'ontologie construite afin de détecter le sujet et l'humeur exprimés dans un tweet.

Au niveau du cinquième chapitre, nous nous intéressons à la surveillance de l'état des patients par rapport à l'épidémie de COVID-19. Afin d'atteindre cet objectif, nous proposons deux méthodes. La première méthode repose sur la surveillance de l'état des patients en utilisant une analyse sémantique permettant de filtrer les tweets inappropriés qui pourraient apparaître au niveau du fil d'actualité de la personne malade afin d'éviter toute source de stress supplémentaire. La deuxième méthode est fondée sur la recommandation des tweets conçus pour aider les personnes malades à surmonter leur anxiété. Les deux tâches sont menées au moyen d'une méthode hybride fondée principalement sur une ontologie et un modèle d'apprentissage profond.

Le sixième chapitre présente les différentes expérimentations menées tout au long de cette thèse. Nous exposons également les résultats d'évaluation de notre système proposé. Nous commençons par une comparaison du style d'écriture indiqué au niveau de notre corpus à celui d'autres corpus afin de valider l'annotation réalisée. Ensuite, nous procédons à la comparaison des résultats de nos méthodes proposées dans cette thèse avec ceux d'autres méthodes présentées dans des travaux antérieurs. De même, nous comparons nos méthodes proposées avec des méthodes de base (baseline), tout en discutant à chaque niveau les résultats présentés.

En conclusion de cette thèse, nous présentons un récapitulatif de l'ensemble de nos contributions ainsi que des perspectives qui peuvent être envisagées à l'issue de ce travail.

LES CONCEPTS DE BASE DE LA DÉTECTION DE LA PERSONNALITÉ



SOMMAIRE

1.1	INTRODUCTION	7
1.2	QU'EST CE QU'UNE PERSONNALITÉ?	7
1.3	QU'EST CE QU'UN TROUBLE DE PERSONNALITÉ?	8
1.4	LES CARACTÉRISTIQUES DE LA PERSONNALITÉ	9
1.4.1	Tempérament	9
1.4.2	Caractère	9
1.4.3	Cognitive	10
1.4.4	Humaniste	12
1.4.5	Psychobiologique	13
1.5	LES TROUBLES DE PERSONNALITÉS EN TUNISIE	13
1.6	MOTIVATIONS POUR L'AUTOMATISATION DE LA DÉTECTION ET DU SUIVI DES PERSONNES AVEC DES TP	14
1.7	LES PRINCIPAUX TRAITS STYLISTIQUES EMPLOYÉS POUR LA DÉTECTION DE LA PERSONNALITÉ	14
1.7.1	Traits lexicaux	15
1.7.2	Traits Morpho-Syntaxiques	15
1.7.3	Traits Sémantiques	16
1.7.4	Traits Numeriques	16
1.8	LE RÔLE DE L'INTELLIGENCE ARTIFICIELLE DANS LA DÉTECTION DE LA PERSONNALITÉ	17
1.8.1	Les outils de TALN	19
1.8.2	Les ontologies	19
1.8.3	Les techniques de fouille de données et de fouille de textes	20
1.8.4	Apprentissage profond	22
1.8.5	Modèle de langage	23
1.8.6	Les techniques de traitement de données massives	24
1.8.7	Vers le Web sémantique 3.0	25
1.9	CONCLUSION	26

1.1 INTRODUCTION

De nos jours, le Web représente une source d'information indispensable en raison du volume et de la diversité du contenu présenté par les internautes (blogs, commentaires, forums). Devant cette abondance de données et de sources, il faut avoir recours aux différents concepts liés aux sciences de l'informatique afin de développer des outils capables d'analyser ces données pour extraire des informations utiles. Dans ce contexte, il existe plusieurs chercheurs qui ont travaillé sur la détection des informations explicites comme l'analyse de sentiments ou d'opinions telle qu'elle est effectuée par [Kaur et Kautish \[2022\]](#). Cependant, au cours des dernières années, il a été observé que la majorité des études visant à analyser les textes produits sur les réseaux sociaux se focalisent sur la détection des informations implicites telles que la détection de l'ironie [Ghanem et al. \[2020\]](#) et du profil utilisateur tel qu'il est réalisé par [Çetinkaya et al. \[2022\]](#). La détection du profil de l'auteur consiste à découvrir à partir des productions textuelles sur les réseaux sociaux l'âge, le genre, les données socioculturelles, la langue d'origine ainsi que la personnalité. En effet, l'étude de la personnalité est désormais un domaine de recherche en pleine évolution. Cette évolution est due par le fait que cette dernière peut donner une justification convaincante des comportements d'une personne face à n'importe quelle situation. Par exemple, elle peut aider les statisticiens à justifier les résultats trouvés à propos des choix des électeurs lors des élections présidentielles (le choix dépend de la personnalité des électeurs vis-à-vis des candidats à l'élection, tel qu'il est démontré par [Pedersen et al. \[2019\]](#)). Le terme de personnalité est un concept très générique qui est fortement lié à d'autres termes ayant un impact très considérable de nos jours comme le charisme, l'esbroufeur, le dévouement, etc. Par conséquent, il est nécessaire de maintenir ce problème dans son cadre et de souligner les concepts associés à ce terme. De ce fait, ce chapitre est structuré comme suit. Dans un premier temps, nous définissons les concepts majeurs liés au terme de la personnalité ainsi que notre motivation. Ensuite, nous présentons les différentes caractéristiques linguistiques, stylistiques et les techniques servant à analyser la personnalité. Enfin, nous clôturons ce chapitre par une conclusion.

1.2 QU'EST CE QU'UNE PERSONNALITÉ ?

Selon le dictionnaire français Larousse¹, la personnalité est l'ensemble des traits physiques et moraux qui englobent les caractéristiques émotionnelles, d'attitudes et de comportements par lesquels on peut distinguer une personne par rapport aux autres. C'est en quelque sorte l'aspect par lequel quelqu'un affirme leur originalité acquise.

Depuis des décennies d'années, plusieurs travaux dans l'éthologie [Gosling \[2008\]](#), [Mehta et Gosling \[2006\]](#) affirment que la personnalité existe aussi chez certains animaux. De même, il a été confirmé par ces travaux que le comportement individuel peut avoir un impact sur le comportement et le rôle du groupe au sein de l'écosystème, ce qui peut affecter l'aptitude d'un groupe à survivre et à évoluer. Plusieurs chercheurs ont pensé à quantifier et à qualifier la personnalité

1. un dictionnaire encyclopédique en langue française édité depuis les années 1980.

animale, au sein de grands groupes des êtres vivants, comme chez les poissons. Cette étude peut donner une idée à propos de l'influence des traits de personnalité chez les animaux sur la conduction à des phénomènes naturels ou d'adaptation, et leurs effets sur l'écoépidémiologie.

Il existe plusieurs techniques qui permettent de définir les types de personnalités. Parmi lesquelles nous citons le terme «Big Five» comme la méthode la plus répandue pour définir les traits de personnalité qui est généralement abrégée par les abréviations suivantes : OCEAN, NEOAC ou CANOE [Salem et al. \[2019b\]](#).

Extravertie : décrit une personne qui exprime facilement ses émotions, notamment envers les autres. Une personne extravertie a un intérêt particulier au développement des relations sociales ainsi que pour une adaptation rapide à toutes les situations.

Stable : démontré par une personne mature capable de gérer des situations difficiles, d'atteindre ses objectifs émotionnels et de réussir professionnellement.

Agréable : c'est une personne qui peut obtenir la confiance et la satisfaction des autres, et elle est toujours prête à résoudre ses divers problèmes avec des différentes manières.

Conscientieux : décrit une personne sérieuse qui peut assumer la responsabilité de ses relations avec les autres aussi avec son travail. Généralement, la personne qui possède ce trait est plus attentive aux détails.

Ouverte : représente une personne non traditionnelle qui est capable d'accepter une nouvelle idée ou une culture différente. D'ailleurs, elle est toujours prête à vivre une nouvelle expérience.

1.3 QU'EST CE QU'UN TROUBLE DE PERSONNALITÉ ?

Selon le manuel MSD², le trouble de personnalité est une maladie qui est caractérisée par des schémas omniprésents et durables de pensées, de perception, de réaction et de relations, entraînant une souffrance notable chez les personnes qui nuisent considérablement ses capacités à servir [Swart \[2016\]](#). D'une manière générale les troubles de personnalité représentent des anomalies dans la façon d'agir et de réagir. De ce fait, le comportement des personnes ayant des troubles de personnalité comprennent l'inhibition, l'agitation, l'agressivité, les troubles obsessionnels compulsifs, les désordres alimentaires (anorexie, boulimie...), l'hyperactivité, les phobies, les différents types d'addictions (à l'alcool, à la drogue, etc.). Selon le DSM-5³, il existe dix types de troubles de personnalité, répartis en trois catégories [Skodol \[2012\]](#) :

Groupe A : parmi les critères communs entre les maladies de ce groupe, on retrouve l'apparence et le comportement inhabituel, excentrique et étrange.

- Paranoïde : méfiance et suspicion.
- Schizoïde : désintérêt envers les autres.
- Schizotypique : idées et comportements bizarres ou excentriques.

Groupe B : parmi les apparences et les comportements partagés entre les maladies de ce groupe, on retrouve l'excès émotif, erratique, instable et dramatique.

2. une des sources d'informations médicales les plus utilisées au monde qui met en disposition les meilleures informations médicales.

3. est un ouvrage de référence pour la diagnostique des troubles mentaux, publié par l'Association américaine de psychiatrie

- Antisociale : irresponsabilité sociale, le mépris envers les autres, la tromperie et la manipulation à des fins personnelles.
- Limite (Borderline) : vide intérieur, peur de l'abandon, relations instables, difficultés à contrôler ses émotions et un comportement impulsif.
- Histrionique : recherche d'attention et avoir un comportement exagéré.
- Narcissique : besoin d'admiration, manque d'empathie et vision exagérée de sa valeur personnelle (mégalo manie).

Groupe C : les traits partagés par les maladies de ce groupe sont l'anxiété, la timidité et la crainte.

- Évitante : éviter les relations personnelles par crainte d'être rejeté.
- Dépendante : soumission et dépendance (par besoin d'être pris en charge).
- Obsessionnelle compulsive : perfectionnisme, rigidité et obstination.

Les tableaux 1.1, 1.2 et 1.3 présentent en détail les différents symptômes liés aux différents types de troubles de la personnalité avec certaines expressions indiquant les signes de chaque type^{4 5 6}.

1.4 LES CARACTÉRISTIQUES DE LA PERSONNALITÉ

Dans la littérature, il existe de nombreux éléments qui peuvent caractériser la personnalité d'une personne. Selon [Magee et Biesanz \[2019\]](#), il y a des caractéristiques qui sont permanentes ou qui peuvent durer longtemps. Ces caractéristiques sont connues sous le nom de **traits de personnalité**. Il y a des caractéristiques qui sont spécifiques à un moment ou à une situation particulière qu'on appelle **états de personnalité**. Parmi les caractéristiques ayant plus d'influence sur le trait et l'état de la personnalité, nous citons le tempérament, le caractère de la personnalité, le concept cognitif, l'aspect humaniste et la psychobiologie.

1.4.1 Tempérament

Le tempérament est un élément présent dès la naissance pour cela on trouve que chaque bébé a sa façon de réagir à son environnement, d'exprimer ses émotions et d'entrer en relation avec les autres, on n'est pas une «cire vierge» [Shiner et al. \[2012\]](#). Quoique ce concept soit déterminé génétiquement, il peut aussi être influencé par le milieu familial et social de l'individu jusqu'au début de l'âge adulte, avant de devenir relativement stable.

Généralement, le tempérament d'un enfant peut influencer son niveau d'activité, d'attention, d'anxiété, de timidité, d'irritabilité et d'adaptabilité aux nouvelles situations. En d'autres mots, le comportement d'un enfant a des effets sur le degré d'intensité de ses émotions et sur son niveau de sensibilité qui dépend en une grande partie sur le tempérament hérité.

1.4.2 Caractère

Le caractère représente la manière avec laquelle une personne réagit habituellement face à une situation quelconque [GOYETTE \[2014\]](#). Par exemple, on appelle

4. <https://www.evolution-101.com/pensees-sur-lestime-de-soi/>

5. <http://marlenefoucheypsychologue.e-monsite.com/blog/les-psychopathologies/>

6. <https://www.evolution-101.com/pensees-sur-le-doute/>

TABLE 1.1 – Comparaison des signes de différents types de troubles de la personnalité pour le groupe A selon DSM-5.

Maladie	Symptômes	Mots clés
Paranoïaque	Méfiance exagérée Le doute incessant Les interprétations négatives des gestes L'agressivité Garde rancune Attendre l'attaque des autres Une surestimation de soi Une sous-estimation des autres	Je me méfie Il n'est pas fiable Ne vous fiez pas à lui Il déjà trompé la confiance de On ne peut pas faire confiance aux gens Il est en train de vous exploiter On s'en servira contre moi Mes expériences contribuent à la richesse
Schizoïde	Repliée sur elle-même Ne montre pas d'intérêt pour les relations Distante/Introvertie/Solitaire Non-compétitive Autonome Froideur de détachement ou d'émoussement de l'affectivité Hypersensible	Je suis à part Je me sent vide J'abandonne tout La vie serait moins compliquée sans les autres Les gens sont source de problèmes Je me suis isolé du monde
Schizotypique	Anxiété excessive Absence d'amis proches Réflexion métaphorique et complexe Comportement et/ou apparence bizarre Idéation méfiante ou persécutrice Préoccuper de phénomènes paranormaux Raconte des événements quotidiens avec plus de conviction	L'important, c'est ce que je sent et ce que je ne vois pas. Il y a des raisons pour tout, rien n'arrive au hasard. On ne peut pas se fier aux apparences.

celui qui aime à faire le plaisir aux autres qu'il est généreux et altruiste. Ce composant n'appartient pas à l'aspect biologique puisque ce n'est pas l'élément avec lequel on arrive à notre vie mais on l'acquies avec le temps. Plusieurs critères qui peuvent intervenir et influencer le caractère de la personne comme par exemple l'éducation, l'environnement vivant, etc. Selon le dictionnaire Larousse⁷, le caractère représente un ensemble de traits émotionnels stables qui constituent l'identité d'une personne (sa personnalité) et déterminent sa façon de réagir face à son environnement. Un caractère peut être autoritaire, hargneux, souple, gai, sournois⁸.

1.4.3 Cognitive

Le concept cognitif est une partie indispensable de la personnalité. Il peut aider les gens à faire des choix et à prendre des diverses décisions fondées sur des situa-

7. Encyclopédie collaborative en ligne, hébergée par la Wikimedia Foundation, organisation américaine à but non lucratif.

8. <https://www.larousse.fr/dictionnaires/francais/caract%C3%A8re/13058>

TABLE 1.2 – Comparaison des signes de différents types de troubles de la personnalité pour le groupe B selon DSM-5.

Maladie	Symptômes	Mots clés
Antisociale	Irritabilité ou agressivité Mépris inconsidéré pour sa sécurité ou celle d'autrui Impulsivité ou incapacité à planifier à l'avance Irresponsabilité persistante Tendance à tromper et incapacité à être stable Le manque d'empathie Tendent à être immorales et cyniques.	Si je veux quelque chose, je l'obtiendrai par n'importe quel moyen Nous sommes dans la jungle et se sont les plus forts qui survivront Ce que pensent les autres de moi n'a guère d'importance Si les autres ne sont pas capables de se défendre, c'est leur problème!
Limite (Borderline)	Instabilité et de multiples transgressions Forte impulsivité et prise de risque Instabilité affective Fortes colères souvent sans raison réelle Troubles de l'identité Sentiment d'ennui, de vide Peur d'être abandonné Réactions démesurées au stress	personne ne m'aime je serai toujours seul personne ne pourra jamais m'accepter je ne peux m'en sortir tout seul, mais qui peut m'aider? on risque de se faire rejeter je n'arriverai jamais à me contrôler
Histrionique	Besoin d'être le centre d'attention Dynamiques, impressionnants, enthousiastes et charmeurs Caractère flou, vague, subjectif Recherchent activement des supports externes	je suis sans attrait. je dois être aimée par tous j'impressionne les gens les gens n'ont pas le droit de me refuser je suis incapable de m'en sortir toute seule
Narcissique	Sens grandiose de sa propre importance Besoin excessif d'être admirée Pense que tout lui est dû Utilise autrui pour parvenir à ses propres fins Fait preuve d'attitudes et de comportements arrogants	J'ai le droit à des privilèges » Je dois être admiré » Ceux qui me critiquent m'en veulent, sont jaloux Se comparent aux autres en terme de supériorité/infériorité, singularité/banalité, puissance/faiblesse, richesse/pauvreté, beauté/laideur.

tions et des incidents déjà vécus et mémorisés [Hoffman \[2014\]](#). Parmi les critères importants relatifs au concept cognitif figure la vision de soi-même, des autres, du monde et du temps. Ces visions constituent la référence liée aux différentes expériences précoces, aux valeurs, aux traumatismes et aux préjugés du milieu culturel et socio-éducatif de la personne. Ces éléments sont latents et inaccessibles d'une manière spontanée, c'est-à-dire qu'ils constituent des hypothèses de la réalité à confirmer ou à nier.

TABLE 1.3 – Comparaison des signes de différents types de troubles de la personnalité pour le groupe C selon DSM-5.

Maladie	Symptômes	Mots clés
Évitant	Hypersensibilité aux critiques ou aux rejets Manque de confiance en soi Isolement social auto-imposé Extrême timidité ou anxiété face à des situations sociales Tentative d'éviter tous Méfiance envers les autres Distance émotionnelle liée à l'intimité	je suis nulle les gens vont me rejeter, je vais les ennuyer
Dépendant	Ne pas être capables de prendre soin d'eux-mêmes Besoin de réconfort et de conseils pour prendre des décisions ordinaires Peur de perdre le soutien ou l'approbation des autres	seul je suis impuissant je ne peux pas m'en sortir tout seul je dois être aidé je ne peux pas vivre sans soutien je suis incapable de décider tout seul je n'y arriverai pas je suis nul, je suis bête les autres ont raison et moi j'ai tort
Obsessionnel-compulsif	Se montre rigide, têtu, consciencieux, scrupuleux Se montre avare Incapacité de jeter des objets Dévotion excessive pour le travail et la productivité Exigences personnelles trop strictes Préoccupations par les détails	je suis responsable de moi même et des autres je sais ce qui est bien il faut faire les choses parfaitement les fautes sont intolérables une personne qui s'est trompée doit être punie je dois contrôler mes émotions

1.4.4 Humaniste

Il s'agit d'un modèle de psychothérapie qui met en lumière la dimension existentielle, volitive et adaptative, ainsi que la valence ontologique spirituelle de la personnalité humaine. En effet, ce concept est lié aux tendances innées de la personne telle que le désir de se réaliser, c'est-à-dire il s'agit de la responsabilisation envers soi-même Olano [2019]. Pour cela, son défi consiste à réunir les forces de la croissance psychologique et à accroître son potentiel dans le but d'atteindre son objectif. Selon le dictionnaire Larousse, le terme humaniste est un adjectif en rapport avec l'humanisme philosophique, qui place l'homme au centre de ses préoccupations⁹.

9. <https://www.larousse.fr/dictionnaires/francais/humaniste/40617>

1.4.5 Psychobiologique

Ce caractère s'articule principalement sur les données de la neurobiologie¹⁰, c'est-à-dire il s'appuie sur des liens entre les données psychologiques et les explications sous-jacentes à caractère neurobiologique La Paglia et Telliez [2021]. La fatigue mentale et le stress sont des états psychobiologiques causées par de longues périodes d'activité cognitive intensive ou excessive. Ces conditions mènent à des temps de réaction plus lents, résultant en un manque d'attention et de concentration, ce qui engendre également une difficulté majeure dans la prise de décision.

1.5 LES TROUBLES DE PERSONNALITÉS EN TUNISIE

Comme dans tous les pays du monde, en Tunisie il y a aussi des personnes ayant des troubles de personnalité notamment suite aux problèmes économiques, sociales et environnementales. Toutes ces conditions peuvent mettre certains citoyens dans une situation critique, qui peut conduire, au fil du temps, à un état alarmant de personnes confrontées à des problèmes psychologiques. Dans ce contexte, le journal en ligne *Leaders*¹¹ a affirmé que la Tunisie a déclaré, en 2014, que 203 des cas ont subi une crise suicidaire (suicides et tentatives de suicide) dont 74.38% chez les hommes et 50% du taux global entre 16 et 35 ans.¹² Ces chiffres sont justifiés par le fait qu'il s'agit d'une société masculine où les hommes ont du mal à accepter l'échec, surtout dans les villes conservatrices comme Kairouan, qui arrive en tête de liste avec 31 cas. Pour imaginer l'ampleur de l'oppression et de la douleur, la pendaison est le moyen le plus fréquemment adopté (58.63%), suivi de l'immolation (15.89%). Ce nombre est en croissance, par exemple en 2015 la Tunisie a recensé 365 cas de suicide (c'est-à-dire un cas par jour)¹³. Malgré les coûts élevés que les personnes ayant des TP peuvent causer à l'état¹⁴, la Tunisie rencontre une difficulté notable dans le suivi et la prévention de ces malades, justifiée par l'absence de ressources financières. Dans ce cadre, Hajji et al. [2016] ont indiqué que seulement 13,7% des personnes atteintes des TP dans la prison de Mahdia ont bénéficié d'un séjour dans un établissement psychiatrique, quoique cette catégorie de personnes relève de la supervision et la responsabilité directe de l'État. Ces chiffres montrent que le nombre de TP en Tunisie augmente de plus en plus. Par conséquent, un intérêt inéluctable de détecter ces cas, ce qui constitue une piste de recherche à explorer.

10. Étude du fonctionnement des cellules et des tissus nerveux.

11. journal en ligne qui présente l'actualité et les News de la Tunisie et du monde :leaders.com.tn

12. <https://www.tuniscope.com/ar/article/62770/actualites/tunisie/ftdes-suicide-203cas-213516>

13. <https://www.leaders.com.tn/article/20973-un-tunisien-se-suicide-chaque-jour-l-immense-detresse-d-une-mal-vie>

14. <https://www.oecd.org/fr/sante/les-troubles-de-la-sante-mentale-representent-un-lourd-fardeau-economique-pour-les-pays-europeens.htm>

1.6 MOTIVATIONS POUR L’AUTOMATISATION DE LA DÉTECTION ET DU SUIVI DES PERSONNES AVEC DES TP

Au cours de ces années, diverses organisations mondiales et divers pays développés ont investi considérablement dans la santé mentale afin de réduire les taux de violence et de suicide. Dans ce contexte, la revue canadienne de psychiatrie a affirmé un investissement de 5 milliards de dollars dans le système de soins de santé mentale au cours des dix prochaines années afin d’offrir des services adaptés et favorables [Palay et al. \[2019\]](#). Cet effort demeure modeste car il n’est pas évident de détecter des personnes souffrant de troubles mentaux et d’interagir avec eux. En effet, il est difficile d’identifier une information sensible que la personne elle-même ne sait pas qu’elle admet. Il est aussi difficile d’interagir avec quelqu’un qui n’est pas dans son état mental normal (toute réaction est prévue). Pour cela, l’automatisation de la tâche d’analyse de la personnalité est un enjeu pris en compte depuis longtemps par les chercheurs en science de l’informatique. Cela s’explique par le nombre effrayant de personnes atteintes de TP dans le monde, ainsi que par le danger qui peut être commis par ces personnes.

Afin d’assurer l’automatisation des tâches liées à la détection et à la prévention des patients atteints de TP, certains chercheurs ont utilisé différents moyens liés à l’IOT, tels que des capteurs électrochimiques [Zamani et al. \[2022\]](#). D’autres chercheurs se sont servis de l’analyse des productions textuelles des individus sur les réseaux sociaux afin de surmonter les problèmes liés à la gestion des objets connectés [Borah et Ganesh Kumar \[2023\]](#). Bien que le traitement de ce genre d’information comporte plusieurs difficultés, puisqu’il s’agit d’information implicite, de nombreux critères peuvent être inclus dans le style d’écriture, tels que l’ironie et l’absence de contexte dans les textes courts. Ces problèmes peuvent être résolus par l’utilisation de nouvelles techniques avancées. Nous pouvons apprécier dans ce cadre l’importance des outils de TALN dans le traitement des données textuelles, en particulier avec l’explosion des données non structurées sur les réseaux sociaux et le Web 2.0.

Tant avec les techniques d’exploration de données, l’extraction de connaissances devient de plus en plus possible. En effet, même avec un psychiatre expérimenté, il n’est pas possible de généraliser les conclusions tirées, étant donné que le nombre de personnes à traiter est toujours limité. De plus, les patients traités par le même médecin ont certainement des points communs.

Par conséquent, il est vivement recommandé de réaliser une analyse stylistique des données personnelles impliquant l’étude du style d’écriture et des relations entre certains éléments clés à l’aide de techniques avancées et de machines puissantes. Cette analyse permet de généraliser les connaissances obtenues et de surmonter les limites des méthodes d’analyse classiques.

1.7 LES PRINCIPAUX TRAITS STYLISTIQUES EMPLOYÉS POUR LA DÉTECTION DE LA PERSONNALITÉ

Une variété de caractéristiques stylistiques ont été identifiées dans la littérature permettant d’effectuer une analyse de la personnalité. En général, l’objectif de l’étude stylistique est de mettre en valeur les spécificités linguistiques qui contri-

buent à identifier un concept psychologique chez une personne en se basant sur l'analyse de leur production textuelle. Nous décrivons ci-dessous les différents aspects identifiés dans une étude linguistique, en prenant en compte les traits lexicaux, morfo-syntaxiques, sémantiques et numériques.

1.7.1 Traits lexicaux

La réalisation de la tâche de détection de la personnalité basée sur des traits lexicaux est essentiellement fondée sur la catégorisation de mots simples et composés (des séquences de mots (n-grammes)) par classe. En effet, cette tâche consiste à associer les termes les plus fréquemment utilisés pour chaque trait de personnalité ou pour chaque type de trouble de la personnalité. Cette mesure fournit des indicateurs qui évaluent la richesse et la diversité du vocabulaire d'un texte. Le résultat de cette technique se présente sous la forme d'une répartition des mots les plus utiles pour chaque classe.

Comme indiqué dans le travail de [Ellouze et al. \[2020\]](#) qui est appliqué au corpus PAN/CLEF 2015 [Stamatatos et al. \[2015\]](#), les mots suivants : *idée, vie, ami, remercier, mensonge, pense, amour, esprit, espérer, manquer, gagner, cadeau, ressentir, avoir besoin, confiance* peuvent aider à distinguer le genre de la personne. Ce travail a démontré aussi que les mots *milliardaire, perfection, canal, extrait, université, percevoir, sentier, enceinte, pivot, récolte, pentecôte, guérir* peuvent aider à établir une distinction entre les catégories d'âge suivantes : 18-24, 25-34, 35-49, 50-XX. Plusieurs outils d'exploration de données ont été utilisés dans ce contexte afin de réaliser l'analyse lexicale des mots par classe donnée, en tirant profit des techniques statistiques et de visualisation permettant d'offrir un aperçu détaillé de la relation du mot avec la classe, comme il est montré dans les deux exemples ci-dessous :



Exemple de mots désignant la domination [Ahmad et Siddique \[2017\]](#).



Exemple de mots désignant la stabilité [Ahmad et Siddique \[2017\]](#).

Cette technique peut être utilisée dans le prétraitement des données en vérifiant le respect des règles lexicales, comme l'emploi de mot en dehors du lexique d'une langue.

1.7.2 Traits Morfo-Syntaxiques

Les traits morfo-syntaxiques regroupent les catégories grammaticales des mots utilisés telles que «noms, adjectifs, verbes, etc.». Ces classes sont aussi connues sous le nom de partie de discours *Part-Of-Speech (POS)* ou encore de catégorie grammaticale. De plus, la syntaxe permet d'étudier la structure de la

phrase, à savoir la disposition des mots dans la phrase. L'étiquetage (*POS tagging*) est le processus qui consiste à attribuer les catégories grammaticales appropriées aux mots d'une phrase. Les critères morpho-syntaxiques peuvent refléter la personnalité de la personne en mesurant les spécificités grammaticales, telles que le taux d'utilisation des verbes, des déterminants, etc. Les exemples suivants illustrent des hypothèses analogiques tirées par des psychiatres et des psychologues, qui ont ensuite été confirmées par plusieurs études d'analyse du style d'écriture [Varshney et al. \[2017\]](#) et [Holtzman et al. \[2019\]](#).

Exemple 1 : La personne impulsive utilise beaucoup les verbes d'action.

Exemple 2 : La personne soupirs en depression utilise fréquemment le temps reflétant le passé.

Exemple 3 : La personne dépendante utilise beaucoup plus le déterminant *ils* ou bien *nous* par rapport aux autres déterminants.

1.7.3 Traits Sémantiques

L'étude de l'aspect sémantique de la phrase s'appuie essentiellement sur une analyse du contexte de la phrase. Ce traitement résout les limites de l'analyse lexicale, étant donné la difficulté de trouver un corpus contenant l'ensemble des mots utilisés pour chaque classe. Il existe plusieurs manières qui assurent la réalisation de cette étape comme le recours à certaines ressources linguistiques telles que les dictionnaires. Cette analyse repose aussi sur l'étude des dépendances sémantiques qui peuvent être assurées par des ressources fondées sur l'ontologie comme Wordnet. Dans ce contexte, les travaux [Hassanein et al. \[2018\]](#), [Ellouze et al. \[2020\]](#) ont montré comment ces ressources ont résolu les problèmes de l'analyse lexicale et comment cela garantit la généralisation du lexique obtenu. De plus, il y a d'autres chercheurs qui ont travaillé sur l'association de lexiques obtenus en fonction du sujet d'étude (sport, famille, etc.), ou de l'humeur et de degré de sensibilité [Schwartz et al. \[2013\]](#), [Baik et al. \[2016\]](#). Dans ce contexte, [Schwartz et al. \[2013\]](#) ont montré que les personnes extraverties utilisent fréquemment des expressions liées au vocabulaire des amis, de la famille, etc. Elles éprouvent également des sentiments plus positifs à l'égard des autres.

1.7.4 Traits Numeriques

Les traits numériques se basent principalement sur une étude statistique des critères stylistiques figurant dans le corpus, qui comprend un ensemble des valeurs numériques. Ces critères comportent un ensemble d'informations variées, telles que le nombre de mots dans la phrase, le nombre de caractères, la longueur moyenne de la phrase, l'utilisation de ponctuations, etc. Ils peuvent donner un aperçu détaillé des différentes particularités linguistiques utilisées par les personnes avec TP notamment lorsque ce traitement est appliqué aux données volumineuses. De nombreux chercheurs ont eu recours à l'utilisation de ces attributs sur la base d'une étude empirique effectuée sur leurs données [Mbarek et al. \[2022\]](#), [Saha et al. \[2021\]](#). Dans ce contexte, [Saha et al. \[2021\]](#) ont extrait plus de 80 attributs numériques de leur corpus, comme illustré dans le tableau 1.4. Aussi, [Mbarek et al. \[2022\]](#) ont appliqué plusieurs attributs numériques répartis sur plusieurs catégories : (i) catégorie grammaticale liée aux critères linguistiques, (ii) catégorie

temporelle liée à la date et à l'heure de publication, (iii) catégorie qui renferme des informations spécifiques aux médias sociaux comme le nombre de publications et le nombre d'amis, (iv) catégorie liée aux objets comme les émoticônes.

TABLE 1.4 – Exemple des traits numériques *Saha et al. [2021]*

Numero d'attribut	Description des attributs
F01	Nombre total de caractères
F02	Nombre total de lettres
F03	Nombre total d'espaces blancs
F04	Nombre total d'espaces de tabulation
F05	Nombre total des caractères spéciaux
F06	Longueur moyenne des mots
F07	Nombre de mots supérieur à 6 caractères
F08	Nombre total des mots courts
F09	Hapax legomena
F10	Hapax disligomena
F11	Mesure Yulies K
F12	Nombre total de simple quote
F13	Nombre total de points-virgules
F14	Nombre total de points
F15	Nombre total de points d'interrogation
F16	Nombre total de points d'exclamations
F17	Nombre total de cõlon
F18	Nombre total de semi-cõlon
F19	Nombre total de lignes
F20	Nombre total de phrases
F21	Nombre total de mots par phrase
F22	Nombre moyen de caractères par phrase
F23..F33	Nombre de mots avec 1..11 caractères
F34	Nombre de mots supérieur à 12 caractères
F35..F60	Nombre de lettres de a..z
F61..F86	Nombre de lettres de A..Z

Pour manipuler les différents traits mentionnés, plusieurs techniques basées sur l'intelligence artificielle s'imposent, permettant une analyse approfondie des publications des internautes susceptibles d'avoir des troubles psychiques, en vue de réaliser une discrimination valide entre les individus.

1.8 LE RÔLE DE L'INTELLIGENCE ARTIFICIELLE DANS LA DÉTECTION DE LA PERSONNALITÉ

L'Intelligence Artificielle (IA) est l'un des ensembles de théories et de techniques servant à fabriquer des machines capables de simuler des tâches qui sont effectuées pour l'instant de façon satisfaisante par des êtres humains [Chikhi \[2019\]](#). Le paradigme de l'IA est un mécanisme qui démontre une capacité d'apprentissage et d'adaptation à de nouvelles situations en exécutant des techniques qui permettent la généralisation, l'abstraction et la découverte d'associations entre les divers composants fondamentaux. L'IA nous permet de : (i) traiter de grandes quantités de données, (ii) discerner des motifs et des modèles indétectables par

un humain, (iii) comprendre et analyser des modèles, (iv) interagir avec les gens, (v) apprendre progressivement. Il existe trois types d'IA : (i) étroit AI : dédié à une tâche, (ii) general IA : se comporter comme un humain, (iii) super AI : intelligente plus qu'humaine.

Les phases suivantes décrivent le processus d'IA dans son ensemble : (i) collecter des données à l'aide de capteurs, (ii) centraliser et analyser les données, (iii) modéliser le schéma de la méthode, (iv) déployer des algorithmes qui apprennent à prévoir ou à catégoriser certains cas. La figure 1.1 montre les différents domaines de l'intelligence artificielle à savoir le traitement automatique des langues naturel, le traitement du signal, l'apprentissage automatique, la robotique, le raisonnement et la représentation des connaissances et l'interaction entre les systèmes multi-agents [De Ganay et Gillot \[2017\]](#).

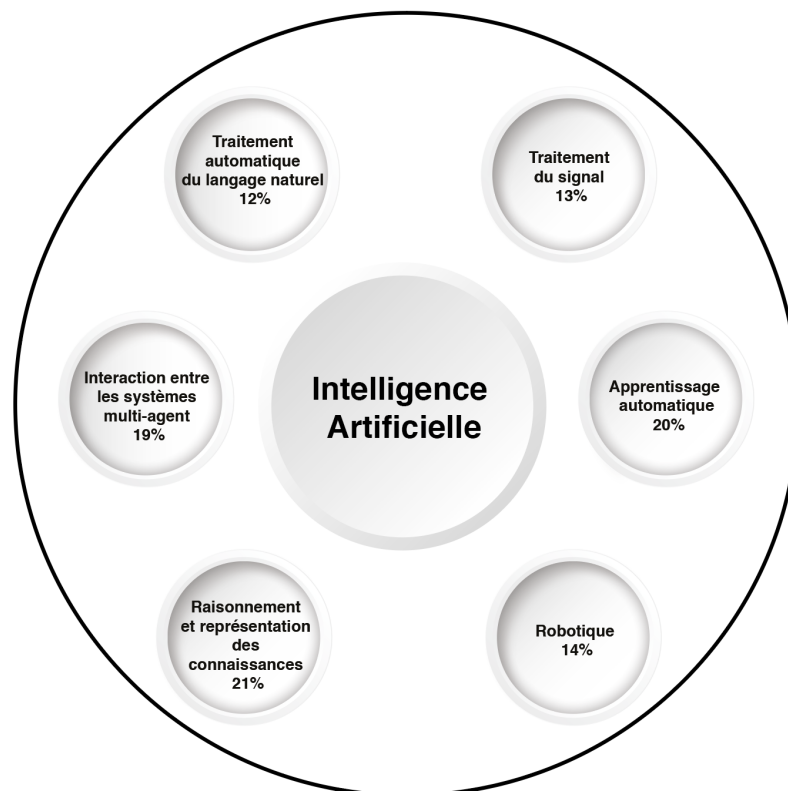


FIGURE 1.1 – Les différents domaines de l'intelligence artificielle [De Ganay et Gillot \[2017\]](#).

Nous présentons, dans la partie qui suit, une vue d'ensemble des diverses techniques d'IA existantes dans la littérature relative aux tâches de détection et de suivi des personnes ayant des TP à partir de l'analyse de leur production textuelle sur les réseaux sociaux.

1.8.1 Les outils de TALN

Le traitement automatique du langage naturel est une branche de la science de l'informatique et de l'intelligence artificielle qui s'intéresse au développement d'outils informatiques capables de manipuler un texte ou une parole en langage naturel [Reshamwala et al. \[2013\]](#). La question principale des techniques de TALN est la manière dont les ordinateurs peuvent être en mesure de comprendre et de réagir en langage naturel. Le but est d'offrir une multitude de services qui peuvent rendre l'utilisation des outils plus agréable et simple en fournissant un cadre d'interaction entre les concepts de langage, les techniques informatiques et l'aspect mathématique. Les applications du traitement du langage naturel touchent plusieurs domaines d'étude, tels que la traduction automatique, le traitement et la synthèse de texte en langage naturel, les interfaces (H/M), etc.

Le TALN a été employé comme un moyen important dans l'analyse des données électroniques de santé liées à la maladie de TP. Par exemple, il permet d'analyser les rapports médicaux afin d'extraire les caractéristiques socio-démographiques et d'estimer leurs impacts sur la durée de séjour à l'hôpital [Barroilhet et al. \[2020\]](#). De même, nous pouvons utiliser les techniques de TALN pour analyser le discours des gens et extraire des informations relatives à leurs pensées et à leurs sentiments quotidiens afin de détecter les symptômes et les signes d'idéation d'automutilation liés au trouble de la personnalité Bordeline [Khazbak et al. \[2021\]](#).

1.8.2 Les ontologies

L'ontologie est une branche de la métaphysique qui s'intéresse à l'existence de l'être et aux catégories fondamentales de l'existant, ce qui veut dire l'étude de l'être et l'expansion de l'existence [Guizzardi et Halpin \[2008\]](#). Le WebOnto¹⁵ distingue quatre types d'ontologies à savoir l'ontologie de domaine, l'ontologie de tâche, l'ontologie de méthode, et l'ontologie d'application¹⁶. Dans la littérature, les chercheurs ont généralement considéré que les ontologies réalisées sur les données des réseaux sociaux s'inscrivent dans la discipline de l'ontologie sociale qui se penche sur la nature et les propriétés de l'univers social. Elle cherche à répondre à des interrogations ontologiques distinctes sur la nature des groupes, des institutions et de tout ce qui peut être qualifié de social [Lauer \[2019\]](#). À l'instar de l'ontologie analytique traditionnelle, qui vise à déterminer ce qui peuple le monde, l'ontologie sociale s'intéresse à ce qui anime le monde social et met en place des méthodes d'analyse et d'évaluation des phénomènes sociaux. Dans ce contexte, [Epstein \[2015\]](#) soutient que l'ontologie sociale peut favoriser le succès de la prédiction et de l'explication en fournissant une image correcte du monde social. Grâce à sa structure extensible et explicative, l'ontologie représente depuis longtemps et jusqu'à nos jours un moyen inéluctable pour modéliser le lexique relatif au trouble de la personnalité [Blashfield et McElroy \[1989\]](#), [Maung \[2021\]](#).

¹⁵. Est un outil qui fournit des services de visualisation, de navigation et d'édition en ligne pour l'élaboration et la maintenance des modèles de connaissances.

¹⁶. <http://dSPACE.univ-tlemcen.dz/bitstream/112/1062/5/ChapitreI.pdf>

1.8.3 Les techniques de fouille de données et de fouille de textes

Le fouille de données (*data mining*) est un processus d'exploration de données visant à extraire toutes les informations cachées, qui sont à la base inconnues et potentiellement utiles, à partir de grandes bases de données Lam [2011]. La figure 1.2 montre en détail le processus général de la démarche de découverte de connaissances à partir d'un ensemble de données (KDD, *knowledge discovery in databases process*). Le processus exploré au niveau de cette figure comporte principalement quatre étapes : (i) la sélection des données cible ou le sous-ensemble d'échantillons de données sur lesquels la découverte doit être effectuée, (ii) la préparation des données qui comporte le nettoyage et le prétraitement des données en prenant des décisions à propos la gestion des champs manquants et la transformation des données en fonction des besoins, (iii) le data mining qui consiste à chercher des modèles sous une forme de représentation particulière (comme des règles, des arbres de classification, etc.), permettant d'extraire des informations, (iv) l'interprétation des informations extraites en vue de fournir les connaissances nécessaires pour la prise de décisions.

L'exploration de données peut également s'expliquer par la recherche de corrélations dans une grande base de données relationnelle en tenant compte des différentes profondeurs d'angles d'analyse. C'est une technique à fort potentiel qui aide les organisations ou les entreprises à augmenter leurs ventes et à tirer d'avantage des informations sur les transactions de leurs clients Agarwal [2013]. En effet, les informations extraites par l'étiquette d'analyse des données ne sont pas explicitement disponibles dans la base de données. L'application de la base de données fournit uniquement les renseignements qui y figurent, avec une capacité de manipulation restreinte. L'extraction de données est donc mieux décrite comme la découverte de connaissances dans les bases de données. Il y a deux tâches supportées par le concept du *data mining* : l'analyse descriptive, qui se concentre sur la présentation des faits et des tendances observés dans les données, et l'analyse prédictive, qui se base sur l'évaluation du passé pour anticiper le futur. Nous citons ci-dessous les techniques les plus utilisées permettant d'effectuer ces types d'analyse Agarwal [2013].

- La classification des données est le processus par lequel on forme un modèle ou un classifieur pour prédire les étiquettes catégoriques qui sont les étiquettes des classes permettant de faire la distinction entre les données. Ces modèles créés servent à attribuer des étiquettes aux nouvelles données d'entrée dont les classes sont inconnues.
- L'exploration de données de prédiction (ou la prévision) est une technique qui est employée pour décrire les données manquantes ou indisponibles. La différence de cette technique par rapport à la technique de classification des données est que les étiquettes des classes de données ne sont pas connues, elles sont à anticiper. Autrement dit, au lieu de déduire le label de classe, la technique de prévision prédira une valeur de classe.
- Le regroupement des données est une technique qui consiste à stocker des données logiquement semblables dans le même groupe. C'est une technique d'exploration de données autonome, car l'objectif est de détecter les

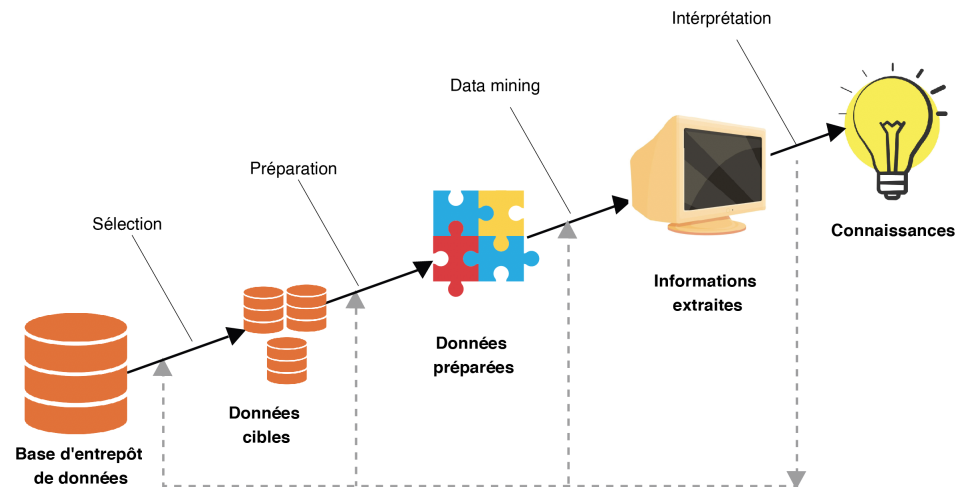


FIGURE 1.2 – Processus KDD dans l'exploration de données *Oprean [2010]*.

classes qui sont également exprimées implicitement. L'agrégation des données est basée sur l'idéologie de l'amplification des similitudes au sein du groupe tout en atténuant les similitudes entre les groupes.

- L'apprentissage en ligne (e-learning) : Parmi les concepts qui permettent d'assurer cette technique, on trouve l'apprentissage par renforcement, dont le principe repose sur l'utilisation d'un système de récompense et de punition pour entraîner l'algorithme. En cela, l'algorithme ou l'agent tire des leçons de son environnement. L'agent obtient des récompenses pour un rendement correct et des sanctions pour un rendement incorrect.
- L'analyse des valeurs aberrantes : Une valeur aberrante est une valeur qui n'est pas disposée à suivre le comportement générique indiqué par les autres valeurs associées à un modèle de données. C'est-à-dire que son comportement est tellement différent qu'il est difficile de croire qu'elle a été générée par le même mécanisme.
- L'extraction de règles d'association (*Association Rule Mining*) : ce type de technique sert à déterminer la pertinence entre les éléments clés de données. Dans un cadre commercial cette technique permet de découvrir la dépendance entre les articles qui sont achetés par les clients. Cela permettra aux entreprises par la suite de planifier leurs stratégies de vente et de gestion de l'espace.
- Les caractéristiques de discrimination : la caractérisation est un processus qui consiste à décrire la classe de manière synthétique. Les caractéristiques utilisées pour cette tâche sont considérées comme la description des concepts clés de la classe.

La fouille de textes (*Text mining*) est une technique qui permet de découvrir de nouvelles informations provenant de données textuelles qui sont auparavant des informations non identifiées ou secrètes *Dang et Ahmad [2014]*. L'exploration de

texte est un domaine multidisciplinaire, portant sur l'analyse de texte, la recherche d'information, l'extraction d'information, la catégorisation, le regroupement, la visualisation, l'exploration de données et l'apprentissage automatique. L'exploration du texte comporte cinq étapes fondamentales : (i) collecter des informations issues de données non structurées, (ii) convertir de l'information obtenue en information structurée, (iii) déterminer le modèle à partir des données structurées, (iv) analyser le modèle, (v) récupérer les informations utiles et les stocker dans la base de données. Dans la littérature, il existe plusieurs outils de fouille de texte qui font l'objet d'une combinaison de différentes disciplines variées, telles que l'aspect linguistique figurant au niveau du corpus, le TALN, et la linguistique computationnelle comme l'outil RapidMiner¹⁷ qui s'occupe à faire l'analyse qualitative assistée par ordinateur. Cet outil est déjà employé dans le cadre de l'analyse de la personnalité des données provenant des réseaux sociaux [Ahmad et Siddique \[2017\]](#).

1.8.4 Apprentissage profond

L'apprentissage profond (*deep learning*) est une classe de l'apprentissage automatique qui fonctionne principalement sur des données non structurées. Les techniques de *deep learning* surpassent les techniques existantes d'apprentissage automatique. Elles permettent aux modèles informatiques d'apprendre progressivement des fonctionnalités à différents niveaux à partir des données. La popularité de l'apprentissage profond a augmenté avec la quantité de données disponibles et l'évolution des ordinateurs puissants [Mathew et al. \[2020\]](#). Les réseaux de deep learning peuvent être utilisés dans de nombreuses applications telles que les voitures autonomes, prévision des taux du marché, le traitement du langage naturel, la reconnaissance visuelle, traduction automatique, l'assistant virtuel de Google. En outre, ils servent aussi à la synthèse et à la génération des textes et des images, la colorisation d'images, la prédiction sismique, l'agrégation d'informations et la détection d'informations frauduleuses. La présente section clarifie quelques architectures de l'apprentissage profond.

- Les réseaux de neurones artificiels (ANN) : ce sont des réseaux de neurones composés de multicouches entièrement interconnectés [Han et al. \[2018\]](#). Ils se composent d'une couche d'entrée, de plusieurs couches cachées et d'une couche de sortie. Chaque noeud d'une couche est relié à l'ensemble des autres noeuds de la couche suivante. Nous approfondissons le réseau par l'augmentation du nombre de couches cachées. Chaque noeud prend la somme pondérée de ses entrées et la transmet à travers une fonction d'activation vers un autre noeud dans la couche suivante. La formation de ce réseau neural profond implique l'apprentissage des poids associés à toutes les arêtes.
- Les réseaux de neurones récurrents (RNN) : ce modèle repose sur la notion de récurrence dont les sorties des états précédents sont alimentées en entrée de l'état courant [Dinarelli et Tellier \[2016\]](#). Cette architecture assure que les couches cachées de RNN mémorisent l'information. RNN peut être utilisé pour prédire les séries temporelles car il peut aussi mémoriser les entrées précédentes, ce qu'on appelle la mémoire à long terme.

¹⁷. est un outil largement utilisé dans plusieurs domaines tels que l'intelligence artificielle, l'exploration de données pour effectuer la tâche de classification et d'analyse prédictive.

- Machines de Boltzmann restreintes (RBM) : les machines restreintes de Boltzmann sont des réseaux neuronaux stochastiques composés de deux couches : les unités visibles et les unités cachées [Fissore \[2022\]](#). Ces réseaux artificiels peuvent apprendre d'une répartition des probabilités à travers un ensemble d'entrées.
- Les réseaux génératifs antagonistes (GAN) : ce type consiste à former simultanément deux modèles de réseaux profonds [Enard et al. \[2020\]](#). Ces réseaux d'apprentissage sont essentiellement en compétition les uns avec les autres - un modèle qui cherche à générer de nouvelles instances ou de nouveaux exemples, appelé générateur. L'autre modèle est appelé le discriminateur, il tente de classer si un cas particulier vient des données d'apprentissage ou du générateur.

Dans le tableau 1.5, nous exposons les divergences en termes de taille et de format des données, ainsi que les variations au niveau du type de l'architecture des algorithmes d'apprentissage et des domaines d'application. Cela vise à démontrer les différents cas d'utilisation entre les algorithmes classiques et les algorithmes profonds selon le site Data Flair ¹⁸.

TABLE 1.5 – Les points de différence entre l'apprentissage classique et l'apprentissage profond selon Data Flair.

	Apprentissage classique	Apprentissage profond
Mode d'organisation des informations	Données structurées	Données non structurées
Taille de données	Taille réduit de volume de données	Grand volume de données
Type d'entraînement	Méthode d'apprentissage basée sur l'intervention humaine	Méthode d'apprentissage autonome
Algorithme	algorithme modifiable	Réseau de neurones
Domaines d'application	Taches simples (marketing)	Taches complexes (assistant vocal, creation de contenu)

1.8.5 Modèle de langage

Pour qu'une machine puisse traiter et comprendre n'importe quel type de texte, il est essentiel de présenter ce texte dans une langue que la machine peut comprendre. Généralement, une machine ne peut fonctionner qu'avec des chiffres, quelles que soient les données que nous lui fournissons : vidéo, audio, image ou texte. Par conséquent, la représentation du texte sous une forme numérique est l'un des enjeux les plus activement recherchés. Par conséquent, nous détaillons ci-dessous les principales techniques appliquées dans notre domaine d'étude.

- Word2Vec est un algorithme qui construit une représentation sémantique distribuée des mots [Church \[2017\]](#). Le concept de Word2Vec est basée principalement sur deux modèles *Bag of Words* et *skip gram*. L'un consiste à

¹⁸. est l'un des meilleurs fournisseurs de formation en ligne sur les certifications Hadoop, Big Data et Spark par le biais d'experts du secteur.

prédire les mots de contexte à l'aide d'un mot central, tandis que l'autre consiste à prédire le mot à l'aide des mots de contexte.

- Sent2Vec : les techniques de *sentence embedding* tiennent compte de toute la phrase et de l'information sémantique qu'elle contient sous forme d'un vecteur numérique [Heo et al. \[2021\]](#). Cette technique peut être considérée comme la moyenne mathématique des représentations vectorielles des mots dans la phrase. Cela permet à la machine de comprendre le contexte, l'intention et d'autres nuances du texte entier.
- Doc2Vec étend l'idée de SentenceToVec ou plutôt Word2Vec, puisque les phrases peuvent également être considérées comme des documents [Kim et al. \[2019\]](#). Il y a deux façons de modéliser le vecteur du paragraphe *PVDM (Distributed Memory version of Paragraph Vector)* et *PVDOBW (Distributed Bag of Words version of Paragraph Vector)*. Pour la première technique, on assigne une phrase vectorielle de paragraphe tout en partageant des vecteurs de mots entre toutes les phrases. Puis on concatène ces vecteurs ensemble pour obtenir la représentation finale du document. La deuxième façon est une extension du type *Skip-gram* où les mots des phrases sont échantillonnés au hasard.

1.8.6 Les techniques de traitement de données massives

De nos jours, le haut usage des objets connectés, du Web 2.0 et des transactions électroniques a généré un volume de données énorme et varié (mega-data). Le principal enjeu dans ce cas est non seulement le stockage de ces données, mais également la manière de les traiter en temps réel afin d'extraire des informations et des connaissances utiles. Par conséquent, un ensemble de données brutes de faible densité est traité et analysé afin d'aider les décideurs à prendre leurs décisions et à réaliser leurs objectifs. À cette fin, un processus en quatre étapes est convenu (Figure 1.3).

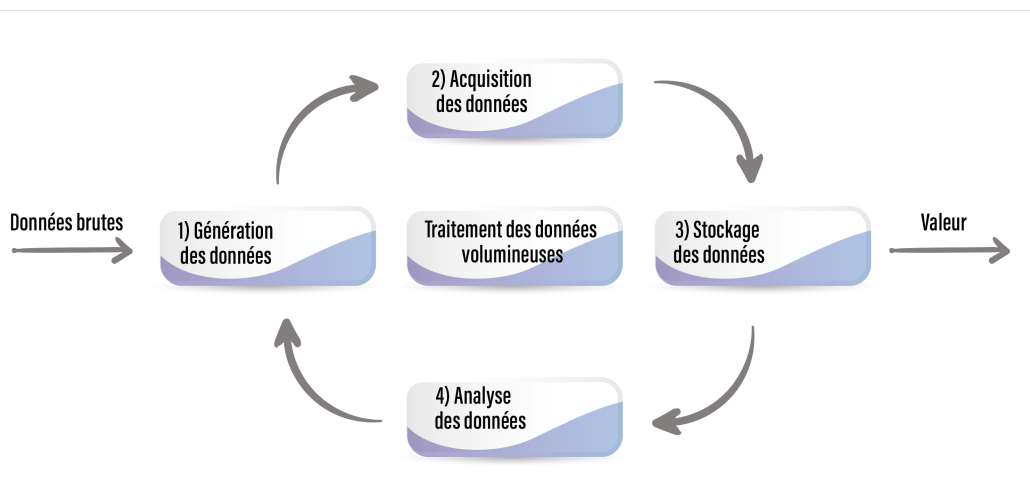


FIGURE 1.3 – Chaîne de valeur du Big Data : de la collecte à l'exploitation des données [Bahri et al. \[2018\]](#).

De même, nous assistons actuellement à l'émergence d'un certain nombre de nouvelles technologies susceptibles de répondre à ce besoin. En l'occurrence, les techniques d'IA sont largement utilisées pour le traitement de ce type de données, en particulier parce que les diverses caractéristiques spécifiques de ces données, telles que le volume, la vitesse et la variété, exigent un traitement intelligent. Par exemple, nous pouvons placer un agent intelligent dans un environnement spécifique tel que les réseaux sociaux et c'est à lui de capter l'information, de la traiter et même d'agir de manière autonome. Les principales caractéristiques de ces données (5v) sont décrites ci-dessous [Chandarana et Vijayalakshmi \[2014\]](#) :

Volume : ce concept désigne la taille exponentielle des données, il s'agit des pétaoctets ou zettaoctets. Ces données nécessitent habituellement des machines avec une capacité considérable pour assurer le stockage. Dans ce contexte, en janvier 2013, Twitter a annoncé que les données générées sur sa plateforme par jour dépassaient les 7 téraoctets et les données sur Facebook dépassaient les 10 téraoctets ¹⁹.

Vélocité : ce concept désigne la vitesse avec laquelle les données sont créées c'est-à-dire que ces données sont à la fois générées, capturées, partagées et mises à jour en quasi-temps réel. Selon le site *SocialMediaToday*, en 2020, 1,7 Mo de données sont créées chaque seconde pour chaque personne dans le monde sur des plateformes en ligne exemple (NETFLIX, Skype, Google, Twitter, etc.) ²⁰.

Variété : il s'agit d'une combinaison de données complexes provenant de sources multiples (web, base de données relationnelles, les objets connectés, etc.). Cette diversité concerne le format de données qui peut être structuré (comme sur les sites Web de commerce électronique), semi-structuré (comme les données des journaux sur le Web) et non structuré (comme les sites Web sociaux contenant des données au format audio, vidéo, image, etc.).

Véracité : ce concept s'attaque au problème des données imprécises découlant du fait que les données sont extraites de sources multiples et qu'elles serviront à la prise de décisions. Pour cela, il faut tenir compte de la fiabilité et de la dimension qualitative de ces données en gérant l'incertitude et les erreurs observées dans certaines données afin de garantir la fiabilité et de réduire les biais. Ceci peut être assuré par diverses techniques, telles que l'analyse des rumeurs dans les réseaux sociaux.

Valeur : l'un des défis du traitement des données volumineuses et variées réside dans la façon dont nous pouvons rassembler ces données afin d'extraire de nouvelles connaissances pour les entreprises, ce qui représente la valeur du Big Data. Exemple : Réalisation de tableaux de bord afin de privilégier les indicateurs les plus qualitatifs permettant d'identifier les influenceurs et les prescripteurs autour d'une marque ou d'un produit.

1.8.7 Vers le Web sémantique 3.0

En 1991, le monde a assisté à l'apparition du Web 1.0, qui se réfère à la première version du Web : la navigation sur Internet en mode lecture seule. Il était connu sous le nom «Informational Mono-directional Web». Le Web 1.0 était une source de renseignements provenant d'un nombre limité d'auteurs et d'un grand

19. https://fr.wikipedia.org/wiki/Big_data#cite_note-45

20. <https://www.socialmediatoday.com/news/how-much-data-is-generated-every-minute-infographic-1/525692/>

nombre d'utilisateurs. Cela signifie qu'il a fonctionné principalement comme une bibliothèque de travaux de référence. Et en 2004, une nouvelle version du Web est mise en place, appelée Social Web 2.0, qui permet aux utilisateurs de créer des liens entre eux. Ces liens leur permettent de communiquer et de collaborer de façon active afin que le contenu puisse être facilement généré et publié. Cela a donné lieu à de nouvelles activités sur Internet. En 2006, John Markoff a suggéré une nouvelle version du Web nommée le web sémantique qui est la version évolutionnaire actuelle du Web. Cette version consiste à transformer les métadonnées en un ensemble de réponses intelligentes pour des opérateurs [Wu et al. \[2015\]](#). En outre, elle a amélioré la gestion des données, soutenu la mondialisation et la collaboration dans le domaine du Web social et elle a accru l'accessibilité de l'Internet mobile. De plus, elle a comporté des propriétés sophistiquées, incitant la machine à comprendre certains renseignements à l'aide de techniques de traitement du langage naturel et d'exploration de données dans le but de rendre le Web lisible par des machines et non seulement par l'homme [Aziz et Madani \[2015\]](#). De manière générale, nous pouvons définir le Web 1.0 comme le Web en mode lecture seule, puisqu'il s'intéresse seulement au contenu créatif des producteurs, contrairement au Web 2.0, qui vise à créer du contenu par les utilisateurs et les producteurs en même temps. La version Web 3.0 est la version qui traite les ensembles de données sémantiquement couplées.

