



**HAL**  
open science

# Détection des humeurs dépressives sur les réseaux sociaux chinois à partir d'une combinaison de plongements lexicaux et de méthodes textométriques

Jinyuan Xu, Pierre Magistry, Mathieu Valette

## ► To cite this version:

Jinyuan Xu, Pierre Magistry, Mathieu Valette. Détection des humeurs dépressives sur les réseaux sociaux chinois à partir d'une combinaison de plongements lexicaux et de méthodes textométriques. JADT 2024 : 17th International Conference on Statistical Analysis of Textual Data, 25-27 juin 2024 Bruxelles (Belgique), Jun 2024, Bruxelles, Belgique. <hal-04687366>

**HAL Id: hal-04687366**

**<https://hal.science/hal-04687366v1>**

Submitted on 4 Sep 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.



Distributed under a Creative Commons CC0 1.0 - Universal - International License

# Détection des humeurs dépressives sur les réseaux sociaux chinois à partir d'une combinaison de plongements lexicaux et de méthodes textométriques

Jinyuan XU<sup>1</sup>, Pierre Magistry<sup>2</sup>, Mathieu Valette<sup>3</sup>

<sup>1</sup>Ertim, Inalco – jinyuan.xu@inalco.fr

<sup>2</sup>Ertim, Inalco – pierre.magistry@inalco.fr

<sup>3</sup>Ertim, Inalco – mathieu.valette@inalco.fr

## Abstract

Studies aimed at detecting users at risk of depression on Chinese social media generally rely on metadata or limited textual characteristics of the users, as well as large language models, foregoing any semantic analysis and ignoring the linguistic expression of emotions. We propose an architecture for extracting and analyzing semantic features based on textual statistics to address this issue. Our corpus is derived from the Sina Weibo Depression Dataset. We will use textual statistics methods to analyze the corpus with the goal of constructing an interpretable semantic model, to which we will then combine machine learning and deep learning methods. The results will be analyzed to determine the contribution of the semantic model to detection and its interpretability.

**Keywords :** detection, classification, specificity calculation, repeated segments, semantic model, machine learning, deep learning, depression, Weibo, Chinese

## Résumé

Les études pour détecter les utilisateurs à risque de dépression sur les réseaux sociaux chinois se basent généralement sur les métadonnées ou les caractéristiques textuelles limitées des utilisateurs et de grands modèles de langues (LLMs). Elles s'affranchissent de toute analyse sémantique et négligent l'expression linguistique des émotions. Nous proposons une architecture d'extraction et d'analyse des caractéristiques sémantiques basée sur les statistiques textuelles pour remédier à ce problème. Notre corpus est extrait du jeu de données *Sina Weibo Depression Dataset*. Nous utiliserons les méthodes de statistiques textuelles pour l'analyse du corpus avec l'objectif d'en construire un modèle sémantique interprétable à partir duquel nous combinons ensuite des méthodes d'apprentissage automatique et d'apprentissage profond. Les résultats seront analysés pour déterminer l'apport du modèle sémantique à la détection et son interprétabilité.

**Mots clés :** détection, classification, calcul de spécificités, segments répétés, modèle sémantique, apprentissage automatique, apprentissage profond, dépression, Weibo, chinois

## 1. Introduction

D'après un rapport de l'Organisation mondiale de la Santé diffusé en mars 2023<sup>1</sup>, près de 300 millions de personnes dans le monde sont affectées par la dépression. En Chine, le nombre de patients souffrant de dépression dépasse les 95 millions, avec une prévalence élevée parmi les jeunes âgés de 15 à 34 ans (Fu et al., 2021). On observe qu'ils privilégient l'expression de leurs états mentaux sur les réseaux sociaux plutôt que le dialogue direct avec leurs proches ou la consultation d'un spécialiste. Aussi le repérage des signes précurseurs de la dépression sur les plateformes s'avère-t-il un enjeu sanitaire et social majeur.

Notre étude vise à combiner une analyse linguistique basée sur le calcul de spécificités de (Lafon 1980) et des plongements lexicaux générés par un modèle de type *Bidirectional*

---

<sup>1</sup> <https://www.who.int/fr/news-room/fact-sheets/detail/depression> consulté le 7 février 2024

*Encoder Representations from Transformers* (BERT) (Devlin et al., 2019) pré-entraîné, *Robustly Optimized BERT Pretraining Approach* (RoBERTa) (Liu et al., 2019), pour détecter les utilisateurs à risque sur *Sina Weibo*, le plus important réseau social en Chine. En associant une méthode de statistique textuelle largement éprouvée avec un modèle à l'état de l'art du Traitement Automatique des Langues (TAL), notre but est double : il s'agit d'une part d'obtenir une vision réflexive et un contrôle sur les objets sémantiques pertinents dans une tâche de détection et, d'autre part, d'amorcer une réduction des coûts énergétiques par rapport aux méthodes basées sur les LLMs actuellement dominantes. Autrement dit, il s'agit de réduire significativement les obstacles techniques pour des recherches interdisciplinaires qui n'ont pas accès à de grandes capacités de calcul, tout en garantissant une certaine capacité d'interprétation des données textuelles et des résultats expérimentaux de bonne qualité.

