

Université de Blida 1 – SAAD DAHLAB



Faculté des sciences

Département d'Informatique

Mémoire présenté par :

MLLE. FERFERA Ahlem Yousra

MLLE. ZAMICHE Yamina

Pour l'obtention du diplôme de Master

Domaine : Mathématiques et Informatique

Filière : Informatique

Spécialité : Traitement Automatique de la Langue

Sujet :

**Conception et Réalisation d'un
ChatBot Psychologue
Multilingue**

Soutenu le : 04/10/2021

devant les membres du jury de thèse :

Mme. Zahra
Mme. CHERIGUENE
Mme M. MEZZI
M. M. LICHOURI

Université de Blida 1
Université de Blida 1
Université de Blida 1
CRSTDLA

Présidente
Examinatrice
Promotrice
Encadreur

Résumé

Un Chatbot est un programme informatique communiquant par texte de manière humaine et qui fournit des services aux utilisateurs humains afin d'accomplir un objectif bien défini. Grâce à des recherches à long terme sur l'intelligence artificielle, le nombre de cas d'utilisation possible des Chatbots augmente constamment.

L'objectif de ce mémoire était de créer un concept de système impliquant une application web afin de déduire la maladie psychologique de l'utilisateur et construire des diagnostics et des conseils. Pour tester l'approche, nous avons eu recours à des algorithmes de Machine Learning et Deep learning et nous avons particulièrement utilisé quatre classifieurs, deux algorithmes de Machine Learning : K-Nearest Neighbors Algorithm et Naive Bayes algorithm et deux algorithmes de Deep Learning : Le réseau neuronal récurrent (RNN), Les réseaux de neurones dits convolutifs (CNN). D'après les résultats obtenus, le meilleur algorithme était l'algorithme de Machine Learning Naive Bayes. C'est pourquoi il a été retenu pour l'application finale qui comprend, un questionnaire, un diagnostic et des Conseils aux patients.

Mots clés : Chatbot, Psychologie, Intelligence Artificielle, Machine Learning, Deep learning.

Abstract

Thanks to long-term research on artificial intelligence, the number of possible use cases for Chatbots is constantly increasing.

The objective of this thesis was to create a system concept involving a web application to deduce the user's psychological malady and construct diagnostics and conseils. To test the approach, we have used Machine Learning and Deep learning algorithms and we particularly used four classifiers, two Machine Learning algorithms: K-Nearest Neighbors Algorithm and Naive Bayes algorithm and two Deep Learning algorithms: Recurrent Neural Network (RNN), Convolutional Neural Networks (CNN). According to the results, the best algorithm was the Naive Bayes Machine Learning algorithm. Thus this later was retained to develop a simple web application that includes a questionnaire, a diagnose section, and gives suggestions.

Keywords: Chatbot, Psychology, Artificial Intelligence, Machine Learning, Deep learning.

ملخص

بفضل البحث طويل الأمد حول الذكاء الاصطناعي، يتزايد باستمرار عدد حالات الاستخدام المحتملة لروبوتات الدردشة.

الهدف من هذه الرسالة هو إنشاء مفهوم نظام يتضمن تطبيق ويب من أجل استنتاج المرض النفسي للمستخدم وإنشاء التشخيصات والحلول. لتطوير الحل الخاص بنا، استخدمنا خوارزميات التعلم الآلي والتعلم العميق واستخدمنا بشكل خاص أربع مصنفات وخوارزميتين للتعلم الآلي: خوارزمية Naive Bayes وخوارزمية K-Nearest Neighbours، وخوارزمتان للتعلم العميق: الشبكة العصبية المتكررة (RNN) وما يسمى بالشبكات العصبية التلافيفية (CNN).

بناءً على النتائج التي تم الحصول عليها، كانت أفضل خوارزمية هي Naive Baye. لذلك تم اعتمادها لتطوير تطبيق ويب، يشتمل على أسئلة وتشخيصات وحلول بسيطة.