1.9 CONCLUSION

Dans ce chapitre, nous nous sommes concentrés sur la présentation de concepts de base ayant une forte relation avec le concept de la personnalité. Pour ce faire, nous avons d'abord défini les termes personnalité et trouble de la personnalité. Nous avons ainsi distingué entre les différents types de personnalité à savoir extravertie, stable, agréable, consciencieux, ouverte ainsi que les catégories du trouble de la personnalité à savoir distant, impulsif et effacé. Ensuite, nous avons exposé le phénomène du trouble de la personnalité en Tunisie et les problèmes rencontrés dans la lutte contre cette maladie. Par la suite, nous avons défini les principaux traits stylistiques dans la littérature pour l'identification de la personnalité à savoir les traits lexicaux, morpho-syntaxiques, sémantiques et numériques. Finalement, nous avons décrit le rôle de l'intelligence artificielle dans la détection des éléments en rapport avec la personnalité. Après avoir exposé notre contexte d'étude, nous allons présenter, dans le chapitre deux les différentes méthodes proposées dans la littérature en vue de résoudre la problématique énoncée qui consiste à détecter et à suivre les personnes atteintes de troubles de la personnalité sur les réseaux sociaux.

ÉTAT DE L'ART DES MÉTHODES D'ANALYSE DE LA PERSONNALITÉ SUR LES RÉSEAUX SOCIAUX

2

SOMMAIRE

2.1	INTRODUCTION	28
2.2	DÉTECTION DES TRAITS DE PERSONNALITÉ	30
2.2.1	Méthodes basées sur les techniques classiques d'apprentissage au- tomatique	30
2.2.2	Méthodes basées sur l'apprentissage profond	31
2.2.3	Méthodes basées sur les techniques de traitement linguistiques	33
2.2.4	Méthodes basées sur des techniques statistiques	34
2.2.5	Méthodes hybrides	34
2.3	DÉTECTION DES TROUBLES DE PERSONNALITÉ	35
2.3.1	Manière explicite	37
2.3.2	Manière implicite	42
2.4	DÉTECTION DES LIENS ENTRE LES ÉLÉMENTS DE LA PERSONNALITÉ	43
2.5	SURVEILLANCE DE L'ÉTAT DU PATIENT	46
2.6	DISCUSSION	47
2.6.1	Détection des traits de personnalité	47
2.6.2	Détection des troubles de personnalité	48
2.7	CONCLUSION	49

2.1 INTRODUCTION

L’intérêt de l’étude des traits de personnalité s’est déclenché depuis longtemps. Cependant à l’aube de l’année 1980, les chercheurs ont réinventé ce concept en tirant profit des progrès de la puissance de calcul des machines. Ces chercheurs ont fondé leurs hypothèses sur une démarche lexicale permettant d’analyser des scores de personnalité à partir d’un ensemble de données textuelles. Pour ce faire, ils ont eu recours à des ensembles d’adjectifs descriptifs reflétant les traits [Soto et Jackson \[2013\]](#). Jusqu’en 1990 plusieurs recherches basées sur l’étude expérimentale ont vu le jour dans le but de profiter de l’intérêt de l’étude de la personnalité dans plusieurs domaines citant dans ce contexte [Davis et al. \[1990\]](#). Ce travail a proposé une étude expérimentale des types de personnalité en réaction avec la stratégie décisionnelle dans un environnement commercial en se basant sur les critères de perception de Myers-Briggs¹ qui regroupent la mesure de sensation, d’intuition, de pensée, de sentiment, de jugement et de perception. Dans le même contexte, [Murray et al. \[1990\]](#) ont proposé une étude empirique sur les effets interactifs des traits de personnalité des étudiants et des différentes méthodes d’enseignement pédagogiques en fonction des résultats des étudiants. Cela est réalisé en mesurant l’indicateur de type Myers-Briggs (MBTI) sur les résultats des examens des étudiants. Cette mesure est réalisée tout en tenant compte d’autres informations telles que l’aptitude et l’attitude académique des étudiants.

Depuis 2005, les chercheurs ont pensé à tirer profit des outils de TALN pour étudier les éléments relatifs à la personnalité à partir des données issues des réseaux sociaux. Jusqu’à présent, plus de 300 000 articles sur Google Scholar comprennent les mots personnalité, social network et TALN. Ce qui rend ce nombre énorme est le fait que de nombreuses combinaisons logiques peuvent être associées entre les différents concepts : (i) l’étude de la différence du sexe sur la personnalité et la connaissance sociale [Pearce et al. \[2019\]](#), (ii) l’exploration des attitudes des usagers de Weibo à l’égard des minorités présentant des orientations différentes [Zheng et al. \[2022\]](#), (iii) la détection de la radicalisation et de l’extrémisme en ligne [Tanoli et al. \[2022\]](#), (iv) l’étude corrélationnelle entre le style de contenu textuel et la réussite des idées [Boghrati et al. \[2022\]](#).

Dans ce chapitre, nous visons à fournir une étude de l’état de l’art des travaux fondés dans ce domaine afin de montrer les points forts et les limites de chaque méthode. Nous détaillons dans la Figure 2.1 notre classification des différents travaux liés à ce domaine (avec AP : Analyse psychologique, DTP : Détection des traits de la personnalité, DPP : Détection des problèmes psychologiques, DCP : Détection des liens entre les composantes de la personnalité, SEP : Surveillance de l’état du patient). La présente classification a pour but de répondre à ces quatre questions principales : (i) quelles sont les méthodes suggérées pour prédire les traits de personnalité à partir des réseaux sociaux?, (ii) quelles sont les tendances de la recherche dans le domaine de la détection des troubles de la personnalité au cours de ces dernières années?, (iii) quelles sont les variables qui se rapportent à l’état de la personne et qui servent à extraire de nouvelles connaissances?, (iv) comment prendre en charge la santé mentale du patient sur les médias sociaux?

1. un outil d’évaluation psychologique qui sert à déterminer le type psychologique d’un sujet parmi seize types différents, selon une méthode proposée en 1962 par Isabel Briggs Myers et Katherine Cook Briggs.

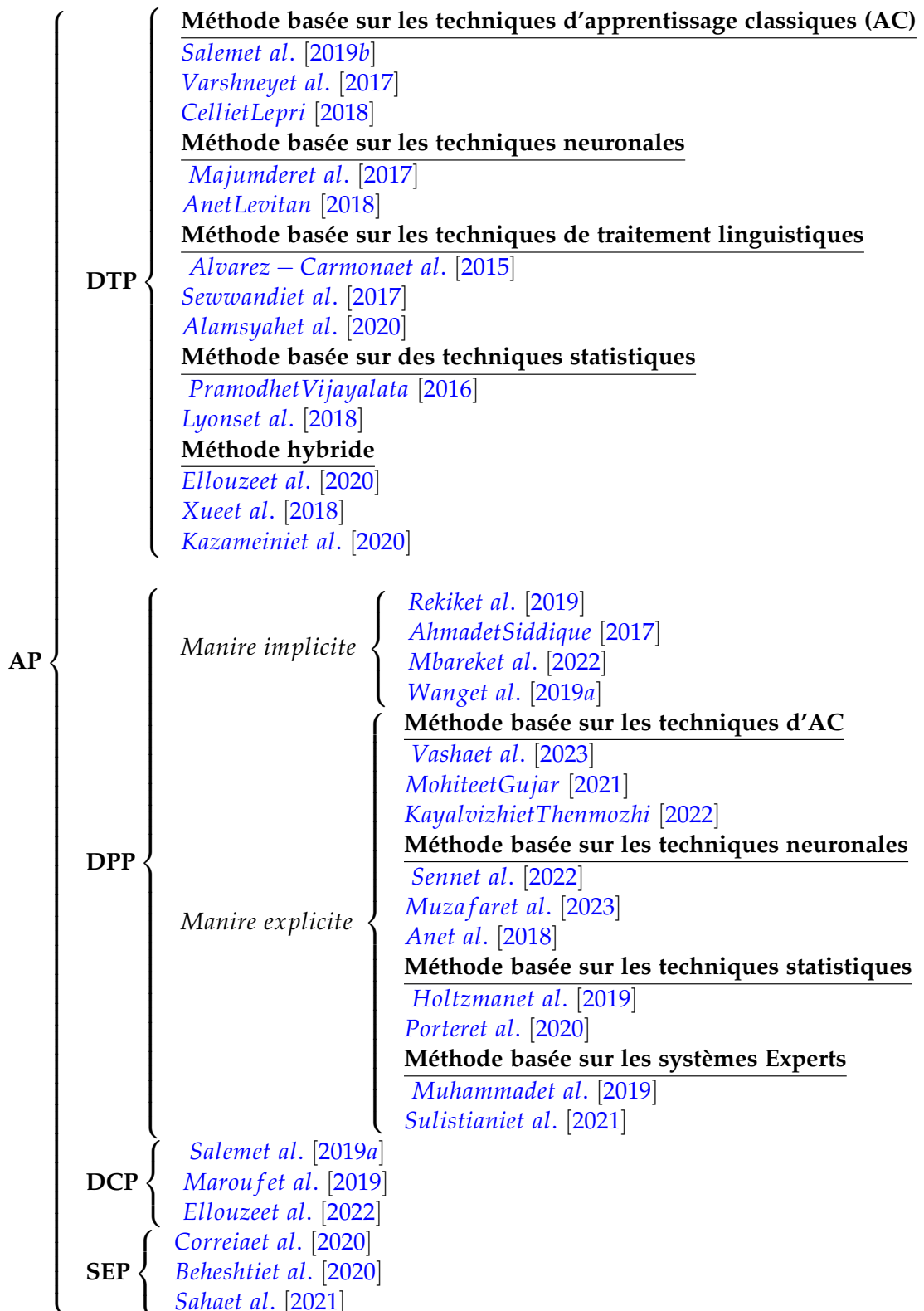


FIGURE 2.1 – Synthèse des méthodes d'analyse des traits et des troubles de la personnalité.

Nous commençons par la présentation des méthodes dédiées à l’étude des traits de personnalité. Puis, nous nous focalisons sur la description des travaux de recherche traitant des problèmes liés à la psychologie ainsi que sur les travaux qui ont envisagé d’assister les personnes atteintes de troubles de la personnalité. De même, nous présentons dans ce chapitre les études qui traitent les relations de dépendance et de corrélation entre les concepts de la personnalité. Enfin, nous clôturons le présent chapitre par une discussion et une conclusion.

2.2 DÉTECTION DES TRAITS DE PERSONNALITÉ

Dans la section de détection des traits de personnalité, nous effectuons un survol des différentes méthodes de la littérature permettant l’identification des traits de personnalité chez les utilisateurs des réseaux sociaux. Nous distinguons principalement les méthodes basées sur les techniques linguistiques, sur l’apprentissage classique, profond et hybride.

2.2.1 Méthodes basées sur les techniques classiques d’apprentissage automatique

Identifier les traits de personnalité à l’aide d’une méthode d’apprentissage automatique est une technique largement utilisée par plusieurs chercheurs. La variance entre leurs travaux se situait essentiellement au niveau : (i) de la source de données, (ii) du langage des données, (iii) des caractéristiques extraites, (iv) de l’algorithme d’apprentissage automatique utilisé. Par exemple, [Salem et al. \[2019b\]](#) se sont focalisés sur la détection des traits de la personnalité des personnes à partir de leur production textuelle en dialecte égyptien sur Twitter. Ce travail a été effectué en calculant le poids de chaque terme (unigramme, bigramme, trigramme) par rapport à chaque trait de personnalité en tirant profit de la mesure TF-IDF². Les poids calculés sont par la suite utilisés par l’algorithme KNN pour effectuer la tâche de classification. Dans le même contexte, [Varshney et al. \[2017\]](#) ont considéré la mesure TF-IDF comme un moyen permettant l’extraction des caractéristiques spécifiques liées à la personnalité. Cependant, la différence avec le travail de [Salem et al. \[2019b\]](#) réside principalement au niveau de l’utilisation des sources multi-variées (Twitter, Facebook, etc.), afin de généraliser le lexique utilisé. La deuxième différence réside dans la combinaison des résultats des trois algorithmes d’apprentissage (SVM, KNN, Naïve Bayes) pour tirer parti des avantages des différentes méthodes et remédier à leurs inconvénients. Dans un contexte similaire [Mohammad et Kiritchenko \[2013\]](#), ont bénéficié des nuances émotionnelles pour identifier la personnalité des personnes. Pour ce faire, ils ont construit un lexique reflétant les émotions de chaque catégorie telle que *excitation*, *culpabilité*, *désir*, *admiration*, etc. Ces caractéristiques ont été par la suite utilisées pour effectuer l’étape de classification en se basant sur l’algorithme SVM. Il est à signaler que les différents travaux présentés ci-dessus ont utilisé seulement les attributs de type lexical. Néanmoins, il existe d’autres travaux qui ont profité d’une variété dans les types de caractéristiques utilisées. Citant dans ce

2. Fréquence terme - fréquence inverse document : est un système de pondération qui priorise le nombre d’apparence d’un terme (terme-fréquence) dans un document.

contexte, [Park et al. \[2015\]](#) qui a tiré l’objet de plusieurs types de caractéristiques (thèmes, fréquence des expressions utilisées, caractéristiques linguistiques) afin de créer un modèle de régression permettant de classer le texte des utilisateurs de Facebook en 5 classes : *Névrosisme*, *Ouverture*, *Agréabilité*, *Conscienciosité*, *Ex-traversivité*. Dans le même contexte, le travail [González-Gallardo et al. \[2015\]](#) est fondé sur la combinaison des critères lexicaux sous forme de n-grammes avec leurs catégories grammaticales, afin de les utiliser par la suite par l’algorithme SVM dans la tâche de classification.

Les informations de métadonnées telles que *le nombre d’amis*, *le nombre de publications*, *etc.*, sont également un type de critère utile pour indiquer la classe des traits de personnalité. À cet effet, plusieurs travaux ont pris en considération ce type de donnée en tant que critère de distinction comme dans [Celli et Lepri \[2018\]](#), qui visait à combiner ce type d’information avec les critères linguistiques : (i) n-grammes, (ii) LIWC³ (*amitié*, *famille*, *sexe*, ...), (iii) les métadonnées (*le nombre de hash-tag*, *le taux de retweet*...). Cette étude ne s’est pas limitée à faire la classification des traits de personnalité uniquement, mais aussi à détecter le type de personnalité (MBTI) afin de connaître le style de communication et la manière dont une personne interagit avec les autres. La classification de Big5⁴ et de MBTI a été faite avec l’algorithme SVM. De même, les auteurs de [Grivas et al. \[2015\]](#) ont travaillé sur la combinaison de plusieurs types de caractéristiques : (i) des critères structurels tels que *le nombre de liens*, *le nombre de hashtags*, *le nombre de mentions d’autres utilisateurs*, *etc.*, (ii) des critères de stylométrie tels que *la longueur de chaque mot*. Puis, en utilisant l’algorithme SVM, ils ont trouvé le meilleur modèle qui permet de combiner tous ces attributs et de prédire le choix adéquat. En plus des méthodes présentées ci-dessus, qui reposent sur les techniques classiques d’apprentissage automatique, il existe d’autres méthodes axées sur les techniques d’apprentissage profond.

2.2.2 Méthodes basées sur l’apprentissage profond

L’apprentissage profond est un type de méthode d’apprentissage automatique qui s’appuie sur un réseau de neurones artificiels inspirés par le fonctionnement du cerveau humain. L’apprentissage profond représente une combinaison de plusieurs couches de traitement qui peuvent apprendre les représentations des données à divers niveaux d’abstraction [Liu et al. \[2020\]](#). La majeure différence de deep learning par rapport aux techniques d’apprentissage classiques réside principalement dans l’autoextraction des critères pertinents depuis les données brutes telles que les données visuelles, audio, textuelles [Mehta et al. \[2020\]](#). L’idée de base est de transférer les caractéristiques extraites depuis le format brute des données en utilisant différents outils tels que *senticNet* [Cambria et al. \[2016\]](#) vers un ensemble de couches successives qui forment un réseau neuronal profond. Dans ce contexte, [Majumder et al. \[2017\]](#) a bénéficié de la technique de plongement lexical (*wordEmbedding* ou *word2Vec*) [Mikolov et al. \[2013\]](#) pour réaliser la tâche d’extraction des attributs ainsi d’une architecture composée de plusieurs couches neuronales pour assurer la classification.

3. Linguistic Inquiry and Word Count.

4. Psychologie des traits de personnalité

Dans le même contexte, une autre étude réalisée par [An et Levitan \[2018\]](#) a présenté deux modèles d’apprentissage profond pour la tâche de reconnaissance de la personnalité à partir des données textuelles provenant de Facebook et des données acoustiques (SAE⁵ et MC⁶). Les deux modèles utilisés sont : (i) MLP (perceptron multicouche) en utilisant la fonction d’activation softmax. La figure 2.2 donne un aperçu détaillé du modèle proposé dans ce contexte, (ii) LSTM (Long Short Term Memory) qui est largement utilisé pour capturer les informations séquentielles [Hochreiter et Schmidhuber \[1997\]](#), [Xian et Tian \[2019\]](#). La figure 2.3 présente en détail le modèle WE-LSTM proposé. Les différentes fonctionnalités utilisées dans cette méthode incluent : (i) LLD (Low-Level Descriptor) qui contient environ 384 caractéristiques prosodiques acoustiques, (ii) LIWC, (iii) DAL⁷ qui est un outil d’analyse lexicale utilisé pour analyser le contenu émotionnel, (iv) Word embedding qui est utilisé pour faire une représentation vectorielle de la partie textuelle, (v) DNN représente la couche du réseau neuronal profond (*deep neural network*) et la couche sortie des deux figures représente les différents traits de la personnalité.

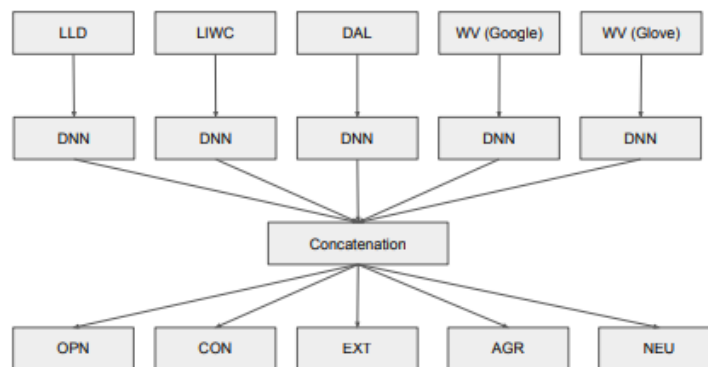


FIGURE 2.2 – Diagramme du modèle MLP pour découvrir les traits de la personnalité [An et Levitan \[2018\]](#).

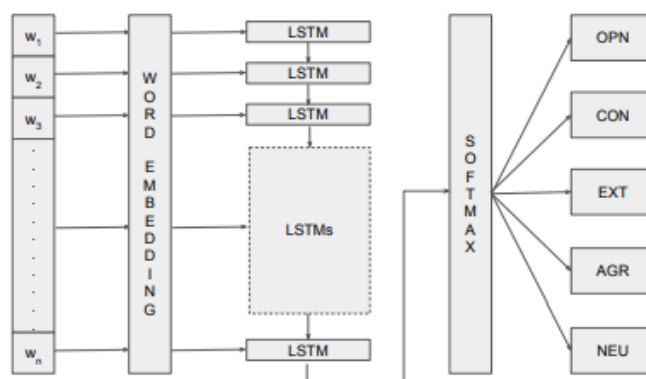


FIGURE 2.3 – Présentation du diagramme du modèle WE-LSTM [An et Levitan \[2018\]](#).

En plus des méthodes fondées sur des techniques d’apprentissage classiques

5. Standard American English
6. Mandarin Chinese
7. Dictionary of Affect in Language

et profonds, plusieurs études ont eu recours à des méthodes linguistiques pour mettre en valeur cet aspect dans la production textuelle des utilisateurs.

2.2.3 Méthodes basées sur les techniques de traitement linguistiques

La linguistique est une discipline axée sur l’étude des particularités d’une langue. Dans ce contexte, [Alvarez-Carmona et al. \[2015\]](#) ont traité la prédiction des traits de personnalité en combinant les deux techniques SOA [López-Monroy et al. \[2013\]](#) et LSA [Daneshvar et Inkpen \[2018\]](#). L’objectif de SOA est de capturer les plus fins détails de la grande diversité dans les documents (les sous-groupes des utilisateurs) afin de détecter l’homogénéité entre les mêmes classes. LSA est une méthode qui est employée pour extraire et représenter la signification contextuelle des mots dans le corpus. L’idée principale de ce travail est de valoriser les propriétés stylistiques et thématiques du document. Une autre technique de reconnaissance de la personnalité des utilisateurs sur les réseaux sociaux est l’ontologie. Plusieurs travaux comme [Sewwandi et al. \[2017\]](#), [Alamsyah et al. \[2020\]](#) ont employé cette technique afin de détecter la personnalité des personnes, en se basant sur la catégorisation des mots par classe. L’ontologie construite par [Alamsyah et al. \[2020\]](#) permet de décrire la personnalité de manière détaillée en assignant à chaque tweet un trait de personnalité et un effet tel que *chaleur*, *grégarité*, etc. L’ontologie citée a été construite à l’aide d’un algorithme qui est basé sur la comparaison entre la similarité des mots par catégorie, ainsi sur l’application d’une mesure de distance statistique. La figure 2.4 présente un extrait de l’ontologie proposée avec certaines catégories de mots.

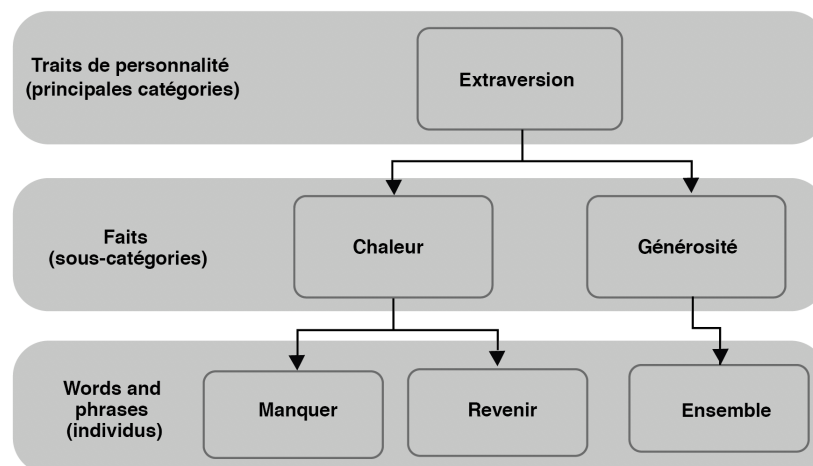


FIGURE 2.4 – Extrait de l’ontologie proposé en vue de la détection des traits de personnalité [Alamsyah et al. \[2020\]](#).

En raison des problèmes associés à la méthode linguistique tels que le manque de ressources et de règles linguistiques, de nombreux chercheurs ont utilisé di-

verses méthodes statistiques pour profiter de l’énorme quantité de données sur les réseaux sociaux.

2.2.4 Méthodes basées sur des techniques statistiques

Les techniques statistiques peuvent également aider à prédire les traits de personnalité. Dans ce contexte, [Pramodh et Vijayalata \[2016\]](#) ont cherché à classer les traits de personnalité d’une manière différente par rapport aux autres travaux présentés. L’idée de base est de construire un lexique positif et un autre négatif pour chaque trait de personnalité, tout en prenant en considération la négation dans ce cas, le mot changera de polarité. Enfin, les auteurs ont effectué un calcul d’appartenance de chaque tweet à chaque classe. De même, dans [Lyons et al. \[2018\]](#) les auteurs ont utilisé une méthode statistique, mais dans ce cas pour extraire des connaissances liées au style d’écriture des personnes ayant des troubles de personnalité. Ce travail a été accompli en appliquant le logiciel SPSS⁸ aux données textuelles des personnes souffrant de troubles de la personnalité. Cette étude a identifié les critères linguistiques, y compris les composantes émotionnelles, cognitives et structurelles, utilisés par les individus ayant des troubles de la personnalité. De même dans [Sagadevan et al. \[2015\]](#), les auteurs ont profité d’une méthode statistique afin de réaliser une étude approfondie sur certains critères psychologiques, démographiques et linguistiques relatifs au style d’écriture en utilisant principalement la mesure Wilk’s lambda⁹. Les résultats de cette étude ont montré qu’en général, les utilisateurs masculins ont attribué plus d’intérêt aux mots d’extraversion que les femmes. En revanche, les femmes ont utilisé beaucoup plus de mots de névrosisme que les utilisateurs masculins. D’autres études de recherche ont été menées dans le but d’effectuer une étude des effets des traces numériques recueillies à partir des réseaux sociaux pour prédire des caractéristiques psychologiques et comportementales à l’aide de coefficients de corrélation, tels que celui de Pearson¹⁰, comme dans l’étude de [Settanni et al. \[2018\]](#). Le résultat de ce travail a montré une véritable signification, en ce qui concerne l’association entre ces traces, la personnalité et le bien-être. Ces associations peuvent être utilisées pour diverses tâches liées au suivi de l’état mental des usagers, notamment pour la détection précoce des personnes à risque de dépression ainsi pour la surveillance de l’évolution de l’état de santé mentale.

L’étude des différentes méthodes présentées montre que chaque méthode comporte des limites telles que la nécessité d’un grand volume de données ainsi que des variétés de données pour les méthodes qui sont basées sur des techniques statistiques. Pour cela, plusieurs chercheurs ont employé dans leurs études plus d’une technique pour combiner leurs bénéfices et surmonter leurs inconvénients.

2.2.5 Méthodes hybrides

La méthode hybride consiste à combiner au moins deux parmi les techniques décrites ci-dessus. Par exemple, les auteurs de [Ellouze et al. \[2020\]](#), [Hassanein](#)

8. Statistical Package for the Social Sciences : est un logiciel sophistiqué d’analyse et de traitement de données conçu par IBM.

9. est une mesure de la capacité de chaque fonction à séparer les cas en groupes.

10. https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

et al. [2018] ont présenté une étude basée sur l’hybridation des deux techniques dans le but de déduire les traits de la personnalité. L’idée de base de ces travaux est de combiner : (i) une méthode lexico-sémantique reposant sur la technique de plongement lexical (*word embedding*) et la base de connaissances WordNet, (ii) une mesure statistique permettant de calculer la similarité entre les vecteurs générés après la phase de prétraitement et d’annotation. Dans Ellouze et al. [2020], les auteurs ont profité de la notion de logique floue pour aborder les limites des techniques d’apprentissage automatique qui reposent sur une méthode déterministe (le degré d’appartenance d’une instance à une classe est de 100%). De même, Poria et al. [2013] ont proposé une méthode hybride pour la détection de traits de la personnalité à partir d’un ensemble de textes en utilisant : (i) la ressource lexicale EmoSenticnetSpace¹¹ qui représente un lexique émotionnel lié à divers aspects (émotionnel, cognitif, social, etc.). Cette ressource est construite à l’aide de la base de connaissances SenticNet Cambria et al. [2016] pour l’analyse des sentiments, la ressource du conseil de recherches médicales (MRC) pour l’étude des terminologies psychologiques, EmoSenticNet Mulholland et al. [2016] pour profiter des ressources lexicales émotionnelles, etc., (ii) un modèle de classification construit à l’aide de l’algorithme d’apprentissage automatique SVM. De même, les auteurs de Xue et al. [2018] ont proposé une méthode basée sur l’apprentissage profond pour la reconnaissance de la personnalité à partir des messages des réseaux sociaux. Ce travail est divisé en trois parties : (i) construction d’une nouvelle structure de réseau neuronal nommée selon les auteurs AttRCNN pour apprendre la représentation sémantique distribuée du message textuel, (ii) modélisation d’un nouveau réseau de neurones hiérarchiques profonds que les auteurs ont renommé AttRCNN-CNNs, ce modèle combine à la fois une structure du modèle AttRCNN et une structure basée sur la couche CNN, (iii) construction d’une méthode de reconnaissance de la personnalité.

Dans le même contexte, Kazameini et al. [2020] ont profité du modèle *deep learning-based Bagged-SVM* qui combine à la fois le modèle de langage BERT Devlin et al. [2018] qui consiste à présenter la partie textuelle dans un ensemble de segments et le classifieur Bagged SVM Mordelet et Vert [2014]. Le tableau 2.1 récapitule les principaux travaux cités dans le cadre de la détection des traits de personnalité. Comme notre aperçu montre, l’étude de la personnalité représente un axe de recherche très important et enrichissant. Mais afin de fournir plus d’intérêt à l’impact social, nous avons procédé à une étude sur les maladies psychologiques liées au trouble de la personnalité.

2.3 DÉTECTION DES TROUBLES DE PERSONNALITÉ

La détection des problèmes psychologiques et leur gravité (*suicide, violence, harcèlement*) sur les réseaux sociaux ont fait l’objet de nombreuses études. Selon notre vue d’ensemble de l’état de l’art, les études analysées dans ce contexte peuvent être réparties en deux volets : voie explicite (détection des troubles psychologiques) et voie implicite (détection des conséquences des troubles psychologiques tels que *la violence et le suicide*).

11. <https://www.gelbukh.com/emosenicnet/>

TABLE 2.1 – La comparaison de différents travaux de détection de la personnalité.

Référence	Données	Méthode	Langue	Attribut	Résultat
Salem et al. [2019b]	Twitter (92 utilisateurs)	Apprentissage automatique pour la classification	Dialecte égyptien	Tf-Idf des n-gramme	32,6%
Varshney et al. [2017]	Facebook, Twitter, Google+ etc.	Apprentissage automatique pour la classification	Anglais	Tf-Idf	-
González-Gallardo et al. [2015]	Twitter (324 utilisateurs)	Apprentissage automatique pour la classification	Espagnol, Anglais, Italien, Néerlandais	N-gramme, POS des n-gram	83,46%
Celli et Lepri [2018]	Twitter (900 utilisateurs)	Apprentissage automatique pour la classification	Anglais, Italien, Espagnol, Néerlandais	N-gramme, LIWC, metadata	61,5%
Alvarez-Carmona et al. [2015]	Twitter (324 utilisateurs)	Linguistique	Espagnol, Anglais, Italien, Néerlandais	LSA+SOA	87,85%
An et al. [2018]	Facebook (9917 status, 250 users)	Neuronale pour la classification	Anglais	LLD, LIWC, DAL, WE ¹²	Accuracy 71%
Alamsyah et al. [2020]	6000 tweets	Linguistique	Anglais	lexicale	Démonstrateur
Ellouze et al. [2020]	Twitter (324 utilisateurs)	Hybride	Anglais	lexicale	Accuracy 82%

2.3.1 Manière explicite

La détection des problèmes psychologiques implique explicitement la détection du trouble de la personnalité lui-même ou la détection de ses symptômes. Pour assurer la détection de ce type d’information, de nombreux chercheurs ont proposé diverses méthodes basées sur différentes techniques telles que l’apprentissage automatique et profond, les techniques statistiques et les méthodes fondées sur des systèmes experts.

2.3.1.1 Méthodes basées sur les techniques classiques d’apprentissage automatique

Parmi les travaux qui ont profité des méthodes inscrites dans les techniques basées sur la notion d’apprentissage classique, il existe ceux de [Stankevich et al. \[2019\]](#) et [Vasha et al. \[2023\]](#). Ces travaux ont été menés dans le but d’assurer la détection de la dépression chez les utilisateurs des réseaux Vkontakte¹³ et Facebook à partir de l’analyse de leurs productions textuelles. Pour ce faire, [Stankevich et al. \[2019\]](#) a exploré plusieurs critères morphologiques tels que *le nombre de verbes, de pronoms, la forme infinitive, etc.* Tandis que [Vasha et al. \[2023\]](#) a profité des métriques statistiques permettant de mesurer l’importance des unités lexicales dans une collection de données. De même, [Kayalvizhi et Thenmozhi \[2022\]](#), [Chiong et al. \[2021\]](#) ont travaillé sur la détection de tromperie dans les réseaux sociaux à l’aide des techniques d’apprentissage supervisé. Les différents attributs utilisés dans ces travaux sont purement textuels tels que (TF-IDF), Matrice terme-document (TDM), Sac de mots (BOW), Word2Vec, etc. Dans le même contexte, [Mohite et Gujar \[2021\]](#) ont introduit une méthode de détection des troubles mentaux sur les réseaux sociaux à l’aide des techniques d’apprentissage automatique. La collection des données a été effectuée manuellement à partir de Facebook. Les différentes caractéristiques utilisées dans ce travail comprennent d’une part les interactions sociales entre les individus et d’autre part les caractéristiques personnelles telles que *l’honnêteté et la sincérité*. L’objectif de ce travail était de classer les profils des personnes en trois classes à l’aide de l’algorithme Naïve Bayes : 1. Dépendance à la relation cybernétique (CR), 2. Coercition nette (NC), 3. Surcharge d’information (IO). Les auteurs de [Singh et al. \[2019\]](#) ont travaillé aussi dans un contexte semblable en proposant une méthode basée sur des techniques de TALN et des techniques d’apprentissage automatique pour détecter les personnes antisociales à partir de leurs comportements sur Twitter. Après une étape de collecte de données à l’aide de la recherche d’un ensemble de mots clés indiquant une probabilité élevée de dépression en utilisant l’API Twitter. Une étape de transformation de ces données textuelles en une matrice est réalisée en tirant parti de la mesure TF-IDF. Ensuite, en utilisant l’algorithme forêts aléatoires (*Random Forest*), chaque tweet est classé comme antisocial ou non antisocial.

D’autres travaux ont porté sur l’extraction de style d’écriture des personnes ayant un problème psychologique à partir de l’analyse de leurs productions textuelles, comme dans le travail de [Holtzman et al. \[2019\]](#). Ce type d’analyse permet d’extraire de nouvelles connaissances sur la spécificité linguistique utilisée

13. le réseau social le plus utilisé en Russie

par les utilisateurs souffrant d’un trouble psychologique. Dans ce contexte, [Celli et al. \[2016\]](#) ont proposé une méthode alternative pour détecter les troubles de personnalité en analysant les données publiées sur les réseaux sociaux afin de : (i) détecter l’humeur d’une personne à l’aide de l’analyse des informations liées aux styles textuels, aux métadonnées et aux traits de personnalité. Ensuite, les auteurs ont combiné l’ensemble des caractéristiques extraites en utilisant une méthode basée sur l’apprentissage automatique, (ii) chercher les différents liens de corrélations entre chaque caractéristique et la classe de classification en utilisant différentes métriques statistiques. D’autres chercheurs ont eu recours aux progrès réalisés dans le domaine de l’apprentissage automatique pour détecter les problèmes psychologiques avec des techniques plus avancées.

2.3.1.2 Méthodes basées sur les techniques neuronales

Plusieurs chercheurs comme [Lin et al. \[2014; 2017\]](#) ont profité d’une méthode neuronale afin de détecter automatiquement le niveau de stress chez les usagers des réseaux sociaux. En commençant par une étape d’extraction des caractéristiques qui comprend : (i) les caractéristiques liées à un seul tweet telles que *le texte, les interactions sociales*, (ii) les caractéristiques numériques relatives à un utilisateur comme *le nombre d’amis et le style d’écriture figurant dans le tweet*. Ensuite, un modèle de réseau neuronal convolutif (CNN) avec des encodeurs automatiques a été élaboré pour combiner le contenu hebdomadaire de bas niveau et générer des attributs pertinents. Enfin, profiter d’un modèle de réseau de neurones (DNN) dans le but d’apprendre les caractéristiques de plus haut niveau et de détecter le degré de stress psychologique chez les utilisateurs. La figure 2.5 illustre l’architecture générale de la méthode proposée pour la détection du stress, comprenant les divers éléments présentés ci-dessus.

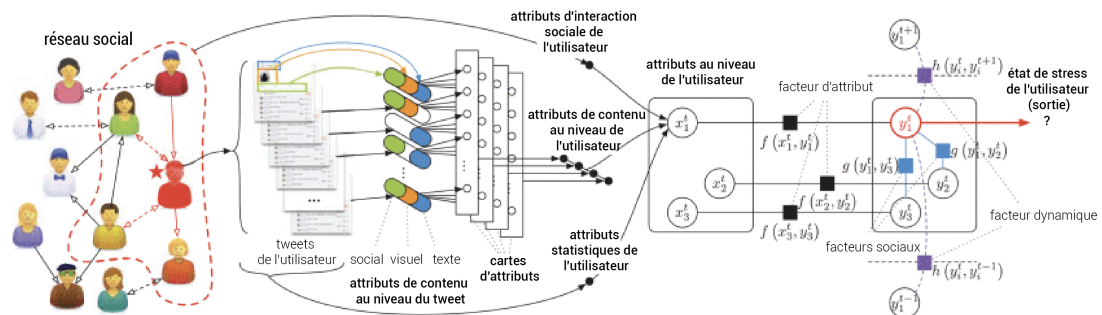


FIGURE 2.5 – Architecture générale pour la détection du stress [Lin et al. \[2017\]](#).

Dans un contexte similaire, [An et al. \[2018\]](#) ont visé à démontrer à quel point les traits de personnalité peuvent être utiles pour prédire la tromperie. Ce travail est effectué en suivant ce processus : (i) premièrement, détecter la déception chez les individus grâce à une méthode hybride qui combine l’output du modèle de perception multicouche (MLP) composé de 256 sorties formée à l’aide des attributs de LIWC, DAL, LLD et WE, avec l’algorithme LSTM. Les résultats obtenus sont ensuite transmis à un nœud de sortie qui applique la fonction d’activation sigmoïde pour prédire la probabilité de présence de la tromperie, (ii) incorporer la personnalité au classifieur de tromperie par les deux voies suivantes : (1) apprendre à prédire conjointement la personnalité et la tromperie du locuteur, la

figure 2.6 illustre en détail l’architecture utilisée pour assurer la prédiction de la personnalité et la déception, (2) alimenter les étiquettes de personnalité dans la couche de sortie du classifieur de tromperie, la figure 2.7 montre l’effet des étiquettes de personnalité sur la prédiction de la déception. La couche d’entrée des deux figures 2.6 et 2.7 contient les attributs qui ont été décrits précédemment. La couche de sortie de la première figure comporte les traits de personnalité et le degré de déception. La couche de sortie de la seconde figure contient uniquement le degré de déception, et les traits de personnalité sont considérés comme une entrée supplémentaire à ce modèle. Dans le même contexte, [Senn et al. \[2022\]](#), [Muzafar et al. \[2023\]](#) ont travaillé sur la détection de problèmes psychologiques tels que la tromperie et la dépression. Les résultats de ces études ont mis en évidence l’utilité de recourir à des modèles sophistiqués de transformation du langage, tels que le modèle de base BERT ou ses variantes telles que Roberta ou DistilBERT. De même, ces résultats démontrent l’intérêt porté à la combinaison de ces modèles afin de préserver autant que possible les particularités linguistiques. Ces études ont également abordé l’importance de l’intégration des couches de LSTM et d’attention pour mettre en évidence les caractéristiques contextuelles.

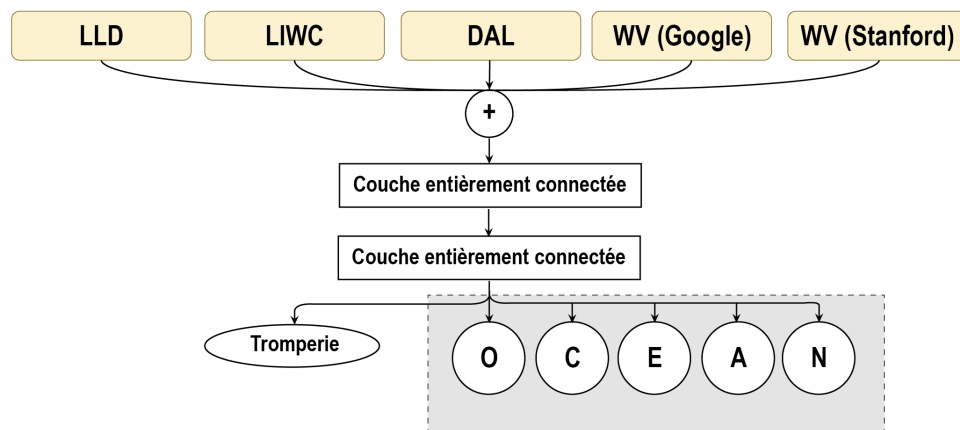


FIGURE 2.6 – Diagramme du modèle MLP pour l’apprentissage multitâches : prédire conjointement la personnalité et la déception [An et al. \[2018\]](#).

En général, plusieurs algorithmes d’apprentissage automatique rencontrent des difficultés sur le plan d’explicabilité fonctionnelle. En effet, certains de ces algorithmes sont considérés comme des boîtes noires. En général, ces algorithmes sont principalement basés sur des mesures statistiques. Pour cette raison, certains chercheurs ont opté pour un niveau d’architecture plus bas de ces méthodes. Selon ces chercheurs, ce choix permet d’atteindre une manipulation plus logique fondée sur des raisonnements spécifiques au domaine d’étude.

2.3.1.3 Méthodes basées sur les techniques statistiques

Dans ce contexte, [Holtzman et al. \[2019\]](#) ont focalisé sur l’identification des marqueurs linguistiques utilisés par les personnes narcissiques. En commençant, par la sélection des caractéristiques de LIWC. Après cela, utiliser la mesure de

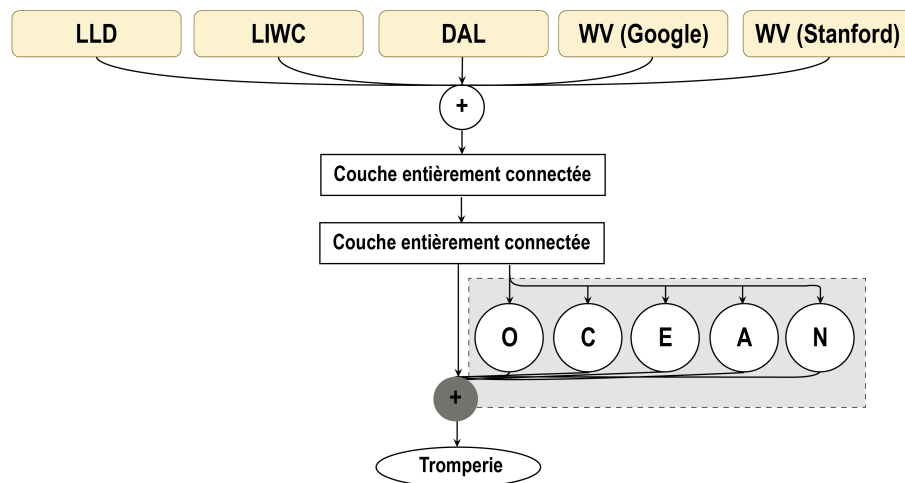


FIGURE 2.7 – Diagramme du modèle MLP pour l’apprentissage multitâches : alimentation des étiquettes de personnalité dans la couche de sortie [An et al. \[2018\]](#).

Pearson pondéré¹⁴ afin de calculer l’effet de l’estimation de chacune des caractéristiques de LIWC sur la présence du trouble narcissique à l’aide du package métaphorique existant dans le langage R. Parmi les résultats démontrés par ce travail, on constate que la personne narcissique emploie des mots plus liés au sport. De plus, ce travail a révélé certaines corrélations négatives entre les caractéristiques du LIWC et le trouble narcissique, comme celle de l’utilisation de termes liés à l’anxiété et à la peur. Les différentes analyses appliquées dans cette étude permettent de distinguer les caractéristiques lexicales utilisées dans un vocabulaire ouvert en fonction du type de personnalité.

Dans un autre travail [Schwartz et al. \[2013\]](#), les auteurs ont tiré parti des mesures statistiques et des techniques de visualisation afin de mesurer et de démontrer la corrélation entre les critères linguistiques tels que le lexique indiquant le topic et la personnalité. Parmi les résultats obtenus par ce travail, on constate que : (1) les personnes les plus nerveux utilisent davantage des expressions liées à la dépression et à la peur comme *malade*, *déprimé*, (2) une vie active implique une stabilité émotionnelle. De plus, les méthodes statistiques ont été utilisées dans d’autres types d’études analytiques, telles que l’arrêt prématuré de la psychothérapie chez les patients présentant un trouble de la personnalité borderline [Gamache et al. \[2018\]](#). L’objectif de cette étude est d’identifier les patients atteints de trouble de la personnalité borderline présentant un risque d’interruption précoce de leur psychothérapie, afin d’adapter des interventions visant à réduire ces interruptions.

D’autres chercheurs ont étudié la distribution des maladies psychologiques selon des critères démographiques. Par exemple, [Porter et al. \[2020\]](#) ont proposé une étude analytique afin de mieux comprendre l’ampleur et la cohérence de l’association entre l’adversité de l’enfance et le trouble de la personnalité borderline. En général, les méthodes statistiques reposent principalement sur l’analyse de l’occurrence de différentes caractéristiques des données et ignorent l’aspect sémantique au niveau du traitement réalisé. Cela peut conduire à des résultats non généralisables, car les différentes interprétations tirées sont principalement

14. est une mesure statistique qui mesure la corrélation linéaire entre deux variables.

liées aux spécificités des données d'entraînement qui dépendent de plusieurs critères tels que le cadre spatiotemporel, les facteurs sociaux, etc. Pour cette raison, plusieurs chercheurs ont choisi d'utiliser des méthodes reposant sur des systèmes experts afin de préserver les liens logiques entre les divers éléments de la psychologie.

2.3.1.4 Méthodes basées sur les systèmes Experts

Plusieurs chercheurs se sont servis d'une méthode basée sur les systèmes experts pour modéliser des règles et des connaissances liées aux concepts de la psychologie. Dans ce contexte, [Muhammad et al. \[2019\]](#) ont défini une architecture basée sur un système expert sous la forme de question/réponse pour le diagnostic des cas présentant un trouble bipolaire. Ce système utilise des codes pour présenter les différents symptômes, les types de maladies et les traitements à suivre. Ensuite, les auteurs de ce travail ont défini un ensemble de règles obtenues à partir d'un ensemble de questions posées à un psychiatre tout en prenant en considération la liste des codes de symptômes. Ces règles permettent de prévoir le diagnostic et de suggérer le traitement à suivre. De même, [Umar et Qamar \[2019\]](#) ont travaillé sur la détection des troubles de la personnalité, à savoir le *trouble dépressif majeur*, le *trouble affectif bipolaire* et la *schizophrénie* à travers la formation d'une série de règles de décision (RST).

L'ensemble de données utilisé dans ce travail a été obtenu auprès de l'AFIMH¹⁵. Ces données contiennent 21 caractéristiques relatives au stress, à l'insomnie et au sommeil, etc. Ensuite, en utilisant l'algorithme Johnson existant dans le système Rosetta¹⁶, ils ont obtenu un ensemble de règles de décision sous la forme par exemple (*sentiment (anxiété) et fatigue (oui) = diagnostic (MDD)*).

Dans le même contexte, [Sulistiani et al. \[2021\]](#) ont souligné l'intérêt d'intégrer les concepts de l'intelligence artificielle (IA) dans les systèmes experts dédiés au diagnostic des troubles de la personnalité. Parmi les concepts recommandés par ces auteurs pour le diagnostic, on retrouve la technique de représentation des connaissances, les règles de raisonnement et la notion de logique floue. Cela permet de garantir un très haut niveau de fonctionnement qui répond aux nombreuses exigences fonctionnelles.

La détection des troubles psychologiques ne se limite pas à identifier uniquement la maladie ou ses symptômes, mais également à identifier les conséquences qui y sont associées. Ces conséquences se manifestent fréquemment dans notre vie quotidienne, telles que le suicide. Bien que nous puissions considérer cette conséquence comme une décision instantanée, elle est en réalité le résultat de l'accumulation de plusieurs problèmes chez une personne instable [Ferlatte et al. \[2015\]](#).

15. Arm Forces Institute of Mental Health

16. est une boîte à outils pour l'analyse de données tabulaires dans le cadre de l'exploration de données et de la découverte de connaissances en se basant sur la théorie des ensembles approximatifs.

2.3.2 Manière implicite

La détection des troubles de la personnalité de manière implicite, consiste à détecter les conséquences des troubles psychologiques tels que *la violence et le suicide*. Dans ce contexte, les auteurs de [Rekik et al. \[2019\]](#) ont profité des techniques d’exploration de données et des techniques de TALN pour détecter les tweets violents à partir du calcul du poids de n-grammes. Pour ce faire, un expert a sélectionné les profils des personnes extrémistes et les profils des personnes non extrémistes. Ensuite, ils ont construit à partir des profils sélectionnés deux types de lexiques. Chaque lexique est constitué d’un ensemble de mots simples et composés sous la forme de n-grammes tout en indiquant le degré d’appartenance d’un mot à ce lexique. Finalement, ils ont utilisé des mesures statistiques pour calculer à quel point un tweet appartient à chacune des classes. Dans un contexte similaire, [Ahmad et Siddique \[2017\]](#) ont pour objectif la détection des tweets inappropriés en les classant en quatre classes : *conformité, dominance, soumission et influence*. Le corpus utilisé dans ce travail est composé d’un ensemble de tweets, qui sont collectés à l’aide de certains mots-clés reflétant les différentes classes sélectionnées. Par la suite, une étape de classification a été effectuée à l’aide de l’outil RapidMiner [Hofmann et Klinkenberg \[2016\]](#). Le résultat de ce travail est une visualisation sous différentes formes de sortie (*graphe, arbre, etc.*), qui montre la distinction des mots simples et composés entre les classes. De manière générale, cette visualisation peut être employée pour diffuser des connaissances.

La détection des personnes ayant des idées suicidaires sur les réseaux sociaux est un axe de recherche évolutif. Plusieurs chercheurs ont travaillé sur ce domaine dans le but d’assurer la surveillance de ces personnes. Nous citons [Mbarek et al. \[2022\]](#) parmi ces travaux. Ce dernier a présenté une méthode visant à détecter ce type de profils de personnes en se basant sur l’analyse de l’historique de leurs publications sur Twitter à l’aide d’un algorithme de classification. Les caractéristiques proposées pour effectuer l’étape de classification sont liées à plusieurs types d’informations tels que : (1) les caractéristiques linguistiques comme *POS, mot fréquent, n-grammes, etc.*, (2) les caractéristiques émotionnelles comme *emojis, termes indiquant la dépression, etc.*, (3) les caractéristiques faciales comme *âge, cheveux, moustache, etc.*, qui sont extraites à partir de la photo de profil de l’utilisateur, (4) les éléments chronologiques comme *le nombre de publications par jour, par mois, etc.*, (5) les informations publiques comme *pays, etc.*

Les émotions peuvent également affecter négativement la personnalité, raison pour laquelle nous avons trouvé divers travaux sur la détection de ce type d’information. Nous pouvons citer le travail de [Wang et al. \[2019a\]](#) comme un exemple parmi de nombreux travaux trouvés dans ce contexte. Ce travail propose un arbre hiérarchique de réseaux neuronaux destiné à la classification des émotions. La première partie consiste à modéliser le document dans un arbre LSTM. Cet arbre rassemble les mots-clés avec leurs poids ainsi les relations entre eux afin de faire une analyse de la dépendance et de montrer la distribution du sujet. Ensuite, une couche de sortie *softmax* est utilisée pour la classification des émotions sociales. Par exemple, *les mots mauvais, en retard, coupable* peuvent déclencher une émotion de tristesse. La figure 2.8 montre en détail l’architecture générale de la méthode proposée, elle contient les différentes informations citées. Le tableau 2.2 présente

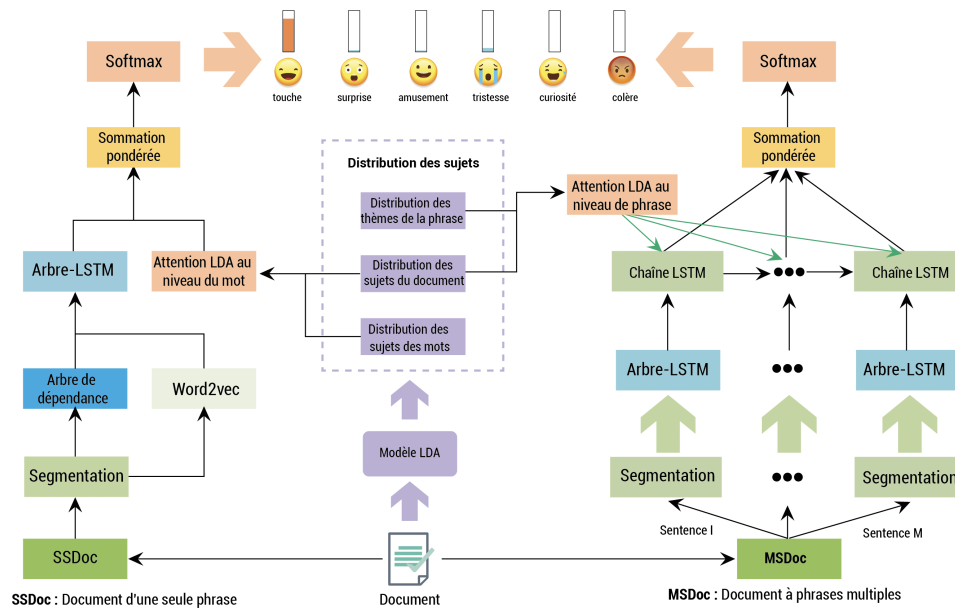


FIGURE 2.8 – Architecture basée sur Tree-LSTM network pour la détection des émotions Wang et al. [2019a].

une synthèse de certains travaux relatifs aux deux voies de détection des troubles psychologiques.

Plusieurs chercheurs qui se sont intéressés à la détection de la maladie, ont utilisé des techniques d'exploration des données dans le but d'extraire des liens de dépendances logiques cachées. Ces liens peuvent par la suite enrichir la base des connaissances des médecins.

2.4 DÉTECTION DES LIENS ENTRE LES ÉLÉMENTS DE LA PERSONNALITÉ

La recherche des liens entre les différents éléments liés aux maladies consiste à identifier les relations de dépendance (cause à effet) et les relations de corrélation entre ces éléments. Cette technique peut donner une idée à propos la raison de l'évolution d'une maladie. De plus, elle peut enrichir la base de connaissances des médecins. Le problème de cette technique repose sur la généralisation des résultats obtenus. À cet effet, l'analyse doit s'effectuer sur des données volumineuses et variées. Il est à signaler que parfois, les liens de dépendance obtenus dans certains travaux ne sont pas crédibles, car d'autres critères liés au pays, à l'âge, au genre, etc., peuvent intervenir et affecter négativement le résultat. Dans ce contexte, l'étude de Salem et al. [2019a] présente une méthode pour la recherche des liens entre les différents éléments liés aux troubles de la personnalité borderline et bipolaire. De même, elle présente une étude pour la prédiction des critères qui influent sur la réadmission. Les données utilisées dans la présente étude proviennent d'une étude menée auprès de 714 adultes, incluant leurs informations

TABLE 2.2 – Comparaison de divers travaux sur la détection de troubles psychologiques.

Référence	Corpus	Méthode	Langue	Attribut	Résultat
Rekik et al. [2019]	Twitter, Youtube	Méthode statistique de détection de la violence	Arabe	poids de n-gramme	96%
Mbarek et al. [2022]	Twitter (287 utilisateurs)	Apprentissage automatique pour la classification des extrémistes	Anglais	linguistique, émotionnel, facial, chronologie, information générale	85%
Stankevich et al. [2019]	application de Vkontakte (1020 profils)	Apprentissage automatique pour la classification de la dépression	Russe	critères morphologiques	71.72%
Holtzman et al. [2019]	Données recueillies auprès de Carey et al. [2015] >4500 échantillons	Méthode statistique pour étudier le style d’écriture des personnes narcissiques	Allemand, Anglais	LIWC	Intervalle de corrélation de confiance 95%
Lin et al. [2017]	DB1 : 492,676 tweets à partir de 23,304 utilisateurs	Neuronal pour la détection du stress	différentes langues	Attributs liés aux tweets	F-mesure 93,40%
Umar et Qamar [2019]	AFIMH (100 patients)	Construction des règles pour la détection des personnes : dépressive, bipolaire, schizophrénie	Anglais	stress, insomnie, sommeil, etc.	96%
Singh et al. [2019]	Twitter (55,000 tweets)	Apprentissage automatique pour la classification des anti-sociaux	Anglais	tf-idf	96.67%
Celli et al. [2016]	Twitter (2000 utilisateurs)	Apprentissage automatique pour la détection de l’humeur	Anglais	trait de personnalité, méta-données, style de communication	61.7%

démographiques et leurs réponses au BPQ¹⁷ pour mesurer la gravité de l'état de chaque patient. L'analyse de ces données consiste à faire un test de corrélation entre les différentes caractéristiques de la maladie. En outre, créer un modèle de régression pour tester les corrélations entre les scores de BPQ, la durée du séjour à l'hôpital et le taux de réadmission pendant une période de 30 jours. De plus, ce travail a été utilisé pour identifier les variables qui peuvent prédire le taux de réadmission du patient. Dans d'autres travaux, [Marouf et al. \[2019\]](#) ont traité la correspondance entre les traits de personnalité et le score de mesure du stress. L'ensemble de données utilisé dans ce travail est collecté à partir des réponses aux questionnaires d'IPIP¹⁸ pour mesurer les traits de personnalité et PSS¹⁹ pour mesurer l'échelle de stress. Ces données sont ensuite classées à l'aide de l'algorithme SVM dans le but de détecter pour chaque personne un score indiquant le degré de stress (faible, modéré, élevé) en fonction de son trait de personnalité. Dans le même contexte [Shen et al. \[2017\]](#) ont cherché à obtenir diverses connaissances sur les associations entre la comorbidité du trouble borderline et celui des autres troubles psychiatriques (troubles dépressifs, troubles bipolaires, ...) et physiques. L'ensemble de données utilisé dans ce travail se présente sous la forme : (i) d'une étude cas-témoins qui a été menée sur des informations provenant de 1460 personnes sélectionnées à partir de la base de données de l'assurance maladie nationale de Taiwan, (ii) d'informations sur la comorbidité physique et psychiatrique, révélées par un diagnostic. Ce travail a bénéficié de : (1) un modèle de régression logistique qui a été utilisé pour calculer les rapports de comorbidité entre les patients avec et sans trouble borderline, (2) une technique ARM²⁰ pour étudier les associations entre le trouble borderline et une ou plusieurs comorbidités psychiatriques.