## 2. État de l'art

De nombreux travaux ont exploré l'application des méthodes d'apprentissage automatique et d'apprentissage profond pour la détection de l'humeur dépressive. Ces études se focalisent sur deux opérations principales : d'une part, la classification binaire de l'humeur dépressive, généralement réalisée à l'échelle individuelle et, d'autre part, la détection des symptômes dépressifs, nécessitant souvent l'expertise de spécialistes en raison de leur association avec des diagnostics pathologiques.

Les rapports de patients hospitalisés sont une source essentielle pour la collecte de corpus d'humeurs dépressifs (Fernandes et al., 2018 ; Bittar et al., 2019), basée sur l'identification des symptômes. Les entretiens enregistrés et les questionnaires de volontaires, utilisés couramment en psychologie (Shen et al., 2022 ; Li et al., 2022), incluent des informations vocales et textuelles, ainsi que des indications sur les symptômes, souvent exploités dans des modèles multimodaux ou pour la fusion des caractéristiques (Shen et al., 2022). Cependant, l'accès à ces données hospitalières sensibles est rendu difficile, notamment pour des raisons de déontologie médicale. Par ailleurs, le recrutement de patients volontaires s'avère coûteux et difficile, et la limitation en termes de volume et de diversité des données peut impacter l'analyse des expressions linguistiques. En revanche, le pseudonymat, la diversité et la quantité des données disponibles sur les réseaux sociaux (Eichstaedt et al., 2018 ; Lin et al., 2020), offrent une des opportunités de constitution de corpus. Ces données, principalement textuelles, sont particulièrement adaptées pour l'analyse sémantique.

Selon le corpus et les axes de recherche, les stratégies de détection des humeurs dépressives varient. Initialement, des caractéristiques statistiques traditionnelles étaient fréquemment utilisées en association avec des algorithmes classiques d'apprentissage automatique, incluant les sacs de mots, les n-grammes, les méta-données (Eichstaedt et al., 2018; Bittar et al., 2019) et le tf-idf. Pour améliorer la détection, de nombreux chercheurs ajoutent des marqueurs linguistiques en se basant sur les travaux antérieurs, tels que l'utilisation de lexiques et de grammaires via l'étiquetage morphosyntaxique, *Linguistic Inquiry and Word Count* (LIWC<sup>2</sup>), les mesures de lisibilité ou la complexité textuelle (Trotzek et al., 2017), tels que *Gunning Fog Index* (Gunning, 1952), *Flesch Reading Ease* (Flesch, 1948), ainsi que divers traits pragmatiques (Yang et al., 2020). Avec l'avènement de l'apprentissage profond, les plongements lexicaux (*word embeddings*) tels que *Word2vec* (Mikolov et al., 2013), sont devenus l'outil privilégié pour la vectorisation des textes. Pour capturer davantage d'informations pour la détection des émotions, l'utilisation du *Long Short-Term Memory* (LSTM) (Sak et al., 2014) et de ses variantes s'est récemment répandue (Li et al., 2022 ; Shen

<sup>2</sup> <https://www.liwc.app/> consulté le 7 février 2024

et al., 2022). La combinaison de *Convolutional Neural Network* (CNN) (Kim, 2014) et LSTM a également été utilisée (Batbaatar et al., 2019). Depuis l'apparition de la structure *Transformer* (Vaswani et al., 2017), les mécanismes d'attention et les plongements contextuels dominent les tâches de classification de textes et de détection d'émotions, souvent réalisées avec BERT (Devlin et al., 2019) ou ses variantes (Tavchioski et al., 2022 ; Lin et al., 2020 ; Kaseb et al., 2022). Actuellement, les LLMs gagnent en notoriété et sont de plus en plus utilisés pour détecter les émotions (Yang et al., 2023 ; Danner et al., 2023). Par ailleurs, des stratégies innovantes sont explorées, telles que l'analyse de similarité (Pérez et al., 2023), l'approche de séries temporelles (Cai et al., 2023), et la combinaison de l'attention avec d'autres réseaux neuronaux (Li et al., 2022).

Dans notre article, nous utiliserons le calcul de spécificités (Lafon, 1980) pour établir une catégorisation sémantique qui intègre des représentations explicables, et nous la combinerons avec des plongements. Initialement appliqué dans le domaine des études littéraires et des sciences humaines (Brunet, 1988 ; Salem et Wu, 2008), le calcul de spécificités s'est avéré efficace dans d'autres domaines, tels que la fouille des opinions et la classification (Eensoo et Valette, 2015 ; Baiocchi, 2019).