الكلمات الرئيسية: الذكاء الاصطناعي، علم النفس، التعلم الآلي، والتعلم العميق، روبوتات الدردشة.

Remerciements

Dieu merci pour la santé, la volonté, le courage et la détermination qui nous ont accompagnés tout au long de la préparation et l'élaboration de ce travail et qui nous ont permis d'achever ce modeste travail.

Le présent travail est non seulement le résultat de notre courage, sacrifice, patience et endurance mais aussi une participation de plusieurs personnes qui nous sont chères. Son succès dépend d'une série d'interventions tant morales qu'intellectuelles.

Nous tenons d'abord à remercier infiniment notre encadreur M.Lichouri pour nous avoir offert l'opportunité d'un stage dans leur centre et pour ses efforts et ses immenses connaissances.

Nos remerciements s'adressent aussi à notre chère promotrice, Dr. Mme. Mezzi, pour son encadrement et ses constantes orientations de notre recherche en y accordant une méticuleuse attention, ainsi que pour ses conseils, sa disponibilité et son extrême amabilité malgré sa grande charge de travail.

Nous remercions également notre cher ami et camarade de promotion, Nadjemi Abdel wahab qui n'a jamais été avare de conseils et qui a toujours été dévouée pour nous aider. Il nous a énormément aidé à réaliser ce travail.

Que tous ceux qui d'une manière ou d'une autre, ont rendu possible la réalisation de ce travail trouvent ici l'expression de notre profonde gratitude.

Dédicaces

Je dédie ce modeste travail à :

À la mémoire de ma grand-mère,

Ce travail est pour moi le fruit de tes prières.

Puisse dieu, le tout puissant t'accueillir dans son vaste paradis.

À ma mère,

Tu m'as donné la vie, la tendresse et le courage pour réussir.

Tout ce que je peux t'offrir ne pourra exprimer l'amour et la reconnaissance que je te porte.

En témoignage, je t'offre ce modeste travail pour te remercier pour tes sacrifices et pour l'affection dont tu m'as toujours entourée.

À mon père,

L'épaule solide, l'œil attentif compréhensif et la personne la plus digne de mon estime et de mon respect.

Aucune dédicace ne saurait exprimer mes sentiments, que Dieu te préserve et te procure santé et longue vie.

À ma sœur douce Yasmine,

Que Dieu te garde, Je t'aime et je te souhaite une vie pleine de succès et de réussite.

À ma chère tante Nadia,

À mon cher binôme, ma sœur, et mon amie Amina.

Yousra.

Dédicaces

Avec l'expression de ma reconnaissance, je dédie ce modeste travail à ceux qui quel que soient les termes embrassés, je n'arriverais jamais à leur exprimer mon amour sincère.

À l'homme, ma précieuse offre de dieu et auquel je dois ma vie, ma réussite et tout mon respect : mon cher père Djelloul.

À la femme qui a souffert sans me laisser souffrir, qui n'a jamais dit non à mes exigences et qui n'a épargné aucun effort pour me rendre heureuse : mon adorable mère Yamina.

À mes chères sœurs : Nassima et Fahima pour leurs encouragements permanents, et leur soutien moral.

À mes chers frères : Youcef, Rafik et Khalil pour leur soutien et leur encouragement.

À mes meilleures Chaima et Abdelouahab et tous mes amis qui m'ont toujours encouragé, et à qui je souhaite beaucoup de succès.

Sans oublier mon binôme et ma perle rare Yousra pour son soutien moral, sa patience et sa compréhension tout au long de ce projet.

À tous ceux que j'aime

Merci d'être toujours là pour moi.

Amina.