Le chevauchement dans le diagnostic cause beaucoup de problèmes dans le domaine médical. Pour cela [Wang et al. \[2019b\]](#), ont tenté de : (i) définir les critères permettant de distinguer les facteurs de risques, (ii) identifier un ensemble d'affections à partir de l'analyse des données de 982 patients. Cela se fait en analysant leurs données avec l'outil SPSS qui affiche graphiquement le chevauchement entre certains éléments-clés figurant dans les rapports. Certaines des caractéristiques et des éléments sont partagés entre les maladies de trouble de la personnalité ainsi qu'entre les traits de personnalité. Cette hypothèse nous a incités à chercher dans un ancien travail les liens de dépendance de 1 à 1 et N à 1 entre les différents traits de personnalité [Ellouze et al. \[2022\]](#), en appliquant des différentes techniques d'apprentissage automatique (arbre de décision, ARM, construction des règles) sur le corpus PAN Clef 2015 (profil d'auteur) [Rangel et al. \[2015\]](#). La détection de problèmes psychologiques sans l'accompagnement d'un module de suivi n'empêche pas les conséquences néfastes de ces problèmes. Pour cela, plusieurs chercheurs ont étudié l'intérêt de l'association d'une opération de surveillance aux différents cas détectés.

17. Borderline Personality Questionnaire

18. Big-Five Factor Markers

19. Perceived Stress Scale

20. Association Rule Mining

2.5 SURVEILLANCE DE L'ÉTAT DU PATIENT

Les réseaux sociaux font partie des moyens offrant la possibilité de surveiller en temps réel les activités et les comportements de ses usagers. Pour cela, nous avons trouvé beaucoup de travaux qui ont traité la tâche de surveillance des utilisateurs sur les réseaux sociaux. Dans ce contexte, le travail [Correia et al. \[2020\]](#) a présenté une étude comparative de quelques travaux afin de montrer l'utilité des réseaux sociaux dans le domaine médical et le rôle qu'ils peuvent jouer en vue d'assurer le bien-être des individus dans le monde. Pour cela, celui-ci a proposé dans ce cadre quelques travaux sur l'importance d'analyser le comportement des personnes sur les réseaux sociaux. Les différents travaux proposés dans ce cadre offrent une surveillance en temps réel et peu coûteuse du comportement humain, en analysant leurs sentiments via des mots-clés ou même des émoticônes afin de contrôler la progression de maladies et d'épidémies telles que le *choléra*. De même, les réseaux sociaux sont utilisés pour apaiser l'anxiété et les rumeurs pendant et après les catastrophes environnementales en profitant de la disponibilité de l'actualité en temps réel. Par ailleurs, les réseaux sociaux peuvent permettre de mesurer le degré de sensibilité des personnes à travers le calcul de scores d'empathie envers les maladies. Dans le même contexte, [Passos et al. \[2019\]](#) ont proposé une revue détaillée sur l'utilisation des techniques de *Big Data* et de *Machine Learning* pour aider les personnes atteintes de trouble bipolaire. Cela se fait en clarifiant les terminologies liées à ce domaine, en décrivant les applications existantes et les défis à venir. Les techniques d'apprentissage automatique utilisées dans ce contexte sont principalement liées à la prédiction de l'humeur et les idées suicidaires chez les usagers souffrant de troubles psychologiques. Aussi, au niveau de la découverte des liens permettant de prédire les conséquences de l'état de chaque patient. Cela permet au médecin d'évaluer plus facilement les risques et de sélectionner les traitements les plus appropriés. La puissance de la notion *Big Data* se manifeste dans la gestion de l'hétérogénéité des données qui proviennent de plusieurs sources (*Smartphone, Web, ...*) l'enjeu majeur se situe dans la partie stockage, intégration, interprétation, analyse, visualisation et inférence. Ce qui permet de mettre en place des outils (tableau de bord, système inférentiel, etc.), permettant de surveiller en même temps l'état d'un groupe d'individus. En outre, ces outils permettent au médecin de mener une étude analytique approfondie aidant à extraire de nouveaux renseignements.

Les comportements et les intérêts des usagers peuvent également fournir un aperçu sur l'état d'esprit des utilisateurs ayant des problèmes psychologiques. Pour cela, [Si et al. \[2019\]](#) se sont concentrés sur l'analyse des intérêts des utilisateurs sur les réseaux sociaux afin de découvrir des liens de dépendance utiles entre leurs centres d'intérêt. Pour ce faire, ce travail a commencé par la collection des données à partir des différents types de réseaux sociaux comme *LinkedIn*, qui contient des info-box permettant d'exprimer *les centres d'intérêt, les loisirs, etc.* Les auteurs ont par la suite calculé les valeurs de métriques indiquant le degré des relations de dépendance à savoir *Confidence* (représente la fréquence d'occurrence de Y dans tous les cas où X apparaît), *Support* (représente la fréquence d'occurrence de X et Y simultanément), *Conviction* (décrit la fréquence d'apparition de Y sans aucun facteur d'influence), *Lift* (représente le nombre d'occurrences de X

affectant l’apparence de Y). Parmi les résultats obtenus, nous citons *intérêt (culture)* -> *intérêt (voyage)*. D’autre part, Beheshti et al. [2020] ont suivi l’évolution des facteurs qui conduisent dans le temps à des troubles de comportement grâce à la construction d’un modèle graphique des comportements (sbGraph). Cela se fait en : (1) enregistrant les données provenant des réseaux sociaux dans une base de données (NoSQL), (2) extrayant des entités telles que mots-clés, entités nommées qui montrent (loisir, travail, etc.), (3) établissant des liens entre des entités nommées grâce à l’existence des bases de connaissances telles que Wikidata, le graphe de connaissances de Google et des exemples de relations définies : *Utilisateur x (suivre)* -> *Utilisateur y*, (4) extrayant des informations pertinentes qui affectent la personnalité sur le plan affectif, cognitif et inquiétant.

D’autres chercheurs ont proposé de suivre l’état actuel des patients déprimé à travers l’analyse de leurs sentiments comme Saha et al. [2021]. Pour ce faire, ces auteurs ont recueilli les données des personnes déprimées depuis plusieurs groupes sur Facebook comme *Depression Support group*. Puis, ils ont procédé à l’annotation manuelle des sentiments exprimés dans ces données grâce à un psychiatre. Ils ont ensuite extrait plus de 80 attributs numériques qui reflètent essentiellement l’aspect du style d’écriture comme *la taille d’une phrase, le nombre d’utilisations de chaque ponctuation, etc.* Après, ils ont procédé à mesurer la relation entre chaque attribut linguistique et la classe indiquant la valeur du sentiment à l’aide de différentes métriques comme l’écart quadratique moyen (RMSE), l’erreur absolue moyenne (MAE) ou avec la courbe ROC (la fonction d’efficacité du récepteur) ou la courbe sensibilité/spécificité (*Receiver Operating Characteristic*). Finalement, ils ont mesuré la performance de chaque algorithme sur le résultat de la classification des sentiments.

2.6 DISCUSSION

Après l’élaboration d’une description détaillée des méthodes proposées pour l’analyse de la personnalité et des maladies psychologiques. À ce niveau, nous procédons à une discussion qui montre les points forts et les points faibles de chaque catégorie de ces travaux.

2.6.1 Détection des traits de personnalité

Sur la base des travaux cités pour la détection des traits de la personnalité, il est à noter que les méthodes les plus utilisées sont fondées sur : (1) l’apprentissage automatique Salem et al. [2019b], Varshney et al. [2017], González-Gallardo et al. [2015], Celli et Lepri [2018], (2) la linguistique Pramodh et Vijayalata [2016], (3) la statistiques Alvarez-Carmona et al. [2015], (4) l’hybridation des méthodes El-louze et al. [2020]. Certains de ces travaux ont utilisé des données obtenues depuis plusieurs sources comme dans les travaux Varshney et al. [2017], Pramodh et Vijayalata [2016] et d’autres qui se sont limités à une seule source de données González-Gallardo et al. [2015], Celli et Lepri [2018], Alvarez-Carmona et al. [2015]. En général, les traits de personnalité sont des informations implicites, pour cela leur détection nécessite le traitement de nombreux types de données. Dans ce contexte, nous avons trouvé ceux qui se sont concentrés sur les caractéristiques de type linguistique Salem et al. [2019b], Varshney et al. [2017], González-Gallardo

et al. [2015], et ceux qui ont combiné les critères linguistiques avec d’autres informations telles que les métadonnées Celli et Lepri [2018]. Cependant, un certain nombre d’études se sont limitées à une méthode lexicale Varshney et al. [2017], Pramodh et Vijayalata [2016].

La méthode lexicale est très limitée car il est nécessaire de disposer d’un corpus d’apprentissage comprenant tout le lexique utilisé par les personnes de la classe ciblée. Nous remarquons aussi que les travaux fondés sur les méthodes d’apprentissage automatique Salem et al. [2019b], González-Gallardo et al. [2015], Celli et Lepri [2018] reposent sur la notion de logique binaire, c’est-à-dire ils ont ignoré le principe de relativité (le degré d’appartenance d’une instance à une classe). En effet, ils se sont reposés sur des méthodes déterministes plutôt que sur des méthodes probabilistes tandis que dans le domaine médical le concept de probabilité est important car il permet de mesurer le degré de chevauchement dans le diagnostic.

2.6.2 Détection des troubles de personnalité

D’après notre étude des travaux menés sur la détection des troubles psychologiques, nous avons noté que certains chercheurs se sont focalisés sur la détection des conséquences des troubles psychologiques Rekik et al. [2019], Ahmad et Siddique [2017], Mbarek et al. [2022] (*suicide*) et d’autres sur la détection des symptômes des troubles de personnalité Celli et al. [2016] comme *l’humeur et les émotions*. Ces symptômes peuvent être instantanés et, avec le temps peuvent produire des maladies. Malgré l’intérêt associé aux problèmes psychologiques, peu d’études se sont concentrées sur ces problèmes Sonkurt et al. [2021], Singh et al. [2019] tels que *l’antisocial et les troubles mentaux*. Même ces travaux ont traité certaines type spécifiques de troubles de personnalité comme *le paranoïde, l’antisocial, etc.*, et non pas le trouble de la personnalité lui-même (tous les types de troubles de personnalité). De plus, nous avons noté la forte utilisation des méthodes basées sur les techniques d’apprentissage automatique Sonkurt et al. [2021], Singh et al. [2019], Celli et al. [2016], bien que ces techniques comportent plusieurs limitations telles que les résultats fournis ne sont pas explicables. Cela signifie que nous risquons d’obtenir des résultats non justificatifs.

La langue anglaise est la langue la plus utilisée dans les travaux cités Singh et al. [2019], Umar et Qamar [2019], Celli et al. [2016] bien que d’autres langues sont largement utilisées dans le monde. Par exemple, selon Poblete et al. [2011], plus de 13% des utilisateurs de Twitter sont des Brésiliens dont la langue maternelle est le portugais. Il y a aussi une grande variance dans l’utilisation des caractéristiques par exemple, il y a ceux qui ont travaillé sur : (1) un seul type d’attribut Vasha et al. [2023], Singh et al. [2019], Rekik et al. [2019] parmi les limites de ce choix est que les problèmes psychologiques sont des informations implicites. Pour cela, les gens n’avoueront pas directement de leur méfiance, leurs problèmes, leurs maladies, etc. Pour cette raison, il est fortement recommandé de tenir compte de nombreuses informations comme *le comportement et le langage utilisés, etc.*, (2) la combinaison de plusieurs types d’attributs Sonkurt et al. [2021], Celli et al. [2016], Mbarek et al. [2022]; la variété de ces attributs peut avoir un effet négatif sur les résultats. En effet, la présence de tous ces attributs peut perturber le bon fonctionnement de l’algorithme de classification. L’annotation du

corpus est également un critère de variation dans les travaux cités. Pour cela, nous avons trouvé ceux qui ont bénéficié d’un expert de domaine [Rekik et al. \[2019\]](#), [Mbarek et al. \[2022\]](#) et ceux qui ont utilisé un questionnaire pour effectuer l’annotation [Stankevich et al. \[2019\]](#). Certains auteurs ont travaillé sur l’analyse d’un seul tweet [Singh et al. \[2019\]](#) et d’autres ont travaillé sur l’analyse de l’historique d’un ensemble de tweets (profil de l’utilisateur) [Stankevich et al. \[2019\]](#), [Rekik et al. \[2019\]](#). Généralement, dans le cas d’une détection de maladie, il est nécessaire d’avoir autant d’informations que possible pour être en mesure de juger convenablement l’état de l’utilisateur.

Les données collectées dans ces études ont été obtenues en mode hors ligne, c’est-à-dire qu’elles n’ont pas été collectées en temps réel, ce qui peut retarder l’étape de détection. Par exemple, nous pouvons trouver une personne qui s’est déjà suicidée et qui n’a pas encore été détectée par les systèmes de ces travaux. De plus, il est fortement recommandé de penser à un autre traitement après l’étape d’identification puisque l’identification seule ne permet pas d’éviter les conséquences dangereuses. Par conséquent, il est essentiel de procéder par la suite à la surveillance de l’état de ces patients. Le tableau 2.3 présente les critères de performance dégagés depuis les différents travaux. Ces critères assurent la généralité et la polyvalence de chaque travail, ce qui assure la capacité d’offrir plus d’opportunités.

2.7 CONCLUSION

Nous nous sommes intéressés, dans ce chapitre, à définir les différentes méthodes proposées pour l’analyse des caractéristiques de la personnalité. Pour ce faire, nous avons commencé par une présentation des différentes techniques utilisées pour la tâche de détection automatique de la personnalité sur les réseaux sociaux. Ensuite, nous avons présenté les principaux travaux ayant une relation avec les troubles psychiques en montrant le processus de détection, de suivi et de recherche de la dépendance entre les divers éléments de la personnalité afin de comprendre les influences de certains éléments sur d’autres. Nous avons également effectué une étude comparative entre les divers travaux présentés. Dans ce cadre, nous avons catégorisé les différentes méthodes en se basant sur des différents niveaux d’abstraction logique ainsi sur un point de vue technique qui affecte l’aspect fonctionnel. Malgré la diversité des méthodes présentées sur le plan originalité et sur le plan impact social, ces travaux nécessitent encore la prise en compte d’autres critères afin d’assurer un service de surveillance de haute qualité pour les utilisateurs des réseaux sociaux. Pour cela, dans le chapitre suivant nous présenterons notre méthode proposée pour le traitement des données issues des réseaux sociaux des personnes susceptibles d’être malades.

TABLE 2.3 – Les critères de comparaison entre les différents travaux de détection des troubles psychologiques.

Référence	Données				Annotation		Attributs		Résultat	
	Source	Langue	Type	Mode d’obtention	Manière d’annotation	Types	Méthode d’extraction	Étiquette	Objectif	
Mbarek et al. [2022]	mono-source	mono-langue	profil d’utilisateur	hors ligne	—	multi-types	personnalisé	mono-étiquette	Détection des extrêmes	
Rekik et al. [2019]	multi-sources	mono-langue	profil d’utilisateur	hors ligne	expert du domaine	mono-type	personnalisé	mono-étiquette	Identification du lexique violent	
Wang et al. [2019a]	multi-sources	multi-langues	mono-publication	hors ligne	—	mono-type	automatique	multi-étiquettes	Détection d’émotion	
Stankevich et al. [2019]	mono-source	mono-langue	profil d’utilisateur	hors ligne	questionnaire	multi-types	personnalisé	mono-étiquette	Détection de la dépression	

DÉTECTION DES PERSONNES SOUFFRANT DE TROUBLES DE PERSONNALITÉ

3

SOMMAIRE

3.1	INTRODUCTION	52
3.2	COLLECTE DU CORPUS POUR LA DÉTECTION DES UTILISATEURS AYANT DES TP	53
3.3	MÉTHODE NEURONALE PROPOSÉE POUR LA DÉTECTION DES PERSONNES AYANT DES TP	55
3.3.1	Prétraitement du corpus collecté	56
3.3.2	Génération des caractéristiques	57
3.3.3	Réseau de neurones convolutifs pour l'extraction des caractéristiques	60
3.3.4	Génération synthétique des données observées	60
3.3.5	Capture des dépendances à long terme entre les unités lexicales pour la tâche de classification des symptômes de TP	62
3.3.6	Classification binaire pour la tâche de détection de trouble paranoïde	63
3.4	MÉTHODE PROPOSÉE BASÉE SUR LES TECHNIQUES D'APPRENTISSAGE CLASSIQUES	64
3.4.1	Prétraitement	65
3.4.2	Extraction des caractéristiques	66
3.4.3	Réduction de la dimensionnalité	69
3.4.4	Classification basée sur la notion de logique floue pour la détection des personnes ayant des TP	70
3.5	CONCLUSION	73

3.1 INTRODUCTION

Dans les chapitres précédents, nous avons exposé les diverses méthodes proposées dans la littérature dédiées à la détection des traits et des troubles de la personnalité chez les utilisateurs des réseaux sociaux. Nous nous sommes principalement intéressés à l'analyse des données dans le but de détecter des informations profondes qui reflètent l'identité de l'individu. Cette analyse est effectuée en se focalisant sur les principaux facteurs contribuant à réaliser cette tâche avec plus d'efficacité y compris les facteurs linguistiques présents dans le texte produit par l'utilisateur et les informations relatives à son comportement sur les plateformes des réseaux sociaux. Cela est important car chaque type de personnalité se caractérise par un style d'écriture et un comportement qui le distinguent des autres types.

Cependant, il existe plusieurs méthodes permettant la détection des éléments liés à la psychologie des utilisateurs des réseaux sociaux, telles que les ontologies, les systèmes experts, les méthodes symboliques, les méthodes neuronales, etc. Nous avons choisi de tirer parti en premier lieu d'une architecture neuronale qui parvient à extraire automatiquement les facteurs lexicaux distinctifs entre les différentes classes de classification. Cette technique prend en compte les particularités linguistiques telles que l'aspect sémantique. Cela permet d'assurer par la suite une amélioration au niveau des performances de la tâche de classification. Dans la littérature, plusieurs chercheurs ont travaillé sur l'application d'une méthode neuronale dans le cadre des traits de personnalité mais très peu de chercheurs ont appliqué cette méthode dans le contexte lié aux troubles psychologiques en dépit de leur impact social. Pour cela, nous nous intéressons dans ce chapitre à la détection du trouble de la personnalité à l'aide d'une méthode neuronale afin de tirer profit des avantages liés à ses architectures. Cette détection est réalisée en passant par la précision des degrés de présence des symptômes liés au trouble à identifier. Nous procédons ensuite à la détection du trouble de la personnalité, à l'aide d'une méthode basée sur la notion d'apprentissage automatique classique fondée sur une technique d'extraction des caractéristiques personnalisées.

Il est à noter que notre première méthode fondée sur une méthode neuronale admet six étapes à savoir, une étape de prétraitement et une autre pour la génération des caractéristiques, suivie par une étape d'extraction automatique des caractéristiques pour capturer uniquement les caractéristiques pertinentes. Après cela, nous procédons à une étape de génération de données afin de pallier aux problèmes liés à la déséquilibration du corpus. Nous passons ensuite à la détection de la présence de symptômes avant la détection de la maladie par une classification binaire. Notre deuxième méthode comprend quatre étapes : une étape de prétraitement du corpus collecté, suivie d'une étape d'extraction des attributs ayant une influence directe sur la personnalité. La dernière étape consiste à réaliser la classification afin d'indiquer la présence ou l'absence d'un trouble de la personnalité, en intégrant des notions de logique floue et de classification par vote.

Dans la suite de ce chapitre, nous détaillerons nos deux méthodes pour la détection des troubles de la personnalité, y compris leurs catégories et leurs types, tout en expliquant comment faire face aux différents enjeux rencontrés.

3.2 COLLECTE DU CORPUS POUR LA DÉTECTION DES UTILISATEURS AYANT DES TP

En raison de l'absence de ressources pour la langue française permettant la détection des personnes ayant des problèmes psychologies, nous avons fait le choix de créer notre propre corpus. Pour ce faire, nous avons effectué d'abord une recherche en temps réel des statuts textuels partagés (publications) pouvant être postées par des personnes ayant des troubles de personnalité. Pour cela nous avons eu recours à certaines expressions comme celles qui figurent dans les tableaux 1.1, 1.2 et 1.3 du chapitre 1 en utilisant l'outil *Apache Spark streaming* utilisé par Aziz et al. [2018]. Ensuite, nous avons procédé au téléchargement des vingt dernières publications des profils de ces utilisateurs. Étant donné que notre domaine d'étude est générique, c'est-à-dire, qu'il n'est pas lié à un domaine d'application spécifique et que nous ne pouvons pas fournir tous les mots clés nous avons également recueilli au hasard les vingt publications récentes de chaque profil qui publient fréquemment sur certaines pages, comme les pages présidentielles. Toutes ces données ont été collectées à partir du site Twitter, qui compte actuellement 436 millions d'utilisateurs, dont 330 millions sont considérés comme actifs¹. Ce réseau est très utilisé en Europe et en Amérique. En effet, il y a 16,1 millions d'utilisateurs actifs en France et 38 millions aux États-Unis. De plus, environ 500 tweets sont publiés chaque jour sur Twitter². Notre corpus comprend une collection de plus de 28 000 tweets de différents domaines (économie, social, politique, etc.). Ce corpus est annoté en fonction du type de trouble, en se basant sur l'existence de symptômes associés à ce trouble. Nous considérons qu'une personne peut présenter un trouble si, dans ses 20 derniers tweets, il y a une redondance d'indicateurs linguistiques reflétant les signes de cette maladie, tels que des informations sémantiques exprimant un trouble ou une crainte grave, comme les expressions suivantes : *mes cheveux se dressent sur la pointe, Je peux à peine respirer, ma gorge se noue, etc.*

Malgré l'existence de certains travaux qui ont affirmé leur capacité à détecter les troubles de la personnalité à partir des données de réseaux sociaux Stankevich et al. [2019], Mohite et Gujar [2021], Singh et al. [2019], dans notre travail nous avons considéré chaque classe indiquant l'existence de la maladie comme susceptible qu'elle reflète la présence du trouble (c'est-à-dire, pas d'une manière absolue). En effet, il reste difficile d'affirmer l'existence d'une maladie de manière virtuelle. Le tableau 3.1 présente la répartition des tweets et des profils d'utilisateurs par classe.

Rappelant que lors de l'annotation, nous nous sommes basés sur l'hypothèse d'intérêt de passer par la détection des symptômes avant la détection de la maladie. En effet, les informations relatives à un symptôme sont plus concrètes que celles de la maladie. De même, nous visons que notre méthode suit le processus logique pour déduire l'existence de la maladie chez les usagers. Dans le tableau 3.2, nous montrons la répartition des instances pour le cas de trouble paranoïaque. Les symptômes sélectionnés dans notre étude sont les plus reconnus par le DSM-5³.

Il est à noter que nous éprouvons parfois des difficultés à recueillir des données. Par exemple, pour les gens qui se croient toujours justes ou isolés, ils n'ont

1. <https://www.blogdumoderateur.com/chiffres-twitter/>

2. <https://www.oberlo.fr/blog/statistiques-twitter>

3. La cinquième édition du Manuel diagnostique et statistique des troubles mentaux

TABLE 3.1 – La répartition des instances par classe.

Classes	Nombre d'instances par utilisateur	Nombre d'instances par tweets
Personne avec un trouble	884 utilisateurs	17680 tweets
Personne sain	531 utilisateurs	10620 tweets
Personnes ayant des personnalités distantes, froides et suspicieuses	422 utilisateurs	8 440 tweets
Personnes ayant des personnalités impulsives, mélodramatiques et susceptibles	580 utilisateurs	11600 tweets
Personnes ayant des personnalités affacées avec troubles anxieux et craintifs	577 utilisateurs	11540 tweets
Personne paranoïde	263 utilisateurs	5260 tweets
Personne schizoïde	97 utilisateurs	1940 tweets)
Personne schizotypique	174 utilisateurs	3480 tweets
Personne antisociale	154 utilisateurs	9000 tweets
Personne borderline	231 utilisateurs	4620 tweets
Personne histrionique	248 utilisateurs	4960 tweets
Personne narcissique	83 utilisateurs	1660 tweets
Personne évitante	79 utilisateurs	1580 tweets
Personne dépendante	258 utilisateurs	5160 tweets
Personne obsessionnelle-compulsive	241 utilisateurs	4820 tweets

pas besoin de parler aux autres et de tenter de les persuader. En particulier pour le premier cas, qui présente une tendance à l'égoïsme. En ce qui concerne les symptômes tels que «lire des significations cachées dans les remarques innocentes» et «soupçons récurrents», chaque élément ou détail peut être un déclencheur et un incitatif pour ces personnes afin d'écrire et de montrer ce qui n'est pas explicitement exprimé de leur point de vue.

Il convient de rappeler que dans notre travail, nous avons recueilli les données en temps réel. C'est la raison pour laquelle les attributs liés aux métadonnées des tweets tels que le nombre de retweets et le nombre de favoris de chaque tweet étaient indisponibles lors de la collecte des données. Mais dans la suite de notre travail, nous avons considéré ces attributs en raison de leur pertinence dans la détection de la maladie. Par conséquent, après un certain temps, nous avons recueilli les métadonnées associées aux tweets et aux utilisateurs à l'aide de Tweepy⁴. Il est à noter que nous avons éliminé de notre corpus les utilisateurs que nous ne pouvons pas récupérer l'un de leurs tweets pour garantir le nombre de 20 tweets par utilisateur. Après la collection de notre corpus, nous précéderons dans la prochaine étape à l'utiliser pour réaliser la phase d'apprentissage dans le but de détecter les personnes ayant des TP.

4. est une bibliothèque Python qui permet aux utilisateurs d'effectuer des requêtes complexes au niveau du raclage de tweets.

TABLE 3.2 – Distribution des instances pour la classification des symptômes.

Symptômes	Nombre de tweets
Agressivité	2420 tweets
Perçoit des attaques	3260 tweets
Soupçons récurrents	5640 tweets
Isolement	920 tweets
Croire qu'ils ont toujours raison	1520 tweets
Lire des significations cachées dans des remarques innocentes	4540 tweets
Mauvaises relations avec les autres	5460 tweets
Douter de l'engagement	3860 tweets
Impitoyables et rancuniers	3740 tweets

3.3 MÉTHODE NEURONALE PROPOSÉE POUR LA DÉTECTION DES PERSONNES AYANT DES TP

Notre méthode neuronale illustrée dans la figure 3.1 permet aux analystes et aux experts d'analyser en temps réel la production textuelle des utilisateurs afin de détecter les personnes atteintes de troubles de la personnalité, sur les plateformes de réseaux sociaux. Cette méthode est conforme au processus de diagnostic mené par les psychiatres, qui consiste à écouter le patient, identifier les différents symptômes et détecter la maladie.

Pour ce faire, nous avons utilisé un nouveau modèle d'apprentissage profond présenté dans la figure 3.2, comprenant : (i) un ensemble de couches convolutionnelles CNN dédiées à la tâche d'extraction automatique des caractéristiques. En effet, même avec l'aide d'un spécialiste de la langue et d'un psychiatre, il reste difficile de gérer les critères de distinction entre les différentes classes [Liang et al. \[2017\]](#), (ii) deux couches LSTM permettant de maintenir les dépendances distantes entre les différentes unités lexicales [Ombabi et al. \[2020\]](#). Le résultat de cette couche est une valeur qui représente le degré de présence de chaque symptôme associé à un type de trouble de la personnalité, (iii) SVM pour détecter la maladie en fonction des degrés des symptômes obtenus à partir de la couche précédente. L'utilisation de SVM est justifiée par leur performance dans les cas de classification binaire. De même, SVM est l'un des algorithmes d'apprentissage les plus configurables [Chen et Zhang \[2018\]](#).

De plus, notre méthode prend en compte d'autres considérations, telles que l'étiquetage multiple, afin de traiter les cas où le patient présente plusieurs types de troubles. Nous abordons également le déséquilibre du corpus à l'aide de la phase de génération de données synthétiques. De plus, notre méthode traite des défis linguistiques tels que le problème de l'ignorance de sens du mot dans la phrase trouvée par la technique de plongement de phrases (*sentence embedding*) [Lin et al. \[2017\]](#).

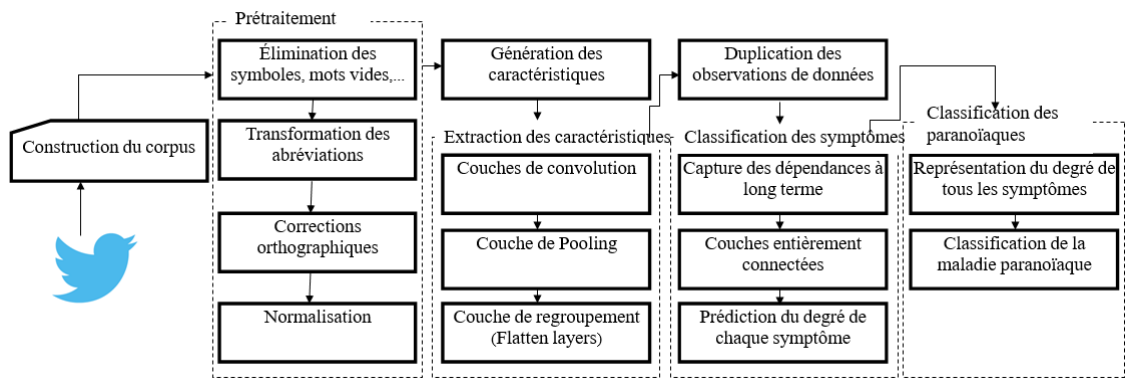


FIGURE 3.1 – Méthode proposée pour la détection de la maladie paranoïaque au moyen de la détection de ses symptômes.

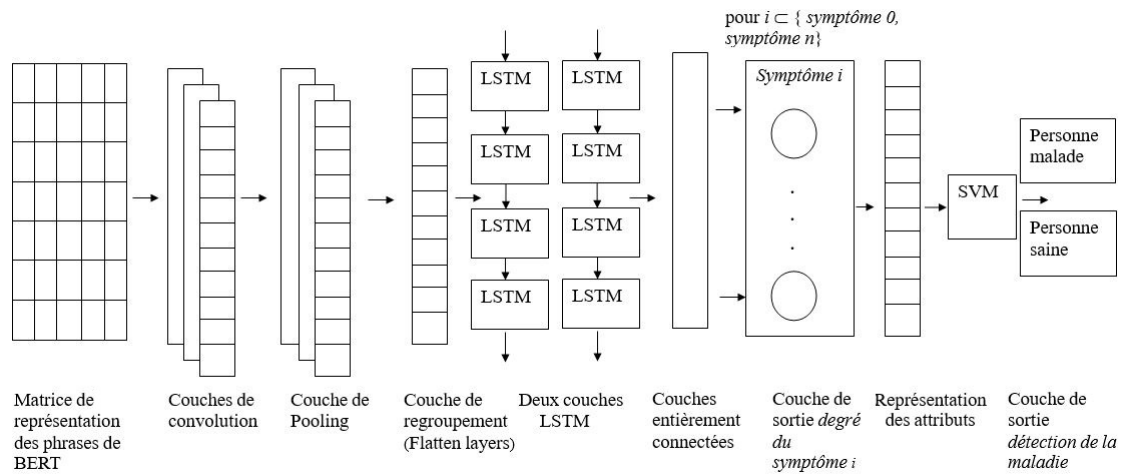


FIGURE 3.2 – Proposition de modèle d'apprentissage profond CNN-BiLSTM-SVM pour la détection de la maladie paranoïde ainsi que ses symptômes.

3.3.1 Prétraitement du corpus collecté

Dans l'étape de prétraitement du corpus collecté, nous nous sommes focalisés sur la préparation de nos données en normalisant les éléments-clés et en retirant les éléments inutiles. Cette préparation est réalisée dans le but de ne pas affecter négativement le traitement ultérieur, notamment que notre travail repose sur des données extraites du Web 2.0⁵. Cette tâche a été réalisée en suivant ces sous étapes :

Suppression des symboles et des mots vides : il s'agit d'enlever les mots vides que nous trouvons dans le corpus indépendamment de la classe. C'est pour cette raison, nous estimons que ces mots n'ont aucune dimension sémantique profonde. Après l'étape de filtrage des mots vides nous passons à l'élimination des différents éléments indiquant les symboles, les liens et les valeurs numériques. En effet, notre travail repose principalement sur des données provenant des réseaux

5. désigne l'ensemble des fonctionnalités communautaires et collaboratives (blogs, avis consommateurs, flux RSS, plateformes d'échanges vidéo, etc.) qui se sont fortement développées sur l'Internet depuis l'année 2005.

sociaux, où les symboles tels que @ pour mentionner une personne ou :D pour exprimer le sourire sont fréquemment utilisés. Pour cette raison, nous avons choisi de se concentrer uniquement sur les détails linguistiques et de négliger les autres éléments. Pour réaliser la sous étape de nettoyage, nous avons utilisé la librairie *NLTK*⁶ avec laquelle nous avons éliminé tous les mots vides (ayant la catégorie grammaticale préposition, pronom ou adverbe). Nous avons utilisé également la bibliothèque d'expression régulière *re*⁷ pour retirer de notre corpus les liens URL et les symboles.

Normalisation : nous visons par l'étape de normalisation à éliminer la nuance trouvée dans la partie textuelle de notre corpus. Cette nuance est due à l'aspect grammatical et conjugal dans l'écriture comme par exemple l'utilisation de majuscules pour indiquer le début d'une phrase ou d'un nom. De même, l'utilisation d'une abréviation pour éviter la répétition. En dépit de l'importance de ces aspects linguistiques pour exprimer la clarté et la signification d'une phrase, ainsi que pour respecter les règles grammaticales, cela peut poser un problème lors du traitement. En effet, d'un point de vue machine, les deux mots *IA* et *intelligence artificielle* ou bien les mots *isolé* et *isolation* ne sont pas les mêmes ce qui peut perturber par la suite le traitement. Pour cela, dans cette partie nous avons d'abord mis l'accent sur la transformation des abréviations en formes ordinaires à l'aide de la ressource *Google Knowledge Graph*⁸. Ensuite, nous nous sommes intéressés à la conversion de tout le texte en minuscules et finalement à une transformation de chaque forme flexionnelle d'un mot en une racine à l'aide de la bibliothèque *NLTK*. Ce traitement permet à notre méthode d'agir de la même manière avec des mots ayant une racine commune et de prendre en considération la valeur des mots, et non pas juste leur forme.

3.3.2 Génération des caractéristiques

L'étape de génération des caractéristiques consiste à convertir les données textuelles de notre corpus en un ensemble de vecteurs numériques pouvant être traités par des algorithmes d'apprentissage automatique. D'après notre étude littéraire, plusieurs méthodes peuvent être utilisées pour cette transformation, comme celle de plongement de mots (*word embedding*). Cependant, le principal problème de cette technique est qu'elle ne conserve pas le sens de la phrase entière, ce qui n'aide pas les algorithmes à décrypter l'intention et le sens du contenu. À cette fin, nous choisissons d'utiliser les techniques de *sentence embedding*. Parmi les techniques du *sentence embedding*, nous citons :

InferSent est un modèle supervisé de *sentence embedding*, lancé par le groupe *Facebook AI Research* en 2018 [Conneau et Kiela \[2018\]](#). Le principe de ce modèle est qu'il est formé sur des données d'inférence du langage naturel, spécifiquement sur l'ensemble de données SNLI (Stanford Natural Language Inference⁹). Il se compose de 570 000 paires de phrases en anglais générées par l'homme, étiquetées manuellement selon l'une des trois catégories suivantes : implication, contradic-

6. Natural Language Toolkit est une bibliothèque logicielle en Python permettant un traitement automatique des langues

7. Le module *re* permet d'utiliser des expressions régulières avec Python.

8. Le graphe de connaissance Google alimente Google en informations précises et utiles.

9. <https://nlp.stanford.edu/projects/snli/>

tion ou neutre. Ce modèle prend en paramètre une paire de phrases et il les encode pour produire un format numérique adéquat.

Universal Sentence Encoder est une technique proposée par Google. Elle figure parmi les techniques de plongements de phrases les plus efficaces [Cer et al. \[2018\]](#). Parmi les avantages de cette technique est sa capacité à être appliquée dans un contexte d'apprentissage multi-tâches telles que l'analyse de sentiments, la classification de textes, la similitude de phrases, etc. Le fonctionnement de cet encodeur s'appuie sur deux modèles d'apprentissage profond, *Transformer*¹⁰ et *Deep Averaging Network (DAN)*¹¹. Ces deux modèles permettent d'assurer la notion de mise-à-jour fréquente des poids à chaque fois que le résultat est retourné au modèle afin de permettre au modèle d'obtenir une meilleure présentation qu'au-paravant.

Language-Agnostic Sentence Representations (Laser) est un ensemble de modèles élaborés par le groupe de recherche de Facebook [Artetxe et Schwenk \[2018\]](#). Ce modèle profite d'une technique de transfert interlangue qui permet de convertir les données en un ensemble de vecteurs indépendants de la langue. Cette tâche est accomplie en développant un modèle dans une langue donnée, puis l'étendre à plus de 100 langues additionnelles. Cette technique permet par exemple à Facebook de classer les revues de films comme favorables ou défavorables.

Simple Sentence EmbeddeR (SISTER) [Delianidi et al. \[2021\]](#) est un modèle qui est composé par 300 dimensions d'incorporations de mots pré-entraînés. Son principe de fonctionnement est de construire l'incorporation de phrases basée sur FastText [Bojanowski et al. \[2017\]](#). FastText¹² est une bibliothèque légère, gratuite et à code source ouvert qui permet aux utilisateurs d'apprendre des représentations et d'appliquer des modèles de classification sur des données textuelles. Ce modèle prend en charge les langues suivantes : anglais, japonais, français.

Bidirectional Encoder Representations of Transformers (Bert) est un modèle de représentation du langage pré-entraîné. Ce modèle a été introduit en 2018 par [Reimers et Gurevych \[2019\]](#). Il a démontré des performances supérieures grâce à son architecture, qui combine un encodeur de transformateur bidirectionnel multicouches avec de multiples têtes d'attention. Le modèle mentionné repose sur un corpus en anglais combinant des ressources obtenus à partir des livres et de l'encyclopédie Wikipédia. Il est entraîné à travers les architectures *Masked Language Modeling*¹³ (MLM) et *Next Sentence Prediction*¹⁴ (NSP). Ces architectures contribuent à régler le problème de la classification au niveau des corpus qui ne sont pas volumineux.

Malgré la différence dans l'architecture de chaque modèle, ces techniques ont démontré une performance élevée lors de la transformation de la partie textuelle en un ensemble de vecteurs numériques [Das et al. \[2021\]](#). Après une étude empirique, nous avons choisi de tirer profit du modèle Sentence Bert puisque ce modèle est formé à partir d'une grande masse de données. De même, il dispose d'une architecture spécifique qui lui permet d'apprendre des représentations bidi-

10. <https://medium.com/analytics-vidhya/transformers-vs-universal-sentence-encoder-980c165bc96d>

11. <https://medium.com/tech-that-works/deep-averaging-network-in-universal-sentence-encoder-465655874a04>

12. <https://fasttext.cc/>

13. <https://towardsdatascience.com/masked-language-modelling-with-bert-7d49793e5d2c>

14. <https://towardsdatascience.com/bert-for-next-sentence-prediction-466b67f8226f>

rectionnelles. En outre, il est prêt à accepter un nombre important de paramètres, ce qui le rend plus flexible et capable de s’adapter aux différentes situations [Eke et al. \[2021\]](#). Cette technique consiste à calculer la similarité entre les phrases en appliquant des couches de *pooling*¹⁵ afin de réaliser le groupement tout en conservant les descripteurs importants. De plus, cette technique fournit comme résultat un ensemble de vecteurs standardisés et normalisés tout en abordant un grand nombre de problèmes reconnues. Ces problèmes sont étroitement liés aux caractéristiques du corpus, en particulier à la taille des phrases présentes dans l’ensemble de données ainsi à la combinaison de vocabulaire au sein du corpus. Ces spécificités sont présentes d’une manière excessive au niveau des données issues des réseaux sociaux, puisqu’elles ne suivent pas des normes lexicales, morphologiques, syntaxiques et parfois sémantiques.

Dans notre travail, nous nous intéressons uniquement aux données textuelles. Pour cela, nous avons exclu les autres types d’informations telles que le nombre de retweets par utilisateur, ou/et le nombre de retweets par tweet. En effet, nous avons travaillé sur des données obtenues en temps réel, de sorte qu’au début, les valeurs de ces attributs sont nulles. De plus notre méthode est basée principalement sur un modèle d’apprentissage profond admettant une architecture spécifique qui offre une assistance pour distinguer automatiquement les attributs pertinents à partir des données brutes. Dans le tableau 3.3, nous avons récapitulé les modèles d’incorporation des phrases *sentence embedding* en fonction des divers paramètres [Das et al. \[2021\]](#).

TABLE 3.3 – Comparaison des différents modèles de *sentence embedding* [Das et al. \[2021\]](#).

Paramètres	word embedding	InferSent	Google’s USE	BERT
Méthode d’apprentissage	non supervisé	supervisé	non supervisé	non supervisé
Architecture	Feed-forward Neural network model (Skip-gram or CBoW)	Bi-directional LSTM with Softmax classifier	Transformer or Deep Averaging Network	Bi-directional Transformer
Données d’apprentissage	Wikipedia	Formé sur GloVe ou FastTextSNL	Wikipedia, actualités du net,web QA	Wikipedia et corpus de livres
Ordre des mots	N’est pas considéré	considéré	considéré	considéré
Relation sémantique	N’est pas considéré	considéré	considéré	considéré
STS Benchmark scores Cer et al. [2017]	70	80.1	87.21	90

15. <https://deeplylearning.fr/cours-theoriques-deep-learning/convolution-pooling-classification/>

3.3.3 Réseau de neurones convolutifs pour l'extraction des caractéristiques

Le réseau de neurones convolutif (CNN) est un type particulier de réseau de neurones admettant une architecture qui diffère par rapport à l'architecture classique du modèle MLP (perceptron multicouche). Cette différence s'articule principalement au niveau de la partie convolutive. L'objectif de cette partie est de réduire la taille brute de la forme d'entrée afin de mettre en évidence les caractéristiques pertinentes. Pour ce faire, le modèle CNN assigne des pondérations et des biais à divers objets dans l'input et il opère ensuite à la distinction de l'un par rapport à l'autre en utilisant des filtres appropriés. C'est pour cette raison, il exige moins de prétraitement par rapport à d'autres algorithmes de classification.

Plusieurs études ont montré l'importance de réseau CNN dans la généralisation et le traitement des problèmes liés aux données d'entrée brutes par rapport aux autres techniques basées sur des descripteurs traditionnels [Alajlan et Saundagar, 2021, Kumar et Sundaram, 2022]. En outre, l'architecture CNN a démontré des performances remarquables dans diverses tâches liées au traitement du langage naturel au niveau de la récupération des éléments syntaxiques et sémantiques. Pour cette raison, le modèle CNN est utilisé dans plusieurs études scientifiques, comme dans Ombabi et al. [2020] pour effectuer certaines analyses tels que l'analyse sémantique, la modélisation des phrases, l'extraction des requêtes de recherche et d'autres défis liés au TALN. Dans notre travail, nous avons appliqué le modèle CNN afin de tirer profit de tous ces avantages.

Dans le modèle CNN, un sous-ensemble d'entrée est connecté à ses couches via une couche convolutive; par conséquent, les couches CNN sont appelées *Features map*. Bien que ces modèles fonctionnent parfaitement, ils exigent l'empilement d'un grand nombre de couches convolutionnelles afin de capturer les dépendances à long terme. Pour cette raison, le modèle CNN utilise une couche *pooling* pour réduire la taille de sortie des couches, ce qui peut contribuer à réduire la complexité du calcul et à conserver que des caractéristiques importantes. Finalement, une couche *flatten* est utilisée pour fournir la sortie de la couche *pooling* et la mettre en correspondance avec les couches suivantes. La figure 3.3 présente en détail l'architecture générale du réseau de neurones convolutifs pour le cas d'extraction des attributs à partir d'une image.

3.3.4 Génération synthétique des données observées

La génération synthétique des données observées vise à maximiser le nombre d'instances, vu la difficulté d'annoter les données et la difficulté d'obtenir des données équilibrées. Pour ce faire, il existe de nombreuses techniques, citons par exemple : *Multiobjective Genetic Sampling for Imbalanced Classification (E-MOSAIC)* Fernandes et al. [2019], *Exploratory Data Analysis (EDA) for handling duplicate records*, *Synthetic Minority Over-sampling Technique (SMOTE)* Chawla et al. [2002], etc. Après une étude empirique, nous avons décidé d'appliquer la technique SMOTE (en français : Technique de Sur-échantillonnage Synthétique des Minorités) sur notre corpus qui n'est pas en rapport avec un domaine spécifique. Ce choix est dû au fait que cette technique montre beaucoup de succès dans une variété d'applications et de domaines Quan et al. [2021], Ishaq et al. [2021]. Cette technique vise

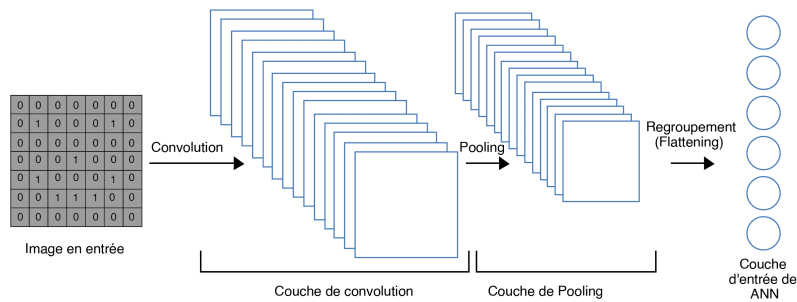


FIGURE 3.3 – Architecture générale du réseau de neurones à convolution appliqué à une image *Haque et al. [2020]*.

à traiter les données non équilibrées par une génération de nouvelles instances au niveau de la classe minoritaire. L'idée principale du fonctionnement de SMOTE consiste à dupliquer des exemples de la classe minoritaire, mais ces exemples ne fournissent pas de nouvelles informations au modèle. Pour ce faire, elle s'appuie sur l'utilisation de l'algorithme k plus proches voisins. La figure 3.4 présente une démonstration du processus de génération d'instances synthétiques en utilisant la méthode SMOTE.

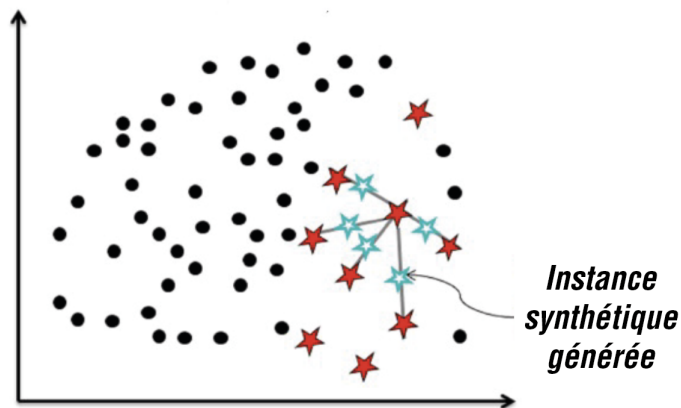


FIGURE 3.4 – Illustration du processus de génération d'instances synthétiques à l'aide de la méthode SMOTE *Soufan et al. [2015]*.

3.3.5 Capture des dépendances à long terme entre les unités lexicales pour la tâche de classification des symptômes de TP

Après la détection des unités lexicales pertinentes, nous visons à ce niveau de maintenir les liens de dépendance entre ces unités afin de préserver le contexte. Cela est dû au fait que généralement les symptômes de TP n'apparaissent pas au niveau des mots mais au niveau des phrases comme la contradiction.

Dans ce contexte, le modèle LSTM a été considéré dans plusieurs travaux de recherche comme la meilleure technique d'apprentissage profond permettant la réalisation de cette tâche [Ombabi et al. \[2020\]](#). Ce modèle est une extension du modèle de base de RNN qui est utilisé pour traiter les problèmes de prédiction de données séquentielles [Kalchbrenner et al. \[2015\]](#). Dans le tableau 3.4, nous avons récapitulé certaines différences entre les différents types de réseaux de neurones ¹⁶.

TABLE 3.4 – Comparaison des types de réseaux de neurones (MLP (ANN) vs. RNN vs. CNN) [thetechplatform \[2021\]](#).

	ANN	RNN	CNN
Données	Données tabulaires	Données séquentielles	Données d'images
Connexion récurrente	Non	Oui	Non
Partage des paramètres	Non	Oui	Oui
Relation spatiale	Non	Non	Oui
Gradient de disparition & Gradient d'explosion	Oui	Oui	Oui

LSTM a été conçue pour résoudre le problème d'explosion et de disparition du gradient rencontré par les RNN. Cela s'explique par le fait que les réseaux de neurones récurrents classiques ne sont efficaces que pour le traitement des dépendances à court terme. Mais d'un point de vue pratique, nous pouvons nous retrouver dans une situation de décalages de durée inconnue entre les différents événements d'une série temporelle [Graves \[2012\]](#). Par conséquent, l'avantage fondamental d'utiliser LSTM est sa capacité à mémoriser *keep in mind* les valeurs passées pendant une durée quelconque et à gérer le flux d'informations dans le réseau, comme les prévisions fondées sur des données de séries chronologiques. Pour ce faire, LSTM applique une exécution récursive du bloc de cellules courant en utilisant l'ancien état masqué et l'entrée courante, tout en utilisant des techniques avancées comme : (i) Dropout : cet élément sert à éviter une adaptation excessive du modèle. Il enlève du réseau des informations étrangères qui ne sont pas utiles pour un traitement ultérieur, ce qui peut également contribuer à améliorer les performances du modèle, (ii) Dense Layer : cette couche associe chaque entrée à chaque sortie avec une telle pondération.

Les LSTM bidirectionnels (BI-LSTM) sont une extension des LSTM traditionnels qui peuvent améliorer les performances du modèle en ce qui concerne les pro-

16. <https://www.thetechplatform.com/post/cnn-vs-rnn-vs-ann-analyzing-3-types-of-neural-networks-in-deep-learning>

blèmes de classification des séquences [Siarni-Namini et al. \[2019\]](#). Pour ce faire, le modèle BI-LSTM utilise deux couches LSTM plutôt qu'une seule sur la séquence d'entrée. La première est appliquée à la séquence d'entrée à l'état actuel, et la deuxième est appliquée à une copie inversée de la séquence d'entrée. Ce type de traitement peut fournir un contexte supplémentaire au réseau et conduire à un apprentissage plus rapide et plus complet du problème. La figure 3.5 présente un exemple du fonctionnement du réseau de neurones BILSTM dans le but de conserver le contexte figuré entre les différentes unités lexicales.

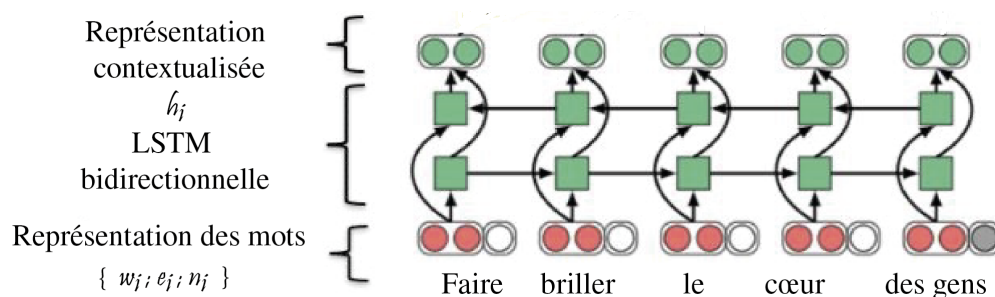


FIGURE 3.5 – Extrait d'une méthode basée sur le réseau de neurones BILSTM [Gao et al. \[2018\]](#).

Pour cela dans notre travail, nous avons concaténé la sortie de la couche convolutive à deux couches LSTM pour un maintien solide de la dépendance entre les différentes unités lexicales. La sortie de ces deux couches est ensuite transmise à une couche composée de 6 neurones permettant de mesurer le degré de chacun des symptômes d'une maladie, comme la méfiance exagérée, l'interprétation négative des gestes d'autrui, le doute incessant, etc.

Dans cette partie, nous avons considéré le type d'apprentissage comme une classification au lieu d'une régression, cela est dû au fait que, dans certains cas, il est difficile d'obtenir un grand nombre d'instances pour certains degrés. Il est à noter que l'étape de classification a été répétée en fonction du nombre de symptômes de la maladie à détecter, étant donné que dans notre cas il s'agit d'un problème de multi-label. En effet, une personne peut avoir plus d'un symptôme en même temps.

3.3.6 Classification binaire pour la tâche de détection de trouble paranoïde

D'une manière générale, la tâche de détection des maladies est très sensible, cette sensibilité est due au manque de règles spécifiques permettant de prendre des décisions déterminantes. Par exemple, il n'est pas nécessaire que la personne présente tous les symptômes pour affirmer qu'elle est malade. De plus, les symptômes peuvent être contradictoires, par exemple, il est difficile de trouver quelqu'un qui a le doute incessant et l'estime de soi à la fois. Il existe également de nombreuses combinaisons des degrés de symptômes. Pour cela, dans cette étape, nous avons profité du degré de tous les symptômes détectés depuis l'étape précédente afin de construire un vecteur qui représente la liste des degrés de symptômes chez une personne. Puis nous avons transféré ce vecteur à la couche SVM

pour la détection de la maladie. Dans cette partie, le nombre de fonctionnalités est réduit pour cela, nous nous sommes limités aux algorithmes de classification classiques tels que SVM, Naïve Bayes, Arbre de décision, etc. Suite à une étude empirique, nous avons décidé d'utiliser l'algorithme SVM, car c'est l'un des algorithmes d'apprentissage les plus configurables, offrant ainsi plus d'opportunités d'obtenir de bons résultats. En outre, beaucoup d'études ont constaté que la meilleure combinaison de l'apprentissage profond avec les algorithmes d'apprentissage traditionnels pour l'analyse textuelle est obtenue par SVM [Chen et Zhang \[2018\]](#), [Ombabi et al. \[2020\]](#), [Thaiyalnayaki \[2021\]](#). De plus, pendant la phase de traitement, SVM tient compte simultanément du taux d'erreur et de la complexité, c'est-à-dire qu'il tente de réduire les deux à la fois [Dilrukshi et al. \[2013\]](#).

Il est à noter que dans le but d'assurer la notion de multi-label permettant de prédire la présence de chaque type de maladie parmi les dix types de TP. Il est possible de reproduire l'exécution du modèle proposé pour chaque type de maladie, passant à chaque fois par une prédiction des degrés de symptômes associés à cette maladie. Malgré l'intérêt de la prédiction de la présence de la maladie passant par la prédiction des symptômes. Ce passage reste difficile à réaliser car cela nécessite un effort considérable pour l'annotation. Particulièrement dans le cas de la détection d'un trouble de la personnalité puisque celui-ci comporte dix types. Pour cette raison, nous avons modifié notre méthode et nous n'avons pas suivi le même processus pour la détection d'autres types de TP. La modification réside dans l'élimination de notre architecture, la phase de prédiction des symptômes. Pour cela, nous avons directement associé la sortie de la couche BiLSTM à la couche SVM, comme il est montré dans la figure 3.6. Le processus suivi pour détecter la présence d'un type de trouble selon notre nouvelle architecture est le suivant : En commençant par la vérification de la présence d'une maladie de TP au sens général, c'est-à-dire quel que soit la nature du trouble. Par la suite, en passant à vérifier l'existence de chaque catégorie parmi les catégories de TP. Enfin, en clôturant notre processus par une vérification de l'existence de chaque type parmi les catégories détectées. Pour détecter les catégories et les types de TP, nous avons utilisé une classification binaire. Ce type de classification permet d'assurer l'aspect multi-label (une instance peut appartenir à plusieurs classes), car plusieurs troubles peuvent survenir simultanément. Après la détection des différents types de TP en utilisant durant les différentes étapes de notre méthode des techniques basées sur la notion d'apprentissage profond. Nous procédons par la suite à la détection de TP, mais en utilisant des techniques d'apprentissage traditionnelles afin d'évaluer l'influence de chaque type d'information sur la personnalité.

3.4 MÉTHODE PROPOSÉE BASÉE SUR LES TECHNIQUES D'APPRENTISSAGE CLASSIQUES

Notre méthode proposée dans la figure 3.7 repose sur une technique d'apprentissage classique permettant d'analyser les comportements des utilisateurs des réseaux sociaux dans le but de détecter les personnes ayant des TP. Elle se compose de deux grandes parties : (i) extraction des attributs : en effet, notre méthode tire profit des informations obtenues à partir de la partie textuelle des tweets en utilisant des techniques de traitement du langage naturel, comme l'analyse

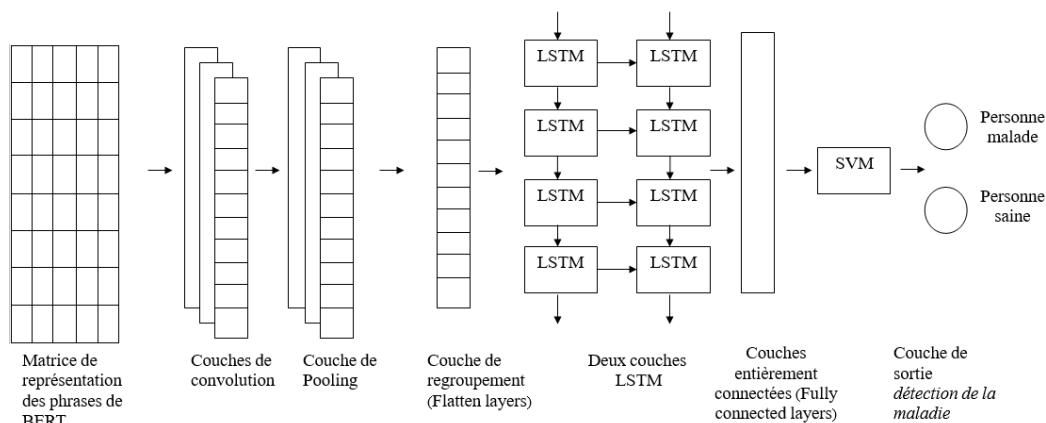


FIGURE 3.6 – Proposition de modèle CNN-LSTM profond pour la vérification de l’existence de la maladie, de sa catégorie et de son type.

morphologique, l’analyse syntaxique, etc. De même, elle tire profit des informations sur l’utilisateur, telles que le nombre d’abonnés et le nombre de publications. En outre, des informations liées au tweet, telles que le nombre de symboles et le nombre d’hashtags. Finalement, des informations relatives à la date de publication des tweets, par exemple le nombre de publications dans la matinée, (ii) classification : après l’extraction des différents types d’attributs, nous visons par la suite à trouver la bonne combinaison de ces attributs afin de prendre les meilleures décisions. Il existe dans la littérature plusieurs algorithmes d’apprentissage classique permettant de construire à partir des données d’entraînement le meilleur modèle permettant d’atteindre l’objectif désiré. Dans notre travail, nous avons pensé à intégrer au fonctionnement d’un algorithme existant la notion de logique floue. L’intégration de ce concept permet de mesurer le degré d’appartenance d’une instance à une classe, ce qui rend notre méthode plus proche de l’esprit de la pensée humaine et naturelle KAY [2016].