Comparé aux autres méthodes d'apprentissage profond et aux LLMs, le calcul de spécificités offre une bonne explicabilité sémantique des caractéristiques textuelles extraites. Il permet notamment une sélection automatique et une "pré-qualification" des unités lexicales non pas basée sur des critères de bas niveau (fréquences absolues, mots trop rares, mots trop fréquents) ou sur des préjugés du TAL comme la suppression des mots dit "vides", lesquels peuvent être porteurs d'un sémantisme pertinent dans l'expression des états mentaux. Par ailleurs, l'intégration de n-grammes, (ou *segments répétés* (Salem, 1986)) apporte un contexte utile, condition minimale de l'interprétation, tout en nécessitant là encore moins de ressources en termes de calcul. La méthode de combinaison peut servir de référence pour les psychologues, en raison de sa grande capacité explicative.

Si la majorité des jeux de données disponibles sont en anglais (Yates et al., 2017), des corpus existent aussi en chinois (Wang et al., 2022 ; Cai et al., 2023), en japonais, en espagnol. La plupart des chercheurs se réfèrent pour annoter leurs données à des référentiels tels que le *Diagnostic and Statistical Manual* (DSM) (Kupfer et al., 2008), le *Beck Depression Inventory* (BDI) (Beck et al., 1996) et le *Patient Health Questionnaire* (PHQ) (Mitchell et al., 2016), ou utilisent comme référence les utilisateurs qui se déclarent eux-mêmes dépressifs.

### 3. Corpus et préparation des données

Notre corpus de recherche provient du jeu de données *Sina Weibo Depression Dataset* (SWDD) (Cai et al., 2023). Il a été extrait de *Sina Weibo*<sup>3</sup> qui compte plus de 500 millions d'utilisateurs actifs. La méthode décrite par les auteurs pour constituer ce corpus inclut une sélection initiale des utilisateurs basée sur des mots-clés relatifs à la dépression, suivie d'une annotation effectuée par des experts. Elle se fonde sur le référentiel DSM-5<sup>4</sup>.

Afin d'adapter les données du SWDD à notre objectif, nous avons effectué l'extraction suivante ; (i) exclusion des utilisateurs ayant publié moins de 80 messages ; (ii) sélection de 1099 utilisateurs autodéclarés dépressifs et étiquetés comme dépressifs dans le jeu de données original ; (iii) choix aléatoire de 1099 utilisateurs qui ne se sont pas déclarés dépressifs (mais

---

<sup>3</sup> <https://s.weibo.com/> consulté le 7 février 2024

<sup>4</sup> <https://www.psychiatry.org/psychiatrists/practice/dsm/educational-resources/dsm-5-fact-sheets> consulté le 7 février 2024

qui peuvent l'être) classés par nous dans une catégorie « Non Dépressif » dans le nouveau jeu de données (*Tableau 1*) ; (iv) suppression de tous les messages dont l'intervalle de publication excède six mois et élimination des balises HTML présentes dans les messages.

La présence de messages publicitaires, caractérisés par leur répétitivité et similarité textuelle, peut induire des biais dans l'analyse statistique des textes. À cet égard, notre calcul de la spécificité de Lafon sur les segments répétés dans les données originales a mis en lumière que de nombreux n-grammes publicitaires affichent aussi des valeurs élevées. Pour l'éviter, nous avons annoté 3500 messages publicitaires et 3500 messages non publicitaires en vue du réglage précis (*fine tuning*) du modèle Sentence-BERT en version chinoise (Reimers et al., 2019). Ce processus a permis d'atteindre un taux d'exactitude de 94% sur le jeu de test pour la détection de messages publicitaires. Nous avons ensuite utilisé ce modèle pour purger l'ensemble du corpus des contenus publicitaires. Bien que notre jeux de données contienne encore une petite partie de tweets publicitaires, la quantité de données disponibles diminue l'impact de ces tweets publicitaires sur notre étude, à telle enseigne qu'ils peuvent être considérés comme non significants.

Dataset	Catégorie	Utilisateurs	Messages
SWDD	Dépressif	3 711	785 689
	Non Dépressif	19 526	4 068 732
	Total	23 237	4 854 421
SWDD-filtred	Dépressif	1 099	194 323
	Non Dépressif	1 099	205 513
	Total	2 198	399 836

*Tableau 1. Jeux de données*

La segmentation du chinois, ou tokenisation, est une étape complexe car le chinois n'utilise pas l'espace comme séparateur de mot, une analyse morphosyntaxique s'avère nécessaire. Nous avons évalué divers outils de segmentation (dont SpaCy<sup>5</sup>, Jieba<sup>6</sup>, LTP<sup>7</sup> et HANLP<sup>8</sup>) et opté pour le mode "fine" de HANLP. Nous avons ensuite constitué un dictionnaire de segments répétés, basé sur la longueur des groupes de tokens dans notre corpus.