Table des matières

Introduction Général	1
1.1 . Contexte global.....	2
1.2 .2. Problématique et Objectif.....	3
1.3 . Organisation du mémoire	4
Chapitre I : Etat de l'Art	6
1. Introduction.....	7
2. Les Chatbots.....	7
2.1. Définition des chatbot.....	7
2.2.Historique des Chatbots	8
2.3.Types de Chatbots	10
2.4. Processus général d'un Chatbot.....	14
2.5. Méthodologies des Chatbots.....	16
3. Les Chatbots psychologiques.....	17
3.1 Principes des chatbots psychologiques.....	17
3.2 Travaux connexes	20
4. Conclusion	26
Chapitre II : Conception et Modélisation de la solution proposée	28
1. Introduction	29
2. Processus global.....	29
3. Constitution du dataset.....	30
3.1. Dataset anglais.....	30
3.2. Dataset arabe	39
5 Spécification et analyse des besoins	40
5.1 Diagramme de cas d'utilisation.....	41
5.2 Diagramme de séquence.....	42
5.3 Diagramme de classes.....	44
5. Conclusion	46
Chapitre III : Tests et Validation de la solution	47
1. Introduction.....	48
2. Bibliothèques utilisées	48
2.1. NumPy	48
2.2. Pandas.....	48
2.3. TensorFlow	49
2.4. Keras.....	50
2.5. Scikit Learn	51
2.6. FastText.....	51
2.7. TextBlob	52

2.8. Seaborn	53
2.9. Stanza	53
2.10. Googletrans	54
3. Chargement des données.....	54
4. Prétraitement.....	56
4.1. Prétraitement surfacique.....	56
4.1.1. Suppression des caractères spéciaux	56
4.1.2. Suppression des mots vides.....	57
4.2. Prétraitement morphologique.....	57
5. Processus de vectorisation.....	62
5.1. Count vectoriser.....	62
5.2. N-gramme	63
5.3. Terme Fréquence-Inverse Document Fréquence (TF-IDF)	65
5.4. Word embedding	67
6. Classification.....	69
6.1 Machine learning	69
6.1.1 Naive Bayes algorithm	69
6.1.1.1 Matrice de confusion	70
6.1.1.2 Rappel et précision.....	72
6.1.2 KNN algorithm	74
6.1.2.1 Matrice de confusion.....	76
6.1.2.2 Rappel et précision.....	78
6.1.3 Performance des différents algorithmes	80
6.2 Deep learning	82
6.2.1 CNN algorithm.....	82
6.2.1.1 Matrice de confusion	83
6.2.1.2 Rappel et précision.....	84
6.2.2 RNN algorithm.....	85
6.2.2.1 Matrice de confusion.....	85
6.2.2.2 Rappel et précision.....	86
6.2.3 Des informations sur chaque algorithme	87
6. Conclusion.....	88
Chapitre IV : Présentation de l'application finale	89
1. Introduction.....	90
2. langages.....	90
3. Environnements	91
4. Outils.....	91
5. Présentations de l'application.....	93
6. Conclusion.....	101
Conclusion générale.....	102
Synthèse.....	103
Perspectives	104