3.4.1 Prétraitement

Le processus de la présente méthode est principalement dédié aux données provenant de Twitter étant donné que c’est le meilleur type de réseau social qui offre plus de souplesse lors de l’extraction de données. Pour cela, dans notre travail, nous nous appuyons sur l’ensemble des attributs fournis par ce réseau afin de détecter la présence de troubles chez un utilisateur. En effet, ce type d’information est souvent implicite, ce qui signifie que parfois, même pour un être humain, il n’est pas facile de la détecter. Dans cette optique, chaque information renvoyée par Twitter peut s’avérer utile et contribuer au processus décisionnel.

Les données issues des réseaux sociaux sont généralement sous le format semi-structuré (XML¹⁷ ou JSON¹⁸). Pour cela nous avons extrait au départ les attributs ayant des valeurs exactes, précises et claires, tels que *le nombre d’amis et le nombre de publications*. Ensuite, pour la valeur de l’attribut *full_text* (comportant le contenu textuel) nous avons appliqué le processus de filtrage et de normalisation men-

17. Extensible Markup Language (XML) qui permet de définir et de stocker des données de manière à pouvoir les partager.

18. (JavaScript Object Notation) est un format de fichier textuel conçu pour l’échange de données.

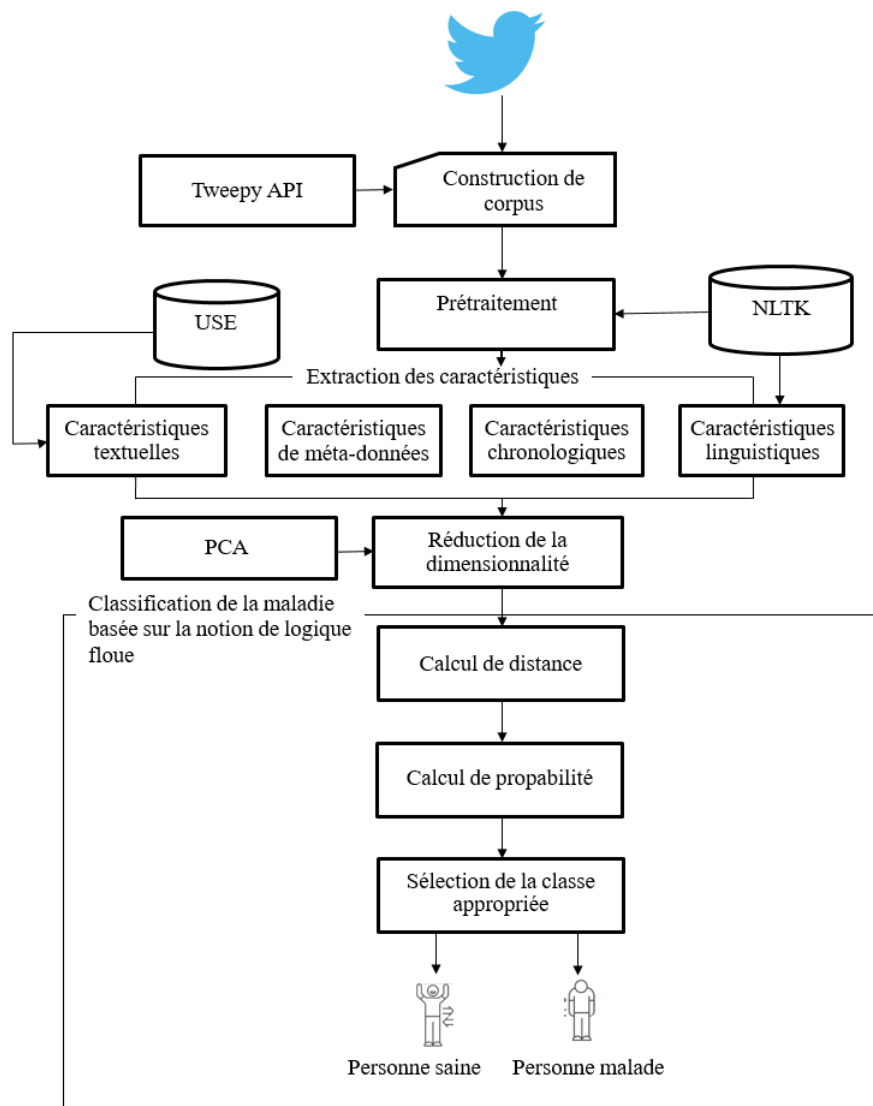


FIGURE 3.7 – Méthode fondée sur la notion de logique floue pour détecter les personnes souffrant de TP sur Twitter.

tionné dans la section 3.2.1 *Prétraitement* visant à supprimer les symboles et les mots vides dans un premier temps et dans un deuxième temps à normaliser le texte en supprimant les spécificités relatives aux règles grammaticales et conjuguales.

3.4.2 Extraction des caractéristiques

L'étape d'extraction des caractéristiques à partir des données est une étape primordiale dans le processus de classification. Pour cela, dans cette étape, nous avons l'intention de faire une collection riche et variée de caractéristiques relatives aux profils des utilisateurs sur Twitter, ce qui conduit à une détection précise des troubles. Pour ce faire, nous avons utilisé différents outils et techniques de fouille de données et de TALN afin de déduire les informations implicites liées au style

d'écriture et à l'état de la personne qui n'ont pas été explicitement indiquées par l'utilisateur.

Dans notre travail, nous avons considéré quatre types de caractéristiques : les caractéristiques du compte utilisateur, les caractéristiques liées aux réactions et au temps associé au tweet et les caractéristiques en rapport avec le style d'écriture. Le choix de ces caractéristiques est effectué sur la recommandation de notre expert du domaine (psychiatre), également en fonction de l'analyse de certains travaux de la littérature [Mbarek et al. \[2022\]](#).

Extraction de caractéristiques textuelles : tout le contenu textuel illustré dans un tweet a été transformé en un vecteur numérique traitable par l'algorithme. De préférence cette transformation doit être exhaustive en intégrant autant d'informations linguistiques que possible, comme les critères sémantiques. Pour cela, au cours de cette phase, nous avons utilisé des techniques avancées basées sur des modèles d'apprentissage profond tels que USE, InferSent, BERT afin de mettre en évidence l'aspect syntaxique et sémantique de la phrase.

Dans cette partie, nous avons choisi d'utiliser la technique de USE, puisque nous disposons déjà d'un grand nombre d'attributs, et USE est l'une des techniques qui transforme le tweet en un vecteur de taille minimale. De même, cette technique est basée sur le réseau *Deep Averaging*¹⁹, assurant des performances élevées en termes de précision et de temps d'exécution [Gardner et al. \[2019\]](#). Il est à noter que cette technique peut être employée dans de nombreuses tâches telles que la mesure de la similarité sémantique, le regroupement, la classification de textes et d'autres tâches liées au traitement du langage naturel.

Extraction des caractéristiques de méta-données : les caractéristiques de méta-données représentent les différentes informations relatives aux tweets ainsi que les informations relatives aux utilisateurs. Ce type d'information est explicitement indiqué sur le profil de l'utilisateur et peut donner un aperçu sur le comportement de l'utilisateur et leur influence sur les autres. Par conséquent, ce type de renseignements peut donner une idée claire de l'état et de la personnalité d'un utilisateur. Par exemple, ces renseignements nous permettent de connaître le degré de sociabilité de l'utilisateur avec les autres. De plus, nous pouvons collecter les messages avec lesquels l'utilisateur interagit via des *likes* ou des *retweets* afin de préciser leur intérêt. Dans notre contexte, ce type d'information peut refléter le trouble chez les personnes. Par exemple, il est connu que parmi les symptômes d'une personne paranoïaque figure la méfiance exagérée. Ce symptôme est généralement présent chez les gens qui éprouvent un sentiment de vide dans leur vie. Ce dernier peut se manifester par de nombreuses caractéristiques, telles que le nombre de suiveurs liés à un utilisateur. En outre, généralement une personne paranoïaque évite de mentionner les gens dans ses tweets, de sorte que l'attribut de mention de l'utilisateur du tweet sera nul dans la plupart des cas. Le tableau ci-dessous montre en détail les attributs retournés par Twitter. Au-delà des informations retenues dans notre étude, qui sont présentées dans le tableau 3.5, l'API de Twitter renvoie d'autres informations telles que la localisation, les informations tronquées et protégées. Dans cette étude, nous avons décidé de ne pas les utiliser car les utilisateurs n'accordent pas à ces attributs un grand intérêt et dans de nombreux cas, l'attribut localisation lié à un tweet n'est pas indiqué.

19. <https://github.com/miyyer/dan>

TABLE 3.5 – Les caractéristiques des métadonnées

Informations relatives à l'utilisateur	Informations relatives au tweet
Nombre de favoris de l'utilisateur	Nombre de retweets par tweet
Nombre d'amis de l'utilisateur	Nombre de favoris par tweet
Nombre d'utilisateurs abonnés	Nombre de hashtags par tweet
Nombre de publications de l'utilisateur	Nombre de symboles par tweet
Nombre d'utilisateurs listés par compte	Nombre d'utilisateurs mentionnés par tweet

Extraction des caractéristiques chronologiques : l'heure et la date de publication de chaque post sur les médias sociaux sont des informations utiles fournies par les API des réseaux sociaux. Ces caractéristiques peuvent être considérées comme des critères importants reflétant la situation de l'utilisateur à un moment spécifique. Essentiellement, dans notre travail, nous nous focalisons sur l'étude des problèmes psychologiques qui peuvent être en relation avec les différentes informations extraites à partir de la caractéristique de l'heure de publication.

Par exemple, d'après [Morera-Fumero et al. \[2017\]](#), les personnes ayant des troubles de la personnalité ont tendance à publier habituellement tard le soir. De même, après une étude empirique, nous avons constaté qu'il y a un rapport entre le symptôme d'agression et le nombre de publications par jour. Pour cette raison, nous avons utilisé l'attribut date de publication fourni par l'API de Twitter afin d'extraire cinq caractéristiques à savoir le nombre de publications par jour, le nombre de publications dans la matinée, le nombre de publications l'après-midi, le nombre de publications dans la nuit et tard dans la nuit. Ces attributs sont présentés dans la table 3.6.

TABLE 3.6 – Description des caractéristiques du plan temporel

Caractéristique temporelle	Description
Poster le matin	Pourcentage de tweets postés de 06h00 à 12h00
Poster l'après-midi	Pourcentage de tweets postés de 12h01 à 18h00
Poster la nuit	Pourcentage de tweets postés de 18h01 à 23h59
Poster tard dans la nuit	Pourcentage de tweets postés de 00h00 à 05h59
Publications par jour	Pourcentage de publications de tweets chaque jour

Extraction des caractéristiques linguistiques : les caractéristiques linguistiques jouent un rôle crucial dans la distinction entre les styles d'écriture associés à chaque type de trouble de personnalité. En effet, de nombreuses études sur l'extraction des connaissances à partir des données des réseaux sociaux se sont focalisées sur les caractéristiques linguistiques pour atteindre l'objectif désiré.

Il semble évident que chaque utilisateur sur les réseaux sociaux dispose d'un style d'écriture unique qui dépend de sa personnalité. Par exemple selon [Carey et al. \[2015\]](#), la personne narcissique utilise plus de pronoms en relation avec le soi comme *je, me, moi, etc.* Selon [Matz et Netzer \[2017\]](#), les personnes extraverties s'intéressent davantage au lexique lié aux hôtels et aux voyages.

Par conséquent, l'extraction de ces caractéristiques peut accroître la probabilité de détecter des profils troublants. Pour cela, dans cette partie, nous avons effectué une étude approfondie du style d'écriture des utilisateurs sur Twitter afin de détecter les spécificités linguistiques permettant d'établir une distinction entre les différentes classes. Notre étude rassemble des critères morphologiques, syn-

taxiques et sémantiques présentés dans la table 3.7 et qui sont extraits à l'aide des bibliothèques liées au TALN.

TABLE 3.7 – Extraction des caractéristiques liées au style d'écriture

Type	Description
Caractéristiques numériques	Nombre de chacune des ponctuations
	Nombre des phrases
	Nombre de mots dans une phrase
	Nombre total de caractères
	Nombre total de lettres
	Nombre total d'espaces blancs
	Nombre total d'espaces de tabulation
	Nombre total des caractères spéciaux
	Longueur moyenne des mots
	Nombre de mots supérieur à 6 caractères
	Nombre total des mots courts
	Nombre total de simple quote
	Nombre total de lignes
	Nombre total de mots par phrase
	Nombre moyen de caractères par phrase
	Nombre de chaque mot de 1 à 11 caractères
	Nombre de mots supérieur à 12 caractères
Nombre de chaque lettre de a..z	
Nombre de chaque lettre de A..Z	
Caractéristiques morphologiques	Nombre de chaque POS
	Nombre de chaque Temps utilisé dans chaque phrase
	Nombre d'entités de genre (masculin/féminin)
	Nombre de formes des entités (singulier/pluriel)
	Nombre de chaque type d'entité nommée
Caractéristique syntaxique	Séquence syntaxique
Caractéristiques sémantiques	Le degré de la polarité du sentiment
	Niveau de subjectivité lié au sentiment
	Nombre de chaque relation sémantique
	Le degré de chaque émotions exprimées
	Nombre de chacune des émoticônes présenté

3.4.3 Réduction de la dimensionnalité

Notre objectif dans l'étape de réduction de la dimensionnalité est de passer d'un espace de grande dimension à un espace de faible dimension afin de réduire la taille des caractéristiques et de préserver seulement les propriétés importantes. Plus particulièrement que dans la présente méthode, nous avons choisi une variété de caractéristiques qui peuvent être utiles selon notre interprétation théorique, mais pas sur un plan pratique. Dans ce contexte, il existe plusieurs études dans la littérature qui ont montré l'effet négatif de l'utilisation d'un grand nombre d'attributs lors de la tâche de classification et l'intérêt de l'équilibrage entre le nombre des instances et le nombre d'attributs [Khalid et al. \[2014\]](#).

Il existe plusieurs techniques permettant d'atteindre cet objectif, par exemple *t-distributed stochastic neighbor embedding* (t-SNE) [Wattenberg et al. \[2016\]](#), *Uniform*

Manifold Approximation and Projection (UMAP) [McInnes et al. \[2018\]](#), *Generalized discriminant analysis* (GDA) [Baudat et Anouar \[2000\]](#), etc. Mais après une étude empirique nous avons choisi de travailler avec la technique *Principal component analysis* (PCA) [Baudat et Anouar \[2000\]](#) puisque c'est la méthode la plus utilisée dans le domaine des sciences sociales [Juera et Paña-Tautho](#).

3.4.4 Classification basée sur la notion de logique floue pour la détection des personnes ayant des TP

Après avoir extrait et transformé les caractéristiques en un ensemble de vecteurs numériques, nous poursuivons leur exploitation dans la partie classification. L'idée fondamentale est de classer la nouvelle instance en comparant la similarité au niveau les caractéristiques de cette instance avec les caractéristiques des autres instances présentes dans le corpus d'entraînement. Dans la littérature, il existe des divers algorithmes de classification, nous citons comme des exemples SVM, Gaussian NB, KNN, Random Forest, XGBoost. Dans le tableau 3.8, nous présentons les divers éléments qui influencent les performances des différents modèles d'apprentissage automatique, ainsi que les moyens permettant d'améliorer les résultats obtenus.

Suite à une étude empirique sur les différents algorithmes de classification, nous décidons d'abord de choisir l'algorithme d'apprentissage automatique KNN [Zhang et al. \[2018\]](#) comme une méthode de base (baseline). Puis de développer un algorithme qui permet d'améliorer le fonctionnement de KNN afin de réduire ses inconvénients (voir algorithme 1). Nous citons dans le paragraphe suivant certains inconvénients du modèle de base de KNN :

- Pour calculer la distance entre les données d'entraînement et les données de test, KNN applique une seule métrique de distance parmi les différentes mesures de distance existantes (distance euclidienne, distance de Minkowski, distance de Mahalanobis, etc.). Selon [Abu Alfeilat et al. \[2019\]](#), la mesure de distance utilisée est celle qui permet à l'algorithme KNN de classer les instances de test avec les meilleurs résultats en termes de métriques d'évaluation telles que la précision et le rappel. Ce processus est appliqué par KNN indépendamment du domaine d'étude. Bien que cela puisse être suffisant pour certains domaines, il peut être limité pour d'autres, en particulier pour le domaine médical où les résultats ont une forte incidence. Par conséquent, des vérifications supplémentaires sont nécessaires.
- Le résultat de l'algorithme KNN de base repose sur une logique discrète, ce qui signifie qu'une instance ne peut appartenir qu'à une seule classe. À cet effet, plusieurs études ont présenté une extension de l'algorithme KNN fondée sur une logique floue [Krishnaiah et al. \[2015\]](#), [Hribar et Duka \[2010\]](#).
- KNN est basé sur le calcul du nombre d'apparitions de chaque classe parmi les k instances les plus proches de la nouvelle instance pour l'étape de classification. C'est-à-dire, si nous voulons classer une instance qui fait déjà partie du corpus d'entraînement avec les détails suivants : (i) cette instance apparaît dans le corpus d'entraînement avec la classe A ; (ii) la valeur de k est fixée à 3 ; (iii) les deux autres instances les plus proches ayant la classe B .

TABLE 3.8 – *Avantages et inconvénients des différents modèles d’algorithmes de machine learning Liu et Lang [2019].*

Algorithmes	Avantages	Inconvénients	Mesures d’amélioration
ANN	Aptitude à traiter des données non linéaires; forte aptitude à s’ajuster.	Possibilité de surapprentissage; risque de rester bloqué dans un optimum local; la formation du modèle demande du temps.	Adoption d’un optimiseur amélioré, de fonctions d’activation et de perte.
SVM	Tirer des informations utiles d’un ensemble de données réduit; Capacité de production élevée.	Fonctionne mal avec les Big Data ou les tâches de classification multiples; Sensible aux paramètres des fonctions du noyau.	Paramètres optimisés par l’algorithme PSO (particle swarm optimization).
KNN	S’applique aux données massives; s’adapte aux données non linéaires; s’entraîne rapidement; robuste au bruit.	Faible précision sur la classe minoritaire; Temps de test long; sensibilité au paramètre K	Équilibrage des séries de données à l’aide de la technique de suréchantillonnage synthétique minoritaire (SMOTE).
Naïve Bayes	Robuste au bruit; Capacité à apprendre progressivement.	N’est pas efficace sur les données des attributs.	Variables latentes importées afin de libérer l’hypothèse d’indépendance.
LR	Simple, peut être formé rapidement; Ajustement automatique des caractéristiques.	Ne fonctionne pas bien sur les données non linéaires; Possibilité de surapprentissage	Régularisation importée afin d’éviter le sur-apprentissage.
Decision tree	Sélection automatique des caractéristiques; interprétation puissante.	Tendances des résultats de la classification selon la catégorie de la majorité; ne pas tenir compte de la corrélation des données.	Équilibrage des ensembles de données avec SMOTE; ajout de variables latentes.
K-means	Simple, peut se former rapidement; très extensible; peut s’adapter aux données volumineuses.	Ne fonctionne pas bien sur des données non convexes; Sensible à l’initialisation; Sensible au paramètre K.	Méthode d’initialisation améliorée.

Dans ce cas, KNN classe cette instance dans la classe B ce qui ne correspond pas à la classification attendue.

- KNN nécessite un coût de calcul élevé pour mesurer la distance relative entre chaque échantillon de test et tous les échantillons d’entraînement.

Par conséquent, nous concevons notre algorithme comme suit :

- Plus configurable que KNN, nous faisons plus d'expériences en combinant le résultat de plusieurs fonctions de calcul de distance. Nous adoptons dans ce contexte le modèle de classification fondé sur le vote afin d'accorder plus de valeur au résultat d'une fonction par rapport à une autre. Par conséquent, cette modification nous donnera plus de flexibilité et nous permettra d'obtenir de meilleurs résultats.
- Plus précis en utilisant la logique floue, étant donné que cette probabilité peut être considérée comme la fiabilité (degré de confiance) de la réponse retournée par l'algorithme de classification. En général, dans le domaine de l'intelligence artificielle, la notion de logique floue est couramment utilisée pour tenter d'imiter le raisonnement du cerveau humain [KAY \[2016\]](#).
- Capable d'assurer le concept de multi label, une instance peut appartenir à plusieurs classes en même temps.
- Capable de sélectionner la classe avec la valeur de probabilité la plus élevée parmi les k éléments les plus proches, plutôt qu'en fonction du nombre d'apparitions de chaque classe.

Algorithm 1 Classification des personnes avec trouble de la personnalité : cas de trouble paranoïde

Result : Prédiction de la personne paranoïaque

Input : Function1, k1, Function2, k2

for each insttest in Xtest **do**

for each insttrain in Xtrain **do**

 tabFct1.append(Function1(insttest,insttrain))

 tabFct2.append(Function2(insttest,insttrain))

 tabIndKMinFct1=indices of minimums k1 elements in tabFct1

 tabIndKMinFct2=indices of minimums k2 elements in tabFct2

for ind_min in tabIndKMinFct1 **do**

if Ytrain[ind_min]==Paranoid **then**

 | class1Fct1=class1Fct1+tabFct1[ind_min]

else

 | class2Fct1=class2Fct1+tabFct1[ind_min]

for ind_min in tabIndKMinFct2 **do**

if Ytrain[ind_min]==Paranoid **then**

 | class1Fct2=class1Fct2+tabFct2[ind_min]

else

 | class2Fct2=class2Fct2+tabFct2[ind_min]

 ProbFct1C1=class1Fct1/(class1Fct1+class2Fct1)

 ProbFct1C2=class2Fct1/(class1Fct1+class2Fct1)

 ProbFct2C1=class1Fct2/(class1Fct2+class2Fct2)

 ProbFct2C2=class2Fct2/(class1Fct2+class2Fct2)

 TotProbC1= (ProbFct1C1+ProbFct2C1)/2

 TotProbC2= (ProbFct1C2+ProbFct2C2)/2

 print("Person is Parnoid with"+TotProbC1)

 print("Person is Not Parnoid with"+TotProbC2)

Il est à noter que dans notre travail, nous nous sommes concentrés exclusivement sur la performance opérationnelle, en faisant abstraction des problèmes liés au temps d'exécution. La première étape de l'algorithme proposé consiste à calculer la distance qui sépare chaque instance dans le corpus de test X_{test} avec toutes les instances existantes dans le corpus d'apprentissage X_{train} en appliquant les deux mesures statistiques : *Fonction1* correspond à *Canberra*²⁰ et *Fonction2* correspond à *Chebyshev*²¹, puis les stocker dans les tables $tabFct1$ et $tabFct2$ respectivement. Ensuite, nous choisissons les nombres d'instances k_1 et k_2 du corpus d'apprentissage les plus proches de l'instance à classer ($inst_{test}$ dans notre cas) à l'aide des distances les plus petites respectivement dans les tables $tabFct1$ et $tabFct2$ et nous les stockons respectivement dans les tables $tabIndKMinFct1$ et $tabIndKMinFct2$. Par la suite, nous calculons pour les résultats de chacune des fonctions la somme des distances par rapport à la classe Paranoïaque ($class1Fct1$, $class1Fct2$) et la somme des distances par rapport à la classe Non Paranoïaque ($class1Fct1$, $class1Fct2$). Puis, nous calculons pour la fonction *Fonction1* les deux probabilités qui indiquent le degré d'appartenance d'une instance à chaque classe : $ProbFct1C1$ et $ProbFct1C2$ et pour la fonction *Fonction2* les deux probabilités : $ProbFct2C1$ et $ProbFct2C2$. Enfin, nous combinons les différentes probabilités des deux fonctions par classe : $TotProbC1$ et $TotProbC2$.

Il est à noter que, nous avons augmenté le nombre de paramètres par rapport à KNN pour rendre notre algorithme plus configurable. De plus, notre algorithme peut être facilement étendu pour inclure d'autres fonctions. En outre, nous pouvons négliger l'effet d'une fonction en lui attribuant un zéro pour son K. Il existe plusieurs manières de combiner les résultats de différentes fonctions, telles que la moyenne harmonique²². Mais le problème avec cette technique est qu'elle fournit uniquement le classement des résultats, pas la probabilité que nous la considérons comme le degré d'appartenance d'une instance à une classe. Il existe de nombreuses formules pour mesurer la distance entre deux vecteurs. Suite à une étude empirique, nous avons opté pour les deux mesures *Canberra* et *Chebyshev*. Le choix de K peut être paramétrable selon la fonction utilisée ainsi selon le résultat obtenu. Il est préférable d'avoir des différentes valeurs de k_1 et k_2 , avec k_1 pair et k_2 impair ou vice versa, afin d'éviter tout chevauchement potentiel. Dans le cas où les décisions de détection d'une maladie se chevauchent (avec un pourcentage de 50% pour chaque classe), il est recommandé de considérer la personne comme étant malade en raison de l'importance du rappel par rapport à la précision dans le domaine médical. Par conséquent, il est préférable de surveiller attentivement la situation de cette personne plutôt que de négliger son cas.

3.5 CONCLUSION

Dans ce chapitre, nous nous focalisons sur la détection des personnes atteintes de troubles de la personnalité sur les réseaux sociaux. Nous nous sommes intéressés en particulier à l'analyse des textes en français produits par les utilisateurs. Pour ce faire, nous avons proposé deux méthodes, la première a focalisé sur une

20. https://en.wikipedia.org/wiki/Canberra_distance

21. https://en.wikipedia.org/wiki/Chebyshev_distance

22. https://fr.wikipedia.org/wiki/Moyenne_harmonique

architecture neuronale profonde pour la classification des personnes ayant des troubles de la personnalité en prenant en compte les particularités linguistiques, tels que le contexte. La deuxième méthode proposée a été fondée sur l'apprentissage classique et la logique floue pour assurer la détection des troubles à l'aide des caractéristiques personnalisées, importantes dans l'analyse des éléments psychologiques. En résumé, ces méthodes offrent des informations complémentaires pour détecter les troubles de la personnalité sur les réseaux sociaux, en combinant des techniques d'apprentissage automatique avancées avec des éléments linguistiques et psychologiques spécifiques. Ces travaux ouvrent de nouvelles perspectives dans la compréhension et la détection précoce de ces troubles, et peuvent contribuer à l'amélioration de la santé mentale en ligne.

Le prochain chapitre exposera notre seconde contribution, qui consiste à détecter les causes des troubles en identifiant les sujets les plus préoccupants pour les utilisateurs ayant des troubles de la personnalité.

IDENTIFICATION DES CAUSES AFFECTANT LES TROUBLES DE LA PERSONNALITÉ

4

SOMMAIRE

4.1	INTRODUCTION	76
4.2	CORPUS UTILISÉ	77
4.3	GESTION DES SPÉCIFICITÉS DES MULTI-VARIATIONS DES DOMAINES	78
4.4	IDENTIFICATION DES SUJETS DES TWEETS	80
4.4.1	Prétraitement des données et traitement des caractéristiques	80
4.4.2	Regroupement des sujets d'étude	81
4.5	CONSTRUCTION AUTOMATIQUE D'ONTOLOGIES DE DOMAINES	84
4.5.1	Sélection, normalisation et enrichissement des concepts	85
4.5.2	Extraction des relations sémantiques	86
4.5.3	Attribution des propriétés aux données extraites	88
4.5.4	Affectation des individus aux données extraites	88
4.6	CONSTRUCTION AUTOMATIQUE DES RÈGLES D'INTERROGATION L'ONTOLOGIE	90
4.6.1	Construction des règles	90
4.6.2	Interrogation de l'ontologie	92
4.7	CONCLUSION	94

4.1 INTRODUCTION

Dans les deux premiers chapitres, nous avons présenté une étude élargie sur les différentes notions liées aux troubles de la personnalité et l'intérêt de réaliser un traitement profond en vue de fournir des résultats détaillés permettant l'extraction de connaissances utiles. Cela garantit la prestation de services appropriés à ces patients afin de répondre à leurs besoins et de renforcer une sensation de bien-être. Ensuite, dans le troisième chapitre, nous avons exposé nos deux méthodes proposées pour détecter les différents types de troubles de la personnalité. En nous basant sur le même principe concernant la capacité des réseaux sociaux par rapport aux systèmes existants basés sur la notion d'IOT (l'internet des objets) [Jayarathna et al. \[2019\]](#), [Sundaravadivel et al. \[2020\]](#), pour surveiller l'état des utilisateurs. Étant donné que les réseaux sociaux sont considérés comme un espace libre et incontrôlé, qui permet à leurs utilisateurs de se sentir à l'aise.

Se limiter à la détection de personnes ayant un TP sur les réseaux sociaux, comme dans [Gamache et al. \[2018\]](#), [Celli et al. \[2016\]](#), ne permet ni d'extraire des connaissances utiles ni d'aider ces personnes à surmonter leurs problèmes. À cet effet, dans ce chapitre, notre objectif est d'étendre notre travail afin d'enrichir les résultats obtenus à partir de l'étape précédente. Cette extension vise à garantir la cohérence des concepts afin de favoriser l'explicabilité et l'intelligibilité des résultats. Notre méthode d'enrichissement, présentée dans ce chapitre, consiste à associer un module permettant d'identifier la raison pour laquelle le patient a été considéré comme malade lors de l'étape précédente. La raison du trouble peut être présentée sous la forme de sujet de tweet à propos duquel le patient a montré un comportement instable. Grâce à cette détection, nous pouvons identifier les sujets les plus préoccupants, susceptibles d'influencer négativement le comportement des individus. Bien que la détection de ce type d'information reste difficile à assurer, en raison de nombreux problèmes linguistiques qui peuvent intervenir. Parmi ces problèmes, nous pouvons citer la variété du lexique, l'évolution du lexique dans le temps, l'appartenance de certaines unités lexicales à plusieurs sujets à la fois, etc. Mais tous ces problèmes peuvent constituer un véritable défi, en particulier avec l'évolution des techniques d'intelligence artificielle.

Pour ce faire, nous entamons ce chapitre par la catégorisation des tweets à l'aide de la technique de regroupement hiérarchique, tout en appliquant de manière récursive une méthode d'apprentissage non supervisée. Ensuite, nous procédons à la modélisation du lexique lié à chaque thème spécifique, extrait de l'étape précédente, à l'aide d'une ontologie construite de manière automatique qui prend en compte le lexique reflétant l'humeur liée à ce sujet. Nous assurons ce processus en combinant des étapes permettant d'extraire d'une manière automatique des concepts, des liens sémantiques, des attributs et des instances. Ce traitement est réalisé en tirant parti d'une variété de techniques visant à mieux gérer les différents défis liés à l'analyse des données provenant des réseaux sociaux. Enfin, nous clôturons notre contribution dans ce chapitre par la construction des règles généralisées de manière automatique à l'aide des techniques statistiques.

4.2 CORPUS UTILISÉ

Dans notre quatrième chapitre, nous avons utilisé des données composées de 6180 tweets obtenus à partir de Twitter en temps réel à l'aide de l'outil Spark Streaming de manière non successive pour avoir plus de diversité au niveau des sujets abordés (le sport, la politique, le Covid-19, etc). Tous ces tweets sont écrits en français et admettent des séquences de mots reflétant les symptômes des troubles, en se basant sur les trois tableaux du premier chapitre 1.1, 1.2 et 1.3.

Par rapport à notre première contribution dans laquelle nous avons annoté la présence de TP sur la base des 20 derniers tweets. Dans cette étape, nous avons réalisé les différentes annotations basées sur un seul tweet, car le sujet et l'humeur correspondent à un tweet. Il est à noter que chaque annotation de tweet est faite indépendamment du fait que ce tweet soit partagé par une personne malade ou non.

D'une manière générale, nous avons réalisé l'annotation des données trois fois pendant cette contribution. La première annotation est effectuée après l'exécution de la première tâche qui consiste à regrouper les tweets de manière récursive. Cette annotation vise à évaluer chaque regroupement effectué. Les autres étapes d'annotations ont été réalisées avant l'exécution des tâches de classification, car il s'agit d'un problème de classification supervisé.

Nous visons, par la deuxième annotation, à détecter les tweets liés au COVID-19, car nous avons décidé de nous focaliser sur ce sujet en raison de son impact social. De plus, COVID-19 est un sujet qui a eu une influence psychologique sur toutes les personnes dans le monde, et pas seulement sur les personnes souffrant de TP. Nous visons par notre troisième annotation à détecter l'humeur des gens associée à chaque concept de sujet de COVID-19. Nous avons choisi de réaliser l'étape d'annotation à chaque fois avec une classification binaire afin de réduire la variation et le degré de chevauchement. Par exemple, est-ce que le tweet porte sur la pandémie de COVID-19 (Oui ou Non)? Le tableau 4.1 donne une description plus détaillée de la répartition de notre corpus à chaque étape.

Généralement, lors de l'annotation de chaque partie du corpus, l'annotateur n'a pas trouvé de véritable chevauchement, ceci est dû au fait qu'au cours de la première annotation l'annotateur a essayé d'être plus diligent puisque les sujets des tweets sont très variés. Et pour la deuxième annotation, le lexique COVID-19 est évident. Pour ce qui concerne la troisième annotation, qui vise à classer l'humeur exprimée au niveau des tweets comme comportement approprié ou inapproprié, ceci présente un certain flou dans le jugement de l'humeur dans certains cas. Les exemples suivants (4.1), (4.2) et (4.3) illustrent bien le flou figuré.

(4.1) *Lors de l'interview du 14 juillet, le président XXXX a semblé très auto-satisfait de ses résultats face au Covid.*

Pour l'annotateur cette phrase n'admet pas l'avis de l'écrivain. Cependant, il a jugé de la mettre dans la classe regroupant un avis neutre et positif.

(4.2) *Le président XXXX. C'est mignon, c'est sympa.*

Malgré les adjectifs qui montrent une opinion positive, cette phrase est considérée comme une phrase avec un comportement négatif, car elle semble être ironique ou sarcastique.

(4.3) *La bonne nouvelle avec la COVID-19, c'est que ça nous a permis de déterminer qui sont de sombres crétins. Au moins ça.*

Dans plusieurs cas, deux adjectifs contradictoires sur le plan de l'opinion interviennent ensemble au niveau d'un tweet, comme *nouvelle, sombres*. En dépit du chevauchement figuré, il semble que cette phrase a été écrite avec une humeur négative.

Dans la suite de ce chapitre, nous présenterons en détail notre méthode en explorant chacune de ses parties.

TABLE 4.1 – La distribution de notre corpus au cours de chaque étape.

Description	Taille du corpus par tweet
Taille pour le premier regroupement	6180 tweets (309 utilisateurs)
Taille pour le deuxième regroupement	3160 tweets
Taille pour le troisième regroupement	1300 tweets
Taille pour la construction de l'ontologie	700 tweet à propos de la COVID-19
Taille du corpus de thème relatif à la COVID-19	525 tweets pour la phase d'entraînement et 175 tweets pour la phase de test
Taille du corpus pour classer l'humeur	525 tweets pour la phase d'entraînement et 175 tweets pour la phase de test

4.3 GESTION DES SPÉCIFICITÉS DES MULTI-VARIATIONS DES DOMAINES

Dans cette contribution, nous proposons une méthode basée principalement sur des techniques d'apprentissage classiques, linguistiques et statistiques comme indiquée dans la figure 4.1. Cette méthode offre à Twitter la possibilité de comprendre les raisons qui mettent les gens dans une situation de trouble. À cet effet, nous nous sommes appuyés sur une méthode en trois volets : (i) Apprentissage non supervisé récursif : Étant donné que notre corpus traite un ensemble de données de différents thèmes, nous visons dans cette étape à fragmenter cet ensemble en sous-thèmes. En utilisant des mesures statistiques et des ressources linguistiques permettant de spécifier les mots en rapport avec chaque domaine. Nous visons, par notre recours à différentes ressources linguistiques, à rendre notre méthode plus générique et à ne pas se limiter aux données figurant dans notre corpus. La technique d'apprentissage non supervisée appliquée dans ce cadre consiste à mesurer la similarité entre les différents tweets. Nous avons appliqué l'apprentissage non supervisé de manière récursive, afin de permettre à notre algorithme de réaliser le regroupement en descendant dans les niveaux, passant du plus général au plus précis. Ce processus conduit à la construction d'un arbre au bout duquel leur feuille représente un sujet particulier. (ii) Construction de l'ontologie : Au cours de cette étape, nous visons à associer une ontologie à chaque feuille de notre arbre. Cette dernière modélise le lexique associé à chaque domaine présent au niveau de chaque feuille. Cette opération est effectuée en suivant automatiquement le processus de construction d'une ontologie. L'ontologie construite garantira la compréhension des concepts et des relations sémantiques qui y sont associées. En outre, cela nous permet d'attribuer aux concepts les valeurs qui les caractérisent, ce qui nous permet ensuite de détecter d'abord le sujet grâce aux mots décrits dans le tweet, puis de détecter l'humeur exprimée dans le tweet en associant à chaque concept deux listes d'adjectifs possibles. Une liste qui démontre que l'humeur ex-

primée dans le tweet est positive et une autre qui démontre que l'humeur est négative. (iii) Construction des règles permettant l'interrogation de notre ontologie : Dans cette étape, nous visons à construire des règles automatiquement en fonction de la vérification de l'existence ou de l'absence de concepts dans l'ontologie, par rapport à l'annotation effectuée pour détecter le sujet du tweet. Ensuite, en fonction des adjectifs associés au concept et de l'annotation effectuée, nous pouvons détecter l'humeur générale exprimée dans le tweet. En général, notre méthode aborde plusieurs enjeux en même temps, en tenant compte des aspects relatifs à la langue et à l'apprentissage automatique afin de remédier au problème du manque de ressources linguistiques et de fournir des résultats explicables. Nous détaillons dans la suite du présent chapitre les différentes parties de notre méthode.

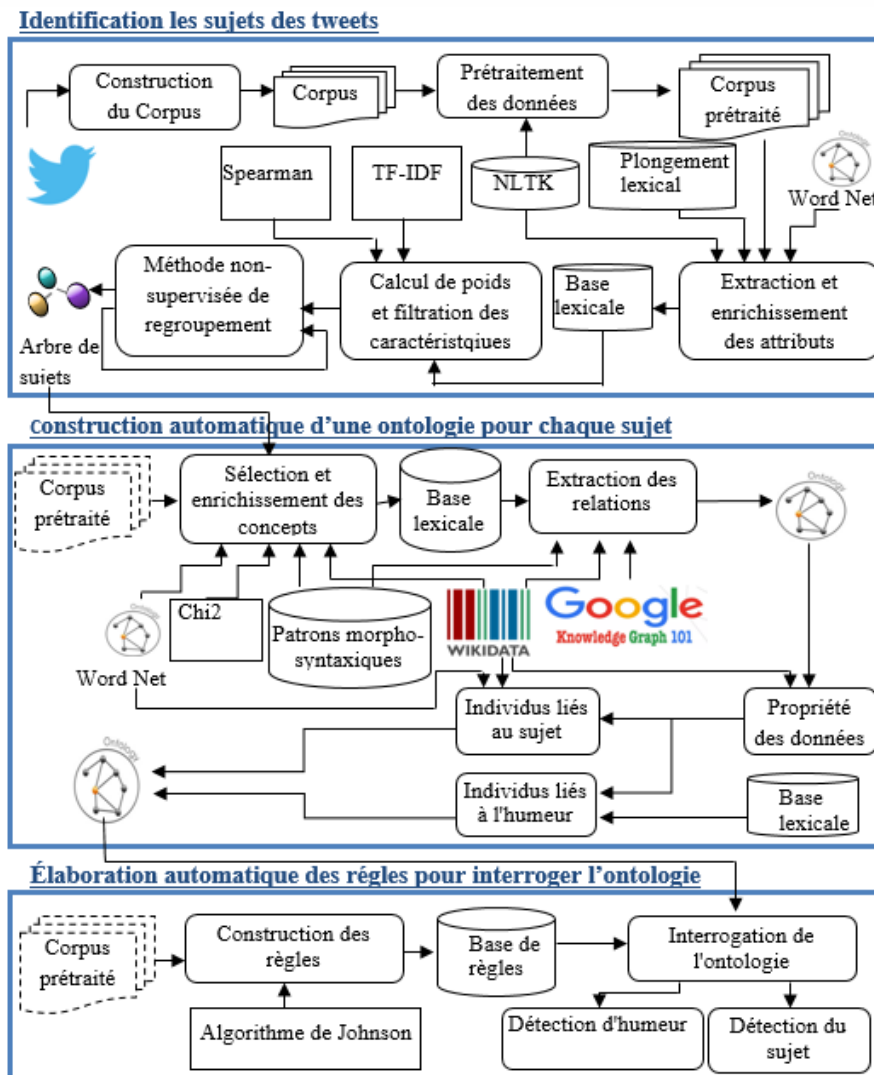


FIGURE 4.1 – Méthode hybride basée sur des techniques linguistiques et statistiques pour la détection des causes provoquant une situation instable.

4.4 IDENTIFICATION DES SUJETS DES TWEETS

L'étape d'identification des sujets des tweets consiste à regrouper hiérarchiquement les catégories de sujets de tweets. Cette partie s'articule autour de deux grandes étapes : une étape de nettoyage et de transformation des données textuelles en données numériques traitables par les algorithmes d'apprentissage automatique. Une deuxième étape qui explore les données numériques afin d'effectuer un regroupement hiérarchique des tweets des patients. Ce processus permet de catégoriser les sujets des tweets à l'aide de l'utilisation des techniques d'apprentissage non supervisées de façon récursive.

4.4.1 Prétraitement des données et traitement des caractéristiques

Dans la phase de prétraitement des données, nous nous concentrons sur l'extraction, l'enrichissement, le filtrage et le calcul des poids des différentes caractéristiques qui représentent dans notre cas, des mots-clés. Pour cette raison, nous avons commencé dans un premier temps à appliquer les diverses techniques classiques de nettoyage de texte, comprenant l'élimination des symboles et des mots vides. En plus, nous avons procédé à la normalisation du texte en transformant les mots en lemmes à l'aide de la bibliothèque NLTK¹. Ensuite, nous avons procédé à la segmentation du document en un ensemble de fragments, qui représente dans notre cas un ensemble de tweets, avec des longueurs invariables. Pour ce faire, nous avons mis en œuvre une technique hybride qui combine à la fois : (i) une technique lexico-sémantique qui consiste à : (a) détecter dans un premier temps des mots simples et composés (1,2,3 Grammes)² qui se répètent plusieurs fois dans le texte et qui ne sont pas des prépositions, (b) enrichir dans un deuxième temps ce lexique en tirant profit des ressources disponibles comme Word embedding [Torregrossa et al. \[2021\]](#) et WordNet [Poli et al. \[2010\]](#). Ce processus consiste à faire correspondre chaque mot qui apparaît dans le lexique avec un ensemble de mots de contexte en utilisant des relations existantes telles que les synonymes, les hyperonymes, les synsets, etc. De plus, nous nous sommes servis de ces ressources afin de ne pas se limiter aux mots qui existent seulement dans notre corpus. Les exemples ci-dessous illustrent l'enrichissement réalisé par ces ressources :

savoir={connaître, compétence, connaissances, cognitif, aptitude, etc.}

dépendance={attachement, interdépendance, adore, attraction, union, etc.}

acceptation={accord, autorisation, permission, compréhension, souplesse, etc.}

(ii) Ensuite, nous avons utilisé la mesure statique TF-IDF³ pour calculer la pondération de chaque mot du corpus. Finalement, nous avons procédé à l'élimination des attributs (lexique) qui n'établissent aucune distinction entre les classes puisque certains mots apparaissent dans le corpus indépendamment de la classe. Cette opération est réalisée à l'aide de la mesure Spearman⁴ en calculant la corrélation entre la pondération d'un mot dans un tweet et le degré de rapport d'autres mots à celui-ci d'une manière approximative, comme il est montré dans les exemples

1. <https://www.nltk.org/>

2. <https://fr.wikipedia.org/wiki/N-gramme>

3. <https://en.wikipedia.org/wiki/Tf%E2%80%93idf>

4. https://fr.wikipedia.org/wiki/Corr%C3%A9lation_de_Spearman

ci-dessous. Il est à noter que le choix de cette mesure a été effectué après une étude empirique.

révolution={*crise, chômage, colère, manifestation, instabilité, etc.*}

économique={*finance, richesse, argent, pouvoir, banque, etc.*}

médecine={*médecin, docteur, médicament, maladie, thérapeutique, etc.*}

Il est à noter que la mesure TF-IDF est une mesure statistique numérique qui est utilisée pour mesurer la signification de l'existence d'un mot dans le corpus. Où tf reflète le nombre de fois que le mot (i) apparaît dans le document (j). Tandis que le composant IDF mesure la signification de l'existence du terme (i) dans tous les documents (n) (voir les équations 4.1 et 4.2).

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{i,j}} \quad (4.1)$$

$$w_{i,j} = tf_{i,j} * \log\left(\frac{N}{df_i}\right) \quad (4.2)$$

L'équation 4.3 représente la mesure de Speaman où $cov(r_{g_X}, r_{g_Y})$ correspond à la covariance des variables de rang, $\sigma_{r_{g_X}}$ et $\sigma_{r_{g_Y}}$ correspondent aux écarts types des variables de rang.

$$r_s = \frac{cov(r_{g_X}, r_{g_Y})}{\sigma_{r_{g_X}} \sigma_{r_{g_Y}}} \quad (4.3)$$

4.4.2 Regroupement des sujets d'étude

L'étape de regroupement des sujets d'étude consiste à combiner les tweets ayant des sujets identiques ou semblables, c'est-à-dire les tweets avec un lexique appartenant au même contexte. Nous mentionnons que parmi les spécificités de notre corpus, il y a la diversité des sujets, ce qui signifie qu'ils ne sont pas reliés à un secteur spécifique. Par conséquent, il est impossible de prédire ni les noms des sujets ni leur nombre.

Pour cette raison, nous avons catégorisé hiérarchiquement ces tweets par sujet en appliquant récursivement des algorithmes d'apprentissage automatique non supervisés. L'apprentissage non supervisé est une branche du concept d'apprentissage automatique qui se caractérise par l'analyse et l'agrégation de données sans étiquetage. Pour ce faire, ces algorithmes apprennent à trouver des modèles qui peuvent créer un ensemble de groupes dans des données hétérogènes sans une intervention humaine. Il existe différents types de méthodes d'apprentissage non supervisées tels que les méthodes de partitionnement, les méthodes hiérarchiques, etc. La figure suivante 4.2 présente les différents types d'apprentissage non supervisé. Nous détaillons dans le tableau ci-dessous 4.2 les fondements de base de chaque type d'apprentissage non supervisé.

Les figures suivantes 4.3, 4.4 et 4.5 présentent des exemples détaillés des résultats obtenus en appliquant les divers types d'apprentissage non supervisé. Ces exemples sont respectivement les résultats des types d'apprentissage non supervisé pour les types à base de partitionnement, de hiérarchique et de densité.

Le résultat de cette technique est une catégorisation hiérarchique basée sur une arborescence qui illustre en détail la sous-catégorisation des thèmes des tweets

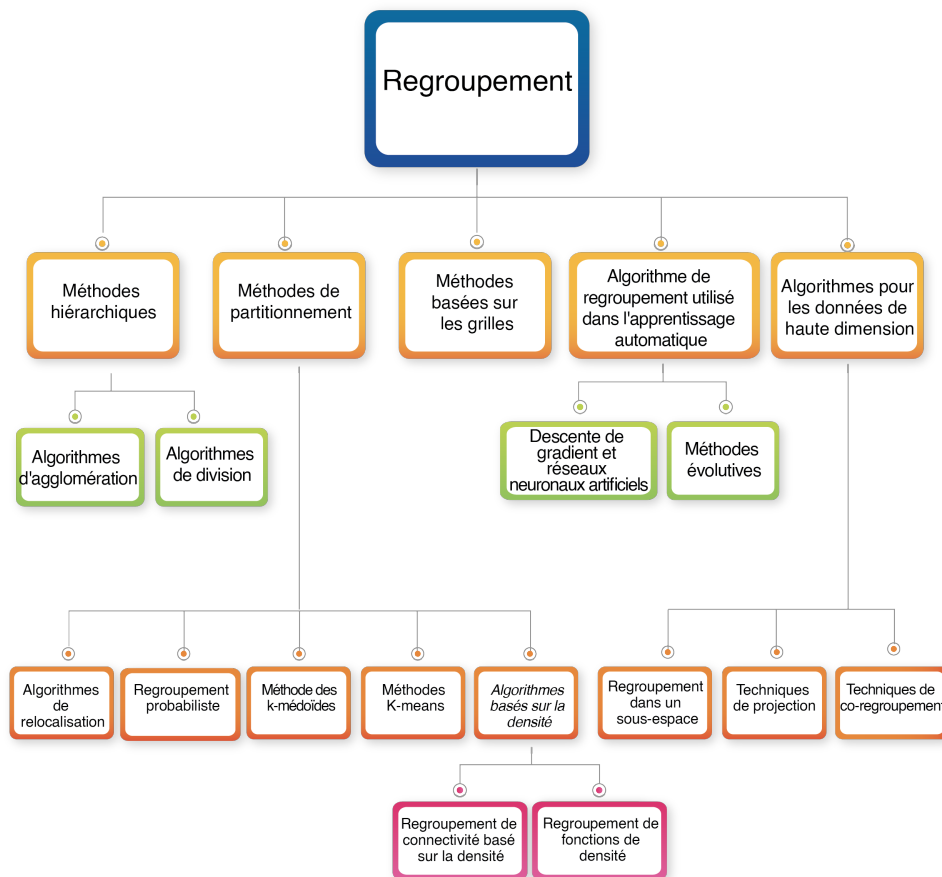


FIGURE 4.2 – Différents types de méthodes d'apprentissage non supervisée Tyagi [2018].

permettant de comprendre les divers niveaux d'abstraction de ces derniers. Les exemples présentés dans la figure 4.6 illustrent la hiérarchie de certains sujets de la typologie des problèmes humains dans le monde.

La catégorisation est effectuée tout en définissant au départ le nombre de classes à chaque niveau hiérarchique, ceci est requis par certains algorithmes tels que k pour l'algorithme k -moyennes. Par ailleurs, il est nécessaire de définir le niveau de profondeur qui représente la condition d'arrêt de la boucle récursive.

Il est préférable que ces nombres ne soient pas élevés afin de réduire au minimum le taux d'erreur commis par l'algorithme. De même, pour éviter un niveau de profondeur élevé notamment que dans notre travail, nous ne sommes pas intéressés par la construction d'une ontologie d'application liée à une organisation spécifique (comme l'hôpital x ou l'équipe de football y).

Pour cela, suite à une étude empirique, nous avons fixé la valeur de K à 2 et la condition d'arrêt en fonction des résultats des scores de dépendance entre les mots de chaque groupe, calculés à l'aide de la ressource Google Knowledge Graph. Mais il est à noter que, dans tous les cas, ce nombre ne doit pas dépasser la valeur de 5. Chaque sujet identifié dans cette partie se distingue par un ensemble d'unités lexicales qui le reflètent. Par conséquent, dans la suite de ce chapitre, nous travaillerons à modéliser le lexique relatif à chaque sujet, de manière à ce que la présence de ces mots dans un tweet puisse permettre d'identifier le sujet.

TABLE 4.2 – Le principe de fonctionnement de chaque méthode d'apprentissage non supervisé [Tyagi \[2018\]](#).

Méthode	Les caractéristiques générales
Méthodes de partitionnement	Trouver des clusters mutuellement exclusifs de forme sphérique Basé sur la distance Peut utiliser la moyenne ou la médiane pour représenter le centre de la grappe. Efficace pour les ensembles de données de petite et moyenne taille
Méthodes hiérarchiques	Le regroupement est une décomposition hiérarchique Ne peut pas corriger les fusions ou les scissions erronées. Peut incorporer d'autres techniques comme le micro-regroupement ou prendre en compte des objets.
Méthodes basées sur la densité	Peut trouver des groupes de formes arbitraires Les groupes sont des régions denses d'objets au sein d'un espace et séparés par des régions de faible densité Densité des groupes : chaque point doit avoir un nombre minimum de points dans son voisinage Peut filtrer les valeurs aberrantes
Méthodes basées sur les grilles	Utilisation d'une structure de données de grille à plusieurs résolutions Temps de traitement rapide (généralement indépendant du nombre d'objets de données, mais il est dépendant de la taille de la grille)

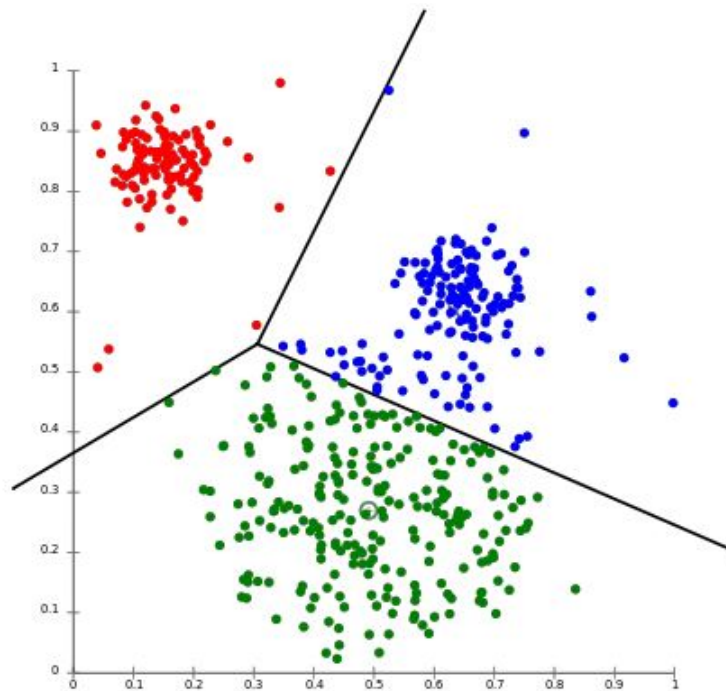


FIGURE 4.3 – Exemple de résultat d'une méthode d'apprentissage non supervisée à base de partitionnement [Gitansh Chadha et Karnin \[2018\]](#).

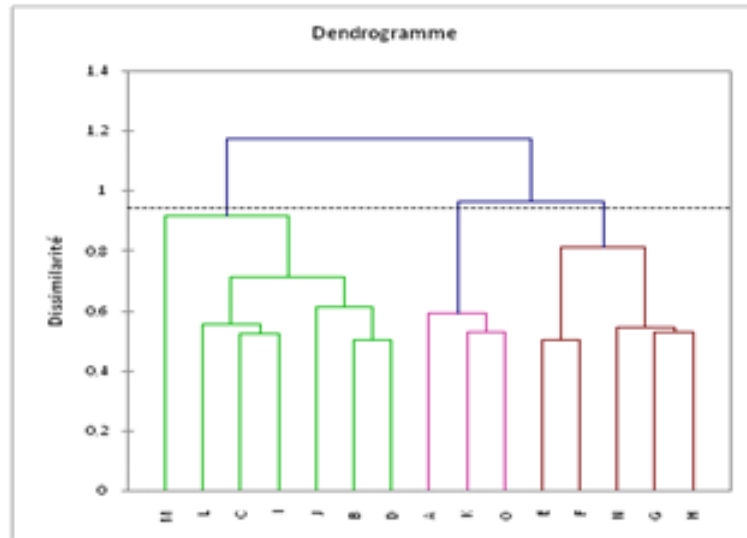


FIGURE 4.4 – Exemple de résultat d’une méthode d’apprentissage hiérarchique non supervisée [Randriamihamison \[2021\]](#).

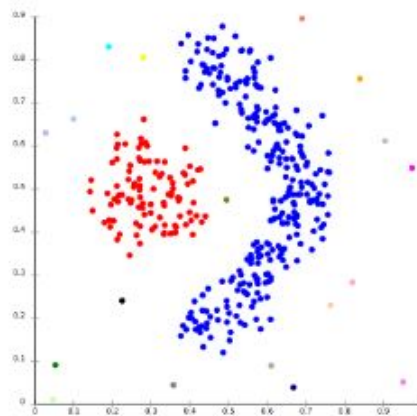


FIGURE 4.5 – Exemple de résultat d’une méthode d’apprentissage non supervisée fondée sur la densité [Campello et al. \[2020\]](#).

4.5 CONSTRUCTION AUTOMATIQUE D’ONTOLOGIES DE DOMAINES

Après le regroupement récursif des différents sujets, nous assumons à ce niveau que chaque noeud dans notre arborescence (l’output de l’étape précédente) représente un sujet. Pour cela, nous adressons après à associer pour chaque thème, une liste de mots qui se rapportent à ce domaine. Cette association est réalisée en développant pour chaque sujet une ontologie de domaine qui inclut le lexique associé à ce domaine. Tout en préservant les relations sémantiques qui les unissent ainsi que les différentes valeurs possibles pour chaque concept. L’idée de base de la méthode proposée est de profiter de certaines ontologies de domaines afin de fournir le lexique utilisé dans chaque thème ainsi que le lexique utilisé pour indiquer l’humeur associée à chaque terme dans ce domaine. Ce processus permettra ensuite de détecter le sujet et l’humeur liés à chaque tweet.

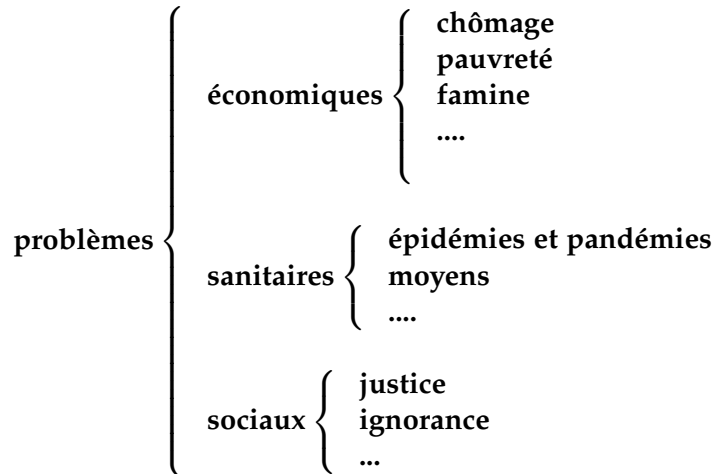


FIGURE 4.6 – Aperçu des exemples liés à la typologie des problèmes humains dans le monde.

4.5.1 Sélection, normalisation et enrichissement des concepts

Un concept représente l'entité ou l'objet définissant l'existence du domaine visé par l'ontologie. Il s'agit de l'unité de base qui sert à décrire le vocabulaire d'une ontologie⁵. Selon Agarwal [2005], un concept peut être abstrait ou concret, élémentaire ou composé, réel ou fictif. D'après une tradition qui remonte au temps d'Aristote, un concept peut être défini comme une entité constituée de trois éléments distincts représentés dans la figure 4.7 : (i) le(s) terme(s) qui exprime le concept dans le langage, (ii) la «notion» ou l'«intension» du concept qui représente la signification du concept, (iii) le(s) objet(s) dénoté(s) par le concept, également appelé(s) «réalisation» ou «extension» du concept.