#### 4. Ingénierie des caractéristiques

Nous réservons les messages de 70 % des utilisateurs pour les données d'entraînement et 30 % pour les données de test. Les statistiques des métadonnées et les indices de spécificité des segments répétés que nous mentionnerons par la suite proviennent des seules données du corpus d'entraînement.

On aura noté que nous considérons l'utilisateur comme l'unité d'analyse principale et non le message. Une analyse statistique préliminaire a permis d'étudier des informations métatextuelles susceptibles d'être porteuses de sémantisme (longueurs de textes, intervalles de publication, heure de la publication, adresse IP et métadonnées disponibles) sans que ces observations n'aient donné à voir des tendances remarquables. La différenciation la plus notable est la distribution horaire des publications, avec une activité accrue pendant la nuit (de 23h00 à 07h00) chez les utilisateurs présentant des humeurs dépressives comparés aux utilisateurs de contrôle, ce que corroborent les études de psychologies liant insomnie et

<sup>5</sup> <https://spacy.io/> consulté le 7 février 2024

<sup>6</sup> <https://github.com/fxsjy/jieba?tab=MIT-1-ov-file> consulté le 7 février 2024

<sup>7</sup> <https://ltp.ai/docs/api/ltp.html> consulté le 7 février 2024

<sup>8</sup> <https://hanlp.hankcs.com/docs/index.html> consulté le 7 février 2024

dépression (Nutt et al., 2008). Par conséquent, la proportion de publications nocturnes est identifiée comme un indicateur.

Des recherches récentes ont mis en évidence que les classificateurs fondés sur RoBERTa surpassent en efficacité d'autres architectures basées sur BERT pour la détection et la classification de troubles de santé mentale (Murarka et al., 2021). Nous avons donc opté pour l'utilisation de ce modèle pour procéder à l'extraction de caractéristiques au sein de notre corpus. Dans notre protocole expérimental, les données en entrée sont des phrases, chaque phrase donne en sortie un plongement de dimension 768, capturant les informations sémantiques. Après avoir obtenu les vecteurs de chaque phrase pour un utilisateur, nous calculons la moyenne sur chacune des dimensions de ces vecteurs pour obtenir une représentation vectorielle unique de dimension 768 reflétant les caractéristiques textuelles propres à cet utilisateur. Cette démarche a pour objectif de synthétiser une représentation sémantique globale des contributions textuelles de l'utilisateur, permettant une exploration approfondie des signes précurseurs de pathologies mentales éventuellement présents dans le contenu produit par l'utilisateur.

Pour améliorer notre objectif de fournir une compréhension sémantique approfondie des données, après avoir examiné les corpus, nous avons opté pour l'analyse des segments répétés jusqu'à une longueur de 7 tokens. L'application de l'indice de spécificité (Lafon, 1980) nous permet d'évaluer la pertinence des liens entre ces segments répétés et le corpus étudié, afin de dégager des perspectives plus précises et réduire les biais causés par les tokens à haute fréquence.

Le calcul de spécificités se formule ainsi.

A : l'événement recensé ; V : l'ensemble des événements possibles (le vocabulaire pour les mots) ; p : la partie considérée ; f : la fréquence de l'événement dans la partie ; F : la fréquence totale de l'événement dans le corpus ; t : le nombre total d'événements ayant lieu dans la partie ; T : le nombre total d'événements ayant lieu dans l'ensemble des parties.

$$Prob_{spécif}(card\{A \in V | A \in p\} = f) = \frac{C_F^f \times C_{T-F}^{t-f}}{C_T^t}$$

Équation 1. Probabilité d'apparition dans une partie

$$Prob_{spécif}(card\{A \in V | A \in p\} \geq f_{obs}) = \sum_{f=f_{obs}}^{card\{A \in V | A \in p\}} Prob_{spécif}(card\{A \in V | A \in p\} = f)$$

Équation 2. Indice de spécificités (Lafon 1980)

Nous avons calculé l'indice de spécificités pour tous les segments répétés de longueur 1 à 7. Afin d'améliorer leur présentation, nous avons appliqué une transformation logarithmique (base 10) à ces indices, une approche similaire à l'implémentation de (Lafon 1980) dans TXM (Heiden et al. 2010). Seuls les segments dont la valeur de pertinence textuelle dépasse 10 ont été retenus, ainsi que ceux présentant des entités nommées relatives à la dépression, comme les noms des médicaments.

En nous appuyant sur des études en psycholinguistique, des questionnaires sur la dépression et des analyses de corpus, et en tenant compte des particularités des médias sociaux et du langage chinois en ligne, nous avons filtré et réparti les segments répétés extraits dans 7 catégories sémantiques définies. (Voir les références dans le *Tableau 2*)

Catégorie Sémantique	Exemples de segments répétés	Nombre
1. Perception subjective (Kupfer et al., 2008 ; Beck et al., 1996 ; Mitchell et al., 2016)		247
1.1. Émotions négatives	抑郁 - « dépression » 焦虑 - « anxiété » 不开心 - « pas content »	97
1.2. État psychologique	做不到 - « incapable de le faire »	16
1.2.1. Perte d'estime de soi		13
1.2.2. Sentiment de culpabilité	对不起 - « désolé »	72
1.2.3. Idées de suicide, d'automutilation et de mort	自杀 - « suicide »	49
1.2.4. Description psychologique négative	绝望 - « désespoir »	
2. Symptômes cliniques (Kupfer et al., 2008 ; Beck et al., 1996 ; Mitchell et al., 2016)		81
2.1. Changement de poids ou d'appétit	厌食 - « anorexie »	4
2.2. Troubles du sommeil	失眠 - « insomnie »	27
2.3. Fatigue corporelle et perte d'énergie	累 - « fatigué »	14
2.4. Malaise et douleur physique	头疼 - « mal de tête »	32
2.5. Hallucinations et manque de concentration	幻听 - « hallucinations auditives »	4
3. Causes externes potentielles		24
3.1. Mariage et relations affectives	离婚 - « divorce »	15
3.2. Famille	原生家庭 - « famille d'origine »	7
3.3. Relations sociales	休学 - « arrêter les études »	2
4. Médical		112
4.1. Liées aux médicaments	舍曲林 - « Sertraline »	43
4.2. Liées au traitement médical	心理咨询 - « consultation psychologique »	28
4.3. Liées à la maladie	重度抑郁 - « dépression majeure »	41
5. Formulation linguistique (Trifu et al., 2017)		128
5.1. Formulation volontaire	不想死 - « ne pas vouloir mourir »	25
5.2. Formulation négative	我不知道 - « je ne sais pas »	57
5.3. Formulation absolue	任何人 - « n'importe qui »	9
5.4. Formulation interrogative	为什么 - « pourquoi »	23
5.5. Discours insultant	傻逼 - « connard »	14
6. Expression grammaticale - structure grammaticale contenant un Pron. (pronom) de la 1ère personne (Bernard et al., 2016)		115
6.1. Sujet (optionnel) + V. (Verbe) + C.O. (complément d'objet) (Pron. de la 1ère personne) ( <i>phrase impérative possible</i> )	救我 - « sauve-moi »	17
6.2. Relation de subordination	我的错 - « mon erreur »	13
6.3. Sujet (Pron. de la 1ère personne) + V. + C.O. (optionnel)	我害怕 - « j'ai peur »	16
6.4. Sujet (Pron. de la 1ère personne) + Négation + V. + C.O. (optionnel)	我不想 - « je ne veux pas »	14
6.5. Sujet (Pron. de la 1ère personne) + V. + C.O. (Verbe ou Nom)	我想死 - « je veux mourir »	6
6.6. (Sujet (Pron. de la 1ère personne) + adv.) ou (adv. + Sujet (Pron. de la 1ère personne)) + (manque des objets)	我真的 - « je suis vraiment »	21
6.7. Conjonction + Sujet (Pron. de la 1ère personne) + (manque des objets)	可是我 - « mais moi »	6
6.8. Sujet (Pron. de la 1ère personne) + Verbe copule + Attribut du sujet	我好难受 - « je me sens très mal »	15
6.9. Sujet (Pron. de la 1ère personne) + V. Int.	我死 - « je meurs »	7

7.	Sans catégorie	喝酒 – « boire de l'alcool »	7
		<b>Au Total</b>	<b>714</b>

Tableau 2. Catégorisation sémantique

## 5. Expérimentations et discussions

Afin de découvrir le potentiel des caractéristiques statistiques explorées pour améliorer les performances des plongements, nous avons sélectionné deux caractéristiques pour analyse : la proportion de messages postés la nuit et les catégories sémantiques induites par l'indice de spécificité de Lafon. Sur cette base, nous avons conçu six combinaisons de caractéristiques :

- 1). Utilisation exclusive des caractéristiques de RoBERTa ;
- 2). Utilisation exclusive des caractéristiques induites par la méthode de Lafon ;
- 3). Combinaison des caractéristiques de RoBERTa, des catégories sémantiques de Lafon et de la proportion de messages nocturnes ;
- 4). Combinaison des caractéristiques de Lafon avec celles de RoBERTa ;
- 5). Combinaison des caractéristiques de Lafon avec la proportion de messages nocturnes ;
- 6). Combinaison des caractéristiques de RoBERTa avec la proportion de messages nocturnes.