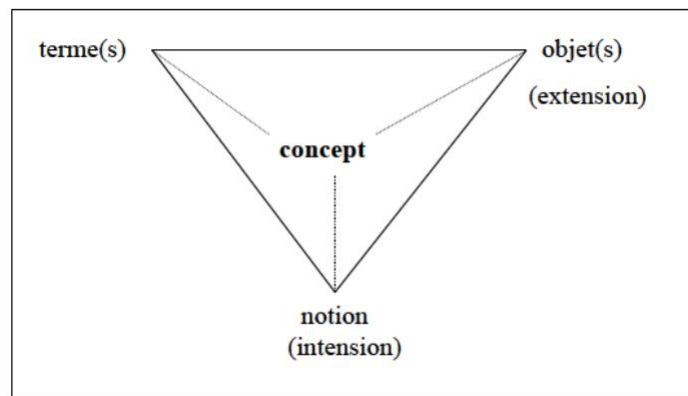


FIGURE 4.7 – les trois principaux composants d'un concept Ogden et Richards [1923].

Notre méthode de sélection des concepts repose sur une technique hybride qui combine à la fois des techniques statistiques et linguistiques. La technique linguistique s'appuie sur la création de patrons morpho-syntaxiques permettant de détecter les mots simples et composés. La technique statistique consiste à détecter des séquences de mots qui apparaissent souvent ensemble en appliquant la me-

5. https://perso.liris.cnrs.fr/alain.mille/enseignements/DEA-ECD/ontologies/notion_ontologie.htm

sure de Chiz⁶ sur les résultats de 2 et 3-grammes. Après l'extraction des concepts trouvés dans le corpus, nous procédons à regrouper ces concepts sur la base de cet exemple (Paris, Marseille par ville) en utilisant la relation d'hyperonymie existante dans la ressource WordNet (WOLF Sagot et Fišer [2008] pour la langue française). Puis en faisant l'enrichissement de notre base lexicale (concepts) par des mots de contexte à l'aide de la ressource Wikidata Vrandečić et Krötzsch [2014]. Par exemple, selon cette ressource, parmi les mots de centexte avec le mot COVID-19 se trouvent les symptômes et signes, médicament ou thérapie utilisé pour le traitement, processus de transmission de la maladie, les facteurs de risque, le nombre de cas, le nombre de morts, etc. Après avoir extrait les concepts lexicaux clés relatifs à un domaine, il est nécessaire de spécifier les liens sémantiques qui les relient afin de préserver l'aspect contextuel du domaine.

4.5.2 Extraction des relations sémantiques

Les relations sémantiques représentées dans l'ontologie sont des associations ou des liens entre des concepts. Une relation s'exprime habituellement par un terme ou par un lien graphique ou plus généralement par un signe. Il s'agit des liens qui structurent les concepts en représentant un type particulier d'interaction entre eux⁷. Selon Mhiri et Gargouri [2009], ces relations peuvent être : (i) une association : relation conceptuelle entre deux concepts, (ii) une agrégation : relation d'ensemble/élément, (iii) une composition : relation de composé/composant, (iv) une généralisation-spécialisation : une relation de classification.

Dans cette étape, nous nous concentrerons sur la recherche des relations sémantiques possibles entre les différents concepts extraits. À cet effet, nous avons eu recours aux trois ressources suivantes :

- **Ressources existantes** : nous avons sélectionné des ressources qui contiennent des relations sémantiques normalisées liées à un domaine spécifique comme Wikidata.

Par exemple dans le contexte de l'épidémie COVID-19 :

- *a des effets* : dans un premier temps, nous considérons ce mot composé comme une relation sémantique en raison de la présence d'un verbe. Ensuite, nous passons à identifier le nom *effets* en tant que concept.
- *les facteurs de risque* : dans ce cas, ce mot composé est considéré comme un concept puisqu'il ne contient pas de verbe. Ensuite, nous nommons la relation qui le relie au concept COVID-19 par le nom *avoir les facteurs de risque*. La même procédure s'applique aux concepts de même cas comme pour le mot *nombre de cas*.
- **Notre corpus** : ceci est en définissant un ensemble de patrons morpho-syntaxiques comme par exemple les patrons suivants :

Règle 1 : *Noun (Concept1) Verb (Relation) Noun (Concept2)*.

Nous citons dans l'exemple suivant deux relations qui sont extraites à l'aide de ce patron :

- *Tous les gens portent les masques chirurgicaux.*
- *Tous les [gens]_{Concept1} [portent]_{Relation} les[masques chirurgicaux]_{Concept2}.*
- *Le masque chirurgical protège certainement la vie.*

6. https://fr.wikipedia.org/wiki/Test_du_CF%87%C2%B2

7. <https://bu.umc.edu.dz/theses/informatique/CHE5271.pdf>

- Le [masque chirurgical]_{Concept1} [protège]_{Relation} certainement la [vie]_{Concept2}.
Règle 2 : [Noun Prép Noun...] (Concept1) [Verb Prép] (Relation) [Noun Prép Noun...] (Concept2) et [Noun Prép Noun...] (Concept3).
- Le vaccin de Janssen contre la Covid-19, est un vaccin élaboré par Janssen Pharmaceutica et le Beth Deaconess Medical Center.
- Le [vaccin]_{Concept1} [élaboré par]_{Relation} [Janssen Pharmaceutica]_{Concept2}.
- Le [vaccin]_{Concept1} [élaboré par]_{Relation1} [Beth Deaconess ...Center]_{Concept3}.
- **Google Knowledge Graph** : cette ressource peut jouer un double rôle ; extraire le nom de la relation entre deux concepts et mesurer le degré de relation entre la paire de concepts pour les cas d'exception. Comme en cas de chevauchement lors de l'application de nos patrons sur notre corps. Les deux exemples qui suivent illustrent en détail l'intérêt d'utiliser cette ressource pour l'extraction des relations entre les concepts.
 - La relation proposée par la présente ressource entre les deux concepts Covid-19 et test est le test diagnostique du SRAS-CoV-2.
 - Les docteurs suivent les états de ses patients ainsi ils ont assuré le soutien moral.

Dans cet exemple, nos règles ne peuvent pas différencier, si la relation *assuré* est entre *docteurs* et *soutien moal* ou entre *patients* et *soutien moral*. Pour cette raison, nous interrogeons la ressource Google Knowledge Graph afin de mesurer le degré de dépendance entre les composants de chaque proposition.

Pour les relations d'héritage, nous avons d'abord utilisé les relations de sous-classes existantes dans Wikidata et dans une seconde étape nous avons défini nos propres règles. Par exemple, si le mot *x* apparaît dans la liste 1-Grammes et apparaît également dans la liste 2-Grammes ou 3-Grammes plus d'une fois (sous la forme de *y* et *z*). Dans ce cas, nous considérerons que les mots *y* et *z* sont des sous-classes du mot *x*. Comme par exemple pour le mot *crise*. Il apparaît dans la liste de 1-Grammes ainsi que dans la liste de 2-Grammes à deux reprises sous le nom *crise économique* et *crise sanitaire*. Dans ce cas, nous considérons que ces deux mots sont deux sous-classes de la classe *crise*.

Si le mot *x* apparaît dans la liste 1-Grammes et apparaît également dans la liste 2-Grammes ou 3-Grammes seulement une fois dans le même format. Dans le cas présent, nous le considérons comme un mot composé et nous maintenons la proposition de 2-Grammes ou celle de 3-Grammes, sans tenir compte de la proposition de 1-Grammes. Cela vise à assurer l'exactitude des concepts de notre ontologie. Comme par exemple, le concept *masque*, celui-ci est apparu dans la liste de 1-Grammes ainsi que dans la liste de 2-Grammes en un format unique avec le nom : *masque chirurgical*. Selon notre hypothèse, il n'existe qu'un seul type de masque ou la différence entre les types de masques n'a pas assez d'importance. À cet effet, nous considérerons que tous les masques sont des masques chirurgicaux. À ce niveau, nous admettons un réseau contenant des concepts et des liens qui représentent l'aspect sémantique lié à un domaine. Mais il convient de noter que chaque concept de domaine possède certaines propriétés linguistiques qui le distinguent des autres. À cette fin, il est préférable d'accorder pour chaque concept une série d'attributs afin d'assurer l'explicabilité de chaque unité lexicale.

4.5.3 Attribution des propriétés aux données extraites

Au niveau de l'étape d'attribution des propriétés aux données extraites, nous visons à définir pour chacun des concepts une série de propriétés explicatives. Les attributs d'un concept en ontologie représentent les propriétés, les fonctionnalités, les caractéristiques ou les paramètres que les objets peuvent posséder et partager. La propriété de données dans l'ontologie permet d'attacher à une instance d'entité une valeur littérale de type de donnée telles que un nombre ou une date. Cette information est une mesure ou une estimation de la propriété de l'objet, qui décompose une donnée composite.

Pour cela dans notre travail, nous avons attribué pour chaque concept la liste des attributs existants dans la ressource Wikidata (comme le label, la description, etc.) ainsi que les attributs de la ressource Google Knowledge Graph (comme *result name*, *result description*, *result type*, *result detailed Description article Body*, etc.). Cette démarche vise à garantir une interprétation explicite des concepts de notre ontologie.

Par exemple, le mot SARS-CoV-2 :

- Description : virus à l'origine de la pandémie de Covid-19.
- Also known as : {2019-nCoV, NCoV-2019, SRAS-CoV-2, coronavirus de Wuhan}
- Label : SARS-CoV-2
- Official name : coronavirus 2 du syndrome respiratoire aigu sévère
- GoogleDescription : Virus
- Type : Thing
- DetailedDescription.articleBody : Le SARS-CoV-2, soit coronavirus 2 du syndrome respiratoire aigu sévère, est le virus responsable de la Covid-19. Son acronyme est parfois partiellement francisé en SRAS-CoV-2.

À ce niveau nous avons un graphe conceptuel qui explique en général les éléments-clés d'un domaine d'étude. Cependant, ce modèle a besoin d'être alimenté par des unités lexicales qui présentent les valeurs réelles possibles pour chaque élément du modèle, ce qui permet à notre modèle d'accomplir ses tâches.

4.5.4 Affectation des individus aux données extraites

L'affectation des individus consiste à assigner des individus (instances) à chaque concept de notre ontologie. Ces instances peuvent inclure des objets concrets tels que des personnes, des animaux, des automobiles, des planètes et même des molécules. Cela peut nous permettre d'atteindre l'objectif de cette contribution qui consiste à détecter : (1) le sujet du tweet, pour cela nous visons à recueillir autant que possible le lexique qui se rapporte à chaque sujet, (2) l'humeur des personnes, et dans ce cas nous avons eu l'idée d'affecter pour chaque concept extrait deux listes d'adjectifs associés à ce concept. La première liste comporte des adjectifs indiquant une mauvaise humeur, tandis que la seconde contient des adjectifs indiquant une humeur positive ou neutre. Notre choix de combiner la liste des adjectifs qui montrent la positivité et la neutralité de l'humeur est justifié par le fait que nous nous concentrerons davantage sur les personnes de mauvaise humeur. En effet, notre objectif global est d'assurer la surveillance subséquente des personnes ayant une maladie psychologique.

À la raison qu'à ce stade, nous nous intéressons à l'étude de l'humeur des malades à l'égard des sujets, cette méthode d'intégration des instances dans l'ontologie est divisée en deux sous-parties :

Sujet : l'assignation des individus liés aux différents sujets se fait à l'aide de : (1) la ressource Wikidata qui contient déjà la liste des noms pour chacun des concepts ainsi que les valeurs attribuées à ce concept, (2) la ressource Word-Net en utilisant les rapports hyperonymiques et la liste de synonymes disponibles afin d'alimenter chaque concept de notre ontologie, (3) notre corpus en intégrant des patrons qui permet d'extraire la valeur d'une instance, comme par exemple les relations : *est un, est une partie de, appelé aussi, est composé par*, etc.

Le tableau 4.3 présente quelques exemples de cas extraits pour des concepts relatifs au sujet de COVID-19, selon la ressource utilisée.

TABLE 4.3 – Exemples d'instances extraites pour chacun des concepts en fonction de la ressource utilisée

Ressource	Concept	Individus
Wikidata	symptômes	toux, fièvre, insuffisance respiratoire, maux de tête, nausée, encéphalite, psychose, etc.
	medicaments	Tocilizumab, favipiravir, ruxolitinib, etc.
	transmission	infection par gouttelettes, transmission aérienne, transmission par contact, infection oculaire, etc.
	facteur de risque	hypertension artérielle, obésité, diabète, hyperglycémie, maladie auto-immune, maladie de l'artère coronaire, etc.
	examen médical	Test antigénique rapide COVID-19, tomographie, réaction en chaîne par polymérase, etc.
WordNet	décès	mort, meurtre, cadavre, décédé, etc.
	surveillance	contrôle, observation, suivi, etc.
	distanciation	éloignement, séparation, etc.
	information	info, renseignement, etc.
	risque	danger, éviter, péril, provoquer, etc.
Corpus	corona	légende urbaine enfaite, partenaire officiel du Medef et des grandes multinationales, etc.
	alerte	info, avertissement, message, etc.
	décès	520 000 décès par coronavirus, 34 décès hospitaliers, 5 décès enregistrés, etc.

Humeur : suite à une étude empirique, nous avons constaté que les adjectifs représentent la meilleure manière de décrire l'humeur des gens. Pour cette raison, nous avons automatiquement attribué à chaque concept deux listes, une liste qui contient des adjectifs de mauvaise humeur et l'autre qui comprend des adjectifs indiquant des humeurs positives et neutres. L'attribution des adjectifs est faite à partir de l'annotation d'humeur réalisée par notre expert, tout en faisant face au problème de la négation, par le fait que nous changeons la polarité de ce terme. De façon générale, le cas d'humeur positive révélé indique qu'il est vrai que la personne présente un trouble de la personnalité et qu'elle a exprimé son opinion au sujet de COVID-19. Cependant, la décision prise concernant leur maladie n'est pas fondée sur cette raison.

Le tableau 4.4 présente quelques exemples de cas sélectionnés illustrant l’humeur exprimée envers chacun des concepts liés à la COVID-19.

TABLE 4.4 – Exemples d’adjectifs extraits associés à chaque concept indiquant le type d’humeur exprimée

Concept	Humeur	Adjectifs
Masque	mauvaise humeur	manque, cher, plus, inutile, dangereux, etc.
	bonne humeur	gratuit, sein, écologique, beau, etc.
Économique	mauvaise humeur	terrifiant, dramatique, difficile, pire, etc.
	bonne humeur	encourageant, intéressant, intelligent, etc.
Confinement	mauvaise humeur	insupportable, général, retard, nerveux, etc.
	bonne humeur	remarquable, interne, sérieux, conscrit, etc.
Vaccin	mauvaise humeur	merde, rejeté, inefficace, grave, allergique, etc.
	bonne humeur	efficace, indispensable, parfait, protégé, sûr, etc.
Gouvernement	mauvaise humeur	danger, désastreuse, mignon, capitaliste, erratique, criminelle, etc.
	bonne humeur	sérieux, auto-satisfait, hauteur, fier, etc.
Virus	mauvaise humeur	élevé, fou, incroyable, infectieux, dérangée, etc.
	bonne humeur	bonne, logique, content, union, etc.

À ce niveau, nous avons un modèle complet qui représente tous les détails du vocabulaire lié à un domaine spécifique. Ce modèle doit être mis en question afin qu’il puisse répondre à nos questions.

4.6 CONSTRUCTION AUTOMATIQUE DES RÈGLES D’INTERROGATION L’ONTOLOGIE

Cette étape a pour but de tirer profit de l’ontologie construite et des instances qu’elle contient afin de fournir des réponses adéquates. Pour ce faire, nous avons divisé cette partie en deux sous-parties :

4.6.1 Construction des règles

Notre objectif dans l’étape de construction des règles est d’établir un ensemble de règles de manière explicite et automatique. Pour cela dans un premier temps, nous avons transformé chaque tweet en un vecteur qui est représenté sous forme de séquence numérique binaire indiquant l’existence ou l’absence de chaque concept de l’ontologie dans le tweet. Ensuite, nous avons automatiquement extrait les règles en profitant de l’algorithme Johnson [Bazan et al. \[1998\]](#) existant dans le système Rosetta⁸ qui prend comme entrée la liste des vecteurs avec leurs labels (selon l’annotation réalisée) et fournit comme résultat une liste de règles logiques et généralisées. Ces règles permettent par la suite de comprendre les liens de dépendance entre les différents concepts et les sujets afin de prendre des décisions concernant l’identification des sujets. Le choix d’utiliser l’algorithme Johnson est justifié par la performance qu’il a démontré dans les problèmes d’optimisation quant à trouver le chemin le plus court dans un graphe. De même, cet algorithme a démontré des performances intéressantes au niveau de la répartition des tâches

8. <http://bioinf.icm.uu.se/rosetta/>

dans les problèmes d’ordonnancement multicritères en assurant la cohérence du processus.

Dans la littérature, il existe plusieurs techniques basées sur la notion d’apprentissage automatique permettant l’extraction automatique de règles à partir des données brutes. Nous pouvons mentionner parmi ces techniques : (i) *Association Rules Mining* (ARM, en français : extraction de règles d’association), (ii) *Set Formation of Decision Rule* (RST, en français : formation d’un ensemble de règles de décision), (iii) les règles obtenues à l’aide de l’algorithme arbre de décision.

La technique d’ARM permet de découvrir des relations cachées dans des bases de données massives. Le but principal de cette technique est d’extraire des associations stimulantes ou causales à partir des données, ceci est en se basant sur la fréquence d’apparition des divers éléments clés. Parmi les métriques les plus appliquées par cette technique pour valider les liens de dépendance, nous citons : (i) Confiance : Il s’agit d’une métrique qui s’occupe à mesurer la fiabilité d’une règle. (ii) Support : elle représente la fréquence des groupes d’éléments figurant ensemble dans les données. (iii) Lift : elle évalue l’importance de la valeur d’une règle. (iv) Conviction : il s’agit de mesurer l’indépendance entre les éléments d’une règle, ce qui signifie quand la règle fournit une prévision erronée? La technique RST consiste à élaborer un ensemble de règles à partir des données historiques. Une règle est composée d’un ensemble de faits liés par des opérateurs logiques. Cette technique représente également des connaissances opérationnelles, permettant de construire un système expert.

La différence entre RST et ARM réside dans le fait que RST reflète un ensemble d’objets qui ne peuvent pas être décrits exactement sur la base des valeurs de leurs ensembles d’attributs. Cependant, ils peuvent être caractérisés par des approximations pour résoudre l’ambiguïté inhérente aux données. Tandis que ARM se base sur la co-occurrence des ensembles d’attributs, qui peuvent être utilisés pour inférer l’existence d’autres attributs. Le tableau 4.5 présente les différences terminologiques entre les deux concepts.

TABLE 4.5 – Comparaison des terminologies relatives à RST et ARM *Gardiner et Gillet [2015]*

Terme	RST	ARM
Entité de base	Système d’information	Tableau d’information
La ligne du tableau correspond à	Objet	Transaction
Colonne	Attribut	Élément
Règle	Règle de décision	Règle d’association
Validation	Précision	Degré de confiance

Nous avons effectué une étude comparative des différentes méthodes utilisées pour extraire les liens de dépendance entre les différents traits de personnalité [Ellouze et al. \[2022\]](#). Nous avons constaté après cette étude qu’en dépit de l’importance de ces techniques sur le plan du traitement et des résultats, il existe des spécificités qui nous incitent à utiliser une technique plutôt qu’une autre. Par exemple, parmi les points qui nous ont incités à utiliser la technique RST basée sur l’algorithme de Johnson par rapport aux règles construites par l’algorithme d’arbre de décision est que : (i) les règles établies par l’algorithme d’arbre de décision sont interreliées. C’est-à-dire que si la règle parentale est erronée, les enfants

auront le même statut. En revanche, les règles établies par l’algorithme Johnson sont indépendantes, (ii) d’après notre étude, le nombre de règles générées par l’algorithme de Johnson est inférieur à celui produit par l’arbre de décision. Il est préférable que ce nombre ne soit pas élevé, cela veut dire que nous pouvons prendre une décision avec un minimum de règles. En plus, notre choix d’appliquer les règles RST que ARM est justifié par le fait qu’ARM cherche à déduire la dépendance entre les attributs en fonction du calcul de la cooccurrence qui est un peu rigide. Tandis que la technique RST offre plus de flexibilité dans la construction des règles, étant donné qu’elle est fondée sur l’approximation des valeurs des attributs.

Le même principe de transformation des données textuelles en un ensemble de vecteurs est appliqué afin de détecter l’humeur des personnes. Ceci est réalisé en y apportant certaines modifications logiques qui sont essentiellement liées à la valeur que chaque concept peut prendre dans chaque règle. Cette valeur est principalement liée à l’adjectif qui y est apparu en relation avec ce concept dans le but de montrer l’humeur exprimée. Pour ce faire, nous procéderons comme suit :

Si le concept apparaît avec un adjectif qui figure dans la liste des humeurs inappropriées, dans ce cas, ce concept sera défini sous la forme d’une valeur négative. Et si le concept apparaît avec un adjectif positif ou neutre, nous lui attribuons une valeur positive et dans le cas où ce concept ne figure pas dans le tweet, nous assignerons la valeur neutre. Et si le concept est apparu mais qu’il n’y a pas d’adjectif associé à celui-ci, nous le considérerons comme une opinion neutre ou positive afin d’éviter le cas dans lequel que chaque tweet ou chaque sujet soit la cause du problème. Comme il est montré dans l’exemple suivant : *N’oubliez pas vos masques pour venir nous voir !*

Les règles établies pendant cette étape ramènent notre ontologie à raisonner dans le but d’assurer les tâches de détection du sujet et d’humeur exprimée par l’utilisateur dans le tweet.

4.6.2 Interrogation de l’ontologie

Au cours de l’étape d’interrogation de l’ontologie, nous visons à transformer les règles généralisées construites en utilisant l’algorithme de Johnson, en règles de raisonnement (SWRL) [Golbreich et Imai \[2004\]](#). SWRL est un langage de règles pour le web sémantique qui combine les ontologies OWL avec des règles de logique, en étendant l’ensemble des axiomes OWL [Bassiliades \[2018\]](#). SWRL représente un choix très populaire pour le développement d’applications basées sur des règles à partir d’une ontologie. Ce langage est supporté par des éditeurs d’ontologie ainsi que par des moteurs de règles et des raisonneurs d’ontologie populaires. Il est à noter que ce langage permet la manipulation des instances par des variables. Mais il ne permet pas de créer des concepts et des relations. C’est à dire il permet seulement d’ajouter des relations selon les valeurs des variables (individus) et la satisfaction de règles. Généralement, les règles SWRL sont construites selon le schéma suivant : Antécédent – > Conséquent. L’antécédent est représenté par des atomes de conjonctions et le conséquent est représenté par un seul atome. Dans ce contexte, un atome peut être soit une instance de concept ou soit une relation OWL. Ces règles permettent d’interroger l’ontologie construite dans le but de prendre des décisions utiles au moment de la détection du sujet du tweet et de

l'humeur de la personne en se basant sur la concordance entre les mots du tweet et les mots de l'ontologie. Comme mentionné précédemment, l'objectif de cette étape est de détecter le sujet du tweet ainsi que l'humeur exprimée en relation avec les concepts de ce sujet.

L'intérêt des règles obtenues par notre méthode par rapport à celles obtenues par d'autres moyens réside dans le fait que : (i) notre méthode tient compte de l'aspect lexico-sémantique propre à un domaine. L'aspect lexical est abordé en tenant compte du cas où un mot peut appartenir à plusieurs sujets en même temps. L'aspect sémantique est pris en compte en prenant en considération la sémantique du contexte ainsi qu'en tenant compte des rapports de synonymie, (ii) notre méthode traite également les défis de détection d'humeur abordés par plusieurs travaux dans la littérature. Étant donné que le même mot peut indiquer une humeur différente dans différents contextes, pour cette raison nous associons l'humeur à un concept spécifique d'un sujet particulier. Tous ces facteurs assureront l'interprétabilité et l'explicabilité des résultats obtenus ultérieurement.

Les exemples qui suivent illustrent l'impact du processus en cours sur les tweets dans le but de détecter à la fois le sujet et l'humeur. Ces exemples démontrent comment notre méthode tient compte des défis auxquels nous avons été confrontés dès le départ : (i) la détection du sujet en utilisant les mots de contexte : dans le deuxième exemple, notre méthode a pu détecter le sujet, qui est le COVID-19, bien que le mot COVID-19 ou l'un de ses synonymes n'existe pas. En effet, la détection se fait au moyen de la combinaison de mots contextuels comme *vaccin* et *confinement*, (ii) la variété de l'humeur exprimée envers l'épidémie : l'humeur exprimée par rapport à la COVID-19 n'est pas nécessairement toujours négative. Mais elle peut être positive comme dans le deuxième exemple ou neutre comme dans le troisième exemple.

— *Est il possible de faire une plainte collective contre le président et son gouvernement pour mise en danger de la vie d'autrui ? Question sérieuse. confinement coronavirus.*

Traitement attendu : *Est il possible de faire une plainte collective contre le président et son [gouvernement]_{Concept 1} pour mise en [danger]_{Adjectif 1} de la [vie]_{Concept 2} d'autrui ? Question srieuse. [confinement]_{Concept3} [coronavirus]_{Concept4}*

— *Oui ce vaccin est super important, oui c'est très probablement la solution qui évitera de se refaire confiner tout l'automne.*

Traitement attendu : *Oui ce [vaccin]_{Concept 1} est [super]_{Adjectif 1} [important]_{Adjectif 2}, oui c'est tres probablement la solution qui evitera de se refaire [confiner]_{Concept 2} tout l'automne.*

— *wasexo Covid19 : le prix du pétrole brut s'écroule, celui du « kpayo » s'envole au Bénin*

Traitement attendu : *wasexo [Covid19]_{Concept 1} le prix du pttrole brut s'ecroule, celui du «kpayo» s'envole au Bénin*

— *urgent | 146 nouveaux cas de coronavirus et un nouveau décès lié à la maladie sont enregistrés en Haïti*

Traitement attendu : *[urgent]_{Adjectif 1} | 146 [nouveaux]_{Adjectif 2} [cas]_{Concept 1} de [coronavirus]_{Concept 2} et un [nouveau]_{Adjectif 3} [deces]_{Concept 3} lie a la [maladie]_{Concept 4} sont [enregistrs]_{Adjectif 4} en Haiti.*

- *Vous pensez qu'il va y avoir un vaccin pour le covid ou on devra vivre avec pour tjrs et s'y habituer.*

Traitement attendu : *Vous pensez qu'il va y avoir un [vaccin]_{Concept 1} pour le [covid]_{Concept 2} ou on devra vivre avec pour tjrs et s'y habituer.*

4.7 CONCLUSION

Dans ce chapitre, nous avons développé notre deuxième contribution, qui vise à enrichir les résultats décelés au cours de notre première contribution en identifiant les sujets qui contribuent à détecter le désordre. Pour ce faire, nous avons proposé dans un premier temps une méthode pour catégoriser les différents sujets de notre corpus dans une arborescence montrant les différents niveaux de nos sujets. Cela résulte en des outputs explicatifs allant du niveau le plus abstrait au niveau le moins abstrait.

Ensuite, dans un deuxième temps, nous avons attribué à chaque feuille de notre arborescence une ontologie qui comprend les différents termes associés à chaque feuille qui représente un sujet. Nous avons également assigné pour chaque concept la liste des adjectifs aidant à présager l'humeur de la personne exprimée envers ce sujet. Finalement, nous avons construit l'ensemble de règles généralisées, permettant d'interroger notre ontologie dans le but de détecter le sujet et l'humeur de l'utilisateur exprimée dans le tweet.

Le chapitre suivant, fera l'objet de notre troisième contribution de notre thèse qui consiste à détailler les principales étapes permettant de garantir la surveillance des utilisateurs malades en manipulant leurs fils d'actualité.

SURVEILLANCE DES PERSONNES ATTEINTES DU TROUBLE DE LA PERSONNALITÉ

5

SOMMAIRE

5.1	INTRODUCTION	96
5.2	CORPUS UTILISÉ	97
5.3	MÉTHODE PROPOSÉE POUR SURVEILLER L'ÉTAT DES PERSONNES AYANT UN TP	99
5.3.1	Méthode basée sur des techniques d'apprentissage supervisé vi- sant à détecter les personnes atteintes de troubles de la personnalité	100
5.3.2	Méthode fondée sur des techniques d'apprentissage supervisé . . .	101
5.3.3	Méthode basée sur les ontologies	103
5.3.4	Méthode basée sur l'apprentissage profond	107
5.4	MÉTHODE POUR SURVEILLER LES UTILISATEURS	112
5.5	CONCLUSION	114

5.1 INTRODUCTION

Dans les deux chapitres qui précèdent, nous avons proposé une méthode pour la détection des différents types de troubles de la personnalité à partir de tweets en français. De même, une méthode visant à identifier automatiquement les causes figurant dans les tweets qui contribuent à l'apparition des troubles. Tout ce travail a été mis en oeuvre dans le but d'assurer le processus commis par le psychiatre lors du diagnostic d'un patient (détecter la maladie, analyser les causes et proposer des solutions sous forme de consignes visant à réduire l'intensité de l'état du patient).

À cet effet, dans ce chapitre, nous voulons nous inspirer de ce processus afin de suivre l'état des gens que nous croyons malades sur Twitter. Cela est accompli en supprimant le malade de la liste des observateurs des tweets non appropriés (ceux qui peuvent apparaître dans leur fil d'actualité et qui peuvent contribuer à faire évoluer leur pression artérielle). En outre, notre processus de suivi suggéré dans cette section permet de recommander des tweets aux utilisateurs malades selon l'état de chacun. Cette recommandation s'effectue en accordant plus de priorité aux différents tweets appropriés afin qu'ils apparaissent dans le fil d'actualité de la personne malade en première position. Si notre méthode se rend compte qu'un tweet pourrait convenir à une personne malade mais qu'il n'est partagé par aucun de ceux qu'elle suit. Dans ce cas, un chatbot peut intervenir pour le transmettre à cet utilisateur sous forme de message privé. Pour ce faire, nous avons commencé notre travail en appliquant notre méthode présentée au premier chapitre qui consiste à identifier les utilisateurs malades à partir de leurs vingt derniers tweets. Ensuite, nous avons détecté le comportement de personnes considérées comme malades depuis l'étape précédente envers un sujet particulier qui représente dans notre cas la pandémie de COVID-19 en tant que choix de domaine d'étude.

Enfin, grâce à une méthode hybride, nous avons passé à assurer le filtrage et la recommandation des tweets pour les usagers malades. Cela passe par la combinaison d'une méthode basée sur l'utilisation d'une ontologie qui préserve l'aspect sémantique ainsi qu'une méthode basée sur la notion d'apprentissage profond qui permet de traiter des cas non couverts par l'ontologie. Le choix d'appliquer notre processus de travail en lien avec COVID-19 est justifié du fait que cette épidémie a progressé rapidement partout dans le monde. De même, elle a eu une incidence négative sur tous les secteurs de ces pays. Par conséquent, elle suscite un degré considérable de crainte, d'inquiétude et de préoccupation de la part du grand public et particulièrement au sein de certains groupes spécifiques tels que les personnes âgées et les personnes souffrant de maladies préexistantes. Dans ce contexte, de nombreuses organisations sanitaires dans le monde entier ont lancé des études sur l'impact psychologique de la crise sanitaire. Cela comprend *France Assos Santé*^{1, 2}, qui a demandé à 10 000 bénévoles de l'aider à mesurer l'impact de la crise de COVID-19 sur la santé mentale de la population. Elle leur a également demandé d'assurer le suivi des états des personnes atteintes d'une maladie chronique pendant cette période d'épidémie.

Dans la suite de ce chapitre, nous visons à détailler les différentes méthodes

1. <https://www.france-assos-sante.org/>

2. <https://www.lavoixdunord.fr/900603/article/2020-11-30/covid-19-un-appel-10-000-volontaires-pour-etudier-l-impact-psychologique-de-la>

pour la tâche de détection des personnes atteintes de troubles de la personnalité ainsi que leurs comportements face à l'épidémie COVID-19. Nous visons également à détailler la partie de l'analyse sémantique fondée sur l'utilisation de l'ontologie et de l'apprentissage profond pour filtrer les tweets inappropriés sur la base de l'état du patient. De plus, nous aborderons la recommandation des tweets qui peuvent être adéquats en fonction de l'état de la personne malade. Cela se fait en présentant des exemples et en détaillant la manière d'interagir face aux différents enjeux rencontrés.

5.2 CORPUS UTILISÉ

Cette étude est dédiée à la proposition d'un agent intelligent qui peut fournir essentiellement deux services aux personnes atteintes de troubles. Le premier service consiste à éliminer les tweets inappropriés qui pourraient empirer la situation d'une personne malade. Le deuxième service est consacré à la génération automatique de réponses aux diverses préoccupations de ces personnes en vue de les aider à surmonter leurs problèmes. Nous avons choisi d'appliquer ces services principalement durant la pandémie de COVID-19 puisque les effets de cette épidémie affectent négativement la stabilité de toutes les catégories des gens même les personnes saines.

Étant donné que l'objectif de notre méthode est de surveiller l'état des personnes ayant un trouble de personnalité envers la pandémie de COVID-19, nous avons entamé cette partie en détectant le comportement de ces personnes à l'égard de cette pandémie. En ce qui concerne la détection des maladies, nous nous sommes appuyés sur le corpus utilisé dans notre première contribution qui est déjà annoté en fonction des différents types et catégories de troubles de la personnalité. Pour la tâche dédiée à la détection de comportement du malade à l'égard du COVID, nous avons réannoté une portion de notre corpus utilisé dans notre deuxième contribution. En effet au cours du traitement de nos données, nous avons distingué plusieurs types de comportement des personnes face à cette épidémie : méfiance exagérée, irresponsabilité totale ou comportement neutre comme l'illustre le tableau 5.1. Ces réactions sont courantes chez les gens qui ont un trouble de la personnalité [Triebwasser et al. \[2013\]](#) puisque parmi les symptômes de cette maladie, nous pouvons mentionner : (i) le doute incessant [Ron et al. \[2016\]](#) qui peut exacerber leur peur et leur attention face au COVID, en particulier compte tenu du nombre élevé de décès, du manque de places à l'hôpital, du manque de médicaments, etc, (ii) la méfiance exagérée [Kirk et al. \[2013\]](#), ce symptôme avec COVID-19 peut amener le malade à considérer que les mesures de sécurité requises par les autorités pour se prémunir contre la propagation du COVID sont exagérées et ne visent qu'à restreindre sa liberté. Au final, notre corpus a été réparti de la façon suivante : Méfiance exagérée (106 tweets), Neutre (74 tweets), Irresponsabilité totale (79 tweets).

Pour la tâche de l'analyse sémantique dédiée à effectuer le filtrage et la recommandation des tweets en se basant sur l'interrogation d'une ontologie. Nous avons ajouté 300 tweets à notre corpus utilisé dans la deuxième contribution qui consiste à détecter les raisons affectant le TP sur les réseaux sociaux composé principalement de 700 tweets afin d'enrichir notre ontologie par des instances

TABLE 5.1 – Exemple de tweets indiquant un éventuel comportement envers COVID-19.

Méfiance exagérée	Neutre	Irresponsabilité totale
Étude Rétrospective de 6493 patients ambulatoires et hospitalisés avec COVID-19.	Dans les hôpitaux, la situation sanitaire du Covid-19 est sous contrôle	Les morts ont pas le covid.
Covid-19 : le rapport choc des pompiers sur la gestion de la pandémie	Après la crise du Covid-19, la tentation du retour au tourisme de masse pour l'industrie du voyage	Sortir affronter le Covid-19
Le nombre de nouvelles infections à la COVID-19 proche des niveaux record	En ce moment avec le Corona logique	Non, mais la c'est une blague ou quoi, ah ok poisson d'avril.
Pour ceux qui doutent encore de l'impact de la COVID-19, voici en vert la courbe de tous les décès enregistrés au Québec	Face au COVID-19, la solidarité entre Européens a sauvé des vies.	On prendra pas ton vaccin de merde.

supplémentaires. Pour l'étape, qui consiste à assurer la tâche de filtrage et de recommandation de tweets basée sur une méthode d'apprentissage profond, nous avons annoté : (i) 775 paires de tweets, pour chaque paire nous indiquons si le contenu du deuxième tweet doit être filtré, (ii) 500 paires de tweets, pour chaque paire, nous indiquons si nous pouvons considérer le deuxième tweet comme un tweet recommandé, en se basant sur le comportement et le sujet exprimés dans le premier tweet. Nous considérons une paire de tweets avec la classe *devrait être filtré*, si le deuxième tweet confirme la préoccupation du premier, ce qui peut provoquer le malade. Nous estimons qu'une paire de tweets appartenant à la classe *devrait être recommandé*, si le deuxième tweet peut jouer le rôle de stabilisateur par rapport à l'inquiétude exprimée au niveau du premier tweet.

L'annotation est effectuée en associant à chaque paire de tweets la classe (à filtrer, à recommander, neutre). Ces comportements ont été choisis suite à une étape d'étude empirique représentant 10% de la taille de notre corpus. Cette étude a également permis d'établir un manuel d'annotation à partir duquel les critères d'annotation sont définis. Il est à noter que durant l'annotation réalisée nous avons pris en considération une certaine relation analogique entre des différents concepts comme *confinement* et *l'ouverture de l'école* dans l'exemple (1) et (2). Malgré qu'il n'y a pas une relation de synonymie entre eux, l'objectif de ce choix est de s'assurer que notre méthode tient compte de l'aspect sémantique lié au domaine d'étude.

(1) À propos du déconfinement : "Les Français ne sont pas des enfants, nous pouvons leur faire confiance".

(2) Mais foutez la paix à ceux qui ont besoin que leurs enfants reprennent l'école!

Le tableau 5.2 présente la répartition de notre corpus à chaque étape de notre contribution.

TABLE 5.2 – La répartition de notre corpus par étape.

Description	Taille du corpus
Personne ayant un comportement de méfiance exagérée envers COVID-19	106 tweets
Personne ayant un comportement d’irresponsabilité totale envers COVID-19	79 tweets
Personne ayant un comportement neutre envers COVID-19	74 tweets
Enrichissement de l’ontologie	1000 tweets
Tweets ayant le même sujet	1000 paires de tweets
Les tweets devraient être filtrés	775 paires de tweets
Les tweets devraient être recommandés	500 paires de tweets

5.3 MÉTHODE PROPOSÉE POUR SURVEILLER L’ÉTAT DES PERSONNES AYANT UN TP

Notre méthode proposée, décrite à la figure 5.1, permet à Twitter d’analyser en temps réel la production textuelle de ses utilisateurs afin d’assurer le suivi des utilisateurs souffrant de troubles de la personnalité sur sa plateforme. La surveillance a été accomplie d’abord en supprimant les messages susceptibles de gêner les utilisateurs malades, puis en interagissant avec eux dans le but de stabiliser leur situation. Cette intervention représente un médiateur pour chaque nouveauté liée au COVID-19. Pour ce faire, nous commençons notre travail en détectant, pour chaque tweet en relation avec cette épidémie, s’il a été publié par une personne ayant un trouble de la personnalité. Cette détection est assurée au moyen d’une architecture basée sur la notion d’apprentissage profond. Si c’est le cas, nous procédons à l’identification du comportement de la personne malade en relation avec l’épidémie au moyen d’une méthode fondée sur l’apprentissage automatique. En fonction des résultats de la détection du comportement, nous effectuons une analyse sémantique fondée dans un premier temps sur une ontologie pour assurer le suivi de cette personne.

Le processus de surveillance proposé implique la suppression des tweets inappropriés susceptibles de transporter le gène à la personne malade à partir de son fil d’actualité. Nous procédons à cette suppression en comparant la similitude sémantique entre chaque tweet publié par la personne malade et tous les autres tweets accessibles par cette personne via un analyseur sémantique. Ensuite, nous poursuivons notre processus de suivi en communiquant avec l’utilisateur malade à l’aide d’un chatbot afin de stabiliser son état psychologique. Pour éviter d’hériter des problèmes liés à la synthèse d’une réponse, nous comptons de répondre à cette personne via un message sous la forme d’un autre tweet basé sur une analyse sémantique réalisée par notre ontologie. Il convient de noter que notre analyseur sémantique proposé pour le filtrage et la recommandation de tweets diffère du concept classique d’analyse sémantique qui est dédié à l’analyse du sens général de la phrase. Cette différence est principalement liée à notre domaine d’étude qui regroupe à la fois le sujet du COVID-19 et celui des troubles psychologiques. Dans notre analyse, nous tenons principalement compte des concepts qui reflètent le sujet et le comportement au niveau d’un tweet, ce qui permet de donner plus de flexibilité au moment de notre réaction avec le patient. En ce qui concerne les

cas non couverts par notre ontologie pour réaliser les deux tâches de filtrage et de recommandation des tweets, nous proposons une méthode basée sur l'apprentissage profond afin d'entraîner l'algorithme à détecter la similarité sémantique entre chaque combinaison de deux tweets. Notre objectif en utilisant cette technique est de permettre à l'algorithme de découvrir des règles détaillées sur les concepts et les comportements liés au COVID-19. En général, notre méthode offre plusieurs opportunités, notamment : (i) le traitement de l'évolution du lexique au fil du temps. Particulièrement avec COVID-19, plusieurs mots apparaissent et deviennent d'actualité pendant une courte période de temps, (ii) l'extraction automatique de caractéristiques de haut niveau à partir de données textuelles brutes, tout en préservant les liens sémantiques entre les différentes unités lexicales, (iii) la prise en compte du sens du mot dans la phrase à l'aide de la technique de plongement de phrases. Nous détaillons dans la suite du chapitre les différentes principales parties permettant la réalisation de notre méthode.

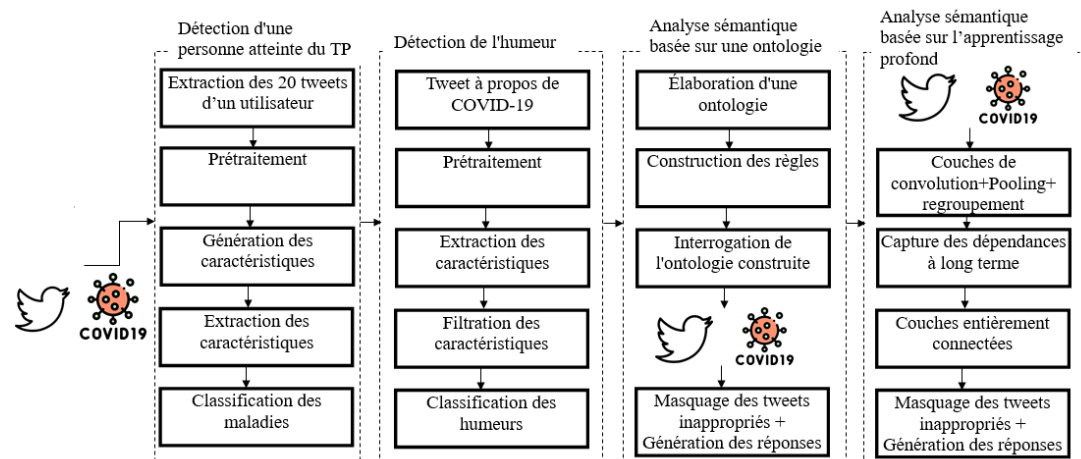


FIGURE 5.1 – Méthode hybride pour assurer la surveillance des personnes souffrant de TP en filtrant et en recommandant des tweets en fonction de leurs états envers COVID-19.

5.3.1 Méthode basée sur des techniques d'apprentissage supervisé visant à détecter les personnes atteintes de troubles de la personnalité

Notre méthode basée sur des techniques d'apprentissage supervisé consiste à détecter les personnes ayant un TP à partir de l'analyse de leurs vingt derniers tweets en se référant à nos travaux antérieurs au niveau de notre première contribution. Il est à noter que durant notre première contribution nous avons réalisé diverses expérimentations : (i) détection des personnes atteintes d'un type de TP à l'aide d'une méthode fondée sur une technique d'apprentissage automatique classique. Cette méthode tire profit des différents types d'informations relatives aux aspects linguistiques et temporels des statuts. En outre, cette méthode prend en considération des informations relatives aux données de Twitter, (ii) détection d'un type de TP à travers la détection du degré de chaque symptôme associé à ce type afin d'assurer l'explicabilité et l'interprétabilité des résultats. Cette tâche est assurée, tout en profitant des techniques d'apprentissage profond permettant l'extraction automatique d'un niveau élevé de caractéristiques. Ces caractéristiques se

rappellent principalement aux unités lexicales du texte de tweet, tout en gardant l'aspect sémantique entre les différentes unités, (iii) détection de la présence de TP en général ainsi que de la catégorie et le type particulier de TP, tout en tenant compte de la notion d'étiquetage multiple. La notion d'étiquetage multiple permet de traiter le cas où une personne peut avoir plus d'un type et d'une catégorie de TP simultanément. Comme dans la deuxième proposition (ii), nous avons choisi de travailler sur des caractéristiques extraites d'une manière automatique à partir des données textuelles brutes. L'extraction des caractéristiques est réalisée avec des techniques avancées fondées sur le concept d'apprentissage convolutionnel, tout en prenant en considérant les dépendances à longue distance entre les différentes unités lexicales à travers une couche BiLSTM. Notre méthode tient également compte du sens du mot dans toute la phrase grâce à une technique de plongement des phrases ainsi que l'historique des publications de chaque utilisateur afin d'identifier l'état du personne à différents moments. De même, notre méthode aborde les problèmes liés à la taille du corpus en utilisant des techniques de génération automatique de données. L'objectif de prendre en compte les mesures présentées est d'accroître la diligence et la précision dans le processus de prise de décision.

Dans cette partie, nous avons choisi d'employer notre troisième proposition pour la tâche de détection de TP afin de maintenir l'aspect de traitement en temps réel par rapport à la première proposition. En outre, ce choix est justifié par le fait qu'il est difficile d'annoter l'ensemble des symptômes de chaque type de TP. De plus, il convient de noter que dans notre troisième proposition, nous avons réalisé l'apprentissage pour la détection des TP en fonction de toutes les catégories et de tous les types de TP. Il est particulièrement important que dans cette contribution, nous visons à fournir une surveillance pour les personnes avec un TP, quel que soit le type de ce trouble. D'une manière générale, chaque patient ou un groupe de patients ont certaines préoccupations liées à des problèmes particuliers. Garantir une supervision adéquate pour tous les cas est une tâche ardue car il est difficile d'adopter une méthode qui prenne en compte toutes les particularités de chaque domaine. Pour cette raison, nous avons focalisé dans notre travail sur un domaine d'application qui tient compte des enjeux qui touchent un grand groupe de personnes dans le monde comme celui du COVID-19.

5.3.2 Méthode fondée sur des techniques d'apprentissage supervisé

COVID-19 est une pandémie d'une maladie infectieuse émergente, appelée maladie à coronavirus 2019, causée par le coronavirus du SRAS-CoV-2³. Le 30 janvier 2020, l'OMS a déclaré une urgence sanitaire mondiale en réponse à cette pandémie, qui a affecté l'ensemble des pays du monde⁴. L'apparition et la progression de cette maladie rapidement rend son traitement difficile. Parmi les problèmes rencontrés dans le traitement de cette maladie, nous pouvons mentionner le comportement des gens à son sujet, qui peut être une méfiance exagérée ou une irresponsabilité totale. Comme le montre la section 5.2 *Corpus utilisé pour surveiller l'état des personnes ayant un trouble de personnalité*, ces deux réactions sont fréquemment présentes chez les individus paranoïaques. Il convient également de

3. https://fr.wikipedia.org/wiki/Pand%C3%A9mie_de_COVID-19

4. https://fr.wikipedia.org/wiki/Pand%C3%A9mie_de_COVID-19

noter que ces réactions peuvent être exacerbées par les effets et les conséquences du COVID-19. En particulier avec le grand nombre de décès dans le monde, le manque de places dans les hôpitaux, ainsi que les mesures de sécurité strictes requises par les autorités. Toutes ces informations peuvent mener à des effets négatifs sur les personnes paranoïaques. Selon [Scocco et al. \[2019\]](#), le doute incessant chez les personnes ayant des problèmes psychologiques peut produire le suicide, la dépression et le deuil compliqué. En outre, la sous-estimation peut accroître le degré d'agressivité chez le patient [Al-Atwi et al. \[2020\]](#). Cette agressivité peut influencer la personne elle-même en la rendant plus susceptible de contracter le virus Corona. De plus, cela peut influencer les autres, car le patient va les décourager d'adhérer aux mesures de sécurité.

Pour cela dans cette phase, nous visons à déterminer le comportement de l'utilisateur ayant un trouble face à l'épidémie de COVID-19. Notre choix de travailler sur la combinaison entre les TPs et l'épidémie de COVID-19 est basé sur notre hypothèse que les divers effets de l'épidémie pourraient affecter les états des personnes ayant des troubles de la personnalité. Pour ce faire, nous avons eu recours à une méthode basée sur une technique d'apprentissage supervisé. Il s'agit de classer chaque tweet portant sur le sujet de COVID-19 en trois classes illustrant le comportement du patient à l'égard de la pandémie (méfiance exagérée/irresponsabilité totale/neutre). Cette distinction est justifiée par le fait que nous voulons prendre en compte les différents symptômes liés à la maladie paranoïaque (doute incessant, interprétations négatives des gestes, la sous-estimation des autres). Cela peut nous ramener à spécifier exactement le type de comportement qui peut être obsessionnel compulsif ou irresponsable. Cette distinction nous permettra d'avoir plus de précision et de flexibilité lors de l'interaction avec la personne malade.

Après avoir mené une étude empirique, nous avons décidé de se limiter dans la présente section à l'utilisation d'une méthode d'apprentissage automatique fondée sur des caractéristiques de bas niveau. Comme indiqué dans la première contribution, les attributs de tweets de type bas niveau sur lesquels nous pouvons fonder notre méthode comprennent essentiellement des types linguistiques, temporels et métadonnées. Par rapport à la première contribution, à cette étape, nous avons décidé de ne pas tenir compte du type de métadonnées (renseignements liés à Twitter). Ce choix a été pris dans le but de maintenir l'aspect traitement des données en temps réel puisque ces attributs au moment du partage sont tous initialisés à zéro. Les différents détails relatifs aux différents types de caractéristiques employés dans cette section sont présentés et justifiés dans les tableaux 3.4, 3.5 et 3.6 figurant dans le troisième chapitre. Nous rappelons qu'en général, la méthode que nous avons proposé offre plusieurs opportunités à la fois, notamment : (i) la mise en évidence des différentes notions existantes dans la littérature, telles que la logique floue ou la classification par vote. Cela permet d'améliorer la performance de KNN et d'intégrer d'autres concepts qui nous semblent utiles dans le domaine médical. Par exemple, nous estimons qu'il est intéressant d'assurer le concept d'étiquettes multiples basées sur la relativité. Cela garantit une meilleure gestion du chevauchement des maladies grâce au degré de confiance associé à la réponse du modèle, (ii) la prise en compte d'une analyse approfondie des données des médias sociaux, notamment le style d'écriture qui figure dans la production textuelle de l'utilisateur. Cela est assurée à l'aide des modèles de langage avancés

admettant des architectures profondes permettant de préserver les détails linguistiques liés à l'aspect sémantique des mots dans la phrase.

Ce travail offre à Twitter plusieurs services lui permettant d'effectuer des traitements intelligents en collectant, analysant et classifiant les données partagées sur sa plate-forme [Casalino et al. \[2017\]](#). Il fournit également une surveillance en temps réel du comportement des personnes paranoïdes. Cela peut aider à réduire le nombre de suicides et à agir rapidement avant d'avoir une incidence négative sur d'autres personnes. De même, ce travail sert à repérer les emplacements géographiques contenant des résidents ayant un comportement inapproprié face à cette épidémie afin de mieux gérer les zones dangereuses. En outre dans notre travail, nous avons envisagé d'associer à Twitter d'autres services supplémentaires qui lui permettent d'agir en toute situation avec le malade en gérant le contenu de son fil d'actualité à l'aide d'un analyseur sémantique.

5.3.3 Méthode basée sur les ontologies

Au cours des dernières années, les analystes des données des médias sociaux ont inventé de nouveaux concepts permettant d'améliorer la performance de leurs travaux. Cette amélioration consiste à adapter l'analyse aux changements au cours du temps ainsi qu'à suivre et même à alerter le responsable ou le décideur. Par exemple dans le domaine de l'analyse de données relatives au commerce électronique sur les réseaux sociaux. Les chercheurs ont commencé à travailler en premier temps sur l'analyse sémantique des avis des consommateurs. Après un certain temps, les décideurs dans le domaine du e-commerce ont montré un besoin spécifique pour certaines tâches qui regroupent : (i) l'analyse des avantages et des inconvénients des produits, (ii) l'analyse des besoins et des attentes des consommateurs, (iii) la détermination des caractéristiques du produit à conserver ou à rénover. Ce qui permet d'établir une stratégie de communication adéquate, mais les chercheurs ont estimé que cela ne suffit pas de : (i) suivre la réputation d'une marque, (ii) mesurer les performances de la stratégie de communication d'une campagne, (iii) assurer une veille concurrentielle. Pour cette raison, ils ont inventé le concept de l'écoute sociale (*Social Listening*) qui consiste à assurer une activité de suivi réalisée sur les réseaux sociaux [Pomputius \[2019\]](#). Cette activité consiste à analyser les informations communiquées par les consommateurs, souvent liées à une marque, sur les réseaux sociaux à l'aide des hashtags, des commentaires ou des mots clés. Par la suite, les chercheurs dans ce domaine ont inventé un autre concept de surveillance sociale (*monitoring social*) qui est plus attrayant que l'écoute sociale [Karafillakis et al. \[2021\]](#). Ce concept examine certaines mentions de marque et envoie des alertes lorsque une marque est mentionnée en ligne. Cela peut aider à réagir rapidement aux sentiments négatifs ou aux plaintes.

Ces notions ont également été appliquées dans le domaine médical par un certain nombre d'organisation. Par exemple l'agence américaine pour le développement international a appliqué des techniques de surveillance des médias sociaux et d'écoute sociale dans quelques pays africains [Silva et al. \[2022\]](#). Cette mesure a été prise dans le but de parvenir à une gestion adaptative de la stratégie d'une campagne qui consiste à améliorer la santé reproductive des jeunes africains. Ces concepts servent à divers services, comme l'élaboration de mesures

de surveillance. De même, ils sont utilisés pour adapter la vision de la campagne en fonction de l'analyse des données des médias sociaux.

Dans notre travail, nous nous sommes inspirés de ces concepts afin de réaliser la démarche entreprise par un psychiatre dans le but de se rapprocher du raisonnement humain. Cette démarche consiste à identifier les préoccupations du patient par une écoute continue. Elle permet également d'assurer à ce patient un soutien constant afin de ne pas provoquer son état et empirer sa situation. Pour ce faire, quand nous détectons un utilisateur souffrant d'un trouble de la personnalité. Nous passons par la suite à identifier la stabilité de son état pendant la période pandémique. Nous considérons une personne ayant un état instable si elle a fait preuve d'un comportement inapproprié au niveau d'un tweet à propos l'épidémie. Dans ce cas, nous proposons de retirer de son fil d'actualité les tweets qui peuvent nuire à son état. Cette tâche peut être accomplie en retirant la personne malade de la liste des abonnés de ce tweet. Le processus de suivi se poursuit par l'ouverture d'une conversation avec le patient au moyen d'un assistant virtuel pour réduire le degré de son anxiété. Les deux tâches du filtrage et de la recommandation s'effectuent à ce niveau à l'aide d'une ontologie.

Rappelons que dans notre deuxième contribution, nous avons proposé une méthode visant à identifier les causes qui rendent les gens dans une mauvaise situation à l'aide d'une analyse de la production textuelle des utilisateurs sur Twitter. Cette méthode s'appuie initialement sur l'application récursive d'un algorithme d'apprentissage automatique non supervisé aux données de Twitter afin de distinguer les différents domaines. Ensuite, pour chaque domaine obtenu à l'étape précédente, nous avons généré automatiquement une ontologie présentant les concepts (lexiques) et les relations sémantiques qui associent ces concepts ensemble. Chaque concept de chaque ontologie créée est relié par deux types de lexique ayant la catégorie grammaticale adjectivale. L'une de ces listes représente une humeur positive ou neutre, tandis que l'autre reflète un état d'esprit négatif, caractérisé par un comportement de doute incessant ou d'irresponsabilité totale. Ce choix repose sur notre hypothèse selon laquelle nous avons considéré que la cause qui affecte le TP fait l'objet du sujet de tweet qui est rédigé avec une mauvaise humeur.

Nous avons tiré profit de l'ontologie proposée au niveau de notre deuxième contribution traitant le sujet de COVID-19. Cette ontologie contient le lexique général avec les liens sémantiques reflétant l'épidémie ainsi que l'humeur associée à chaque concept. Il convient de noter que les deux listes d'états d'esprit sont établies en fonction du contexte. Cela signifie que, selon l'exemple (5), l'adjectif «heureux» sera ajouté à la liste qui montre un état d'esprit qui n'est pas approprié. Avant d'appliquer l'ontologie, nous avons effectué quelques modifications au niveau des individus de l'ontologie et plus particulièrement dans les deux listes reflétant l'humeur. Les modifications sont dues au fait que lors de l'association de chaque concept avec les deux listes d'adjectifs reflétant le type d'humeur, nous n'avons pas pris en compte les spécificités de l'état d'esprit (neutre ou positif, doute ou irresponsabilité). En effet, au niveau du quatrième chapitre nous avons fait une distinction qui vise à spécifier l'humeur de la personne par rapport à un sujet de façon générale.

Pour cette raison, nous avons apporté des modifications afin que notre ontologie réponde à notre besoin spécifique relatif à cette étape. Ces modifications

consistent à réaliser une distinction entre les adjectifs exprimant le comportement de l'utilisateur à savoir un comportement neutre, positif, irresponsable ou douteux à partir des deux listes de comportements. La première distinction réalisée au niveau du comportement inapproprié (irresponsable et douteux) est effectuée à l'aide de notre méthode de détection des comportements des utilisateurs envers l'épidémie Covid-19. La deuxième distinction entre les comportements positifs et neutres est établie manuellement à cause du manque de données pour la phase d'apprentissage.

La tâche de filtrage consiste à retirer du fil d'actualité de la personne malade qui a montré un comportement de doute à l'égard du COVID-19, les tweets discutant du même sujet avec le même comportement pour ne pas accroître le niveau d'anxiété et de peur chez le patient. Comme dans le cas du tweet (2) par rapport au tweet (1). Nous retirons également du fil d'actualité du malade, qui a manifesté un comportement irresponsable envers la pandémie de COVID-19, les tweets exprimant le même état d'esprit envers cette épidémie. Ceci est afin de ne pas encourager la personne malade à ne pas respecter les mesures prises par les autorités responsables. L'algorithme 2 décrit le processus de fonctionnement de notre analyseur sémantique basé sur une ontologie pour le filtrage des tweets inappropriés.