Pour évaluer de manière exhaustive l'utilité de ces combinaisons de caractéristiques, des expériences ont été menées sur cinq algorithmes d'apprentissage automatique, chacun configuré avec ses meilleurs paramètres pour chaque combinaison : *Support Vector Machine* (SVM) (Cortes et al., 1995), *eXtreme Gradient Boosting* (XGBoost) (Chen et al., 2016), *Random Forest* (RF) (Breiman, 2001), *K-Nearest Neighbors* (KNN) (Peterson, 2009), Régression Logistique (LR) (Yu & al., 2011) et un réseau de neurones profond (*Deep Neural Network* (DNN) avec trois couche cachés, configuré avec : optimiseur Adam, taux d'apprentissage de 0.001 ; fonction de perte de l'entropie croisée binaire ; 30 époques ; taille de batch de 32) (LeCun et al., 2015). La méthode de validation croisée k-fold a été utilisée pour assurer la précision et la fiabilité des résultats. Les résultats des expérimentations avec l'évaluation d'exactitude et de F1-mesure sur l'ensemble de test sont résumés au *Tableau 3*.

SVM	XGBoost	RF	KNN	LR	DNN	<b>Exactitude</b>
89.75%	88.80%	86.89%	<b>85.52%</b>	90.03%	89.21%	RoBERTa
81.28%	80.87%	81.42%	78.14%	81.15%	81.12%	Lafon_spécificité ( LS )
<b>91.26%</b>	<b>91.12%</b>	<b>88.25%</b>	<b>85.52%</b>	<b>91.39%</b>	<b>89.94%</b>	RoBERTa + LS + horaire nocturne ( HN )
83.74%	82.10%	82.38%	80.60%	82.79%	83.43%	LS + HN
89.21%	89.48%	87.16%	84.43%	89.89%	89.17%	RoBERTa + HN
90.30%	90.71%	87.57%	84.97%	90.16%	89.71%	RoBERTa + LS
SVM	XGBoost	RF	KNN	LR	DNN	<b>F1-score</b>
89.57%	88.71%	86.29%	<b>84.86%</b>	90.07%	93.07%	RoBERTa
81.31%	81.28%	81.87%	78.32%	81.60%	84.24%	LS
<b>91.28%</b>	<b>91.01%</b>	<b>87.75%</b>	84.77%	<b>91.43%</b>	93.02%	RoBERTa + LS + HN
83.85%	82.18%	82.50%	80.60%	82.93%	87.22%	LS + HN
89.19%	89.26%	86.57%	83.53%	89.92%	<b>93.19%</b>	RoBERTa + HN
90.29%	90.58%	86.98%	84.15%	90.19%	92.53%	RoBERTa + LS

Tableau 3. Résultats de caractères différents dans les divers algorithmes

Quoi qu'honorables, les résultats du calcul de spécificités apparaissent sensiblement moins bons que ceux de RoBERTa – on ne s'en étonnera guère, près d'un demi-siècle sépare ces deux méthodes mathématiques. On constate toutefois que la combinaison des caractéristiques

obtenues par calcul de spécificités, horaire nocturne et les plongements issus du modèle RoBERTa pré-entraîné contribue à une amélioration du score d'exactitude (*accuracy*) du modèle et la combinaison de certains indicateurs classiques et les plongements permettent d'atteindre ou de se rapprocher du meilleur score F1 du modèle.

## 6. Conclusion et perspectives

L'objectif de notre recherche est de développer un système de détection qui soit à la fois économe en ressources de calcul et doté d'une meilleure explicabilité. Les classificateurs basés sur BERT nécessitent un déploiement assez complexe pour faire le fine-tuning ; les LLMs basés sur le *Generative Pre-trained Transformer* (GPT) requièrent un bon support matériel et ont un coût significatif en termes de calcul. Notre approche se distingue par sa simplicité technique et ses faibles exigences matérielles, rendant les outils accessibles et économiques pour les chercheurs en humanités numériques, tout en assurant une interprétabilité suffisante du modèle. Notre approche, en associant textométrie et catégories sémantiques, ne se contente pas d'améliorer la précision de détection des thématiques de la dépression mais facilite aussi l'exploration sémantique du corpus. Dans nos futures recherches, nous envisageons d'améliorer notre modèle sémantique par une analyse approfondie des segments répétés et de combiner les catégories sémantiques avec le mécanisme d'attention (Vaswani et al., 2017) pour optimiser l'extraction des caractéristiques, adaptant ainsi mieux notre modèle à la diversité des plateformes de médias sociaux chinois.

## Bibliographie

- Baiocchi M. (2019). *Diversité et recommandation : une investigation sur l'apport de la fouille d'opinions pour la distinction d'articles d'opinion dans une controverse médiatique*, PhD thesis, Université de Montréal, 509 p.
- Brunet E. (1988). Le vocabulaire de Victor Hugo (*tome 1, 484 p.; tome 2, 637 p.; tome 3, 556 p. + 27 microfiches contenant l'Index synoptique de Hugo [6878 p.]*). Slatkine, Champion. (Numéro 38)
- Beck A. T., Steer R. A., Ball R. and Ranieri W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in psychiatric outpatients. *Journal of Personality Assessment*, 67 (3), 588-597.
- Breiman L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Bernard J. D., Baddeley J. L., Rodriguez B. F. and Burke P. A. (2016). Depression, Language, and Affect: An Examination of the Influence of Baseline Depression and Affect Induction on Language. *Journal of Language and Social Psychology*, 35 (3), 317–326.
- Batbaatar E., Li M. and Ryu K. H. (2019). Semantic-Emotion Neural Network for Emotion Recognition From Text. *IEEE Access*, 7, 111866-111878.
- Bittar A., Velupillai S., Roberts A. and Dutta R. (2019, August). Text Classification to Inform Suicide Risk Assessment in Electronic Health Records. *In MedInfo*, 40-44).
- Cortes C. and Vapnik V. (1995). Support-vector networks. *Machine Learning*, 20 (3), 273–297.
- Chen T. and Guestrin C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, San Francisco, 785-794.
- Cai Y., Wang H., Ye H., Jin Y. and Gao W. (2023). Depression detection on online social network with multivariate time series feature of user depressive symptoms. *Expert Systems with Applications*, 217, 119538.
- Devlin J., Chang M.-W., Lee K. and Toutanova K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the NAACL: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, Minnesota, 4171–4186.

- Danner M., Hadzic B., Gerhardt S., Ludwig S., Uslu, I., Shao P., Weber T., Shibani Y. and Ratsch M. (2023). Advancing Mental Health Diagnostics : GPT-Based Method for Depression Detection. In *2023 62nd Annual Conference of the Society of Instrument and Control Engineers*, 1290-1296.
- Eensoo E. and Valette M. (2015). Une méthodologie de sémantique de corpus appliquée à des tâches de fouille d'opinion et d'analyse des sentiments : Étude sur l'impact de marqueurs dialogiques et dialectiques dans l'expression de la subjectivité. *Actes de la 22e conférence sur le TALN*, Caen.
- Eichstaedt J. C., Smith R. J., Merchant R. M., Ungar L. H., Crutchley P., Preotiuc-Pietro D., Asch D. A. and Schwartz H. A. (2018). Facebook language predicts depression in medical records. In *Proceedings of the National Academy of Sciences*, 115 (44), 11203-11208.
- Flesch R. (1948). A new readability yardstick. *Journal of applied psychology*, 32 (3), 221-233.
- Fernandes A. C., Dutta R., Velupillai S., Sanyal J., Stewart R. and Chandran D. (2018). Identifying Suicide Ideation and Suicidal Attempts in a Psychiatric Clinical Research Database using Natural Language Processing. *Scientific Reports*, 8 (1), 7426.
- Fu X., Zhang K., Chen X. and Chen, Z. (2021). Report on the Development of National Mental Health in China (2019–2020) [M]. *Beijing: Social Sciences Academic Press*.
- Gunning R. (1952). The technique of clear writing. New-York, McGraw-Hill.
- Heiden S., Magué J. P. and Pincemin B. (2010, June). TXM: Une plateforme logicielle open-source pour la textométrie-conception et développement. In *10th JADT 2010*, Vol.2 (3), 1021-1032.
- Yu H.-F., Huang F.-L. and Lin C.-J. (2011). Dual coordinate descent methods for logistic regression and maximum entropy models. *Machine Learning* 85 (1-2), 41-75.
- Kupfer D. J., First M. B. and Regier D. A. (Eds). (2008). *A research agenda for DSM V*.
- Kim Y. (2014). Convolutional Neural Networks for Sentence Classification Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on EMNLP*, Doha, Qatar, 1746–1751.
- Kaseb A., Galal O. and Elreedy D. (2022). Analysis on Tweets Towards COVID-19 Pandemic : An Application of Text-Based Depression Detection. *2022 4th Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, Giza, 131-136.
- Lafon P. (1980). Sur la variabilité de la fréquence des formes dans un corpus. *Mots*, 1 (1), 127-165.
- LeCun Y., Bengio Y. and Hinton G. (2015). Deep learning. *Nature*, 521 (7553), 436-444.
- Liu Y., Ott M., Goyal N., Du J., Joshi M., Chen D., Levy O., Lewis M., Zettlemoyer L. and Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. DOI: 10.48550/arxiv.1907.11692
- Lewis P., Perez E., Piktus A., Petroni F., Karpukhin V., Goyal N. and Kiela D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.
- Lin C., Hu P., Su H., Li S., Mei J., Zhou J. and Leung, H. (2020). SenseMood : Depression Detection on Social Media. In *Proceedings of the 2020 International Conference on Multimedia Retrieval*, Dublin, 407-411.
- Li M., Xu H., Liu W. and Liu, J. (2022). Bidirectional LSTM and Attention for Depression Detection on Clinical Interview Transcripts. *2022 IEEE 10th ICICN*, 638-643.
- Mikolov T., Chen K., Corrado G. and Dean J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Mitchell A. J., Yadegarfar M., Gill J. and Stubbs B. (2016). Case finding and screening clinical utility of the Patient Health Questionnaire (PHQ-9 and PHQ-2) for depression in primary care: A diagnostic meta-analysis of 40 studies. *BJPsych Open*, 2 (2), 127-138.
- Murarka A., Radhakrishnan B. and Ravichandran, S. (2021). Classification of mental illnesses on social media using RoBERTa. In *Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis*, pages 59–68. DOI: 10.48550/arxiv.2011.11226.
- Nutt D., Wilson S. and Paterson L. (2008). Sleep disorders as core symptoms of depression. *Dialogues in clinical neuroscience*, 10 (3), 329-336.