(1) *Eliminer des personnes âgées au rivotril est aussi délibérée. Tout ça a causé un grand nombre de morts considérable.*

(2) *Le nombre de morts du Covid-19 serait « probablement plus élevé » que le bilan officiel aux Etats-Unis.*

Algorithm 2 Analyse sémantique basée sur une ontologie permettant de masquer les tweets inappropriés du fil d'actualité d'une personne

- 1: **Définir** les règles de raisonnement SWRL pour extraire la liste des **concepts** représentés dans un tweet, tout en gérant les relations (synonymie, hyponymie, hyperonymie)
 - 2: **Définir** des règles de raisonnement SWRL pour extraire le **comportement** de l'utilisateur envers chaque concept
 - 3: **Pour chaque** nouveau tweet₁ susceptible d'apparaître dans le fil d'actualité de la personne malade
 - 4: **Interroger** l'ontologie pour extraire la liste des **concepts** et des **comportements** liés à chaque concept
 - 5: **Pour** chaque tweet₂ partagé par la personne malade
 - 6: **Interroger** l'ontologie afin d'extraire la liste des **concepts** et des **comportements** liés à chaque concept
 - 7: **Si** tweet₁ [conceptX] == tweet₂ [conceptY] et tweet₁ [comportementX] == tweet₂ [comportementY] **alors**
 - 8: **Masquer** le nouveau tweet du fil d'actualité de la personne malade
 - 9: **Break**
 - 10: **FinSi**
 - 11: **FinPour**
 - 12: **FinPour**
-

Pour la tâche de recommandation des tweets qui consiste à envoyer un message de clavardage au patient en raison d'un tweet posté à propos du COVID-19 avec un comportement négatif. Cette tâche consiste à répondre à ce tweet par un autre

tweet qui aborde le problème soulevé au sujet du tweet de la personne malade mais avec un comportement positif. Nous nous sommes limités à répondre à un tweet par un autre tweet afin de résoudre les questions relatives à la génération de réponse. Parmi les défis à relever pour générer une réponse, nous mentionnons l'aspect lexical, morpho-syntaxique et sémantique liés aux différents mots de réponse. De même, les enjeux liés à l'évolution du lexique au fil du temps, particulièrement dans le contexte du COVID-19, certaines situations impliquent le changement d'une réponse. Par conséquent, nous employons notre ontologie liée au COVID pour effectuer une analyse sémantique entre le tweet posté par le malade et les tweets qui sont postés par d'autres utilisateurs de Twitter. Cette analyse nous permet de sélectionner le tweet à recommander pour la personne malade suivant son état. Notre analyse sémantique dans ce cadre consiste à rechercher un tweet avec des concepts semblables au tweet du malade, mais avec un différent comportement. En effet, si l'utilisateur a posté un tweet au sujet de sa crainte d'ouvrir des écoles. Nous procédons à chercher un autre tweet avec un concept similaire lié à l'ouverture de l'école mais avec un bon état d'esprit. Cette comparaison a été effectuée à l'aide de notre ontologie. Généralement, ce processus propose divers tweets à recommander. Notre choix du tweet qui sera envoyé au malade est basé sur le tweet le plus récent contenant l'antonyme du comportement exprimé par le malade. La détection du rapport d'antonyme est assurée par la ressource Wordnet. Cela permet d'aborder l'aspect sémantique, comme le démontrent les tweets (3) et (4).

(3) *Le nombre de morts du Covid-19 serait « probablement plus élevé » que le bilan officiel aux Etats-Unis.*

(4) *Le taux de mortalité due au Covid-19 est minime Meme le WHO ne recommande pas le port du masque en dehors des hôpitaux.*

Dans le cas où le patient a agi de façon irresponsable, nous cherchons un tweet avec un concept similaire, mais avec un état d'esprit exprimant une humeur craintive dans le but d'effrayer le patient, comme dans (5) et (6).

(5) *Je suis le seul à être content pour le corona ?*

(6) *Il y a trop de choses très fastidieuses. Et j'ai vraiment peur d'être infecté COVID-19. Ma mère aussi. Je ne peux pas m'endormir à cause de ça.*

Notre méthode proposée à ce niveau prend en considération les différents aspects linguistiques, ce qui permet d'assurer la cohérence et la consistance des résultats obtenus. En outre, notre méthode porte sur les spécificités liées à un domaine d'étude particulier ayant un impact social significatif : la surveillance des personnes ayant des troubles de la personnalité durant la période du COVID-19. Aussi notre méthode assure une surveillance personnalisée qui dépend de l'état du patient. En effet, elle considère un tweet est inapproprié en fonction de l'état d'un patient particulier. Il convient également de noter que notre méthode prend en compte les différents problèmes liés au traitement linguistique : (i) l'évolution du lexique au fil du temps ce qui rend notre méthode applicable à toute période, (ii) les problèmes de génération automatique de réponses exigeant un effort lexical, grammatical et sémantique. En dépit des détails pris en compte par notre méthode en matière d'analyse, il est à noter que l'énorme nombre de spécificités fait l'existence de certains cas non couverts par notre méthode. Pour cette raison, les méthodes numériques peuvent être une véritable solution dans notre contexte.

5.3.4 Méthode basée sur l'apprentissage profond

Comme il est montré dans le tableau 5.3 les différentes méthodes fondées sur des techniques linguistiques présentent de nombreux avantages en termes d'explicitabilité et d'intelligibilité du traitement et des résultats obtenus. Toutefois, elles ne sont pas généralisées parce que ce n'est pas évident de disposer toujours des ressources qui s'appliquent à tous les domaines d'application. De ce fait, il est impossible d'obtenir des règles précises pour résoudre tous les cas spécifiques. Pour cela à ce niveau, nous visons à atteindre l'objectif de l'étape précédente, mais cette fois à l'aide d'une technique d'apprentissage profond. Cette démarche permet de donner à la machine la capacité d'apprendre la meilleure façon d'extraire des règles contribuant à atteindre des résultats appropriés en analysant un volume important de données.

TABLE 5.3 – Les critères de comparaison entre les méthodes numériques et linguistiques [Maâloul \[2012\]](#).

	Méthode numérique	Méthode linguistique
Avantages	Mise en œuvre facile Portables Applicables à tout type de corpus Temps de développement réduit	Prise en compte des catégories des mots Prise en compte de la structure des phrases et du corpus
Inconvénients	Besoin de larges corpus d'apprentissage Incohérence et inconsistance des résultats obtenus	Manque d'outils d'analyse de textes Temps de développement important Besoin de ressources et de connaissances linguistiques

Notre processus de travail décrit dans la figure 5.3 consiste à : (i) filtrer les tweets inappropriés du fil d'actualité de la personne malade, (ii) sélectionner des tweets à recommander par notre chatbot. Pour ce faire, nous avons commencé ce processus avec une phase de prétraitement. Cette phase consiste à supprimer les éléments inutiles, comme les mots vides, tout en gardant les éléments indiquant la négation et les chiffres. En effet, nous travaillons dans le contexte du COVID-19, où les chiffres représentent un critère important pour montrer le nombre de décès et de personnes infectées. Après cela, nous avons procédé à la conversion des données textuelles en un ensemble de vecteurs numériques qui peuvent être traités par des algorithmes d'apprentissage automatique. Après une étude empirique, nous avons décidé de tirer parti du modèle de plongement de phrases, basé sur une représentation d'encodeur bidirectionnel à partir de réseaux autoattentifs Bert (*Bidirectional Encoder Representations from Transformers*) : (i) le modèle est entraîné sur une grande quantité de données combinant des ressources provenant de certains livres et de Wikipédia, (ii) le modèle maintient la dépendance sémantique existante entre les diverses unités lexicales comme il est montré dans la figure 5.2, (iii) le modèle Bert accepte un grand nombre de paramètres, ce qui le rend mieux

adapté aux différents besoins, (iv) le modèle cité au cours de son entraînement tente de modéliser le langage masqué (MLM, *Masked Language Modeling*), ainsi de prédire la prochaine séquence de mots (NSP, *Next Sentence Prediction*). Ces dernières techniques permettent de résoudre le problème de la classification dans un corpus réduit. Le fonctionnement global de cette technique consiste à calculer la similarité entre les phrases en utilisant des couches de mise en correspondance, tout en conservant les descripteurs importants. Le modèle BERT offre un ensemble de vecteurs normalisés et standardisés tout en abordant un grand nombre de problèmes reconnus comme la combinaison de vocabulaire du corpus.

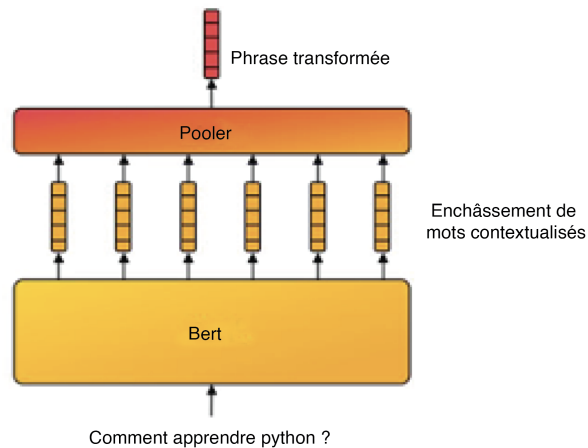


FIGURE 5.2 – Exemple d'entraînement d'un modèle d'intégration de phrases [Reimers et Gurevych \[2019\]](#).

Au cours de ces dernières années, de nombreuses extensions de BERT ont été développées dans le but d'améliorer leur fonctionnalité en citant à titre d'exemple :

RoBERTa [Murarka et al. \[2021\]](#) : par rapport à BERT, le modèle RoBERTa repose sur un corpus plus large qui intègre des données textuelles plus informelles au cours de la phase d'entraînement, comme les données provenant de Reddit. La méthodologie d'entraînement de RoBERTa intègre une technique étendue du modèle MLM plutôt que la technique de NSP.

DistilBERT [Sanh et al. \[2019\]](#) : l'objectif de DistilBERT est de réduire la taille du modèle de base BERT, cela le rend moins complexe et plus attractif en termes de temps d'exécution. Cependant, le modèle DistilBERT préserve la majorité des techniques de compréhension du langage ainsi que la formation des données pour un usage général.

XLM BERT Base [Velankar et al. \[2022\]](#) : le modèle XLM est une version multilingue de distilroberta-v1, qui a été entraîné sur des données parallèles obtenues à partir de plus de 50 langues qui sont composées de millions de fragments de phrases.

XLNet [Das et al. \[2021\]](#) : ceci est un modèle linguistique autorégressif généralisé (AR) qui permet la modélisation linguistique par permutation (PLM) pour l'apprentissage de contextes bidirectionnels. XLNet s'inspire du modèle de langage d'encodeur automatique et du modèle de langage autorégressif, tout en évitant leurs limites. De plus, XLNet dépasse BERT en vingt défis, souvent avec une

marge importante, y compris l'inférence du langage naturel, l'analyse des sentiments, la réponse aux questions et le classement des documents.

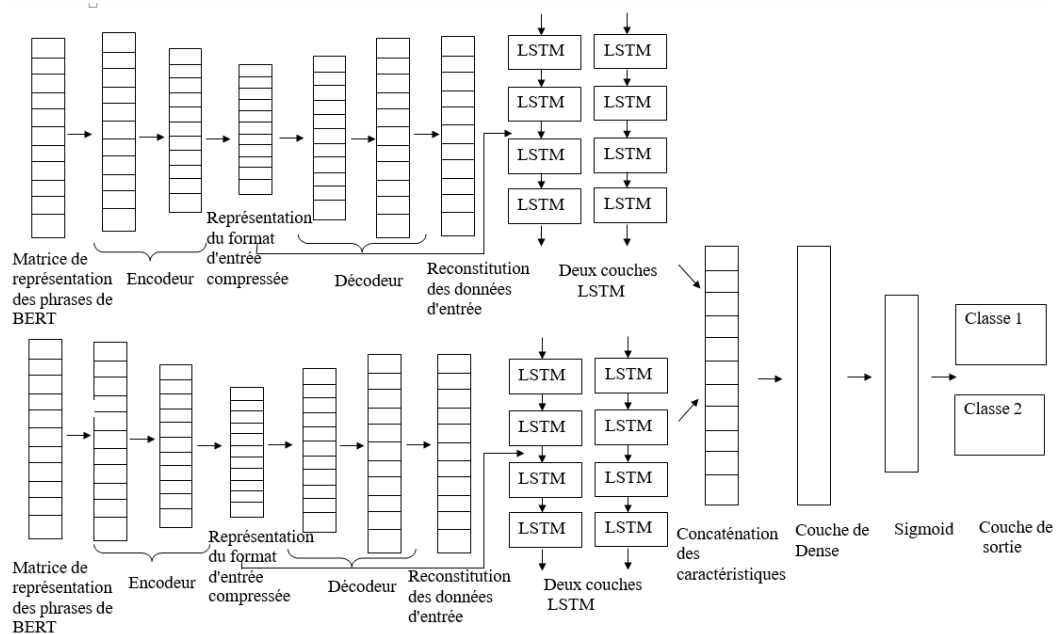


FIGURE 5.3 – Réseau neuronal ConvAE-LSTM-SVM pour la tâche de filtrage ou de recommandation des tweets.

En dépit de la différence dans l'architecture de chaque modèle, en général ces techniques ont prouvé une haute performance. En effet, de nos jours ces modèles sont les plus utilisés pour transformer la partie textuelle en un ensemble de vecteurs numériques Das et al. [2021]. Après l'étape de transformation, nous procédons à l'extraction de caractéristiques de haut niveau à partir de données textuelles brutes. Pour ce faire, nous avons effectué une étude de la littérature afin de déterminer les meilleurs moyens d'accomplir cette tâche. En général, il existe des techniques basées sur des mesures statistiques comme les techniques de réduction de la dimension ou les techniques de corrélation. Il existe d'autres techniques fondées sur des méthodes neuronales telles que la machine Boltzmann restreinte⁵, Autoencodeur, CNN, etc. Plusieurs travaux ont effectué des études expérimentales sur la différence au niveau de l'utilisation des techniques neuronales pour extraire les attributs Wajeed et Sreenivasulu [2019], Hedayati et al. [2021]. Parmi les conclusions tirées, l'auto-encodeur peut être considéré comme un bon choix si nous disposons d'un volume limité de données. En effet, il exige moins de paramètres et moins de données pour la généralisation. De même les auto-encodeurs peuvent apprendre des fonctionnalités plus abstraites et de niveau supérieur que celles apprises par CNN. En effet, le réseau CNN a tendance à se concentrer sur des caractéristiques de bas niveau comme les contours et les textures dans le contexte de l'analyse d'image. Pour cela, l'auto-encodeur peut être utile si nous voulons extraire des fonctionnalités qui capturent davantage la structure globale du contenu d'entrée en particulier si la forme générale des entrées n'est pas complexe. Sur la base des raisons citées, nous avons choisi d'exécuter la

5. https://fr.wikipedia.org/wiki/Machine_à_boltzmann_restreinte

tâche d'extraction des attributs à l'aide des auto-encodeurs. Ce choix s'explique également par l'intérêt accordé à ce modèle pour les tâches de traitement automatique du langage naturel en rapport avec le domaine médical [Pandey et al. \[2021\]](#). Un auto-encodeur est un modèle de réseau de neurones qui peut être appliqué pour apprendre une représentation compressée de données brutes. Comme il est montré dans la figure 5.4 un auto-encodeur est composé de trois fonctions : une fonction d'encodage, une fonction de décodage et une fonction de distance permettant de mesurer la quantité d'informations perdues entre les représentations compressée et non compressée des données.

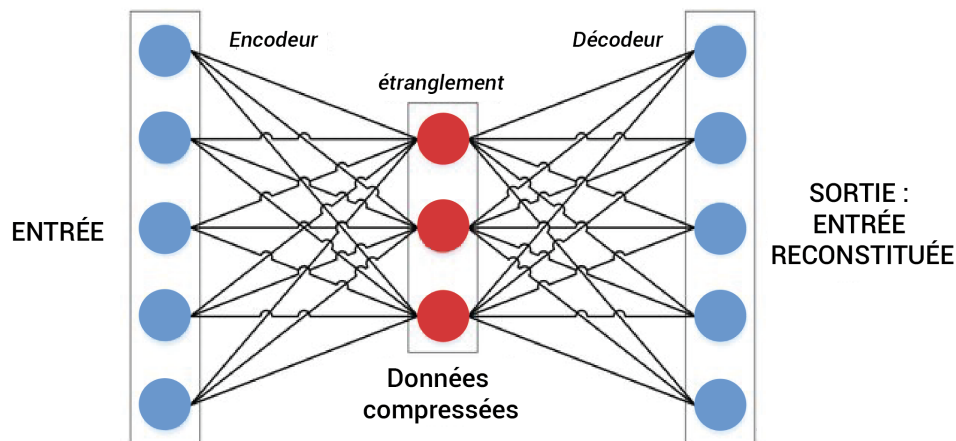


FIGURE 5.4 – Architecture de base d'un auto-encodeur monocouche [Sublime et Kalinicheva \[2019\]](#).

Dans la littérature, il existe de nombreuses extensions de la version simple de l'auto-encodeur basée sur une couche entièrement connectée (*fully-connected layer*). Nous mentionnons parmi ces extensions, un auto-encodeur clairsemé [Rao et al. \[2021\]](#) (*a sparse autoencoder*), un auto-encodeur profond avec des couches entièrement connectées [Pan et al. \[2020\]](#) (*a deep fully-connected autoencoder*), un auto-encodeur convolutif profond [Ali et al. \[2020\]](#) (*a deep convolutional autoencoder*), etc.

Dans notre travail, nous avons appliqué les trois différents modèles d'auto-encodeurs suivants : un auto-encodeur simple basé sur une couche entièrement connectée, un auto-encodeur profond basé sur des couches entièrement connectées et un auto-encodeur convolutif profond. Nous nous sommes limités à cette sélection, puisque les autres versions sont plus adaptées au traitement d'image.

La principale distinction entre les auto-encodeurs simples et profonds est que l'auto-encodeur simple possède une couche cachée entre l'entrée et la sortie. Tandis que l'auto-encodeur profond comporte de multiples couches cachées en fonction de la configuration désirée. En ce qui concerne l'architecture de l'auto-encodeur à convolution profonde, elle repose sur des opérations spécifiques de convolution et de pooling. La première couche de ce modèle consiste à trouver les composants clés, la deuxième couche les combine pour trouver des formes. La dernière couche combine tous ces éléments pour dériver des objets entiers. Ensuite, deux couches LSTM sont utilisées par notre modèle pour mettre en évidence les dépendances à longue distance entre les différentes unités lexicales de la partie textuelle. Les deux sorties de chaque dernière couche correspondant à la base de chaque tweet sont ensuite combinées pour détecter s'il existe une relation

sémantique entre les deux tweets afin d'assurer les tâches de filtrage et de recommandation. Pour cette raison, la dernière couche de chaque tâche est composée de deux neurones qui classifient chaque paire de tweets selon qu'ils doivent être supprimés (Oui/Non) ou recommandés (Oui/Non).

Au niveau de la tâche de filtrage, notre modèle compare chaque tweet publié par le malade à chaque tweet qui peut apparaître dans son fil d'actualité. C'est-à-dire avec chaque tweet qui est posté par l'une des personnes qu'il suit leurs actualités. Si les deux tweets sont liés au même sujet *COVID-19*, nous passons à une classification pour décider si le deuxième tweet (tweet à observer par le malade) doit être affiché ou masqué. En ce qui concerne la tâche de recommandation, celle-ci est divisée en deux parties : la première consiste à recommander un tweet du fil d'actualité de la personne malade pour qu'il apparaisse en premier lieu, c'est-à-dire au niveau la tête de la liste. La deuxième consiste à analyser le tweet posté par le malade et chaque tweet posté sur Twitter dans un ordre chronologique. Si le deuxième tweet de la paire a été classé comme recommandé, nous suggérons à un agent d'envoyer ce tweet au patient sous la forme d'un message.

Notre choix d'utiliser la technique d'apprentissage profond à ce niveau est justifié par l'intérêt qu'elle a montré dans plusieurs travaux relatifs à l'identification d'informations implicites à partir d'un texte. Nous citons dans ce contexte l'analyse de l'ironie et du langage évasif [Apolinario-Arzube et al. \[2020\]](#), [Ghanem et al. \[2020\]](#). Cela se fait en tenant compte des enjeux linguistiques liés à la langue des médias sociaux, comme l'absence de contexte, la présence de termes en dehors du vocabulaire, etc. En outre, l'utilisation de la technique d'apprentissage profond peut offrir plus de flexibilité dans le processus décisionnel comparativement à la technique fondée sur l'ontologie. Cela malgré le fait que nous nous sommes limités à l'étude d'un domaine spécifique et que nous avons tenté de regrouper autant que possible ses concepts. Cela est dû à la difficulté de formaliser des règles linguistiques qui tiennent compte de l'ensemble des spécificités de ce domaine. Nous mentionnons ci-dessous certains cas non couverts à l'étape précédente par l'ontologie pour la tâche de filtrage des tweets inappropriés.

Cas 1 : Par exemple, la flexibilité sémantique est toujours limitée par notre ontologie. Le couple de tweets suivant (7) montre un degré de correspondance sémantique qui doit être pris en considération au niveau du filtrage des tweets inappropriés afin de ne pas encourager la personne à adopter un comportement irresponsable.

(7) *Ah mais moi ça me d erange pas d'avoir le corona. /Jusqu''a mtn je m'en foutais a mort du corona.*

Cas 2 : L'ontologie proposée s'appuie sur la catégorie grammaticale *adjectif* pour déterminer le comportement exprimé dans le tweet. L'exemple suivant (8) montre un tweet indiquant un comportement inapproprié et aucun de ses mots fait partie de la catégorie grammaticale adjectivale.

(8) *Cette crise nous l'a montré, l'UE ne sert à rien, il n'y a ni union ni solidarité.*

Cas 3 : Généralement, les méthodes fondées sur des règles sont rigides et difficiles à les appliquer ; surtout que d'un point de vue sémantique, l'être humain est toujours diligent. Malgré l'évidence et l'importance de la notion de diligence il est difficile de la modéliser par notre ontologie car cette notion n'est pas vraiment tangible. L'exemple ci-dessous (9) montre la diligence humaine impliquée dans la

détection du rapport sémantique entre les deux tweets sur laquelle l'annotateur a accordé à ces deux tweets la classe à masquer.

(9) *Je vais sûrement pas signer cette pétition! Je soutiens notre président et notre gouvernement et j'assume. / À noter d'ailleurs que c'est le premier chef d'État à être atteint par le coronavirus félicitations.*

Il convient de noter que lors de la phase d'entraînement, nous avons utilisé des tweets redondants dans différents couples. Ce choix permettra à l'algorithme d'apprendre les relations de dépendance entre les différents concepts liés à un domaine. Par exemple, la première paire est constituée d'exemples (10),(11) et la deuxième paire est constituée d'exemples (10),(12). Par conséquent, nous poussons l'algorithme à déterminer les relations entre les concepts des exemples (11) et (12) :

(10) *@gouvernementXX Cette décision est criminelle! Covid-19.*

(11) *@ministreXX Vous avez déjà lâché tout le monde avec votre gestion ératique.*

(12) *@présidentXX L'espoir sera le jour ou tu ne sera plus! Covid-19.*

Dans notre travail, nous avons considéré que la tâche de recommandation des tweets est plus délicate que la tâche de filtrage des tweets. En effet, si nous faisons une erreur en filtrant un tweet qui n'est pas inapproprié, l'utilisateur ne sera pas dérangé. De plus dans le cas, où ce tweet ne convient pas à une personne spécifique, mais notre méthode ne l'élimine pas. Cela ne change rien étant donné que ce tweet apparaît dès le début dans le fil d'actualité sans notre intervention. En revanche, en ce qui concerne la recommandation de tweets, commettre une erreur est plus préjudiciable. En effet au départ le tweet à recommander n'est pas affiché, mais notre chatbot à inciter la diffusion de l'information. Le pire se présente au cas où ce tweet va aggraver la situation de la personne malade. Les exemples suivants illustrent certains des problèmes évoqués.

Cas 1 : La sensibilité au moment de sélectionner un tweet à envoyer au patient. Par exemple, il est inapproprié d'envoyer le tweet suivant (13) au patient.

(13) *Les Français ne sont pas des enfants, nous pouvons leur faire confiance.*

Cas 2 : L'importance de l'ordre des tweets dans le couple, comme il est montré dans le couple de tweets suivant (14).

(14) *Je mon fous je vais embrasser mon ex / si vous continuez à trop faire circuler le virus, vous pouvez tuer des gens.*

Cas 3 : La généralisation du contexte dans l'analyse sémantiques des concepts, comme il est illustrée dans l'exemple (15).

(15) *Bon le président XX voulait déconfinement général alors qu'un déconfinement très progressif par région aurait été beaucoup plus approprié / N'aie pas peur continue à appliquer les gestes barrières, ça ira*

Bien qu'il soit intéressant de lier notre travail à un domaine spécifique comme celui de COVID-19, notre méthode reste limitée. En effet, elle est très liée à un cadre spatial et temporel spécifique. Pour cette raison, proposer un travail destiné à donner un usage répandu est très gratifiant dans notre cas.

5.4 MÉTHODE POUR SURVEILLER LES UTILISATEURS

Notre méthode vise à profiter des données partagées sur les médias sociaux pour assurer l'intégralité du processus de diagnostic entrepris par un psychiatre. Cette mesure vise à assurer un suivi en temps réel de la situation des personnes

atteintes de TP envers l'épidémie COVID-19 dans les médias sociaux. Cet objectif est atteint en tirant profit des techniques principalement liées à l'intelligence artificielle combinant des méthodes de TALN et d'apprentissage automatique.

Nous pouvons étendre la méthode proposée de sorte qu'elle soit généralisée et non pas liée à un domaine précis comme celui de COVID-19. Pour ce faire, nous commençons notre travail en déterminant si un tweet a été partagé par un utilisateur qui a parmi ses abonnés quelqu'un avec un trouble de la personnalité. Dans ce cas, nous comparerons le thème du tweet partagé par cet utilisateur avec celui de chaque tweet de la liste des 20 derniers tweets de l'utilisateur suspecté d'être malade. Cette comparaison peut être réalisée grâce à l'ontologie de chaque domaine proposée au niveau de notre deuxième contribution. Nous nous rappelons dans ce contexte qu'au niveau de cette contribution nous avons proposé une méthode automatique permettant la construction d'une ontologie à un domaine spécifique. Par conséquent, si nous trouvons une paire de tweets ayant le même sujet et que le comportement exprimé par le malade dans le tweet est négatif. Nous procédons à mesurer la distance sémantique entre les deux tweets en utilisant la technique de plongement de phrases. Si la valeur obtenue excède le seuil, nous retirons cet utilisateur de la liste des abonnés de ce tweet pour ne pas le provoquer. La figure 5.5 présente l'extension de la figure 5.1, afin qu'elle puisse être appliquée aux autres domaines d'études.

Il est à noter que l'intérêt du croisement thématique consiste à assurer la préservation du contexte. Par exemple, les deux exemples suivants (16) et (17) présentent une certaine similarité sémantique, mais en réalité, ils n'ont pas la même signification sémantique.

(16) *Je ne veux pas de ce nourriture.*

(17) *Je ne veux pas de ce sport.*

Il est également à noter que le choix d'utiliser des métriques de calcul de distance de ressemblance entre des tweets est justifié par le fait que les domaines sont très étendus. Nous prenons l'exemple de ces deux tweets (18) et (19) qui abordent le même thème, mais qui ne reflètent pas le même objet.

(18) *Je suis fier de notre championne de natation.*

(19) *Je suis très en colère de voir mon équipe de football perdre le match.*

Dans cette section, nous nous focalisons principalement sur le sujet commun des deux tweets et l'humeur exprimée par le patient (dans le premier tweet). L'ignorance de l'humeur indiquée au niveau du deuxième tweet est justifiée par les deux exemples suivants (20) et (21) :

(20) *Je suis déçu, mon équipe XX a fait la troisième défaite consécutive en une semaine.*

(21) *L'équipe YY a gardé la première place du classement pendant les deux derniers mois.*

Selon notre hypothèse, le deuxième tweet sera supprimé car il aborde le même sujet que le premier et exprime une mauvaise humeur. Ce choix est appuyé par l'argument que les deux équipes YY et XX pourraient être des concurrentes. Cependant, si nous appliquons ces critères dans le cadre de l'éclosion de COVID-19, il n'est pas nécessaire de retirer le deuxième tweet. Au contraire, nous pouvons le recommander au patient. Nous admettons que dans notre travail, nous avons utilisé plusieurs filtres, mais cela a été fait dans le but d'éviter les effets néfastes du surfiltrage.

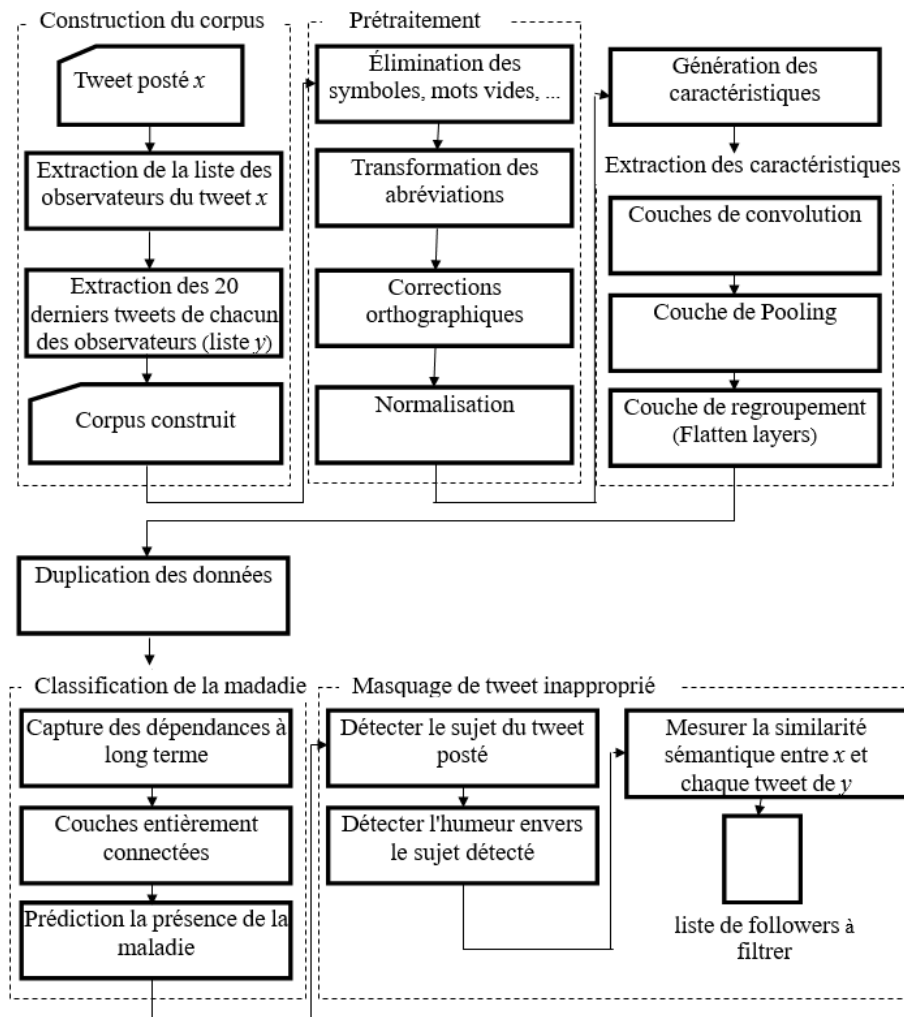


FIGURE 5.5 – Méthode proposée pour la détection et le suivi des personnes souffrant de TP.

5.5 CONCLUSION

Dans ce chapitre, nous avons développé notre troisième contribution, qui vise à assurer une surveillance des personnes ayant des TP sur les réseaux sociaux. Pour ce faire, nous avons proposé une première méthode basée sur une ontologie qui permet de manipuler le fil d’actualité de la personne malade en filtrant des tweets et en recommandant d’autres tweets au moyen d’une analyse sémantique. Cette analyse consiste à établir une comparaison entre les tweets publiés par le patient et ceux qui peuvent apparaître dans son fil d’actualité. Par la suite, nous avons proposé une deuxième méthode qui nous permet d’accomplir l’objectif similaire à celui de la tâche précédente. Mais cette fois l’analyseur sémantique proposé a été fondé sur un apprentissage profond au lieu d’une ontologie. En effet, ce choix permet de résoudre les conflits linguistiques délimités au moyen d’une analyse profond du style d’écriture. Il est à note que nous avons fait appel aux principales contributions développées au cours de cette thèse qui consistent à la détection des personnes ayant des TP ainsi que les sujets qu’elles perturbent. Ces contribu-

tions contribuent à assurer le bon fonctionnement de la méthode proposée pour surveiller les gens avec TP sur les réseaux sociaux.

Dans le sixième chapitre, nous exposerons en détail les expériences qui mettent en oeuvre les méthodes proposées dans les chapitres précédents. Nous décrirons également les évaluations que nous avons menées sur le corpus élaboré à l'aide d'une discussion sur les résultats obtenus.

EXPÉRIMENTATIONS ET DISCUSSION DES RÉSULTATS OBTENUS

6

SOMMAIRE

6.1	INTRODUCTION	118
6.2	ENVIRONNEMENT DE PROGRAMMATION	118
6.3	CORPUS DE RÉFÉRENCE	119
6.3.1	Dataset pour l'analyse de la dépression sur Twitter Infamouscoder [2022b]	119
6.3.2	Dataset pour l'analyse du suicide et de la dépression Komati [2021]	120
6.3.3	Dataset pour l'analyse de la dépression sur Reddit Infamouscoder [2022a]	120
6.3.4	Dataset pour l'analyse de la santé mentale Namdari [2023]	120
6.3.5	Dataset pour l'analyse du comportement anti-social sur Twitter pour le public indonésien Astuti [2021]	121
6.3.6	Dataset provenant d'un groupe de témoignage sur Facebook : Trouble Personnalité Borderline	121
6.4	VALIDATION DU CORPUS COLLECTÉ POUR LA DÉTECTION DES PERSONNES AYANT DES TP	122
6.5	DÉTECTION DES PERSONNES SOUFFRANT DES TROUBLES DE LA PERSONNALITÉ SUR LES RÉSEAUX SOCIAUX	125
6.5.1	Résultats des méthodes de détection des troubles	125
6.5.2	Évaluation des méthodes de détection des troubles	127
6.5.3	Discussion des méthodes de détection des troubles	130
6.6	IDENTIFICATION DES CAUSES AFFECTANT LA MALADIE DE TP	132
6.6.1	Résultats de la méthode d'identification des causes du TP	133
6.6.2	Évaluation de la méthode d'identification des causes du TP	136
6.6.3	Discussion de la méthode d'identification des causes du TP	137
6.7	SURVEILLANCE DES PERSONNES ATTEINTES DU TP SUR LES RÉSEAUX SOCIAUX : ÉTUDE DE CAS COVID-19	138
6.7.1	Résultats de la méthode de surveillance des cas des personnes ayant un trouble	138
6.7.2	Évaluation de la méthode de surveillance des cas des personnes ayant un trouble	139

6.7.3	Discussion de la méthode de surveillance des cas des personnes ayant un TP	142
6.8	CONCLUSION	147

6.1 INTRODUCTION

Dans les chapitres précédents, nous avons détaillé nos trois contributions proposées pour la détection et la surveillance des personnes ayant des troubles de personnalité sur Twitter. La première était basée sur des algorithmes d'apprentissage classiques avec une méthode d'extraction des attributs personnalisés. Aussi elle comporte une autre méthode basée sur des algorithmes d'apprentissage profond en profitant d'une méthode d'extraction des attributs d'une manière automatique à partir des données brutes. Notre deuxième contribution consiste à dégager les causes des troubles de la personnalité qui sont considérées comme les sujets de tweets qui sont écrits avec une humeur négative. Cette contribution est effectuée en se basant sur une méthode hybride qui combine à la fois des méthodes d'apprentissage automatique, des méthodes linguistiques ainsi que des méthodes statistiques. Notre troisième contribution consiste à éliminer les tweets inappropriés et à recommander d'autres tweets appropriés en fonction de l'état de l'utilisateur malade. Cette contribution est réalisée en utilisant une méthode hybride qui combine à la fois des techniques linguistiques et des méthodes d'apprentissage profond. Le présent chapitre est consacré à l'expérimentation et à l'évaluation de chacune de ces méthodes, y compris celles développées dans d'autres études au niveau du même contexte.

En outre, ce chapitre est réparti comme suit : nous présentons dans un premier temps l'environnement de programmation utilisé, le langage de programmation utilisé et les différentes bibliothèques exploitées. Ensuite, nous exposons les différents corpus avec lesquels nous avons évalué les différentes parties de notre travail. Puis, dans un deuxième temps, nous présentons les expérimentations et les résultats des différentes méthodes de détection et de suivi avec une évaluation de chaque méthode appliquée. Cela se fait en comparant à chaque fois les résultats de notre méthode avec les résultats obtenus dans le même contexte. La dernière section est consacrée à une discussion mettant en relief les facteurs qui ont contribué à obtenir les résultats de chaque méthode.

6.2 ENVIRONNEMENT DE PROGRAMMATION

Dans la section environnement de programmation, nous présentons le langage de programmation Python et les différentes bibliothèques employées au cours de nos expérimentations. Le langage Python est un langage qui a été développé en 1989, il possède les caractéristiques suivantes :

- Il est open source, ce qui signifie que son usage est libre ainsi que ses fichiers sources sont disponibles et éditables.
- Il dispose d'une large base de bibliothèques d'API, qui sont offerts gratuitement et peuvent être facilement intégrés. Ces bibliothèques permettent de couvrir un certain nombre de cas d'utilisation des données.
- Il est dédié à plusieurs cas d'utilisation tels que le calcul scientifique, les statistiques, la visualisation, etc.
- Il est hautement portable, ce qui signifie qu'il est autonome du système d'exploitation (linux, Windows, MacOS).
- Il est orienté objet, ce qui garantit le concept de modularité afin de mieux intégrer des variables et des fonctions complexes.

- Il comporte un module de typage dynamique, c'est-à-dire que le type est attribué automatiquement au cours de l'exécution du programme.
- Il prend en charge l'intégration des autres langues.

Le tableau suivant 6.1 présente les principales bibliothèques employées dans notre contexte pour l'implémentation de notre méthode afin d'analyser les textes provenant des réseaux sociaux.

TABLE 6.1 – Liste des principales bibliothèques utilisées de python.

Package	Version	Description
nlTK	3.5	Pour appliquer les différents analyseurs linguistiques.
os	3.7	Pour accéder aux espaces de travail.
re	3.7	Pour le prétraitement.
pandas	1.1.4	Pour le stockage des résultats et la génération du rapport d'analyse.
numpy	1.19.4	Pour le traitement des données statistiques.
math	0.0.1	Pour appliquer des mesures mathématiques.
emoji	2.2.0	Pour la détection des émoticôns.
NRCLex	4.0	Pour mesurer l'effet émotionnel d'un corps textuel.
TextBlob	0.17.1	Pour l'analyse des sentiments d'un segment textuel.
sklearn	1.1.3	Pour la manipulation des algorithmes d'apprentissage.
matplotlib	3.6.3	Pour la gestion des interfaces graphiques.
gensim	4.3.0	Pour mesurer la similarité sémantique.
sentence-transformers	2.2.2	Pour la représentation numérique de la partie textuelle.
PyTorch	1.10.0	Pour la création et l'entraînement de réseaux neuronaux.
Hugging Face	4.12.2	Pour le traitement du langage naturel à l'aide de l'IA.

6.3 CORPUS DE RÉFÉRENCE

Pour la détection des personnes présentant des troubles de la personnalité sur les réseaux sociaux, nous avons identifié, dans la littérature, plusieurs corpus rédigés dans différentes langues. De manière générale, ces corpus sont dédiés à la détection d'informations spécifiques liées au contexte de notre étude (comme la détection de symptômes ou de conséquences de problèmes psychologiques). Dans ce qui suit, nous présentons les corpus utilisés pour l'évaluation de notre corpus collecté, ainsi que les corpus employés pour valider la performance de notre méthode de détection des utilisateurs susceptibles d'être malades.

6.3.1 Dataset pour l'analyse de la dépression sur Twitter [Infamouscoder \[2022b\]](#)

Les données de ce corpus sont recueillies de Twitter via son API à l'état brut ce qui signifie qu'elles ne sont pas nettoyées. Les tweets collectés sont rédigés princi-

palement en anglais. Le corpus est composé de 20000 tweets d'utilisateurs étiquetés comme dépressifs et non-dépressifs. L'objectif général de ce corpus consiste à classer la santé mentale de l'utilisateur au niveau du tweet en mesurant le degré de dépression. Mis à part le contenu textuel des tweets, ce corpus intègre pour chaque tweet : (i) la classe de classification qui indique si l'utilisateur est en dépression ou non, (ii) 9 autres types d'attributs regroupant les métadonnées de Twitter telles que le nombre des favoris, les amis, les followers, etc.

6.3.2 Dataset pour l'analyse du suicide et de la dépression [Komati \[2021\]](#)

L'ensemble de données de ce corpus représente une collection de 232074 messages collectés à partir des groupes de *SuicideWatch* et de *depression* de la plateforme Reddit. Les différents messages de ce corpus sont rassemblés au moyen de l'API Pushshift. Tous les messages qui ont été postés sur *SuicideWatch* ont été collectés du 16 décembre 2008 (date de création) au 2 janvier 2021. Tandis que les messages du groupe *depression* ont été collectés entre le 1er janvier 2009 et le 2 janvier 2021. Ce corpus est composé principalement de deux attributs le premier représente les messages qui sont écrits en anglais et le deuxième représente la classe de classification : suicide et non-suicide. L'ensemble de données collecté peut être utilisé pour détecter le suicide et la dépression au niveau d'un message.

6.3.3 Dataset pour l'analyse de la dépression sur Reddit [Infamouscoder \[2022a\]](#)

L'ensemble de données brutes de ce corpus est collecté par le phénomène de *web scrapping* à partir des *Subreddits*. Un *subreddit* (*sous-reddit*) est une sous-partie du site Reddit consacrée principalement à un thème spécifique où les utilisateurs peuvent poster un contenu relatif à ce thème. Ces données sont ensuite nettoyées grâce à plusieurs techniques de TALN.

L'ensemble de données comprend essentiellement 7650 lignes, dont chacune est constituée par deux attributs. Le premier attribut correspond à la partie textuelle rédigée en anglais. Le deuxième attribut représente la classe de classification indiquant si l'utilisateur est déprimé. Ce qui permet d'atteindre l'objectif principal qui consiste à la classification du contenu du texte en fonction de l'état mental.

6.3.4 Dataset pour l'analyse de la santé mentale [Namdari \[2023\]](#)

Le corpus de la santé mentale représente une collection de textes relatifs aux personnes souffrant d'anxiété, de dépression et d'autres problèmes de santé mentale. La totalité du corpus est de 27972 lignes, dont chaque ligne est composée de deux colonnes : l'une contenant les commentaires, et l'autre contenant des étiquettes indiquant si les commentaires sont jugés toxiques ou non. Un commentaire est considéré comme toxique s'il reflète la dépression et l'anxiété.

Les données du présent corpus peuvent être utiles aux chercheurs et aux professionnels en santé mentale. Elles peuvent servir à différentes tâches, telles que l'analyse des sentiments, la détection du langage toxique, l'analyse et la compréhension du langage relatif à la santé mentale.

6.3.5 Dataset pour l'analyse du comportement anti-social sur Twitter pour le public indonésien Astuti [2021]

L'ensemble de données de ce corpus comprend principalement 1251 tweets qui sont tous rédigés en indonésien. Chaque ligne de ce corpus possède 12 attributs comportant des informations sur l'utilisateur comme son identifiant. Aussi elle comporte des informations relatives aux métadonnées de Twitter comme le nombre d'utilisation de l'émoticône *j'aime* ainsi le nombre de status partagés. L'attribut indiquant la classe peut avoir la valeur d'une des valeurs suivantes : non-conformité avec les normes sociales et le comportement légal, irritabilité et agressivité, négligence téméraire envers la sécurité, absence de remords, une classe indiquant que la personne est non antisociale. Cette diversité peut offrir plus de précision par rapport à d'autres corpus. Cette précision peut nous permettre de comprendre les spécificités de son état mental afin d'y réagir plus efficacement.

6.3.6 Dataset provenant d'un groupe de témoignage sur Facebook : Trouble Personnalité Borderline

Nous avons collecté des données relatives aux individus qui ont admis souffrir d'un trouble de personnalité de type borderline dans le groupe *Trouble Personnalité Borderline* sur le réseau de Facebook. L'objectif de ce groupe est de donner l'opportunité aux personnes ayant la trouble Borderline de discuter ensemble à propos certaines de ses préoccupations afin de trouver des solutions tangibles. Les données de ce groupe peuvent être une ressource importante pour étudier le style d'écriture, les principaux soucis et les sujets nerveux pour les personnes malades. En effet, Facebook donne plus de flexibilité sur la taille d'un statut (pas comme Twitter 280 caractères par tweet). Ces données sont collectées du 01/01/2021 au 01/10/2022, elles sont écrites en français et abordent différents thèmes. Après la collection nous avons éliminé dans un premier temps les données de multimédia comme les vidéos et les images. De même nous avons éliminé les statuts qui sont très réduits afin de garantir que chaque statut comporte une idée complète. Le tableau 6.2 présente une description générale de la taille du corpus collecté.

TABLE 6.2 – Aperçu général sur la taille du corpus collecté à partir Facebook.

Description	Taille
Le nombre de statuts	1419
Le nombre des utilisateurs	304
La moyenne des publications pour chaque utilisateur	4.67
Le nombre des utilisateurs avec plus de 5 publications	87
Le nombre des utilisateurs avec plus de 10 publications	30
Le nombre de publications avec plus de 280 caractères	654

Le tableau 6.3 montre les principales différences entre les divers corpus de référence sélectionnés.

TABLE 6.3 – Aperçu général sur les corpus de références sélectionnés.

Référence	Source	Volume	Langue	Attributs	Objectif
Infamouscoder [2022b]	Twitter	20000 tweets	anglais	11 attributs (texte+metadata+classe)	Détecter les utilisateurs en dépression
Komati [2021]	Reddit	232074 messages	Anglais	2 attributs (texte+classe)	Déterminer si un usager a des idées suicidaires.
Infamouscoder [2022a]	Reddit	7650 messages	Anglais	2 attributs (texte+classe)	Déterminer si un utilisateur est déprimé.
Namdari [2023]	—	27972 commentaire	Anglais	2 attributs (texte+classe)	Détecter la production textuelle des personnes souffrant d’anxiété et de dépression
Astuti [2021]	Twitter	1251 tweets	Indonésien	12 attributs (texte+metadata+classe)	Détecter le type d’antisociale exprimé
Groupe de témoignage	Facebook	1419 publications	Français	1 attribut (texte)	Analyser les styles d’écriture des personnes ayant des TP de type borderline

6.4 VALIDATION DU CORPUS COLLECTÉ POUR LA DÉTECTION DES PERSONNES AYANT DES TP

Dans la partie de validation du corpus, nous visons à vérifier la justesse de notre corpus élaboré pour la détection des personnes ayant des TP. Nous avons mentionné dans les différentes phases de construction de corpus que nous avons eu recours à un seul annotateur au cours de la phase d’annotation. Ce choix est dû à la difficulté de la tâche d’annotation en terme de temps et d’effort. Le processus de validation est entamé par le calcul du degré d’interdépendance entre les éléments liés au style d’écriture et la présence de la maladie en profitant de la mesure χ^2 ¹. Ensuite, ce processus est suivi d’une comparaison entre les degrés d’interdépendance identifiés à partir de notre corpus et ceux qui sont identifiés à partir des différents corpus de référence. Nous présentons dans le tableau ci-dessous 6.4 la variation de l’interdépendance figurant au niveau : (i) de notre corpus, (ii) des deux corpus issus de la littérature [Astuti \[2021\]](#) et [Infamouscoder \[2022a\]](#) qui sont respectivement dédiés à l’identification de personnes antisociales et dépressives, (iii) le corpus de témoignage extrait à partir du groupe de Facebook Borderline. L’hypothèse nulle est généralement rejetée lorsque la métrique P-valeur est inférieure à 0,05.

Le choix de ces corpus est justifié par leurs variétés en termes de sources et de langage de données. Ils abordent aussi plusieurs éléments liés à un trouble. Afin de mieux exploiter ces données, nous avons apporté quelques modifications

1. https://fr.wikipedia.org/wiki/Test_du_chi%C2%B2

TABLE 6.4 – Analyse du style d'écriture pour les différents corpus.

	Notre Corpus		Corpus As-tuti [2021]		Corpus Infamous-coder [2022a]		Corpus de témoignage	
	χ^2 test	P-value	χ^2 test	P-value	χ^2 test	P-value	χ^2 test	P-value
ponctuation : point-virgule	4.17	0.38	3.91	0.27	0.0	1.0	1.63	0.89
POS : conjonction, pronom, déterminant et préposition	104.66	0.04	81.66	2.32e-05	70.16	0.99	4.46	0.99
ponctuation : points d'exclamation et d'interrogation	21.74	0.17	10.70	0.05	5.00	0.54	7.31	0.83
POS : singulier & entité nommée	66.03	0.05	102.37	0.5e-09	121.79	0.99	121.49	0.99
POS : adjectif & adverbe catégories grammaticales	99.51	0.01	57.1	0.00045	79.13	0.96	18.72	0.99
analyse des sentiments : sentiment négatif	4.71	0.58	0.54	0.46	200.29	1.00	17.78	0.99
relation sémantique : conséquence & explication	10.02	0.05	11.07	0.006	8.28	0.98	1.65	0.89
relation sémantique : liaison & addition	22.07	0.01	12.81	0.05	30.35	0.99	12.30	0.99
relation sémantique : opposition	4.86	0.08	0.07	0.78	15.75	0.94	0.79	0.67

sur ces corpus. En effet, nous avons enrichi le corpus de témoignage de Facebook avec d'autres statuts partagés par des personnes ayant un comportement stable. L'objectif de cet enrichissement est de pouvoir mesurer la différence entre les spécificités des deux profils en termes de styles d'écriture. Pendant cette partie, nous nous concentrons uniquement sur la validation de notre corpus développé au cours du troisième chapitre puisque cette partie a plus de sensibilité par rapport aux autres contributions. En effet, la détection de la maladie diffère de celle du sujet ou de l'humeur en termes de complexité et d'intérêt. Finalement, nous avons clôturé la tâche de validation du corpus développé par une comparaison de nos résultats avec les hypothèses mentionnées par l'association américaine de parole, de langage et d'audition (ASHA)². Les hypothèses qui nous intéressent

2. <https://www.asha.org/Practice-Portal/Clinical-Topics/Written-Language-Disorders/Signs-and-Symptoms-of-Written-Language-Disorders/>

dans notre travail portait essentiellement sur les spécificités relatives aux opérations d'écriture et de compréhension du malade. Le tableau ci-dessous 6.5 montre la validation des hypothèses d'ASHA par rapport aux données des corpus sélectionnés. La mesure de cette validation consiste à évaluer la fréquence d'apparition de chaque élément lié au style d'écriture en fonction de la présence du trouble.

TABLE 6.5 – Comparaison des résultats de notre analyse du style d'écriture avec les résultats d'ASHA.

ASHA	Correspondance au niveau style d'écriture	Notre corpus	Corpus de Astuti [2021]	Corpus de Infa-mous-coder [2022a]	Corpus de témoignage
Tendance à utiliser des fragments de phrase courts et une structure de phrase moins complexe.	Le taux d'utilisation de la ponctuation point-virgule est notable.	✓	✓	✗	✗
Le discours explicatif est peu employé.	Le recours à des relations sémantiques pour illustrer les explications est réduit.	✓	✓	✓	✓
Le discours narratif n'est pas bien organisé; il a du mal à concrétiser ses idées.	L'utilisation de relations sémantiques pour mettre en évidence les connexions et les conséquences est réduite.	✓	✓	✗	✓
Incapable d'écrire à partir de différents points de vue.	Réduction du taux de recours à la relation d'opposition.	✓	✗	✗	✓
Souffre de difficultés morphologiques rigides, ce qui influence négativement sur la structure de la phrase	Le déterminant POS, la conjonction et la préposition sont moins utilisés.	✓	✓	✗	✓
Difficulté à comprendre et à interpréter des textes écrits par autrui.	Les points d'exclamation et les points d'interrogation sont fréquemment employés, tout comme les phrases avec un ton négatif.	✓	✓	✗	✓
Utilisation d'un langage moins abstrait.	Le recours au singulier plutôt qu'au pluriel est assez impressionnant. Utilisation fréquente des adjectifs et des entités nommées.	✗	✗	✓	✓

Généralement, une personne souffrant de TP admet un style de pensée pessimiste (rejet, les gens vont l'agresser, etc.). Pour cette raison, nous notons qu'une personne malade n'utilise pas régulièrement l'opposition comme un rapport sémantique. De plus, son sentiment est généralement négatif, ce qui peut engendrer une situation d'incompréhension de ce malade vis-à-vis d'autrui. Ce choix peut également justifier l'usage fréquent de points d'exclamation et d'interrogation. Le manque d'empathie, l'impulsivité et l'agressivité du malade ont affecté de manière considérable son style d'écriture. Cela peut se manifester par le non-respect des critères morphologiques ainsi par l'utilisation insuffisante des relations sémantiques comme l'explication. Il convient de noter que le style d'écriture peut varier en fonction du type de trouble. En effet, une personne obsessionnelle compulsive tente généralement de donner une réponse détaillée afin de décrire correctement la situation et les conséquences. Par contre une personne évitante n'a pas généralement l'audace de donner une réponse détaillée.

6.5 DÉTECTION DES PERSONNES SOUFFRANT DES TROUBLES DE LA PERSONNALITÉ SUR LES RÉSEAUX SOCIAUX

Nous présentons à ce niveau les résultats de nos méthodes dédiées à la tâche de détection des personnes atteintes de TP au niveau des réseaux sociaux. Après la présentation des résultats de ces méthodes, nous avons procédé à les évaluer afin de montrer la performance d'une méthode par rapport aux autres. Finalement, nous clôturons cette section par une discussion à propos des résultats obtenus afin de mettre en valeur les éléments pertinents.

6.5.1 Résultats des méthodes de détection des troubles

L'objectif principal de notre première contribution est d'élaborer une méthode permettant de détecter les personnes ayant des TP à partir de leurs productions textuelles sur les réseaux sociaux. Nous nous rappelons qu'au niveau de cette partie nous avons appliqué trois méthodes basées sur des différentes techniques permettant de mieux gérer les données obtenus. La première méthode repose sur la notion d'apprentissage profond qui comporte une étape d'extraction automatique des attributs à partir des données brutes. Cette méthode est réalisée en appliquant la transformation des données textuelles en des données numériques, tout en conservant le contexte du discours. La contribution majeure de cette méthode réside dans sa capacité à détecter la présence de la maladie en évaluant les niveaux de symptômes qu'elle admet. Ce passage permet d'apporter des réponses plus détaillées et mieux justifiées aux décisions qui ont été prises. Le tableau 6.6 présente un extrait de nos résultats pour la détection du degré de chaque symptôme de la maladie paranoïde.

La deuxième méthode de cette contribution consiste à réinventer les mêmes concepts que la première en ce qui concerne le traitement des caractéristiques linguistiques portant sur la transformation et l'extraction des descripteurs à partir des données textuelles. La différence réside dans la détection directe du trouble sans passer par la détection des symptômes. La majeure contribution de cette méthode est qu'elle permet de détecter initialement le trouble, puis de procéder à

TABLE 6.6 – Exemple de détection de degrés de symptômes liés à la maladie paranoïaque.

Tweets de l'utilisateur	Douter de l'engagement	Impitoyable et garder rancune	Isolement
1. Il y a des gens vous êtes vrmt pourri jusqu'à l'âme. 2. Imagine t'aimes pas provoquer? quelle vie fade. 3. Depuis que je suis ma seule priorité jsuis beaucoup plus heureuse, ça paye pas de faire passer les autres avant soi. 4. Cette nuit j'aurais préféré ne pas rêver. 5. Y'a d comportement que jsupporte plus. 6. C'est quoi cette nouvelle mode des meufs de se vanter de parler à 1939101 mecs c'est pas une fierté et ça le sera jamais. 7. Maintenant être fidèle c'est devenue une qualité alors que ça devrait être normal. 8. Y'en a vous êtes culotté c'est incroyable. 9. T'as qu'a imaginer que c'était un rêve et que tous ce qu'on a vécu c'était pas vrai. 10. L'ingratitude est la pire des choses sah tu donnes tout à des gens qui se foutent de ta gueule. 11. Rencontrer quelqu'un avec le même état d'esprit que vous est rare. 12. Le dégoût sa aide beaucoup a oublier. 13. Arrête de croire que tout le monde te considère comme tu les considérer! 14. C'était des grandes paroles en l'air.	3	2	2

la détection des catégories de trouble et finalement à la détection des types de trouble, tout en conservant l'aspect de l'étiquetage multiple. Le choix d'intégrer le concept d'étiquetage multiple est justifié par le fait qu'en général, une personne peut souffrir de plus d'un trouble en même temps. En outre, à cause de l'existence de certaines similitudes entre les symptômes de ces troubles, et en analysant seulement 20 tweets sans discuter avec le patient, il n'est pas possible de distinguer de manière définitive le recoupement entre ces troubles. Le tableau 6.8 ci-dessous illustre un exemple de chevauchement identifié par notre méthode lors de la détection du trouble, sa catégorie et son type. Pour appliquer cette méthode, nous avons utilisé trois couches de convolution avec 320 caractéristiques et une fonction d'activation (AF) ReLU pour chaque couche, suivies de trois couches de mise en commun (pooling layers) avec une taille de (1,9) pour chaque couche. Ensuite, nous avons utilisé deux couches LSTM composées de 250 neurones pour la première couche et de 100 neurones pour la deuxième couche, associées à une couche cachée avec une fonction d'activation *softmax*. La sortie de cette couche est associée à la dernière couche de l'architecture composée de 2 neurones (représentant la présence ou l'absence de cette maladie). Le modèle d'entrée et de sortie du CNN avec les paramètres utilisés sont présentés dans la figure 6.1.

La table 6.7 présente certaines techniques qui peuvent être appliquées dans le but d'améliorer et d'optimiser le modèle.