- Peterson L. E. (2009). K-nearest neighbor. *Scholarpedia*, 4 (2), 1883.
- Pérez A., Warikoo N., Wang K., Parapar J. and Gurevych I. (2023). Semantic Similarity Models for Depression Severity Estimation. In *Proceedings of the 2023 Conference on EMNLP*, Singapore, 16104– 16118.
- Trifu R. N., Nemes B., Bodea-Hategan C. and Cozman D. (2017). Linguistic indicators of language in major depressive disorder (MDD). An evidence based research. *Journal of Evidence-Based Psychotherapies*, 17 (1), 105-128.
- Reimers N. and Gurevych I. (2019). Sentence-bert: Sentence embeddings using siamese bertnetworks. In *Proceedings of the 2019 Conference on EMNLP and the 9th IJCNLP*, Hong Kong, 3982– 3992.
- Salem A. (1986). Segments répétés et analyse statistique des données textuelles. *Histoire & Mesure*, 1 (2), 5-28.
- Salem A. and Wu L. C. (2008). Essai de textométrie politique chinoise. In Salem A. et Fleury S. (Eds), *Lexicométrica–Explorations textométriques*, 1
- Sak H., Senior A. and Beaufays F. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modeling. *Interspeech 2014*, 338-342.
- Shen T., Jia J., Shen G., Feng F., He X., Luan H., Tang J., Thanassis T., Chua T.-S. and Hall W. (2018). Cross-domain depression detection via harvesting social media. In *Proceedings of the 27th international joint conference on artificial intelligence*, Stockholm,. 1611–1617.
- Shen Y., Yang H. and Lin, L. (2022). Automatic Depression Detection : An Emotional Audio-Textual Corpus and A Gru/Bilstm-Based Model. *2022 IEEE ICASSP*, 6247-6251.
- Trotzek M., Koitka S. and Friedrich C. M. (2017). Linguistic metadata augmented classifiers at the CLEF 2017 task for early detection of depression. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum. Dublin, Ireland, September 11-14, 2017*.
- Tavchioski I., Koloski B., Škrlj B. and Pollak S. (2022). E8-IJS@LT-EDI-ACL2022—BERT, AutoML and Knowledge-graph backed Detection of Depression. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, Dublin, 251-257.
- Vaswani A., Shazeer N., Parmar, N., Uszkoreit J., Jones L., Gomez A. N., Kaiser L. and Polosukhin I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Wang X, Chen S, Li T, Li W, Zhou Y, Zheng J, Zhang Y. and Tang B. (2019). Assessing depression risk in Chinese microblogs: a corpus and machine learning methods. *2019 Presented at IEEE ICHI; June 10-13, 2019; Xi'an*.
- Wang, X., Chen, S., Li, T., Li, W., Zhou, Y., Zheng, J., Chen, Q., Yan, J., & Tang, B. (2020). Depression Risk Prediction for Chinese Microblogs via Deep-Learning Methods : Content Analysis. *JMIR Medical Informatics*, 8(7), e17958.
- Wang Y., Wang Z., Li C., Zhang Y. and Wang, H. (2022). Online social network individual depression detection using a multitask heterogenous modality fusion approach. *Information Sciences*, 609, 727–749.
- Yates A., Cohan A. and Goharian, N. (2017). Depression and self-harm risk assessment in online forums. In *Proceedings of the 14th conference on EMNLP Copenhagen*, 2968–2978.
- Yang X., McEwen R., Ong L. R. and Zihayat, M. (2020). A big data analytics framework for detecting user-level depression from social networks. *International Journal of Information Management*, 54, 102141.
- Yang K., Ji S., Zhang T., Xie Q., Kuang Z. and Ananiadou, S. (2023). Towards Interpretable Mental Health Analysis with Large Language Models. In *Proceedings of the 2023 Conference on EMNLP*, 6056-6077.