Notre troisième méthode proposée dans le cadre de la détection du trouble

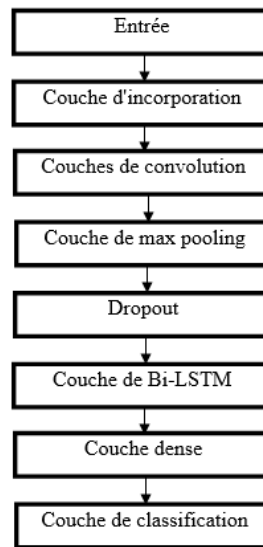


FIGURE 6.1 – Architecture du modèle proposé.

repose essentiellement sur le concept d'apprentissage classique en intégrant la notion de logique floue. Cette notion peut fournir d'autres concepts tels que le degré de confiance de l'algorithme pour chacune des réponses retournées. La différence entre cette méthode et les méthodes basées sur le concept d'apprentissage profond réside au niveau de la manière de sélection des attributs. Dans cette partie, nous avons fait le choix de sélectionner des attributs personnalisés, ce qui permet de mieux évaluer l'apport de chaque type d'attributs. Le tableau 6.9 présente un exemple de résultat de la tâche de classification des tweets en fonction du degré de présence du trouble paranoïaque.

Après la présentation de nos résultats concernant les diverses méthodes permettant l'identification des personnes ayant des TP. Nous nous intéresserons dans la prochaine phase à l'évaluation des résultats obtenus par chacune des méthodes.

6.5.2 Évaluation des méthodes de détection des troubles

L'évaluation des méthodes proposées dans notre première contribution consiste à comparer le résultat fourni par chacune des méthodes avec le résultat de l'annotation correspondante. Cette évaluation est réalisée à l'aide des métriques d'évaluation classiques : rappel, précision, F-mesure, accuracy qui permettent de mettre en évidence les instances faussement estimées (bruit) et les instances qui n'ont pas été estimées (silence). Ces mesures peuvent être extraites de la matrice de confusion qui met en valeur le nombre de vrais positifs, de faux négatifs, de faux positifs et de vrais négatifs. L'évaluation de la première méthode proposée comporte deux parties : l'évaluation de la classification des symptômes et l'évaluation de la classification des maladies. Pour la classification des degrés de symptômes, nous avons appliqué les métriques mentionnées aux résultats de la première couche de sortie de notre modèle qui reflète le degré de chaque symptôme de la maladie paranoïaque. En ce qui concerne l'évaluation de la classification de la maladie paranoïaque, nous avons appliqué les mêmes mesures mentionnées sur les résultats de la dernière couche de notre modèle.

TABLE 6.7 – Comparaison des méthodes d'apprentissage profond *Mathew et al. [2020]*.

Méthode	Description	Avantages	Inconvénient
Backpropagation	Utiliser dans le problème d'optimisation	Pour le calcul du gradient	Sensible aux données bruyantes
Stochastic gradient descent	Trouve le minimum optimal dans les problèmes d'optimisation	Évite le piégeage dans le minimum local	Temps de convergence plus long, calcul coûteux
Learning rate decay	Réduire progressivement le taux d'apprentissage	Augmente les performances et réduit le temps d'entraînement	Coûteux en termes de calcul
Dropout	Supprime les unités/connexions pendant l'apprentissage	Évite le surajustement	Augmente le nombre d'itérations nécessaires pour converger
Max pooling	Applique un filtre max	Réduit les dimensions et les coûts de calcul	Ne considère que l'élément maximum qui peut conduire à un résultat inacceptable dans certains cas
Batch normalization	Normalisation par lots de l'entrée à une couche	Réduit le décalage covariant, augmente la stabilité du réseau, le réseau s'entraîne plus rapidement, permet des taux d'apprentissage plus élevés	Frais généraux de calcul pendant l'apprentissage
Skip-gram	Utilisé dans les algorithmes d'intégration de mots	Peut travailler sur n'importe quel texte brut et nécessite moins de mémoire	La fonction Softmax est coûteuse en calcul et le temps d'apprentissage est élevé.
Transfer learning	La connaissance du premier modèle est transférée au deuxième problème	Améliore les performances, progression rapide dans la formation du deuxième problème	Fonctionne uniquement avec des problèmes similaires

Il convient de noter que l'évaluation de la classification de la maladie englobe également l'évaluation des résultats de détection des degrés de symptômes. Les deux tableaux 6.10 et 6.11 offrent une évaluation plus détaillée concernant les résultats de chaque partie.

Pour réaliser l'évaluation de la deuxième méthode qui consiste à détecter la présence d'un TP, les catégories et les types associés en utilisant la même architecture de la première méthode mais sans passer par la détection des symptômes. Nous avons eu recours aux mêmes métriques d'évaluation de la première méthode afin de les appliquer pour tester la performance de chaque type de classification (trouble de la personnalité, catégorie, type). Les différents

TABLE 6.8 – Résultats de la détection du trouble, des catégories et des types de troubles.

Extrait de l'historique des tweets d'un utilisateur	Troubles de la personnalité	Catégorie	Type
<p>1. Du grand délire! Une hystérie incommensurable.</p> <p>2. Le Professeur Raoult confirme la fin de l'épidémie sur Radio Classique et règle ses comptes avec les sorciers prévisionnistes de la catastrophe.</p> <p>3. Ils sont payés au PV, dommage, les FDO perdent toute crédibilité!</p> <p>4. Second vague couplée à d'autres nouvelles pandémies donc confinement à perpétuité jusqu'à se que les populations ne se rappellent plus ce que signifie les mots liberté et contestation ...</p> <p>5. Très vindicative Ruth Elkrief envers Michel Onfray. Insupportables ces journalistes serviteurs du gouvernement.</p> <p>6. Nous avons beaucoup de variétés de penis de chiens.</p> <p>7. Attention : âmes sensibles, ne regardez pas! Resto chinois, suite et fin (2/2) Question bonus : Qu'avons-nous de commun avec cette culture ?</p> <p>8. Aujourd'hui, pour les poulets, c'est un peu comme l'ouverture des soldes, faut remplir un maximum avant minuit, alors ils sont en mode racket intensif. Ils doivent toucher une prime pour être d'aussi "bons serviteurs de l'état".</p>	Oui	Groupe A : Distant + Groupe B : Impulsif	Paranoïaque + Borderline

tables 6.12, 6.13, 6.14, 6.15, 6.16 , 6.17 présentent en détail l'évaluation des différents niveaux de la méthode de détection des maladies en fonction des diverses techniques de plongement de phrases. Dans les différentes parties d'évaluation de notre méthode, nous avons appliqué la technique de validation croisée (*cross-validation*).

Le tableau 6.18 présente une analyse détaillée de la variation de la performance entre les deux méthodes proposées pour la détection du troubles paranoïaque : détecter le trouble directement sans passer par la détection des symptômes et détecter le trouble en passant par la détection des symptômes.

Les mêmes mesures d'évaluation (rappel, précision et f-mesure) ont été appliquées lors de l'évaluation de la troisième proposition visant à détecter la présence d'un type de TP à l'aide d'une méthode basée sur l'extraction des attributs personnalisés. Cette méthode permet également d'intégrer le concept de logique floue au fonctionnement d'un algorithme d'apprentissage classique. Notre objectif dans ce contexte consiste à : (i) mesurer l'influence de chaque type d'attribut sur la personnalité, (ii) évaluer l'importance de la logique floue dans le contexte médical tel

TABLE 6.9 – Résultats de classification du trouble paranoïaque en se basant sur la notion de logique floue.

Extrait des Tweets	Paranoïaque	Non Paranoïaque
1. Toutes ces choses que le rail cache et ne prend pas en compte pour tenter de se donner une image vertueuse pour tenter de masquer ce qu'il coûte réellement aux français;) Au fait ces motrices et wagons Bombardier, elles arrivent d'où 2. C'est exactement ce que je dis... Du coup avec une population si faible, quel phénomène a bien pu faire monter aussi vite le niveau des mers, quel phénomène à fait brutalement monter les températures ? Des voitures diesels ? Des avions ? 3. Le taux de co2 commence à être nocif à 8%.... on est loin avec 0,0004% non ? Toutes les ressources carbonés de la Terre n'y suffirait pas. Allez parlez science pour de vrai, avec des chiffres. 4. De Janvier à Mars 2012 ce sont des rebels qui clashen le NTC à Benghazi. En Sept 2012 ce sont bien les factions islamiques qui tuent l'ambassadeur Américain et 3 autres citoyens...	0.67	0.33

que la classification fondée sur le degré de confiance. Le tableau 6.19 présente un aperçu détaillé de l'influence de chaque type de combinaison de caractéristiques sur les résultats. Dans le tableau 6.20, nous comparons les résultats obtenus par à notre méthode avec les résultats de l'algorithme de base de KNN.

Dans le tableau 6.21, nous présentons les résultats obtenus par notre méthode ainsi que par d'autres méthodes proposées dans le contexte de la détection de troubles psychologiques. Ces méthodes ont été appliquées aux différents corpus de références élaborés pour la détection des troubles sur les réseaux sociaux. Dans le tableau 6.22, nous présentons une comparaison entre les résultats des différentes méthodes proposées pour la détection du trouble paranoïaque au niveau de notre première contribution. Dans l'ensemble, nos résultats sont satisfaisants par rapport aux autres travaux du même contexte [Saha et al. \[2021\]](#), [Sampath et Durairaj \[2022\]](#).

6.5.3 Discussion des méthodes de détection des troubles

Après la présentation de nos résultats pour les différentes méthodes visant à identifier les personnes atteintes de TP sur Twitter. Nous procédons à présenter une discussion expliquant l'influence des avantages de chaque méthode sur les résultats obtenus. En général, les méthodes proposées dans cette contribution prennent en considération un ensemble important de critères. En effet, elles traitent la persistance du symptôme au cours du temps en suivant le fil d'actualité, ainsi elles abordent la combinaison de symptômes pour inférer qu'il y a vraiment un problème psychologique. C'est-à-dire, il ne s'agit pas d'une déduction qu'une personne est déprimée ou non en analysant un seul tweet. Pour atteindre notre objectif, nous avons pris en compte plusieurs catégories d'informations, en parti-

TABLE 6.10 – Comparaison des résultats obtenus pour la classification des symptômes liés à la maladie paranoïaque à l'aide du modèle CNN+BiLSTM.

Symptômes	Rappel (%)	Précision (%)	F-mesure (%)
Agressivité	80	73	76
Perçoit des attaques	68	68	68
Souçons récurrents	61	59	60
Isolement	86	77	81
Croire qu'il a toujours raison	85	78	81
Lire les sens cachés dans les remarques innocentes	69	65	67
Mauvaises relations avec les autres	58	56	57
Douter de l'engagement	72	66	69
Impitoyables et rancuniers	80	75	77

TABLE 6.11 – Comparaison des résultats obtenus pour la détection de la maladie paranoïaque sur la base des symptômes.

	Rappel (%)	Précision (%)	F-mesure (%)
Softmax	60	61	60
Boosting de gradient ³	60	57	58
KNN	62	62	62
Adaptive boosting ⁴	65	60	62
Forêts aléatoires ⁵	61	60	60
SVM	66	64	65

culier celles liées aux aspects linguistiques comme le contexte de la phrase. Cela est réalisé en employant des techniques de représentation sémantique des phrases comme celle de plongement des phrases. De même, notre méthode tient compte de l'aspect des étiquettes multiples au niveau des classes de classification ainsi que le traitement des données en temps réel. Nous avons également traité les problèmes de déséquilibre entre les instances de chaque classe du corpus à l'aide d'une technique de génération de données. Les résultats de nos méthodes comportent plus de détails par rapport aux méthodes proposées dans la littérature, ceci est assuré par la détection des symptômes ou par une détection hiérarchisée du TP. De même, ils sont plus précis grâce à la notion de logique floue. En ce qui concerne les résultats de classification des personnes paranoïaques, nous notons que nous avons obtenu les meilleurs résultats avec l'algorithme KNN, ce qui se justifie par le fait que nos instances sont très proches. Pour les résultats obtenus au niveau des tables 6.19 et 6.20, nous remarquons que les meilleures caractéristiques qui ont une influence sur la personnalité sont celles liées aux métadonnées et au type linguistique. Il est possible que cela soit justifié en se basant sur certaines hypothèses, par exemple considérer qu'une personne est antisociale si elle n'a pas beaucoup d'amis. Autre hypothèse dans le même contexte, un utilisateur peut être décrit comme évitant, s'il ne publie pas fréquemment. Pour l'intérêt des attributs de type linguistique, cela est justifié par les tableaux 6.4 et 6.5 qui montrent l'effet du style d'écriture sur la personnalité. Notre méthode proposée dans ce contexte a démontré une amélioration des performances de l'algorithme KNN de base. Cette amélioration est due au fait que notre méthode est plus paramétrable et flexible au niveau du traitement. Pour l'analyse des résultats des trois méthodes proposées,

TABLE 6.12 – Comparaison des résultats obtenus pour la classification des TP.

	BERT	LASER	SISTER	USE	Doc2Vect
BILSTM	80	81	81	81	87
LSTM	78	82	86	85	84
LSTM+SVM	83	85	81	76	84
BILSTM+SVM	84	81	85	85	84

TABLE 6.13 – Comparaison des résultats obtenus pour la classification des TP selon la catégorie suspecte.

	BERT	LASER	SISTER	USE	Doc2Vect
BILSTM	74	65	66	67	51
LSTM	77	64	68	59	48
LSTM+SVM	72	71	65	62	58
BILSTM+SVM	77	76	63	76	51

nous avons obtenu les meilleurs résultats de classification en utilisant la méthode d'extraction des attributs personnalisés. Cependant, il convient de noter que cette méthode n'est pas appliquée en temps réel et qu'elle traite plusieurs types de renseignements, tandis que nos méthodes d'apprentissage profond ne traitent que du texte. Cela vaut dire que si nous effectuons une comparaison sur la base des attributs textuels, les méthodes d'apprentissage profond excelleront par rapport aux techniques d'apprentissage classiques. En ce qui concerne l'évaluation des résultats des méthodes basées sur la notion d'apprentissage profond, nous constatons que la méthode qui consiste à détecter la maladie à travers les symptômes a montré plus de performance par rapport à la méthode de base. D'après le tableau 6.10, les résultats les plus satisfaisants pour la classification des symptômes (F-mesure égale à 81%) sont les suivants : *croyance qu'ils ont toujours raison* et *isolement*. En revanche, les résultats les plus médiocres sont obtenus au niveau de la classification du symptôme *mauvaises relations avec les autres* (F-mesure égale à 57%) en dépit qu'il admet plus de données par rapport au symptôme *croyance qu'ils ont toujours raison*. Cette différence peut s'expliquer par les facteurs suivants : (i) il existe de nombreux phénomènes linguistiques comme l'ironie et la négation qui sont plus liés à un symptôme par rapport à un autre, (ii) il existe plusieurs niveaux de gravité d'un symptôme par rapport à un autre. Par exemple, dans le cas d'une personne évitante, il est très difficile d'avoir cinq niveaux différents de 0 à 5, (iii) il existe plusieurs obstacles liés à l'utilisation des caractéristiques qui représentent un texte, tels que le lexique général, la possibilité d'exprimer une idée de plusieurs façons, l'absence de contexte, etc.

6.6 IDENTIFICATION DES CAUSES AFFECTANT LA MALADIE DE TP

Après avoir présenté les expérimentations et les résultats de la méthode de détection des personnes ayant des TP sur les réseaux sociaux. Nous visons à présenter à ce niveau les expérimentations et les résultats de la méthode dédiée à l'identification des causes affectant ces troubles. Nous visons également à mener une discussion spécifique sur les différents résultats obtenus.

TABLE 6.14 – Comparaison des résultats obtenus pour la classification des TP selon la catégorie émotionnelle.

	BERT	LASER	SISTER	USE	Doc2Vect
BILSTM	50	61	46	56	48
LSTM	77	64	68	85	59
LSTM+SVM	60	55	57	53	60
BILSTM+SVM	57	62	61	52	52

TABLE 6.15 – Comparaison des résultats obtenus pour la classification des TP selon la catégorie anxieuse.

	BERT	LASER	SISTER	USE	Doc2Vect
BILSTM	57	46	61	59	51
LSTM	55	53	55	52	51
LSTM+SVM	55	46	58	50	58
BILSTM+SVM	56	61	56	46	60

6.6.1 Résultats de la méthode d’identification des causes du TP

L’objectif principal de notre deuxième contribution est de développer une méthode capable à détecter les causes qui affectent des troubles de la personnalité. Nous nous rappelons que nous avons considéré une cause est le sujet du tweet qui a été écrit par la personne malade avec une mauvaise humeur. Pour ce faire, nous avons proposé une méthode hybride qui combine à la fois des techniques linguistiques et des techniques numériques basées sur l’apprentissage non supervisé et les mesures statistiques. Le résultat de la première partie qui consiste à appliquer des techniques d’apprentissage non supervisé d’une manière récursive est une arborescence qui hiérarchise le sujet et ses branches. La figure ci-dessous 6.2 présente le résultat de l’arborescente élaborée.

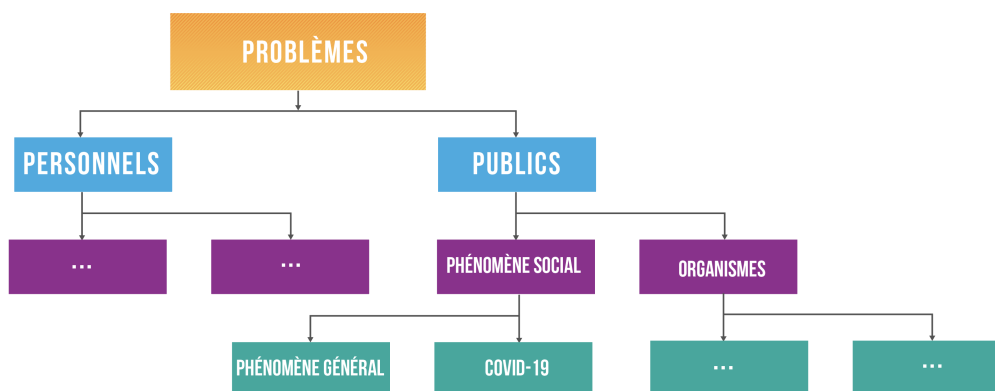


FIGURE 6.2 – Extrait du résultat de l’application récursive de notre méthode d’apprentissage automatique non supervisée.

Le deuxième résultat de cette méthode est sous la forme d’une base de connaissances (une ontologie) qui permet de présenter les lexiques liés à chaque sujet

TABLE 6.16 – Comparaison des résultats obtenus pour la détection des types de TP à l'aide du modèle LSTM+SVM.

	BERT	LASER	SISTER	USE	Doc2Vect
Antisocial	55	49	54	52	52
Borderline	74	77	64	69	60
Compulsif	62	66	56	62	55
Dépendant	76	67	61	65	58
Évitant	75	74	71	73	80
Histrionique	55	52	48	58	48
Narcissique	55	70	51	61	57
Paranoïaque	56	62	51	61	56
Schizoïde	55	59	71	61	57
Schizotypique	56	58	57	50	50

TABLE 6.17 – Comparaison des résultats obtenus pour la détection des types de TP à l'aide du modèle BiLSTM+SVM.

	BERT	LASER	SISTER	USE	Doc2Vect
Antisocial	59	54	57	57	58
Borderline	76	79	74	77	83
Compulsif	66	72	67	85	61
Dépendant	65	66	72	86	67
Évitant	73	81	79	88	71
Histrionique	60	62	52	54	60
Narcissique	60	74	76	66	53
Paranoïaque	60	69	56	68	61
Schizoïde	70	66	68	75	65
Schizotypique	53	67	54	58	52

TABLE 6.18 – Comparaison des résultats obtenus pour la détection du trouble paranoïaque (sans ou avec le passage par la détection des symptômes).

	Rappel (%)	Précision (%)	F-mesure (%)
Méthode de base (sans passer par la détection des symptômes)	63	57	60
Notre proposition (en passant par la détection des symptômes)	66	64	65
Amélioration	3	7	5

d'étude. Le choix des différents sujets d'étude a été effectué selon les résultats des feuilles de l'arborescente développée. De même, cette base de connaissances permet de montrer les relations sémantiques et les valeurs possibles de chacun des concepts. La figure 6.3 illustre un extrait du résultat de la construction de l'ontologie pour le sujet du COVID-19. Cette ontologie peut d'abord être utilisée pour détecter les sujets de tweet en comparant le lexique d'un tweet et le lexique de l'ontologie. Ensuite elle peut être utilisée pour l'analyse de l'humeur exprimée au niveau du tweet. En effet, l'apport de l'ontologie construite par notre méthode est qu'elle tient compte du lexique de l'humeur. En effet, chaque concept de notre ontologie est associé par deux listes : (i) une liste lexicale qui reflète une attitude

TABLE 6.19 – Comparaison des résultats obtenus pour la détection du trouble paranoïaque en se basant sur la combinaison des type d’attributs.

Attributs	XG Boost ⁶ (%)	Forêts aléatoires ⁷ (%)	Gaussien de Naïve Bayes ⁸ (%)	SVM (%)	Boosting adaptatif ⁹ (%)	KNN (%)
Tous types (831 attributs)	59	65	67	65	59	69
Texte (513 attributs) + Méta-données (9 attributs)	50	54	54	45	64	65
Texte (513 attributs) + Linguistique (304 attributs)	60	67	45	39	51	59
Texte (513 attributs) + Temporel (5 attributs)	52	67	51	57	56	52
Texte (513 attributs)	49	60	51	54	61	56

TABLE 6.20 – Comparaison des résultats obtenus par notre méthode et la méthode baseline pour la classification des personnes paranoïaques.

Attributs	KNN de base (%)	Notre méthode (%)	Amélioration (%)
Tous types d’attributs	69	74	5
Texte+Metadonnées	65	69	4
Texte+Linguistique	59	71	12
Texte+Temps	52	60	8
Texte	56	59	3

TABLE 6.21 – Évaluation de notre méthode par rapport à d’autres méthodes proposées pour la détection des personnes ayant des TP.

	Infamouscoder [2022b]	Komati [2021]	Namdari [2023]
Saha et al. [2021]	68	71.5	80
Stankevich et al. [2020]	68	73	79
AlSagri et Ykhlef [2020]	64	84	85
Notre méthode	93	79	85

TABLE 6.22 – Évaluation des diverses méthodes proposées pour la détection du trouble paranoïaque.

	Rappel (%)	Précision (%)	F-mesure (%)
Méthode de base (sans passer par la détection des symptômes)	63	57	60
Apprentissage profond (en passant par la détection des symptômes)	66	64	65
Apprentissage avec logique floue	75	74	74

positive ou neutre, (ii) une deuxième liste lexicale dont l'apparition de chacun de ses mots avec ce concept dénote une mauvaise humeur.

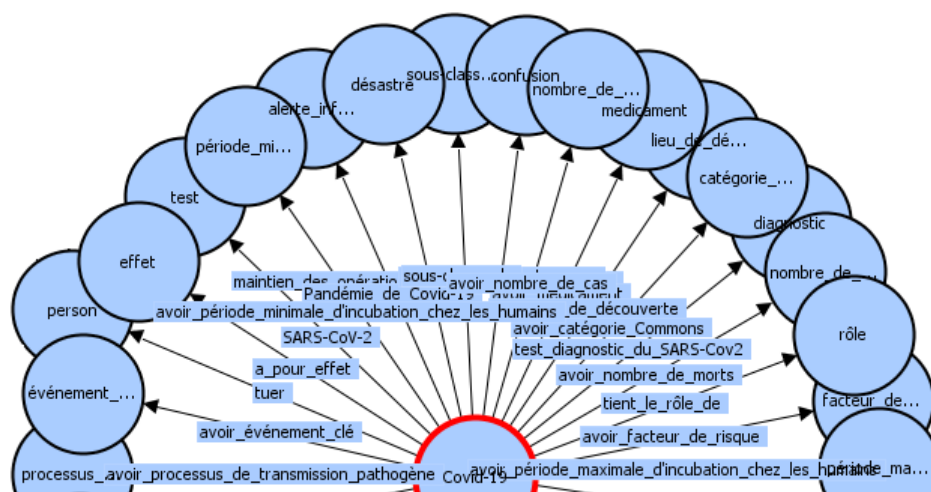


FIGURE 6.3 – Notre ontologie élaborée pour le sujet Covid-19.

6.6.2 Évaluation de la méthode d'identification des causes du TP

L'évaluation de notre méthode d'identification des causes du TP consiste d'abord à évaluer le résultat de chaque regroupement réalisé par la partie de l'apprentissage non supervisé. Cette évaluation est assurée en mesurant la similitude sémantique entre les différents groupes en fonction de la répartition des sujets qui ont été jugés par notre expert. La deuxième évaluation est réalisée en évaluant les règles de SWRL construites, c'est-à-dire que pour chaque tweet nous exécutons les règles qui consistent à détecter le sujet. Au niveau de cette partie, nous sommes limités au sujet du COVID-19. Par conséquent, chaque fois que nous appliquons ces règles, c'est dans le but de déterminer si le sujet de tweet concerne le sujet COVID-19 ou non. Dans un deuxième temps, nous nous sommes focalisés sur la détection de l'humeur exprimée au niveau du tweet. Pour ce faire, nous avons appliqué les règles SWRL spécifiquement conçues à cet effet. Après cela, nous évaluons les règles obtenues en comparant les résultats de ces règles avec les résultats de l'annotation au niveau de l'humeur. Nous évaluons la performance de notre méthode en utilisant deux méthodes; la première méthode repose sur l'évaluation opérationnelle. Dans ce contexte, la performance de notre système est mesurée à l'aide de trois métriques d'évaluation standard : rappel, précision et f-mesure. Le tableau 6.23 et 6.24 présentent les diverses valeurs obtenues à chaque phase du processus de travail de notre deuxième contribution selon une méthode d'évaluation opérationnelle (avec R : Rappel, P : Précision, F : F-mesure).

La deuxième méthode d'évaluation repose sur le principe de l'évaluation basée sur l'expertise humaine. Après avoir construit notre ontologie, nous avons demandé à notre expert de l'évaluer sur la base des critères de Gruber [Rector et al. \[2019\]](#). Notre expert a affirmé que notre ontologie est claire puisqu'elle est représentée par des termes issus d'un domaine d'étude réel et reconnu. De même que notre ontologie est extensible par le fait qu'elle est prête à toute sorte de modification et d'extension. En outre, l'engagement minimal de l'ontologie (*minimal*

TABLE 6.23 – Évaluation de la tâche d'apprentissage récursif non supervisée.

	1er regroupement			2e regroupement			3e regroupement		
	R	P	F	R	P	F	R	P	F
*) K-moyennes	71	72	71	58	63	60	75	86	80
*) Classification Hiérarchique sur Composantes	69	71	70	66	66	66	74	86	80
*) Partitionnement spectral	57	56	56	48	46	47	52	81	66

TABLE 6.24 – Évaluation de la tâche d'identification du sujet et d'humeur.

	Rappel	Précision	F-mesure
*) Détection du sujet de Covid-19	81	65	73
*) Détection de l'humeur des gens face à la Covid-19	97	88	92

ontological commitment) est assuré par la présence de la plupart des termes couvrant le domaine étudié. De même que la cohérence et la déformation minimale de l'encodage de notre ontologie sont également garanties par l'ensemble des liens sémantiques acquis. L'ensemble de ces critères assure la possibilité de partager la base de connaissances relatives à ce domaine.

6.6.3 Discussion de la méthode d'identification des causes du TP

Notre méthode d'identification des causes du TP proposée s'appuie sur un concept hybride qui nous permet de bénéficier de tous les avantages des différentes méthodes numériques et linguistiques, tout en évitant leurs inconvénients respectifs. Parmi les inconvénients de la méthode numérique nous citons l'absence de l'explicabilité et de l'interprétabilité au niveau des résultats obtenus. Pour ce qui concerne les inconvénients de la méthode linguistique nous citons l'absence des règles traitant tous les cas possibles. Les résultats de notre méthode montrent qu'à un certain niveau ces problèmes ont été traités. L'apport de notre méthode s'applique aussi au niveau du traitement de la spécificité linguistique : (1) au niveau lexical, c'est-à-dire je peux poster un tweet qui porte sur le sujet de Covid-19 sans citer ce nom, (2) sur le plan flou, dans le cas où un mot peut appartenir à plusieurs sujets en même temps. Comme le mot crise économique, ce terme peut apparaître dans l'ontologie de covid ainsi que dans une autre ontologie économique ou sociale, (3) au niveau contextuel, par exemple le mot content : (i) si ce mot apparaît dans le contexte où l'utilisateur est content par le décroissement au niveau du taux de cette maladie. Notre méthode considère que cette personne se comporte bien face à cette épidémie, (ii) si ce terme apparaît dans un contexte de haine et de rancœur, notre méthode considère que la personne qui l'utilise a un comportement inapproprié.

Pour l'interprétation des résultats obtenus au niveau du regroupement récursif, nous remarquons qu'aucun algorithme n'a excellé par rapport aux autres. Mais les algorithmes k-moyennes et CAH ont donné des résultats acceptables en comparaison avec les autres. Pour l'application des règles, nous avons obtenu les meilleurs résultats en appliquant les règles pour identifier l'humeur, car il y a moins de chevauchements dans le lexique utilisé pour l'exprimer. En revanche, dans la détection du sujet COVID-19, certains termes apparaissent dans plus d'un

groupe, comme le terme crise économique. Nous pouvons accepter cette différence étant donné que, dans notre cas, le processus de déclenchement de la surveillance dépendra principalement du résultat de la détection de l'humeur. Pour le sujet détecté, ceci est considéré comme un enrichissement du résultat. Ce qui peut nous offrir plus de flexibilité et de précision lors de la surveillance.

6.7 SURVEILLANCE DES PERSONNES ATTEINTES DU TP SUR LES RÉSEAUX SOCIAUX : ÉTUDE DE CAS COVID-19

Notre objectif dans la section de surveillance des personnes atteintes du TP est d'exposer les résultats et les évaluations de la méthode présentée dans le cinquième chapitre, qui consiste à surveiller les utilisateurs des réseaux sociaux. Nous visons également à présenter une partie de discussion afin d'expliquer les résultats obtenus et les points forts de notre méthode proposée.

6.7.1 Résultats de la méthode de surveillance des cas des personnes ayant un trouble

Nous nous rappelons que notre objectif principal dans la troisième contribution est de proposer à Twitter un module qui permet d'assurer la surveillance des personnes ayant des TP. Cette proposition est réalisée en offrant à Twitter la possibilité de filtrer des tweets et de recommander d'autres en fonction de l'état du patient envers l'épidémie Covid-19.

Notre méthode proposée dans ce contexte (voir le cinquième chapitre), est basée principalement sur quatre parties. Le résultat de la première partie qui consiste à détecter les gens ayant des troubles est présenté dans la sous-section (6.5 détection des personnes souffrant des troubles de la personnalité sur les réseaux sociaux). Pour cette raison, nous entamons par la présentation de résultats de la partie dédiée à la détection de l'humeur de la personne malade envers l'épidémie COVID-19. L'objectif principal de cette méthode est de classer les tweets liés à l'épidémie, publiés par une personne malade, en trois classes distinctes. Ces classes visent principalement à donner un aperçu général du comportement du malade (méfiance exagérée/irresponsabilité totale/neutre). Le tableau 6.25 ci-dessous montre les résultats de cette partie en appliquant une méthode d'apprentissage supervisé basée sur la notion de logique floue sur quelques tweets.

TABLE 6.25 – Résultats de la partie de détection du comportement des personnes envers Covid-19.

Covid-19	Irresponsabilité totale	Méfiance exagérée	Neutre
[ALERTE VIRUS] Des chercheurs ont découvert une souche de virus de grippe porcine en Chine présentant toutes les caractéristiques capables de provoquer une future pandémie.	0	0.86	0.14
C'est le scoop du jour celle là! Tout le monde se doute sauf les politiciens! Ils ont vraiment tous fait l'école des clowns	0.46	0.44	0

Dans la suite de cette sous-section, nous présentons les résultats des différentes parties impliquées dans les tâches de filtrage et de recommandation des tweets liés à l'épidémie de COVID-19. Nous nous rappelons que ces tâches sont réalisées en se basant sur une méthode hybride combinant : (i) une méthode linguistique basée sur l'interrogation d'une ontologie à travers des requêtes, (ii) une méthode basée sur le concept d'apprentissage profond qui prend en entrée deux tweets et génère en sortie une classification parmi les trois suivantes : à recommander ou à filtrer ou neutre. Les résultats des deux méthodes de filtrage et de recommandation des tweets sont présentés dans les deux tableaux suivants 6.26 et 6.27, avec l'indication de la méthode responsable de chaque décision.

Pour l'élaboration de notre architecture d'apprentissage profond, nous nous sommes appuyés sur un modèle à deux entrées (2 tweets). Chaque entrée est transmise à un modèle d'auto-encodeur constitué de deux couches pour l'encodage et de deux couches pour le décodage afin de mieux extraire les attributs pertinents. La couche contenant les données compressées est ensuite transmise à deux couches LSTM comprenant 350 neurones pour la première couche et 150 neurones pour la deuxième couche. Chaque dernière couche est associée à une couche cachée avec : (i) la fonction *sigmoïde* comme une fonction d'activation, (ii) la fonction *Binary crossentropy* comme une fonction de coût, (iii) *Adam* comme un algorithme d'optimisation. La couche de sortie est constituée de deux neurones (représentant la classe de classification).

TABLE 6.26 – Résultats de la tâche de filtrage des tweets inappropriés en fonction des préoccupations du malade.

Tweet 1	Tweet 2	Décision concernant le filtrage des tweets
Pour ceux qui doutent encore de l'impact de la COVID-19, voici en vert la courbe de tous les décès enregistrés au Québec.	le nombre de morts du Covid-19 serait « probablement plus élevé » que le bilan officiel aux États-Unis.	Oui sur la base de l'ontologie
@PrésidentXX super , on prendra pas ton vaccin de merde , dégage de ton poste! Covid-19	Le taux de mortalité due au Covid-19 est minime Meme le WHO ne recommande pas le port du masque en dehors des hôpitaux.	Non sur la base de l'ontologie
Clairement un homme dans le métro sans masque se met à tousser fort et bien entendu pas dans le coude hein.	J'ai détesté ma journée, c'était full stressant, vraiment les gens respectent rien, ni le port du masque, ni la distance.	Oui sur la base du modèle d'apprentissage

6.7.2 Évaluation de la méthode de surveillance des cas des personnes ayant un trouble

L'évaluation de notre troisième contribution consiste à évaluer les quatre objectifs de cette méthode : (i) évaluer la méthode qui consiste à détecter les personnes ayant des problèmes psychologiques à partir de leurs productions textuelles sur les réseaux sociaux, (ii) évaluer la méthode de détection de l'humeur des patients à

TABLE 6.27 – Résultats de la tâche de recommandation des tweets appropriés en fonction des pré-occupations du malade.

Tweet 1	Tweet 2	Décision concernant la recommandation des tweets
Le nombre de morts du Covid-19 serait « probablement plus élevé » que le bilan officiel aux Etats-Unis.	Le taux de mortalité due au Covid-19 est minime Meme le WHO ne recommande pas le port du masque en dehors des hôpitaux.	Oui sur la base de l'ontologie
@PrésidentXX super , on prendra pas ton vaccin de merde , dégage de ton poste! Covid-19	Bravo, ce vaccin sera une protection efficace et indispensable pour tous les Européens. Une fois tout les Européens vaccinés, nous serons tous immunisés contre le virus, et notre avenir sera plus sûr.	Oui sur la base de l'ontologie
Je vais finir en dépression à cause de ce virus!; plus de foot, plus d'école, plus sortir!	N'aie pas peur continue à appliquer les gestes barrières, ça ira on va s'en sortir.	Oui sur la base du modèle d'apprentissage

l'égard de l'épidémie de COVID-19, en analysant leurs tweets contenant des mots qui font partie du lexique du COVID, (iii) évaluer la méthode d'analyse de la combinaison de deux tweets portant sur l'épidémie (un tweet partagé par la personne malade et un autre qui est considéré comme un tweet susceptible d'apparaître dans le fil d'actualité de la personne malade). Le but de cette méthode d'analyse est d'éliminer les tweets ennuyants pour la personne malade, (iv) évaluer la quatrième méthode qui consiste également à comparer la combinaison de deux tweets au sujet de Covid. Cependant, dans ce cas, pour déduire les tweets recommandés à être observés par le malade. Pour la première méthode (i), nous avons déjà évalué cette partie dans la section (6.5 Détection des personnes souffrant des troubles de la personnalité sur les réseaux sociaux). Pour la deuxième partie de notre méthode qui consiste à détecter le comportement de l'utilisateur malade envers Covid-19 en intégrant la notion de logique floue à un algorithme d'apprentissage supervisé. Nous avons appliqué les trois mesures statistiques (rappel, précision et F-mesure). Le tableau 6.28 présente les résultats obtenus par la mise en œuvre de quelques exemples d'algorithmes d'apprentissage automatique. Le tableau 6.29 met en évidence l'amélioration apportée par notre méthode par rapport à la méthode de base.

De même nous avons appliqué les trois mesures classiques mentionnées pour le cas de filtrage et de recommandation. Il convient de noter que les résultats obtenus pour les deux tâches de filtrage et de recommandation combinent la tâche de l'analyse sémantique fondée sur l'ontologie et l'apprentissage profond. Les tableaux 6.30 et 6.31 illustrent les différents résultats obtenus pour la tâche de filtrage des tweets inappropriés. Ces résultats sont obtenus en fonction de l'application des différents modèles d'incorporation de phrases ainsi qu'en fonction des différents algorithmes d'apprentissage. En outre, les résultats présentés sont obtenus respectivement à partir d'un modèle d'auto-encodeur convolutif profond

TABLE 6.28 – Comparaison des résultats obtenus pour la classification du comportement lié à la crise COVID-19.

Comportement à l'égard du Covid-19	XG Boost (%)	Ran dom Forest (%)	Gaus sian NB (%)	SVM (%)	Ada Boost (%)	KNN (%)
Tous les attributs	51	55	42	44	53	47
Textuel + Métadonnées	53	55	45	46	54	48
Textuel + Linguistiques	57	57	52	41	55	62
Textuel	51	53	53	54	57	58

TABLE 6.29 – Comparaison des résultats obtenus par notre méthode et la méthode baseline pour la classification du comportement.

Comportement à l'égard du Covid-19	Méthode de base de KNN (%)	Notre architecture (%)	Amélioration (%)
Tous les attributs	47	70	23
Textuel+ Métadonnées	48	64	16
Textuel+ Linguistiques	62	70	8
Textuel	58	65	7

et d'un modèle d'auto-encodeur profond qu'ils sont appliqués pour la tâche de sélection des attributs importants.

Afin de réaliser la combinaison des résultats de différents modèles de langage tels que (Bert+Roberta+Distilbert) dans le tableau 6.31, nous avons appliqué les trois méthodes suivantes : combinaison par concaténation, combinaison par moyenne et combinaison par vote. La figure 6.4 fournit une illustration détaillée du processus de chaque méthode.

Nous avons comparé les résultats de notre méthode hybride conçue pour filtrer les tweets inappropriés qui combine à la fois des techniques basées sur les ontologies et sur l'apprentissage profond par rapport aux résultats : (i) de notre méthode basée uniquement sur l'ontologie, présentée dans la section 5.4 du cinquième chapitre, (ii) d'autres méthodes proposées dans le cadre de la mesure de similarité entre les phrases [Kim et al. \[2021\]](#), [Shajalal et Aono \[2019\]](#). Les résultats de l'application de notre ensemble de données sur les différentes méthodes citées sont présentés dans la figure 6.5.

Nous présentons dans les tables 6.32 et 6.33 les résultats obtenus concernant la partie de recommandation des tweets adaptés aux personnes malades. Ces résultats sont établis en fonction des différentes techniques d'incorporation de phrases et d'un ensemble d'algorithmes de classification.

Pour assurer l'évaluation qualitative des réponses retournées par notre méthode, nous avons demandé à un expert d'évaluer les résultats de notre chatbot par rapport à ceux de chatGPT. Ensuite, l'expert a interprété ces résultats en fonction des critères établis par [Guesmia \[2019\]](#). Ces critères ont été définis pour évaluer la

TABLE 6.30 – Comparaison des résultats obtenus pour le filtrage des tweets inappropriés sur la base d'un auto-encodeur convolutif profond.

	BERT	XLNET	XLM	Roberta	Distil bert	Roberta +Distil bert	Bert+ Roberta+ Distilbert	Bert+ Distil bert
BILSTM+SVM	64	63	57	62	65	57	58	60
LSTM+SVM	60	61	56	62	64	59	60	61
LSTM+ Sigmoid	54	58	58	55	55	52	50	49
BILSTM+ Sigmoid	58	59	59	57	56	50	42	44
BILSTM+ Random forest	57	58	57	55	56	48	49	44
BILSTM+ xg-boost	59	59	61	56	58	50	52	55

TABLE 6.31 – Comparaison des résultats obtenus pour le filtrage des tweets inappropriés sur la base d'un auto-encodeur profond.

	BERT	XLNET	XLM	Roberta	Distil bert	Roberta +Distil bert	Bert+ Roberta+ Distilbert	Bert+ Distil bert
BILSTM+SVM	71	59	67	66	72	64	61	62
LSTM+SVM	71	60	65	61	70	63	59	63
LSTM+ Sigmoid	66	56	64	66	65	60	57	60
BILSTM+ Sigmoid	70	60	65	66	65	59	55	58
BILSTM+ Decision tree	47	48	47	45	46	40	42	43
BILSTM+ Gaussian NB	59	59	62	59	61	58	55	56

qualité des agents conversationnels et des chatbots en se référant aux normes d'utilisation. Un extrait de la comparaison est présenté dans le tableau 6.34. Afin d'obtenir des réponses de chatGPT en rapport avec le contexte du Covid-19, comme le montre le tableau 6.34. Nous avons demandé à ChatGPT d'agir comme s'il était en période de pic de Covid-19 et de fournir une réponse concise à une requête d'une personne en état instable.

6.7.3 Discussion de la méthode de surveillance des cas des personnes ayant un TP

Au niveau de notre troisième contribution, nous avons proposé une méthode intelligente qui combine à la fois : (i) des méthodes numériques basées sur des techniques d'apprentissage automatique classique et profond afin d'assurer les tâches de classification, (ii) des méthodes linguistiques fondées sur l'utilisation des analyseurs de TALN et de ressources linguistiques disponibles pour la modélisation des connaissances. L'objectif de cette méthode est de tirer profit des

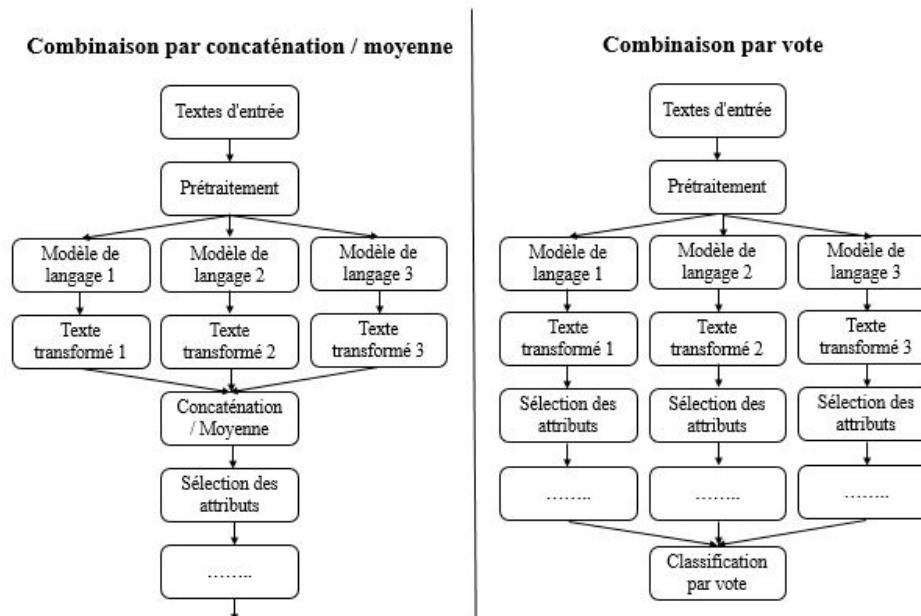


FIGURE 6.4 – Comparaison des différentes méthodes utilisées pour intégrer les résultats des différents modèles de langage.

données partagées sur les médias sociaux afin d’assurer l’intégralité du processus de diagnostic effectué par un psychiatre. Cette mesure vise à offrir un suivi en temps réel des malades sur les plateformes de médias sociaux pendant la pandémie de COVID-19, en supprimant les tweets inappropriés selon l’état de chaque utilisateur malade. De plus, le processus de surveillance implique l’envoi d’une réponse à cet utilisateur en fonction des préoccupations exprimées dans ses tweets. Ce processus s’effectue en débutant par la tâche de détection des personnes atteintes d’une maladie psychologie ainsi que de leur comportement lié à l’épidémie. Cette méthode aborde plusieurs problèmes identifiés dans des études antérieures, notamment : (i) la priorisation de la détection des maladies par rapport à l’ensemble du processus (détection et surveillance), (ii) l’ignorance de l’évolution du lexique au fil du temps. Particulièrement au cours de la période de COVID-19, de nombreux mots apparaissent et deviennent d’actualité dans un court laps de temps, (iii) la négligence des inconvénients liés à la méthode linguistique, comme l’absence de ressources et de règles précises, (iv) l’ignorance des particularités linguistiques du texte au cours du traitement des données brutes, par exemple les liens sémantiques entre les différentes unités lexicales, le contexte du discours, etc., (v) les difficultés linguistiques liées à la formulation d’une réponse, y compris le choix des mots, de leur ordre et de leur conjugaison.

Nous avons obtenu un résultat de F-mesure égal à 70% pour la tâche de détection du comportement des utilisateurs sur Twitter. Ces résultats sont efficaces par rapport à certains travaux qui se concentrent sur l’analyse du comportement des utilisateurs sur les réseaux sociaux [Zhang et al. \[2020\]](#). De même, ils sont moins performants par rapport à ceux obtenus par d’autres études de recherche menées dans le même contexte [Gaind et al. \[2019\]](#). En dépit de la performance de la dernière étude citée en relation avec notre travail, il convient de noter que notre méthode aborde la notion de logique floue avec ses avantages en termes de

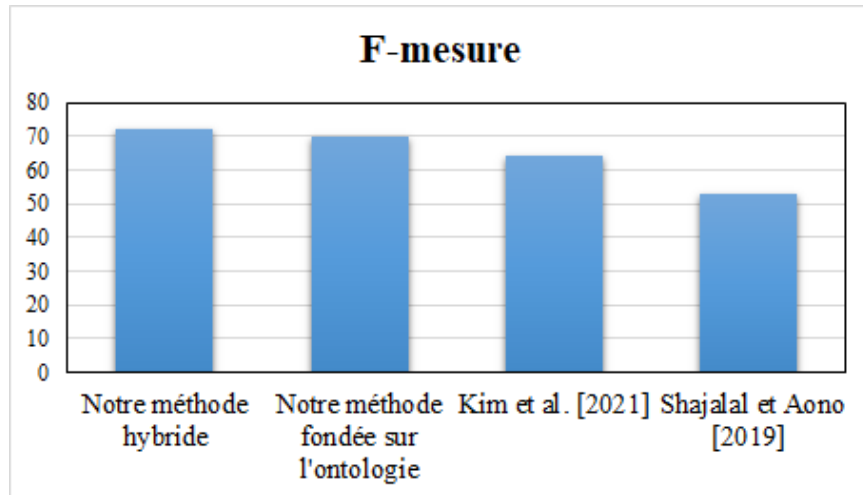


FIGURE 6.5 – Évaluation des performances des différentes méthodes pour le filtrage des tweets inappropriés.

TABLE 6.32 – Comparaison des résultats obtenus pour la recommandation des tweets appropriés sur la base d'un auto-encodeur convolutif profond.

	BERT	XLNET	XLM	Roberta	Distilbert	Roberta +Distilbert	Bert+ Roberta+ Distilbert	Xlnet +Roberta
BILSTM+SVM	63	70	75	65	75	62	62	68
LSTM+SVM	61	73	71	63	74	66	67	70
LSTM+ Sigmoid	60	66	65	58	68	59	59	61
BILSTM+ Sigmoid	60	67	68	57	70	56	55	63
Random forest	57	60	60	55	65	58	59	55
Xgboost	59	59	61	58	59	51	54	57
SVM	57	57	59	58	62	53	54	56

précision dans les réponses fournies. De même, notre méthode tient compte du contexte du discours en rapport avec un domaine particulier. Nous avons obtenu respectivement un F-mesure qui égale à 72% et 75% pour l'évaluation de la méthode hybride fondée sur l'ontologie et l'apprentissage profond pour classer les tweets : devant être filtrés, pouvant être recommandés ou neutre. Nous constatons que d'une manière générale, ces résultats peuvent être considérés comme intéressants compte tenu des problématiques rencontrées en termes de spécificités liées à notre domaine d'étude. Ce qui peut valider notre hypothèse sur la performance du modèle d'attention de BERT au niveau de son architecture pour préserver le contexte linguistique. Nous remarquons à partir des résultats obtenus en utilisant différents modèles de transformation, que les résultats sont variés et qu'aucun modèle ne surpasse les autres. Cependant, dans plusieurs cas, les modèles Roberta et Distilbert ont surpassé le modèle Bert. Cela démontre l'importance de la partie textuelle informelle prise en compte par Roberta. De plus, les spécificités des couches d'attention de Distilbert qui permettent au modèle d'être plus attractif que Bert. Pour la différence des résultats obtenus en utilisant le modèle auto-codeur

TABLE 6.33 – Comparaison des résultats obtenus pour la recommandation des tweets appropriés sur la base d'un auto-encodeur profond.

	BERT	XLNET	XLM	Roberta	Distil bert	Roberta +Distil bert	Bert+ Roberta+ Distilbert	Bert+ Distil bert
BILSTM+SVM	59	69	52	68	66	62	64	61
LSTM+SVM	59	69	49	65	66	63	65	63
LSTM+ Sigmoid	53	60	50	62	59	55	55	58
BILSTM+ Sigmoid	53	62	52	61	60	52	48	58

profond ou le modèle convolutif profond. De nombreuses études comme [Wajeed et Sreenivasulu \[2019\]](#), [Hedayati et al. \[2021\]](#) ont démontré que la couche convolutive donne de meilleur résultat quand il y a plus de détails et de spécificités dans les données à analyser. Il est reconnu aussi que l'auto-encodeur profond est plus performant quand nous prenons une décision basée sur le contexte général. Il convient de noter que dans la sous-section 3.4, nous avons mentionné que la tâche de recommandation de tweets est plus sensible que la tâche de filtrage de tweets. Cette hypothèse peut démontrer que dans la tâche de filtrage, le modèle autoencodeur profond a produit des meilleurs résultats par rapport à la tâche de recommandation.

Pour les résultats obtenus concernant la tâche de filtrage des tweets inappropriés, notre méthode a obtenu de meilleurs résultats de performance par rapport aux autres travaux [Shajalal et Aono \[2019\]](#), [Kim et al. \[2021\]](#). Pour ce qui concerne la méthode [Shajalal et Aono \[2019\]](#), basée sur une méthode lexico-sémantique, elle a atteint un taux de précision de 53%. En effet, les termes relatifs à l'épidémie de Covid-19 ne sont pas intégrés dans les ressources linguistiques utilisées par cette méthode. Le problème de la méthode [Kim et al. \[2021\]](#) réside dans le fait qu'elle aborde les relations sémantiques entre les phrases de manière générale, tandis que l'analyse sémantique requise dans notre travail est très spécifique. En effet, elle exige la prise en compte de plusieurs critères et de cas exceptionnels. En ce qui concerne les résultats de notre méthode basée sur l'ontologie, conçue pour être applicable à tous les domaines, nous avons obtenu des résultats proches (-2%) à ceux de notre méthode hybride. Il convient de noter que l'évaluation de ce travail n'est effectuée que lorsque l'ontologie interagit avec la requête, ce qui implique qu'elle renvoie un résultat indépendamment de son exactitude. Par ailleurs, ce travail tient principalement compte du sujet partagé par les deux tweets et se concentre uniquement sur l'humeur exprimée par le malade. Cela signifie qu'il ignore l'humeur exprimée dans le deuxième tweet, ce choix est justifié par l'exemple suivant :

Je suis déçu, mon équipe XX a subi sa troisième défaite consécutive en une semaine./L'équipe YY a conservé la première place du classement au cours des deux derniers mois.

Par conséquent, le second tweet sera supprimé car il traite du même sujet que le premier et exprime une mauvaise humeur. Ce choix s'appuie sur l'argument principal que les deux équipes YY et XX pourraient être concurrentes.

Pour ce qui concerne l'évaluation qualitative des réponses renvoyées par notre

TABLE 6.34 – Analyse comparative de la qualité des réponses fournies par notre chatbot et chatGPT.

Les critères de performance des réponses fournies	
Efficacité	
Performance	Traitement robuste des entrées
Requête	Quand tu vois comment les gens portent leurs masques, ya de quoi se poser des questions.
Réponse de ChatGPT	Il est vrai que certains peuvent ne pas porter correctement leur masque, mais il est important de se concentrer sur notre propre comportement responsable pour protéger notre santé et celle des autres.
Réponse de notre Chatbot	La police continue à contrôler le port du masque.
Effectivité	
Fonctionnalité	Interprétation précise des ordres donnés
	Utilisation adéquate des degrés de formalité et de précision linguistique
	Engagement dans la résolution de problèmes
Requête	Lutte contre la Covid-19 : un masque acheté à plus de 2 millions francs!!!
Réponse de ChatGPT	Il est important de se rappeler que le prix des masques peut varier, mais l'essentiel est de se protéger et de protéger les autres en les portant correctement.
Réponse de notre Chatbot	Un amendement de La France Insoumise proposant la gratuité des masques, soumis au projet d'état d'urgence sanitaire.
Humanité	Interaction convaincante, satisfaisante et naturelle
	Intégration des erreurs afin de rendre la communication plus réaliste
Requête	Je vais finir en dépression à cause de ce virus! ; plus de foot, plus d'école, plus sortir!
Réponse de ChatGPT	Il est normal de ressentir de la frustration face aux perturbations causées par la pandémie, mais il est important de chercher des moyens de prendre soin de sa santé mentale et de rester connecté avec ses proches malgré les restrictions.
Réponse de notre Chatbot	Si on prend toutes les précautions, on va pouvoir retrouver une vie plus normale.
Satisfaction	
Aspect affectif	Exprimer la chaleur et les émotions à travers le ton, l'inflexion et l'expressivité
Requête	Grâce à moi, le virus a progressé. Et il restera là. Covid-19
Réponse de ChatGPT	Il est essentiel de se rappeler que la lutte contre le virus est une responsabilité collective, et nous devons tous nous soutenir mutuellement pour surmonter cette situation difficile.
Réponse de notre Chatbot	Une nouvelle vague de COVID? C'est comme ça qu'on remercie nos soignants! Un peu de bon sens...

chatbot et chatGPT dans le but d'assurer la tâche de recommandation de tweets selon les standards proposés par Guesmia [2019]. Selon notre expert, notre chatbot a été plus spontané sur le plan affectif, montrant plus d'émotion et de chaleur que chatGPT, qui est basé sur un langage formel articulé sur des raisons logiques et raisonnables. De plus, les réponses de ChatGPT concernant le Covid-19 sont interprétées comme une série de lignes directrices et de conseils, qui contiennent des informations plus abstraites. En revanche, les résultats de notre chatbot sont jugés plus précis et incluent des informations plus concrètes.

Il est vrai que notre principale méthode hybride prend en compte de nombreux détails par rapport à d'autres travaux présentés. Cependant, au niveau des résultats obtenus, nous avons identifié un échec spécifique lié aux circonstances particulières suivantes :

Cas 1 : Tweet1 : La mère de mon amie a battu Corona, des mois en soins intensifs, deux semaines dans le coma. Je suis si heureuse pour elle. Bonne nouvelle, merci à Dieu et aux médecins.

Tweet2 : Suis-je le seul à être heureux pour Corona ?

Notre méthode a chevauché dans la classification de ces deux tweets. Le mot heureux apparaît simultanément dans la liste d'humeurs appropriées et inappropriées. En effet, selon nos experts, le premier tweet est classé comme un comportement approprié avec une humeur positive. Tandis que le deuxième tweet est classé comme un comportement inapproprié.

Cas 2 : Coronavirus : la pandémie "loin d'être terminée", le Canada se prépare "à une deuxième vague"

Dans le contexte du COVID-19, « terminé » signifie une humeur positive, mais dans le cas présent, il a été catégorisé comme une humeur négative.

Cas 3 : Tweet1 : Le corona est parti et la chloroquine est guérie. Ne mettez pas de masque, à la longue c'est nocif.

Tweet2 : La corona est toujours d'actualité, gardez votre masque buccal et coupez vos voyages ou vous créez une 2ème vague.

Dans ce cas, le deuxième tweet est jugé comme recommandé pour la personne qui a publié le premier tweet, mais la partie concernant le voyage n'a aucune signification dans ce contexte.

Cas 4 : Le taux de mortalité dû au Covid-19 est minime Même l'OMS ne recommande pas l'utilisation de masques en dehors des hôpitaux.

Ce tweet est considéré comme recommandé aux personnes qui craignent le taux de mortalité. Cependant, dans ce cas, il est nécessaire de vérifier la validité des informations.

6.8 CONCLUSION

Dans ce chapitre, nous avons présenté les différentes expérimentations, que nous avons menées dans le but de détecter et de surveiller les utilisateurs ayant un trouble psychologique. Dans un premier temps, nous avons procédé à une validation du corpus élaboré. Cette validation consiste à comparer le style d'écriture présent dans notre corpus avec celui des autres corpus extraits à partir de la littérature. Les corpus sélectionnés sont de différentes langues et traitent des différents types de troubles psychologiques, ce qui permet d'assurer la généralisation

des connaissances interprétées. Dans un deuxième temps, nous avons comparé les différents styles d'écriture tirés des différents corpus aux hypothèses de l'ASHA. Nous avons ensuite procédé à la validation des contributions de notre travail. Pour ce faire, nous avons comparé nos résultats aux résultats de certains travaux de la littérature. De même, nous avons illustré l'amélioration obtenue grâce à nos contributions par rapport aux méthodes de base. Les principaux résultats obtenus de la métrique de F-mesure par rapport à chaque contribution sont 74%, 92%, 75%. Nous avons également tenu une discussion expliquant les résultats obtenus pour chaque contribution, ainsi que les avantages et les inconvénients par rapport à d'autres travaux de la littérature.

PUBLICATIONS SCIENTIFIQUES

- ELLOUZE, Mourad, MECHTI, Seifeddine, et BELGUITH, Lamia. A hybrid approach based on linguistic analysis and fuzzy logic to ensure the surveillance of people having paranoid personality disorder towards Covid-19 on social media. *International Journal of General Systems*, 251-274, 2023. «Scimago Quartile : **Q2**, Impact Factor :**2.4**»
- ELLOUZE, Mourad et BELGUITH, Lamia. Semantic analysis based on ontology and deep learning for a chatbot to assist persons with personality disorders on Twitter. *Behaviour & Information Technology*, 1-20, 2023. «Scimago Quartile : **Q2**, Impact Factor :**3.7**»
- ELLOUZE, Mourad et BELGUITH, Lamia. A hybrid approach for the detection and monitoring of people having personality disorders on social networks. *Social Network Analysis and Mining*, vol. 12, no 1, p. 67, 2022. «Scimago Quartile : **Q2**, Impact Factor : **2.8**»
- ELLOUZE, Mourad, MECHTI, Seifeddine, et BELGUITH, Lamia. A Comparative Study on the Extraction of Dependency Links Between Different Personality Traits. *SN Computer Science*, vol. 3, no 6, p. 495, 2022. «Scimago Quartile : **Q2**»
- ELLOUZE, Mourad, et al. A deep learning approach for detecting the behaviour of people having personality disorders towards COVID-19 from Twitter. *IJCSE*, vol. 25, no 4, p. 353-366, 2022. «Scimago Quartile : **Q3**»
- ELLOUZE, Mourad et BELGUITH, Lamia. Approach based on bayesian network and ontology for identifying factors impacting the states of people with psychological problems from data on social media. *MEDI*, 128-141, 2023. «CORE 2023, **Rank C**»
- ELLOUZE, Mourad et BELGUITH, Lamia. A data warehouse model for analyzing the behavior and writing style of people with personality disorders on social media. *ISDA*, 2023. «CORE 2023, **Rank C**»
- ELLOUZE, Mourad et BELGUITH, Lamia. A Deep Learning Architecture Based on Advanced Textual Language Models for Detecting Disease Through Its Symptoms Associated with a Reinforcement Learning Algorithm. *ICSOFT*, 207-229, 2022. «CORE 2021, **Rank B**»
- ELLOUZE, Mourad, MECHTI, Seifeddine, et BELGUITH, Lamia. Deep Learning CNN-LSTM Approach for Identifying Twitter Users Suffering from Paranoid Personality Disorder. *ICSOFT*, 2022. «CORE 2021, **Rank B**»
- ELLOUZE, Mourad, MECHTI, Seifeddine, et BELGUITH, Lamia. Approach based on ontology and machine learning for identifying causes affecting personality disorder disease on twitter. *KSEM*, 2021. «CORE 2021, **Rank B**»
- ELLOUZE, Mourad, MECHTI, Seifeddine, et BELGUITH, Lamia. Automatic profile recognition of authors on social media based on hybrid approach. *KES*, vol. 176, p. 1111-1120, 2020. «CORE 2020, **Rank B**»

CONCLUSION GÉNÉRALE

Les recherches effectuées dans cette thèse s'inscrivent dans le cadre de la détection et la surveillance des utilisateurs des réseaux sociaux ayant des troubles de personnalité. Le but était d'analyser le comportement et le style d'écriture interactif entre les individus au niveau des statuts partagés en français, couvrant divers domaines (tels que la politique, la société, l'économie, etc.).

Pour ce faire, nous avons défini dans le premier chapitre le concept de la personnalité ainsi que les différents types et caractéristiques en rapport avec les traits et les troubles de personnalité. Notons que dans le cadre de cette thèse, nous nous sommes intéressés principalement aux troubles de la personnalité en raison de leur impact social par rapport à l'étude des caractéristiques des traits de la personnalité. Dans ce même chapitre, nous avons exposé une vue d'ensemble des caractéristiques linguistiques qui aident à reconnaître les particularités liées à la personnalité, notamment les caractéristiques lexicales, morpho-syntaxiques, sémantiques et numériques. À la fin de ce chapitre, nous avons examiné en détail les principales techniques d'intelligence artificielle utilisées dans l'analyse de la personnalité.

Dans un deuxième temps, notre but était de présenter une vue d'ensemble des différentes méthodes proposées dans la littérature pour le traitement des concepts liés à la personnalité. Ces informations ont été présentées au niveau du deuxième chapitre. Nous avons commencé ce chapitre par une présentation des méthodes appliquées pour la détection des traits de personnalité. Ces méthodes se basent sur des techniques numériques, telles que l'apprentissage automatique et les statistiques, ainsi que sur celles qui se basent principalement sur les aspects linguistiques. En outre, nous avons présenté les principales méthodes fondées sur l'hybridation, c'est-à-dire la combinaison de deux ou plusieurs techniques. Ensuite, nous avons présenté les travaux dédiés à la détection implicite ou explicite des troubles psychologiques sur les réseaux sociaux, en utilisant les diverses techniques appliquées à la détection des traits de personnalité. Nous avons également présenté les méthodes dédiées à la détection de différents types de relations entre les concepts liés à la personnalité telle que les liens corrélatifs et causals au moyen de diverses techniques telles que les techniques numériques et linguistiques. Par la suite, nous avons présenté les moyens permettant de surveiller les utilisateurs présentant une forte probabilité d'être malades, dans le but de contrôler la stabilité de leur état de santé. Cette tâche est accomplie, tout en discutant en détail les points forts et les points faibles des différentes méthodes présentées dans ce chapitre. Notre analyse critique des méthodes existantes nous a permis d'aborder la problématique de la tâche de l'analyse de la personnalité et d'étudier sa complexité. À l'issue de cette étude, nous avons proposé dans le cadre du présent travail trois contributions pour détecter et surveiller l'état sanitaire des patients en temps réel.

La première contribution qui a été présentée au niveau du troisième chapitre consiste à détecter les troubles de la personnalité sur les médias sociaux en suivant différentes manières. La première est fondée sur l'intérêt du passage par la détection du degré de symptômes avant la détection du trouble. Cet objectif est accompli à l'aide d'une méthode basée sur l'apprentissage profond prenant en considération la transformation, l'extraction, la génération et la classification. La seconde méthode applique les mêmes techniques que la première pour la transformation, l'extraction, la génération et la classification, mais elle ne nécessite pas la détection du degré des symptômes pour détecter le trouble. L'avantage de la seconde méthode par rapport à la première est qu'elle traite tous les types et les catégories de troubles de la personnalité. La troisième méthode repose sur l'extraction d'attributs personnalisés ainsi que sur la classification par relativité et par vote dans le but d'accroître la précision. Dans cette méthode, nous avons profité de plusieurs ressources linguistiques ainsi que des techniques statistiques pour l'extraction et la classification. Les résultats de la métrique de F-mesure pour ces trois méthodes sont respectivement 65%, 60% et 74%.

La deuxième contribution décrite dans le quatrième chapitre vise à présenter une méthode pour identifier les causes du trouble (les raisons qui ont fait preuve du trouble). Cette méthode comporte trois tâches : la première consiste à regrouper récursivement les tweets ayant les mêmes sujets à l'aide d'une technique d'apprentissage non supervisée. La seconde consiste à créer automatiquement une ontologie en utilisant des ressources et des règles linguistiques ainsi qu'une métrique statistique pour modéliser le lexique et l'humeur associés à chaque sujet détecté. La troisième repose sur l'élaboration automatique d'un ensemble de règles visant à interroger une ontologie dans le but de détecter le sujet et l'humeur exprimés au niveau d'un tweet. Cette tâche est réalisée en utilisant une méthode statistique. Dans ce contexte, nous avons obtenu une F-mesure de 73% pour la détection du sujet des tweets relatifs au COVID-19, ainsi qu'une F-mesure de 92% pour la détection de l'humeur associée à ces tweets.

La troisième contribution illustrée au niveau du cinquième chapitre consiste à assurer une surveillance de l'état des patients en lien avec l'épidémie de COVID-19. Pour atteindre cet objectif, nous avons proposé une méthode hybride basée sur l'analyse sémantique, combinant les résultats d'une ontologie et d'un modèle d'apprentissage profond. Nous avons employé cette méthode pour recommander des tweets aux patients afin de les aider à mieux gérer leur anxiété. Nous avons également utilisé cette méthode pour mieux gérer le fil d'actualité d'un patient, en filtrant les tweets inappropriés en fonction de l'état de santé. Nous avons obtenu respectivement une F-mesure de 72% et 75% pour l'évaluation de la méthode hybride fondée sur l'ontologie et l'apprentissage profond permettant de classifier les tweets selon trois classes (doit être filtré, peut être recommandé ou neutre).

Enfin, dans le sixième chapitre, nous avons décrit et discuté les différentes expérimentations menées tout au long de cette thèse. De même, nous avons établi une comparaison avec des méthodes et des corpus existants. La tâche d'évaluation consiste à comparer à chaque niveau les résultats de nos méthodes avec les résultats des annotations. Toutes les méthodes proposées et les diverses expérimentations effectuées dans ce contexte ont démontré des résultats très encourageants. Néanmoins, plusieurs opportunités prometteuses suggèrent des axes de recherche pour des études ultérieures.

Dans la continuité à court terme de notre thèse, nous envisageons, dans un premier temps, d'explorer les données et les résultats de nos méthodes à l'aide d'un tableau de bord. Ce dernier permet de visualiser le style d'écriture, le comportement et l'emplacement géographique des personnes atteintes de troubles de la personnalité. Dans un deuxième temps, nous prévoyons d'explorer les liens de dépendance et de corrélation entre les différents concepts associés à ces troubles, afin de mieux comprendre leurs complexités et leurs interrelations.

À moyen terme, nous prévoyons de détecter d'autres types de troubles, tels que la maladie d'Alzheimer, la nymphomanie et les phobies. De plus, nous envisageons de tester notre méthode sur d'autres modèles de langage tels que GPT-3, ainsi que d'utiliser des architectures avancées d'apprentissage profond telles que GAN dans le cadre de la génération de données pour la tâche de recommandation de tweets.

À long terme, notre objectif est de généraliser notre méthode de détection de troubles et de surveillance des patients. Cette généralisation permet d'assurer l'indépendance au niveau de la langue, des domaines d'études et des formats de données, y compris le contenu multimédia incluant des images et du son pour les vidéos.

Nous envisageons aussi d'intégrer une étape supplémentaire à notre méthode d'analyse, qui permettra de traiter non seulement des données en ligne (sur le Web), mais également des données hors ligne, telles que les rapports médicaux stockés localement.

BIBLIOGRAPHIE

- Chapter 2 - fundamentals of type-1 fuzzy logic theory. Dans Erdal Kayacan et Mojtaba Ahmadi Khansar, éditeurs, *Fuzzy Neural Networks for Real Time Control Applications*, pages 13 – 24. Butterworth-Heinemann, 2016. ISBN 978-0-12-802687-8. URL <http://www.sciencedirect.com/science/article/pii/B9780128026878000025>.
- Haneen Arafat Abu Alfeilat, Ahmad BA Hassanat, Omar Lasassmeh, Ahmad S Tarawneh, Mahmoud Bashir Alhasanat, Hamzeh S Eyal Salman, et VB Surya Prasath. Effects of distance measure choice on k-nearest neighbor classifier performance : a review. *Big data*, 7(4) :221–248, 2019.
- Pragya Agarwal. Ontological considerations in giscience. *International Journal of Geographical Information Science*, 19(5) :501–536, 2005.
- Shivam Agarwal. Data mining : Data mining concepts and techniques. Dans *2013 international conference on machine intelligence and research advancement*, pages 203–207. IEEE, 2013.
- Nadeem Ahmad et Jawaid Siddique. Personality assessment using twitter tweets. *Procedia computer science*, 112 :1964–1973, 2017.
- Amer Ali Al-Atwi, Yahua Cai, et Joseph Amankwah-Amoah. Workplace ostracism, paranoid employees, and service performance : A multilevel investigation. *Journal of Managerial Psychology*, 2020.
- Shatha AbdulAziz AlAjlan et Abdul Khader Jilani Saudagar. Machine learning approach for threat detection on social media posts containing arabic text. *Evolutionary Intelligence*, 14(2) :811–822, 2021.
- Andry Alamsyah, Rafa Syafiq Bastikarana, Alya Rysda Ramadhanti, et Sri Wi-diyani. Recognizing personality from social media linguistic cues : A case study of brand ambassador personality. Dans *2020 8th International Conference on Information and Communication Technology (ICoICT)*, pages 1–5. IEEE, 2020.
- Hazrat Ali, Ahsan Ullah, Talha Iqbal, et Shahid Khattak. Pioneer dataset and automatic recognition of urdu handwritten characters using a deep autoencoder and convolutional neural network. *SN Applied Sciences*, 2(2) :1–12, 2020.
- Hatoon S AlSagri et Mourad Ykhlef. Machine learning-based approach for depression detection in twitter using content and activity features. *IEICE Transactions on Information and Systems*, 103(8) :1825–1832, 2020.

- Miguel A Alvarez-Carmona, A Pastor López-Monroy, Manuel Montes-y Gómez, Luis Villasenor-Pineda, et Hugo Jair-Escalante. Inaoe's participation at pan'15 : Author profiling task. *Working Notes Papers of the CLEF*, page 103, 2015.
- Guozhen An et Rivka Levitan. Lexical and acoustic deep learning model for personality recognition. Dans *INTERSPEECH*, pages 1761–1765, 2018.
- Guozhen An, Sarah Ita Levitan, Julia Hirschberg, et Rivka Levitan. Deep personality recognition for deception detection. Dans *INTERSPEECH*, pages 421–425, 2018.
- Óscar Apolinario-Arzube, José Antonio García-Díaz, José Medina-Moreira, Harry Luna-Aveiga, et Rafael Valencia-García. Comparing deep-learning architectures and traditional machine-learning approaches for satire identification in spanish tweets. *Mathematics*, 8(11) :2075, 2020.
- Mikel Artetxe et Holger Schwenk. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. corr abs/1812.10464 (2018). *arXiv preprint arXiv :1812.10464*, 2018.
- Andri Astuti, Firtri. Antisocial behaviour public twitter indonesia, 2021.
- Hassan A Aziz et Aisha Madani. Evolution of the web and its uses in healthcare. *Clinical Laboratory Science*, 28(4) :245–249, 2015.
- Khadija Aziz, Dounia Zaidouni, et Mostafa Bellafkih. Real-time data analysis using spark and hadoop. Dans *2018 4th international conference on optimization and applications (ICOA)*, pages 1–6. IEEE, 2018.
- Safa Bahri, Nesrine Zoghlami, Mourad Abed, et João Manuel RS Tavares. Big data for healthcare : a survey. *IEEE access*, 7 :7397–7408, 2018.
- Jongbum Baik, Kangbok Lee, Soowon Lee, Yongbum Kim, et Jayoung Choi. Predicting personality traits related to consumer behavior using sns analysis. *New Review of Hypermedia and Multimedia*, 22(3) :189–206, 2016.
- Sergio A Barroilhet, Amelia M Pellegrini, Thomas H McCoy, et Roy H Perlis. Characterizing dsm-5 and icd-11 personality disorder features in psychiatric inpatients at scale using electronic health records. *Psychological Medicine*, 50(13) : 2221–2229, 2020.
- Nick Bassiliades. Swrl2spin : A tool for transforming swrl rule bases in owl ontologies to object-oriented spin rules. *arXiv preprint arXiv :1801.09061*, 2018.
- Gaston Baudat et Fatiha Anouar. Generalized discriminant analysis using a kernel approach. *Neural computation*, 12(10) :2385–2404, 2000.
- J Bazan et al. A comparison of dynamic and non-dynamic rough set methods for extracting laws from decision tables. *Rough sets in knowledge discovery*, 1 : 321–365, 1998.

- Amin Beheshti, Vahid Moraveji-Hashemi, Shahpar Yakhchi, Hamid Reza Motahari-Nezhad, Seyed Mohssen Ghafari, et Jian Yang. personality2vec : Enabling the analysis of behavioral disorders in social networks. Dans *Proceedings of the 13th international conference on web search and data mining*, pages 825–828, 2020.
- Roger K Blashfield et Ross A McElroy. Ontology of personality disorder categories, 1989.
- Reihane Boghrati, Jonah A Berger, et Grant M Packard. Style, content, and the success of ideas. *Available at SSRN 4004771*, 2022.
- Piotr Bojanowski, Edouard Grave, Armand Joulin, et Tomas Mikolov. Enriching word vectors with subword information. *Transactions of the association for computational linguistics*, 5 :135–146, 2017.
- Trinayan Borah et S Ganesh Kumar. Application of nlp and machine learning for mental health improvement. Dans *International Conference on Innovative Computing and Communications*, pages 219–228. Springer, 2023.
- Erik Cambria, Soujanya Poria, Rajiv Bajpai, et Björn Schuller. Senticnet 4 : A semantic resource for sentiment analysis based on conceptual primitives. Dans *Proceedings of COLING 2016, the 26th international conference on computational linguistics : Technical papers*, pages 2666–2677, 2016.
- Ricardo JGB Campello, Peer Kröger, Jörg Sander, et Arthur Zimek. Density-based clustering. *Wiley Interdisciplinary Reviews : Data Mining and Knowledge Discovery*, 10(2) :e1343, 2020.
- Angela L Carey, Melanie S Brucks, Albrecht CP Kufner, Nicholas S Holtzman, Mitja D Back, M Brent Donnellan, James W Pennebaker, Matthias R Mehl, et al. Narcissism and the use of personal pronouns revisited. *Journal of personality and social psychology*, 109(3) :e1, 2015.
- Gabriella Casalino, Ciro Castiello, Nicoletta Del Buono, et Corrado Mencar. Intelligent twitter data analysis based on nonnegative matrix factorizations. Dans *International Conference on Computational Science and Its Applications*, pages 188–202. Springer, 2017.
- Fabio Celli, Arindam Ghosh, Firoj Alam, et Giuseppe Riccardi. In the mood for sharing contents : Emotions, personality and interaction styles in the diffusion of news. *Information Processing & Management*, 52(1) :93–98, 2016.
- Fabio Celli et Bruno Lepri. Is big five better than mbti? a personality computing challenge using twitter data. Dans *CLiC-it*, 2018.
- Daniel Cer, Mona Diab, Eneko Agirre, Inigo Lopez-Gazpio, et Lucia Specia. Semeval-2017 task 1 : Semantic textual similarity-multilingual and cross-lingual focused evaluation. *arXiv preprint arXiv :1708.00055*, 2017.
- Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, et al. Universal sentence encoder. *arXiv preprint arXiv :1803.11175*, 2018.

- yusuf mucahıt Çetinkaya, mesut GÜrlek, ismail hakki toroslu, et pinar Karagöz. Twitter account classification using account metadata : organizationvs. individual. *Turkish Journal of Electrical Engineering and Computer Sciences*, 30(4) :1404–1418, 2022.
- Parth Chandarana et M Vijayalakshmi. Big data analytics frameworks. Dans *2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA)*, pages 430–434. IEEE, 2014.
- Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, et W Philip Kegelmeyer. Smote : synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16 :321–357, 2002.
- Yuling Chen et Zhi Zhang. Research on text sentiment analysis based on cnns and svm. Dans *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 2731–2734. IEEE, 2018.
- Salim Chikhi. Computational intelligence : Before and beyond. Dans *JERI*, 2019.
- Raymond Chiong, Gregorius Satia Budhi, Sandeep Dhakal, et Fabian Chiong. A textual-based featuring approach for depression detection using machine learning classifiers and social media texts. *Computers in Biology and Medicine*, 135 : 104499, 2021.
- Kenneth Ward Church. Word2vec. *Natural Language Engineering*, 23(1) :155–162, 2017.
- Alexis Conneau et Douwe Kiela. Senteval : An evaluation toolkit for universal sentence representations. *arXiv preprint arXiv :1803.05449*, 2018.
- Rion Brattig Correia, Ian B Wood, Johan Bollen, et Luis M Rocha. Mining social media data for biomedical signals and health-related behavior. *Annual review of biomedical data science*, 3 :433–458, 2020.
- Saman Daneshvar et Diana Inkpen. Gender identification in twitter using n-grams and lsa. Dans *Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018)*, 2018.
- Shilpa Dang et Peerzada Hamid Ahmad. Text mining : Techniques and its application. *International Journal of Engineering & Technology Innovations*, 1(4) :22–25, 2014.
- Souvick Das, Novarun Deb, Agostino Cortesi, et Nabendu Chaki. Sentence embedding models for similarity detection of software requirements. *SN Computer Science*, 2(2) :1–11, 2021.
- Donald L Davis, Stephen J Grove, et Patricia A Knowles. An experimental application of personality type as an analogue for decision-making style. *Psychological Reports*, 66(1) :167–175, 1990.
- Claude De Ganay et Dominique Gillot. Pour une intelligence artificielle maîtrisée, utile et démystifiée. *Report, Office parlementaire d'évaluation des choix scientifiques et technologiques*, 2017.

- Marina Delianidi, Konstantinos Diamantaras, George Chrysogonidis, et Vasileios Nikiforidis. Student performance prediction using dynamic neural models. *arXiv preprint arXiv :2106.00524*, 2021.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, et Kristina Toutanova. Bert : Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv :1810.04805*, 2018.
- Inoshika Dilrukshi, Kasun De Zoysa, et Amitha Caldera. Twitter news classification using svm. Dans *2013 8th International Conference on Computer Science & Education*, pages 287–291. IEEE, 2013.
- Marco Dinarelli et Isabelle Tellier. Étude de réseaux de neurones récurrents pour l'étiquetage de séquences. Dans *Traitement Automatique des Langues Naturelles (TALN)*, 2016.
- Christopher Ifeanyi Eke, Azah Anir Norman, et Liyana Shuib. Context-based feature technique for sarcasm identification in benchmark datasets using deep learning and bert model. *IEEE Access*, 9 :48501–48518, 2021.
- Mourad Ellouze, Seifeddine Mechti, et Lamia Hadrich Belguith. Automatic profile recognition of authors on social media based on hybrid approach. *Procedia Computer Science*, 176 :1111–1120, 2020.
- Mourad Ellouze, Seifeddine Mechti, et Lamia Hadrich Belguith. A comparative study on the extraction of dependency links between different personality traits. *SN Computer Science*, 3(6) :1–12, 2022.
- Marie-Agnès Enard, Pascal Guitton, et Thierry Viéville. C'est l'histoire d'un gan. *Binaire*, 2020.
- Brian Epstein. *The ant trap : Rebuilding the foundations of the social sciences*. Oxford Studies in Philosophy 0, 2015.
- Olivier Ferlatte, Joshun Dulai, Travis Salway Hottes, Terry Trussler, et Rick Marchand. Suicide related ideation and behavior among canadian gay and bisexual men : a syndemic analysis. *BMC public health*, 15(1) :1–9, 2015.
- Everlandio RQ Fernandes, Andre CPLF de Carvalho, et Xin Yao. Ensemble of classifiers based on multiobjective genetic sampling for imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 32(6) :1104–1115, 2019.
- Giancarlo Fissore. Modélisation générative : physique statistique des machines de boltzmann restreintes, apprentissage avec informations manquantes et apprentissage scalable des flux linéaires. 2022.
- Bharat Gaind, Varun Syal, et Sneha Padgalwar. Emotion detection and analysis on social media. *arXiv preprint arXiv :1901.08458*, 2019.
- Dominick Gamache, Claudia Savard, Sophie Lemelin, Alexandre Côté, et Évens Villeneuve. Premature termination of psychotherapy in patients with borderline personality disorder : A cluster-analytic study. *The Journal of nervous and mental disease*, 206(4) :231–238, 2018.

- Ge Gao, Eunsol Choi, Yejin Choi, et Luke Zettlemoyer. Neural metaphor detection in context. *arXiv preprint arXiv :1808.09653*, 2018.
- Eleanor J Gardiner et Valerie J Gillet. Perspectives on knowledge discovery algorithms recently introduced in chemoinformatics : rough set theory, association rule mining, emerging patterns, and formal concept analysis. *Journal of chemical information and modeling*, 55(9) :1781–1803, 2015.
- Andrew Gardner, Neshat Elhami, et Rastko R Selmic. Classifying unordered feature sets with convolutional deep averaging networks. Dans *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 3447–3453. IEEE, 2019.
- Bilal Ghanem, Jihen Karoui, Farah Benamara, Paolo Rosso, et Véronique Moriceau. Irony detection in a multilingual context. Dans *European Conference on Information Retrieval*, pages 141–149. Springer, 2020.
- Piali Das Gitansh Chadha et Zohar Karnin. K-means clustering with amazon sagemaker, 2018. URL <https://aws.amazon.com/fr/blogs/machine-learning/k-means-clustering-with-amazon-sagemaker/>.
- Christine Golbreich et Atsutoshi Imai. Combining swrl rules and owl ontologies with protégé owl plugin, jess, and racer. Dans *7th International Protégé Conference, Bethesda, MD*, 2004.
- Carlos E González-Gallardo, Azucena Montes, Gerardo Sierra, J Antonio Núñez-Juárez, Adolfo Jonathan Salinas-López, et Juan Ek. Tweets classification using corpus dependent tags, character and pos n-grams. Dans *CLEF (working notes)*, 2015.
- Samuel D Gosling. Personality in non-human animals. *Social and Personality Psychology Compass*, 2(2) :985–1001, 2008.
- NANCY GOYETTE. Le bien-être dans l’enseignement : Étude des forces de caractère chez des enseignants persévérants du primaire et du secondaire dans une approche axée sur la psychologie positive. 2014.
- Alex Graves. Long short-term memory. Dans *Supervised sequence labelling with recurrent neural networks*, pages 37–45. Springer, 2012.
- Andreas Grivas, Anastasia Krithara, et George Giannakopoulos. Author profiling using stylometric and structural feature groupings. Dans *CLEF (Working Notes)*, 2015.
- Seyf Eddine Guesmia. Identification, validation et importance relative des dimensions de qualité des chatbots tels que perçus par les utilisateurs. 2019.
- Giancarlo Guizzardi et Terry Halpin. Ontological foundations for conceptual modelling. *Applied Ontology*, 3(1-2) :1–12, 2008.

- Kilani Hajji, Sami Khammouma, Ilyes Marrag, et Mohamed Nasr. Troubles psychiatriques en milieu pénitentiaire tunisien. *L'information psychiatrique*, 92(6) : 474–482, 2016.
- Su-Hyun Han, Ko Woon Kim, SangYun Kim, et Young Chul Youn. Artificial neural network : understanding the basic concepts without mathematics. *Dementia and Neurocognitive Disorders*, 17(3) :83–89, 2018.
- Md Amaan Haque, Abhishek Verma, John Sahaya Rani Alex, et Nithya Venkatesan. Experimental evaluation of cnn architecture for speech recognition. Dans *First international conference on sustainable technologies for computational intelligence*, pages 507–514. Springer, 2020.
- Mariam Hassanein, Wedad Hussein, Sherine Rady, et Tarek F Gharib. Predicting personality traits from social media using text semantics. Dans *2018 13th International Conference on Computer Engineering and Systems (ICCES)*, pages 184–189. IEEE, 2018.
- Rohollah Hedayati, Mohammad Khedmati, et Mehran Taghipour-Gorjikotaie. Deep feature extraction method based on ensemble of convolutional auto encoders : Application to alzheimer's disease diagnosis. *Biomedical Signal Processing and Control*, 66 :102397, 2021.
- Tak-Sung Heo, Jong-Dae Kim, Chan-Young Park, et Yu-Seop Kim. Sentence similarity evaluation using sent2vec and siamese neural network with parallel structure. *Journal of Intelligent & Fuzzy Systems*, 40(4) :7735–7744, 2021.
- Sepp Hochreiter et Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8) :1735–1780, 1997.
- Martin L Hoffman. Empathy, social cognition, and moral action. Dans *Handbook of moral behavior and development*, pages 299–326. Psychology Press, 2014.
- Markus Hofmann et Ralf Klinkenberg. *RapidMiner : Data mining use cases and business analytics applications*. CRC Press, 2016.
- Nicholas S Holtzman, Allison M Tackman, Angela L Carey, Melanie S Brucks, Albrecht CP Kufner, Fenne Große Deters, Mitja D Back, M Brent Donnellan, James W Pennebaker, Ryne A Sherman, et al. Linguistic markers of grandiose narcissism : A liwc analysis of 15 samples. *Journal of Language and Social Psychology*, 38(5-6) :773–786, 2019.
- Lovre Hribar et Denis Duka. Software component quality prediction using knn and fuzzy logic. Dans *The 33rd International Convention MIPRO*, pages 402–408. IEEE, 2010.
- Infamouscoder. Depression : Reddit dataset (cleaned), 2022a.
- Infamouscoder. Depression : Twitter dataset + feature extraction, 2022b.
- Abid Ishaq, Saima Sadiq, Muhammad Umer, Saleem Ullah, Seyedali Mirjalili, Vaibhav Rupapara, et Michele Nappi. Improving the prediction of heart failure patients' survival using smote and effective data mining techniques. *IEEE access*, 9 :39707–39716, 2021.

- Titus Jayarathna, Gaetano D Gargiulo, et Paul P Breen. Polymer sensor embedded, iot enabled t-shirt for long-term monitoring of sleep disordered breathing. Dans *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pages 139–143. IEEE, 2019.
- Louie C Juera et Marie Antonette Paña-Tautho. An exploratory study on the dimensions of agricultural technology adoption : A modification of technology acceptance model.
- Nal Kalchbrenner, Ivo Danihelka, et Alex Graves. Grid long short-term memory. *arXiv preprint arXiv :1507.01526*, 2015.
- Emilie Karafillakis, Sam Martin, Clarissa Simas, Kate Olsson, Judit Takacs, Sara Dada, Heidi Jane Larson, et al. Methods for social media monitoring related to vaccination : systematic scoping review. *JMIR public health and surveillance*, 7(2) : e17149, 2021.
- Ramandeep Kaur et Sandeep Kautish. Multimodal sentiment analysis : A survey and comparison. *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines*, pages 1846–1870, 2022.
- S Kayalvizhi et D Thenmozhi. Data set creation and empirical analysis for detecting signs of depression from social media postings. *arXiv preprint arXiv :2202.03047*, 2022.
- Amirmohammad Kazameini, Samin Fatehi, Yash Mehta, Sauleh Eetemadi, et Erik Cambria. Personality trait detection using bagged svm over bert word embedding ensembles. *arXiv preprint arXiv :2010.01309*, 2020.
- Samina Khalid, Tehmina Khalil, et Shamila Nasreen. A survey of feature selection and feature extraction techniques in machine learning. Dans *2014 science and information conference*, pages 372–378. IEEE, 2014.
- Mohamed Khazbak, Zeyad Wael, Zahwa Ehab, Maria Gerorge, et Essam Eliwa. Mindtime : Deep learning approach for borderline personality disorder detection. Dans *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, pages 337–344. IEEE, 2021.
- Donghwa Kim, Deokseong Seo, Suhyoun Cho, et Pilsung Kang. Multi-co-training for document classification using various document representations : Tf-idf, lda, and doc2vec. *Information Sciences*, 477 :15–29, 2019.
- Taeuk Kim, Kang Min Yoo, et Sang-goo Lee. Self-guided contrastive learning for bert sentence representations. *arXiv preprint arXiv :2106.07345*, 2021.
- Hannah Kirk, Andrea Gilmour, Robert Dudley, et Deborah M Riby. Paranoid ideation and assessments of trust. *Journal of Experimental Psychopathology*, 4(4) : 360–367, 2013.
- Nikhileswar Komati. Suicide and depression detection, 2021.

- V Krishnaiah, G Narsimha, et N Subhash Chandra. Heart disease prediction system using data mining technique by fuzzy k-nn approach. Dans *Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1*, pages 371–384. Springer, 2015.
- Vineet Kumar et Suresh Sundaram. Offline text-independent writer identification based on word level data. *arXiv preprint arXiv :2202.10207*, 2022.
- Daria J Kuss et Mark D Griffiths. Social networking sites and addiction : Ten lessons learned. *International journal of environmental research and public health*, 14(3) :311, 2017.
- V La Paglia et F Telliez. Les facteurs de risque en matière de troubles neuromusculo-squelettiques : vers une définition socio-psycho-biologique. 2021.
- Nam Quang Lam. A machine learning and compiler-based approach to automatically parallelize serial programs using openmp. 2011.
- Richard Lauer. Is social ontology prior to social scientific methodology? *Philosophy of the Social Sciences*, 49(3) :171–189, 2019.
- Hong Liang, Xiao Sun, Yunlei Sun, et Yuan Gao. Text feature extraction based on deep learning : a review. *EURASIP journal on wireless communications and networking*, 2017(1) :1–12, 2017.
- Huijie Lin, Jia Jia, Quan Guo, Yuanyuan Xue, Qi Li, Jie Huang, Lianhong Cai, et Ling Feng. User-level psychological stress detection from social media using deep neural network. Dans *Proceedings of the 22nd ACM international conference on Multimedia*, pages 507–516, 2014.
- Huijie Lin, Jia Jia, Jiezhong Qiu, Yongfeng Zhang, Guangyao Shen, Lexing Xie, Jie Tang, Ling Feng, et Tat-Seng Chua. Detecting stress based on social interactions in social networks. *IEEE Transactions on Knowledge and Data Engineering*, 29(9) : 1820–1833, 2017a.
- Zhouhan Lin, Minwei Feng, Cicero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, et Yoshua Bengio. A structured self-attentive sentence embedding. *arXiv preprint arXiv :1703.03130*, 2017b.
- Hongyu Liu et Bo Lang. Machine learning and deep learning methods for intrusion detection systems : A survey. *applied sciences*, 9(20) :4396, 2019.
- Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, et Matti Pietikäinen. Deep learning for generic object detection : A survey. *International journal of computer vision*, 128(2) :261–318, 2020.
- A Pastor López-Monroy, Manuel Montes-y Gómez, Hugo Jair Escalante, Luis Villasenor-Pineda, et Esaú Villatoro-Tello. Inaoe’s participation at pan’13 : Author profiling task. Dans *CLEF 2013 evaluation labs and workshop*, 2013.
- Minna Lyons, Nazli Deniz Aksayli, et Gayle Brewer. Mental distress and language use : Linguistic analysis of discussion forum posts. *Computers in Human Behavior*, 87 :207–211, 2018.

- Mohamed Hédi Maâloul. *Approche hybride pour le résumé automatique de textes. Application à la langue arabe. (Hybrid approach for the automatic abstract of texts. Application with the Arab language)*. PhD thesis, University of Provence, Aix-en-Provence, Marseille, France, 2012. URL <https://tel.archives-ouvertes.fr/tel-00756111>.
- Carly Magee et Jeremy C Biesanz. Toward understanding the relationship between personality and well-being states and traits. *Journal of personality*, 87(2) :276–294, 2019.
- Navonil Majumder, Soujanya Poria, Alexander Gelbukh, et Erik Cambria. Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, 32(2) :74–79, 2017.
- Ahmed A Marouf, Adnan F Ashrafi, Tanveer Ahmed, et Tarikuzzaman Emon. A machine learning based approach for mapping personality traits and perceived stress scale of undergraduate students. *International Journal of Modern Education & Computer Science*, 11(8), 2019.
- Amitha Mathew, P Amudha, et S Sivakumari. Deep learning techniques : an overview. Dans *International Conference on Advanced machine learning technologies and applications*, pages 599–608. Springer, 2020.
- Sandra C Matz et Oded Netzer. Using big data as a window into consumers’ psychology. *Current opinion in behavioral sciences*, 18 :7–12, 2017.
- Hane Htut Maung. Psychiatric euthanasia and the ontology of mental disorder. *Journal of Applied Philosophy*, 38(1) :136–154, 2021.
- Atika Mbarek, Salma Jamoussi, et Abdelmajid Ben Hamadou. An across online social networks profile building approach : Application to suicidal ideation detection. volume 133, pages 171–183. Elsevier, 2022.
- Leland McInnes, John Healy, et James Melville. Umap : Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv :1802.03426*, 2018.
- Pranjal H Mehta et Samuel D Gosling. How can animal studies contribute to research on the biological bases of personality. *Biology of personality and individual differences*, pages 427–448, 2006.
- Yash Mehta, Navonil Majumder, Alexander Gelbukh, et Erik Cambria. Recent trends in deep learning based personality detection. *Artificial Intelligence Review*, 53(4) :2313–2339, 2020.
- Mohamed Ben Ahmed Mhiri et Faïez Gargouri. Méthodologie de construction des ontologies pour la résolution de conflits des systèmes d’information. *Tech. Sci. Informatiques*, 28(10) :1263–1287, 2009.
- Tomas Mikolov, Kai Chen, Greg Corrado, et Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv :1301.3781*, 2013.

- Saif Mohammad et Svetlana Kiritchenko. Using nuances of emotion to identify personality. Dans *Seventh International AAAI Conference on Weblogs and Social Media*, 2013.
- Ms Surabhi S Mohite et Satish Gujar. Mining online social networking data for detecting mental disorders. Rapport technique, EasyChair, 2021.
- Fantine Mordelet et J-P Vert. A bagging svm to learn from positive and unlabeled examples. *Pattern Recognition Letters*, 37 :201–209, 2014.
- Armando L Morera-Fumero, Estefania Diaz-Mesa, Pedro Abreu-Gonzalez, Lourdes Fernandez-Lopez, et Maria del Rosario Cejas-Mendez. Low levels of serum total antioxidant capacity and presence at admission and absence at discharge of a day/night change as a marker of acute paranoid schizophrenia relapse. *Psychiatry research*, 249 :200–205, 2017.
- A Muhammad, B Hendrik, et R Iswara. Expert system application for diagnosing of bipolar disorder with certainty factor method based on web and android. Dans *Journal of Physics : Conference Series*, volume 1339, page 012020. IOP Publishing, 2019.
- Eleanor Mulholland, Paul Mc Kevitt, Tom Lunney, et Karl-Michael Schneider. Analysing emotional sentiment in people’s youtube channel comments. Dans *Interactivity, Game Creation, Design, Learning, and Innovation*, pages 181–188. Springer, 2016.
- Ankit Murarka, Balaji Radhakrishnan, et Sushma Ravichandran. Classification of mental illnesses on social media using roberta. Dans *Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis*, pages 59–68, 2021.
- Harry G Murray, J Philippe Rushton, et Sampo V Paunonen. Teacher personality traits and student instructional ratings in six types of university courses. *Journal of educational psychology*, 82(2) :250, 1990.
- Danish Muzafar, Furqan Yaqub Khan, et Mubashir Qayoom. Machine learning algorithms for depression detection and their comparison. *arXiv preprint arXiv :2301.03222*, 2023.
- Rihanna Namdari. Mental health corpus, 2023.
- Charles Kay Ogden et Ivor Armstrong Richards. The meaning of meaning : A study of the influence of thought and of the science of symbolism. 1923.
- Marc Olano. Les pionniers de la psychologie humaniste. *Les Grands Dossiers des Sciences Humaines*, 55(6) :11–11, 2019.
- Abubakr H Ombabi, Wael Ouarda, et Adel M Alimi. Deep learning cnn-lstm framework for arabic sentiment analysis using textual information shared in social networks. *Social Network Analysis and Mining*, 10(1) :1–13, 2020.

- OMS. Les investissements dans le traitement de la dépression et de l'anxiété rapportent quatre fois leur valeur. <https://www.who.int/fr/news/item/13-04-2016-investing-in-treatment-for-depression-and-anxiety-leads-to-fourfold-return>, 2016. Accessed : 2023-04-30.
- Jinnie Ooi, John Michael, Sakari Lemola, Stephen Butterfill, Cynthia SQ Siew, et Lukasz Walasek. Interpersonal functioning in borderline personality disorder traits : a social media perspective. *Scientific reports*, 10(1) :1068, 2020.
- Cristina Oprean. État de l'art sur les aspects méthodologiques et processus en knowledge discovery in databases. 2010.
- Joshua Palay, Tamara L Taillieu, Tracie O Afifi, Sarah Turner, James M Bolton, Murray W Enns, Mark Smith, Alain Lesage, Jeffrey A Bakal, Brian Rush, et al. Prevalence of mental disorders and suicidality in canadian provinces. *The Canadian Journal of Psychiatry*, 64(11) :761–769, 2019.
- Yiteng Pan, Fazhi He, et Haiping Yu. Learning social representations with deep autoencoder for recommender system. *World Wide Web*, 23(4) :2259–2279, 2020.
- Babita Pandey, Devendra Kumar Pandey, Brijendra Pratap Mishra, et Wasiur Rhmann. A comprehensive survey of deep learning in the field of medical imaging and medical natural language processing : Challenges and research directions. *Journal of King Saud University-Computer and Information Sciences*, 2021.
- Gregory Park, H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Michal Kosinski, David J Stillwell, Lyle H Ungar, et Martin EP Seligman. Automatic personality assessment through social media language. *Journal of personality and social psychology*, 108(6) :934, 2015.
- Ives C Passos, Pedro L Ballester, Rodrigo C Barros, Diego Librenza-Garcia, Benson Mwangi, Boris Birmaher, Elisa Brietzke, Tomas Hajek, Carlos Lopez Jaramillo, Rodrigo B Mansur, et al. Machine learning and big data analytics in bipolar disorder : a position paper from the international society for bipolar disorders big data task force. *Bipolar Disorders*, 21(7) :582–594, 2019.
- Eiluned Pearce, Rafael Wlodarski, Anna Machin, et Robin IM Dunbar. Genetic influences on social relationships : sex differences in the mediating role of personality and social cognition. *Adaptive Human Behavior and Physiology*, 5(4) : 331–351, 2019.
- Rasmus T Pedersen, Jens Olav Dahlgard, et Manuele Citi. Voter reactions to candidate background characteristics depend on candidate policy positions. *Electoral Studies*, 61 :102066, 2019.
- Barbara Poblete, Ruth Garcia, Marcelo Mendoza, et Alejandro Jaimes. Do all birds tweet the same ? characterizing twitter around the world. Dans *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1025–1030, 2011.
- Roberto Poli, Michael Healy, et Achilles Kameas. *Theory and applications of ontology : Computer applications*. Springer, 2010.

- Ariel Pomputius. Can you hear me now ? social listening as a strategy for understanding user needs. *Medical reference services quarterly*, 38(2) :181–186, 2019.
- Soujanya Poria, Alexandar Gelbukh, Basant Agarwal, Erik Cambria, et Newton Howard. Common sense knowledge based personality recognition from text. Dans *Mexican International Conference on Artificial Intelligence*, pages 484–496. Springer, 2013.
- Carly Porter, Jasper Palmier-Claus, Alison Branitsky, Warren Mansell, Helen Warwick, et Filippo Varese. Childhood adversity and borderline personality disorder : A meta-analysis. *Acta Psychiatrica Scandinavica*, 141(1) :6–20, 2020.
- Kasula Chaithanya Pramodh et Y Vijayalata. Automatic personality recognition of authors using big five factor model. Dans *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, pages 32–37. IEEE, 2016.
- Yinghui Quan, Xian Zhong, Wei Feng, Jonathan Cheung-Wai Chan, Qiang Li, et Mengdao Xing. Smote-based weighted deep rotation forest for the imbalanced hyperspectral data classification. *Remote Sensing*, 13(3) :464, 2021.
- Nathanaël Randriamihamison. Hierarchical agglomerative clustering under contiguity constraint for hi-c data analysis. 2021.
- Francisco Rangel, Paolo Rosso, Martin Potthast, Benno Stein, et Walter Daelemans. Overview of the 3rd author profiling task at pan 2015. Dans *CLEF*, page 2015. sn, 2015.
- K Narayana Rao, K Venkata Rao, et Prasad Reddy PVGD. A hybrid intrusion detection system based on sparse autoencoder and deep neural network. *Computer Communications*, 180 :77–88, 2021.
- Udaipurwala Rashida et K Suresh Kumar. Social media mining to detect mental health disorders using machine learning. Dans *Sentiment Analysis and Deep Learning : Proceedings of ICSADL 2022*, pages 923–930. Springer, 2023.
- Alan Rector, Stefan Schulz, Jean Marie Rodrigues, Christopher G Chute, et Harold Solbrig. On beyond gruber :“ontologies” in today’s biomedical information systems and the limits of owl. *Journal of Biomedical Informatics*, 100 :100002, 2019.
- B Dinesh Reddy, Eali Stephen Neal Joshua, N Thirupathi Rao, et Debnath Bhattacharyya. Mental disorder detection in social networks using svm classification : An improvised approach. Dans *Smart Technologies in Data Science and Communication : Proceedings of SMART-DSC 2022*, pages 241–250. Springer, 2023.
- Nils Reimers et Iryna Gurevych. Sentence-bert : Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv :1908.10084*, 2019.
- Amal Rekik, Salma Jamoussi, et Abdelmajid Ben Hamadou. Violent vocabulary extraction methodology : Application to the radicalism detection on social media. Dans *International Conference on Computational Collective Intelligence*, pages 97–109. Springer, 2019.

- Alpa Reshamwala, Dhirendra Mishra, et Prajakta Pawar. Review on natural language processing. *IRACST Engineering Science and Technology : An International Journal (ESTIJ)*, 3(1) :113–116, 2013.
- Omri Ron, Ela Oren, et Reuven Dar. The doubt-certainty continuum in psychopathology, lay thinking, and science. *Journal of behavior therapy and experimental psychiatry*, 53 :68–74, 2016.
- Saravanan Sagadevan, Nurul Hashimah Ahamed Hassain Malim, et Mohd Heikal Husin. Sentiment valences for automatic personality detection of online social networks users using three factor model. *Procedia Computer Science*, 72 :201–208, 2015.
- Benoît Sagot et Darja Fišer. Building a free french wordnet from multilingual resources. Dans *OntoLex*, 2008.
- Ananna Saha, Ahmed Al Marouf, et Rafayet Hossain. Sentiment analysis from depression-related user-generated contents from social media. Dans *2021 8th International Conference on Computer and Communication Engineering (ICCCCE)*, pages 259–264. IEEE, 2021.
- Haitham Salem, Ana Ruiz, Sarah Hernandez, Kareem Wahid, Fei Cao, Brandi Karnes, Sarah Beasley, Marsal Sanches, Elaheh Ashtari, et Teresa Pigott. Borderline personality features in inpatients with bipolar disorder : impact on course and machine learning model use to predict rapid readmission. *Journal of Psychiatric Practice*[®], 25(4) :279–289, 2019a.
- Marwa S Salem, Sally S Ismail, et Mostafa Aref. Personality traits for egyptian twitter users dataset. Dans *Proceedings of the 2019 8th International Conference on Software and Information Engineering*, pages 206–211, 2019b.
- Kayalvizhi Sampath et Thenmozhi Durairaj. Data set creation and empirical analysis for detecting signs of depression from social media postings. Dans *International Conference on Computational Intelligence in Data Science*, pages 136–151. Springer, 2022.
- Victor Sanh, Lysandre Debut, Julien Chaumond, et Thomas Wolf. Distilbert, a distilled version of bert : smaller, faster, cheaper and lighter. *arXiv preprint arXiv :1910.01108*, 2019.
- H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. Personality, gender, and age in the language of social media : The open-vocabulary approach. *PloS one*, 8(9) :e73791, 2013.
- P Scocco, A Preti, S Totaro, PW Corrigan, C Castriotta, SOPROXI Team, et al. Stigma, grief and depressive symptoms in help-seeking people bereaved through suicide. *Journal of affective disorders*, 244 :223–230, 2019.
- Saskia Senn, ML Tlachac, Ricardo Flores, et Elke Rundensteiner. Ensembles of bert for depression classification. Dans *2022 44th Annual International Conference*

- of the IEEE Engineering in Medicine & Biology Society (EMBC), pages 4691–4694. IEEE, 2022.
- Michele Settanni, Danny Azucar, et Davide Marengo. Predicting individual characteristics from digital traces on social media : A meta-analysis. *Cyberpsychology, Behavior, and Social Networking*, 21(4) :217–228, 2018.
- Dilini Sewwandi, Kusal Perera, Sajith Sandaruwan, Oshani Lakchani, Anupiya Nugaliyadde, et Samantha Thelijagoda. Linguistic features based personality recognition using social media data. Dans *2017 6th National Conference on Technology and Management (NCTM)*, pages 63–68. IEEE, 2017.
- Md Shajalal et Masaki Aono. Semantic textual similarity between sentences using bilingual word semantics. *Progress in Artificial Intelligence*, 8 :263–272, 2019.
- Cheng-Che Shen, Li-Yu Hu, et Ya-Han Hu. Comorbidity study of borderline personality disorder : applying association rule mining to the taiwan national health insurance research database. *BMC medical informatics and decision making*, 17(1) : 1–10, 2017.
- Rebecca L Shiner, Kristin A Buss, Sandee G McClowry, Samuel P Putnam, Kimberly J Saudino, et Marcel Zentner. What is temperament now ? assessing progress in temperament research on the twenty-fifth anniversary of goldsmith et al.(.). *Child Development Perspectives*, 6(4) :436–444, 2012.
- Huayou Si, Jiayong Zhou, Zhihui Chen, Jian Wan, Neal N Xiong, Wei Zhang, et Athanasios V Vasilakos. Association rules mining among interests and applications for users on social networks. *IEEE Access*, 7 :116014–116026, 2019.
- Sima Siami-Namini, Neda Tavakoli, et Akbar Siami Namin. The performance of lstm and bilstm in forecasting time series. Dans *2019 IEEE International Conference on Big Data (Big Data)*, pages 3285–3292. IEEE, 2019.
- Martha Silva, Jonathan Walker, Erin Portillo, Leanne Dougherty, et al. Strengthening the merci mon héros campaign through adaptive management : Application of social listening methodology. *JMIR Public Health and Surveillance*, 8(6) : e35663, 2022.
- Ravinder Singh, Jiahua Du, Yanchun Zhang, Hua Wang, Yuan Miao, Omid Ameri Sianaki, et Anwaar Ulhaq. A framework for early detection of antisocial behavior on twitter using natural language processing. Dans *Conference on Complex, Intelligent, and Software Intensive Systems*, pages 484–495. Springer, 2019.
- Andrew E Skodol. Personality disorders in dsm-5. *Annual review of clinical psychology*, 8 :317–344, 2012.
- Harun Olcay Sonkurt, Ali Ercan Altınöz, Emre Çimen, Ferdi Köşger, et Gürkan Öztürk. The role of cognitive functions in the diagnosis of bipolar disorder : a machine learning model. *International Journal of Medical Informatics*, 145 :104311, 2021.
- Christopher J Soto et Joshua J Jackson. *Five-factor model of personality*. Oxford University Press, 2013.

- Othman Soufan, Wail Ba-Alawi, Moataz Afeef, Magbubah Essack, Valentin Rodionov, Panos Kalnis, et Vladimir B Bajic. Mining chemical activity status from high-throughput screening assays. *PLoS one*, 10(12) :e0144426, 2015.
- Efstathios Stamatatos, Martin Potthast, Francisco Rangel, Paolo Rosso, et Benno Stein. Overview of the pan/clef 2015 evaluation lab. Dans *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 518–538. Springer, 2015.
- Maxim Stankevich, Andrey Latyshev, Evgenia Kuminskaya, Ivan Smirnov, et Oleg Grigoriev. Depression detection from social media texts. Dans *Data Analytics and Management in Data Intensive Domains : 1 International Conference DAMDID/RCDL'2019, 2019.*–496 , page 352, 2019.
- Maxim Stankevich, Ivan Smirnov, Natalia Kiselnikova, et Anastasia Ushakova. Depression detection from social media profiles. Dans *Data Analytics and Management in Data Intensive Domains : 21st International Conference, DAMDID/RCDL 2019, Kazan, Russia, October 15–18, 2019, Revised Selected Papers 21*, pages 181–194. Springer, 2020.
- Jérémie Sublime et Ekaterina Kalinicheva. Automatic post-disaster damage mapping using deep-learning techniques for change detection : Case study of the tohoku tsunami. *Remote Sensing*, 11(9) :1123, 2019.
- Heni Sulistiani, Kurnia Muludi, et Admi Syarif. Implementation of various artificial intelligence approach for prediction and recommendation of personality disorder patient. Dans *Journal of Physics : Conference Series*, volume 1751, page 012040. IOP Publishing, 2021.
- Prabha Sundaravadivel, Vividha Goyal, et Lakshman Tamil. i-rise : An iot-based semi-immersive affective monitoring framework for anxiety disorders. Dans *2020 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–5. IEEE, 2020.
- Inette Swart. New developments in neuroscience can benefit the learning and performance of music. *Muziki*, 13(1) :113–136, 2016.
- Irfan Tanoli, Sebastião Pais, João Cordeiro, et Muhammad Luqman Jamil. Detection of radicalisation and extremism online : A survey. 2022.
- K Thaiyalnayaki. Classification of diabetes using deep learning and svm techniques. *International Journal of Current Research and Review*, 13(01) :146, 2021.
- thetechplatform. Cnn vs. rnn vs. ann – analyzing 3 types of neural networks in deep learning. <https://www.thetechplatform.com/post/cnn-vs-rnn-vs-ann-analyzing-3-types-of-neural-networks-in-deep-learning>, 2021. Accessed : 2021-04-30.
- François Torregrossa, Robin Allesiardo, Vincent Claveau, Nihel Kooli, et Guillaume Gravier. A survey on training and evaluation of word embeddings. *International Journal of Data Science and Analytics*, 11(2) :85–103, 2021.

- Joseph Triebwasser, Eran Chemerinski, Panos Roussos, et Larry J Siever. Paranoid personality disorder. *Journal of personality disorders*, 27(6) :795, 2013.
- Piyush Tyagi. Clustering, 2018. URL <https://medium.com/@pytyagi/clustering-83f210299e47>.
- Anila Umar et Usman Qamar. Detection and diagnosis of psychological disorders through decision rule set formation. Dans *2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA)*, pages 33–37. IEEE, 2019.
- Vanshika Varshney, Aman Varshney, Tameem Ahmad, et Asad M Khan. Recognising personality traits using social media. Dans *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, pages 2876–2881. IEEE, 2017.
- Zannatun Nayem Vasha, Bidyut Sharma, Israt Jahan Esha, Jabir Al Nahian, et Johora Akter Polin. Depression detection in social media comments data using machine learning algorithms. *Bulletin of Electrical Engineering and Informatics*, 12(2) :987–996, 2023.
- Abhishek Velankar, Hrushikesh Patil, et Raviraj Joshi. Mono vs multilingual bert for hate speech detection and text classification : A case study in marathi. *arXiv preprint arXiv :2204.08669*, 2022.
- Denny Vrandečić et Markus Krötzsch. Wikidata : a free collaborative knowledge-base. *Communications of the ACM*, 57(10) :78–85, 2014.
- Mohammed Abdul Wajeed et Vallamchetty Sreenivasulu. Image based tumor cells identification using convolutional neural network and auto encoders. *Traitement du Signal*, 36(5), 2019.
- Chang Wang, Bang Wang, et Minghua Xu. Tree-structured neural networks with topic attention for social emotion classification. *IEEE Access*, 7 :95505–95515, 2019a.
- JunJie Wang, Zheng Chen, Wei Sun, YuChen Zheng, Li Hui, XiaoHua Liu, JiJun Wang, HaiSu Wu, Yi Qiao, et TianHong Zhang. A strategy to address high comorbidity of personality disorders in a chinese population : A principal and subordinate diagnostic model. *Psychiatry*, 82(3) :272–282, 2019b.
- Martin Wattenberg, Fernanda Viégas, et Ian Johnson. How to use t-sne effectively. *Distill*, 1(10) :e2, 2016.
- W Wu, LJ KINGWenxia Wu, et Leslie J King. Artificial intelligence and elearning 4.0. *Emerging Web 3.0/semantic Web applications in higher education : growing personalization and wider interconnections in learning*, page 81, 2015.
- Yang Xian et Yingli Tian. Self-guiding multimodal lstm—when we do not have a perfect training dataset for image captioning. *IEEE Transactions on Image Processing*, 28(11) :5241–5252, 2019.

- Di Xue, Lifa Wu, Zheng Hong, Shize Guo, Liang Gao, Zhiyong Wu, Xiaofeng Zhong, et Jianshan Sun. Deep learning-based personality recognition from text posts of online social networks. *Applied Intelligence*, 48(11) :4232–4246, 2018.
- Marjon Zamani, Tatum Wilhelm, et Ariel L Furst. Perspective—electrochemical sensors for neurotransmitters and psychiatrics : Steps toward physiological mental health monitoring. *Journal of The Electrochemical Society*, 169(4) :047513, 2022.
- Shichao Zhang, Debo Cheng, Zhenyun Deng, Ming Zong, et Xuelian Deng. A novel knn algorithm with data-driven k parameter computation. *Pattern Recognition Letters*, 109 :44–54, 2018.
- Xiao Zhang, Wenzhong Li, Haochao Ying, Feng Li, Siyi Tang, et Sanglu Lu. Emotion detection in online social networks : a multilabel learning approach. *IEEE Internet of Things Journal*, 7(9) :8133–8143, 2020.
- Quan Zheng, Ying Guo, Zhen Wang, Frank Andrasik, Ziyi Kuang, Junyi Li, Sheng Xu, et Xiangen Hu. Exploring weibo users' attitudes toward lesbians and gays in mainland china : A natural language processing and machine learning approach. *Computers in Human Behavior*, 127 :107021, 2022.

Résumé Le trouble de la personnalité est une maladie qui prend de l'ampleur au quotidien. Son impact négatif sur les personnes atteintes, ainsi que sur leur entourage et même sur les économies des pays, est une source d'agacement pour tous les acteurs impliqués. Plusieurs outils existants permettent d'aborder cette problématique à travers une analyse du comportement des patients dans un environnement réel et virtuel. Néanmoins, les outils actuellement disponibles pour les environnements virtuels sont limités à la détection de quelques symptômes ou troubles psychologiques de manière indépendante, et se concentrent uniquement sur la détection plutôt que sur la surveillance. Dans cette thèse, nous avons proposé une approche visant à détecter et surveiller les personnes souffrant de troubles psychologiques en analysant leurs productions textuelles en français, sur les réseaux sociaux. Dans la première contribution, nous avons proposé trois méthodes permettant la détection des TP sur les réseaux sociaux en combinant des techniques linguistiques et numériques. Dans la deuxième contribution, nous avons proposé une ontologie permettant de détecter pour chaque malade les différents facteurs qui ont influencé négativement son comportement. La troisième méthode consiste à filtrer des tweets et à recommander d'autres en fonction de l'état du malade en utilisant un analyseur sémantique basé sur des techniques symboliques et d'apprentissage profond. Les résultats obtenus en termes de F-mesure par rapport à chaque méthode sont respectivement 74%, 92%, 75%.

Mots-clés Trouble de la personnalité, Réseaux sociaux, Apprentissage profond, Traitement du langage naturel, exploration du texte, Analyse sémantique

Abstract A personality disorder is a sickness that is becoming more common on a daily basis. Their negative impact on the affected persons, as well as on their entourage and even on the economies of the countries, is a source of annoyance for all the actors involved. There are several technologies available to solve this issue by analyzing patient behavior in both a real and virtual environment. Nevertheless, the tools currently available for virtual environments, they are restricted to the diagnosis of a few psychiatric symptoms or illnesses, and they are primarily focused on detecting rather than monitoring. In this thesis, we proposed an approach to detect and monitor people suffering from psychological disorders by analyzing their French textual production on social networks. In the first part, we proposed three methods allowing the detection of PD on social networks by combining linguistic and numerical techniques. For the second part, we proposed a method based on an ontology allowing to detect for each patient the different factors that negatively influenced his behavior. The third method consists in filtering tweets and recommending others according to the patient's state using a semantic analyzer based on symbolic and deep learning techniques. The main results obtained from the F-measure metric are respectively 74%, 92%, 75%.

Keywords Personality disorder, Social networks, Deep learning, Natural language processing, Text mining, Semantic analysis