Liste des figures

<i>Figure 1-1 : Schéma récapitulatif sur les Chatbots</i>	10
<i>Figure 1.2 : Les types de chatbots</i>	11
<i>Figure 1.3 : Processus général d'un Chatbot</i>	15
<i>Figure 1.4 : Représentation graphique de la mise en œuvre technique des chatbots</i>	19
<i>Figure 1.5 : Conversation avec Woebot avec différents types de réponses</i>	21
<i>Figure 1.6 : Conversation avec Wisa sur les troubles du sommeil</i>	22
<i>Figure 1.7 : Test de dépression avec l'application Mind Diagnostics</i>	24
<i>Figure 1.8 : Anxiety Test et les resultats de test</i>	25
<i>Figure 2.1 : Processus global de notre application</i>	29
<i>Figure 2.2 : Capture de la première version du corpus.</i>	31
<i>Figure 2.3 : Capture de la première version du corpus d'arabe</i>	39
<i>Figure 2.4 : Exemple de traduction de la version finale du dataset anglais</i>	40
<i>Figure 2.5: La version finale du dataset arabe</i>	40
<i>Figure 2.6 : Diagramme de cas d'utilisation</i>	41
<i>Figure 2.7 : Diagramme de séquence de notre système</i>	43
<i>Figure 2.8 : Diagramme de classes des différentes BDs de notre système</i>	45
<i>Figure 3.1: Logo de la bibliothèque NumPy</i>	48
<i>Figure 3.2 : Logo de Pandas..</i>	49
<i>Figure 3.3 : Logo de la bibliothèque TensorFlow</i>	50
<i>Figure 3.4 : Code d'importation de la bibliothèque Kears..</i>	50
<i>Figure 3.5 : Quelques opération de prétraitement avec Kears.</i>	50
<i>Figure 3.6 : Logo de Scikit Learn..</i>	51
<i>Figure 3.7 : Classificateurs de Sklearn..</i>	51
<i>Figure 3.8 : Sklearn Feature Code.</i>	51
<i>Figure 3.9 : Sklearn Metrics Code.</i>	51
<i>Figure 3.10 : Logo de FastText.</i>	51
<i>Figure 3.11 : Logo de TextBlob.</i>	53
<i>Figure 3.12 : Logo de Seaborn..</i>	53
<i>Figure 3.13 : Logo de Stanza.</i>	54
<i>Figure 3.14 : Organigramme de chargement des données</i>	55
<i>Figure 3.15 : Fonction de suppression des caractères spéciaux..</i>	57
<i>Figure 3.16 : Fonction de suppression des mots vides..</i>	57
<i>Figure 3.17 : Fonction de lemmatisation, POS et NER.</i>	58
<i>Figure 3-18 : Implémentation de la fonction de lemmatisation, POS et NER</i>	59
<i>Figure 3-19 : Fonction de lemmatisation de phrase</i>	59
<i>Figure 3-20 : Exemple de lemmatisation de phrase.</i>	59
<i>Figure 3-21 : Fonction de POS de phrase</i>	60
<i>Figure 3-22 : Exemple de POS de phrase.</i>	60
<i>Figure 3-23 : Fonction de NER de phrase</i>	61
<i>Figure 3-24 : Exemple de NER de phrase</i>	61

<i>Figure 3-25 : Organigramme du Prétraitement Morphologique</i>	62
<i>Figure 3-26 : Code de count vectoriser</i>	63
<i>Figure 3-27 : Fréquence de quelques termes</i>	63
<i>Figure 3-28 : Le code de bi-gramme avant la suppression des mots vides</i>	63
<i>Figure3-29 : La distribution des premiers bi-gramme avant la suppression des mots vides</i>	64
<i>Figure3-30 : La distribution des premiers bi-gramme après la suppression des mots vide</i>	65
<i>Figure3-31 : Le code de tf-idf</i>	66
<i>Figure 3-32 : Schéma du fonctionnement de l'opération TF-IDF</i>	67
<i>Figure 3-33 : Code de la tokenisation</i>	68
<i>Figure3-34 : Code de Embedding matrix</i>	68
<i>Figure3-35 : Schéma de la création de la matrice d'intégration</i>	68
<i>Figure3-36 : Code de L'algorithme MNB</i>	69
<i>Figure3-37: Matrice de confusion Crp1- one hot encoding-</i>	70
<i>Figure3-38 : Matrice de confusion Crp2 -TF-IDF word n-gram-</i>	71
<i>Figure3-39 : Matrice de confusion Crp3 -one hot encoding-</i>	72
<i>Figure3-40 : Courbe rappel et précision Crp1 -one hot encoding-</i>	73
<i>Figure3-41 : Courbe rappel et précision Crp2 -TF-IDF word n-gram-</i>	74
<i>Figure3-42 : Courbe rappel et précision Crp3 one hot encoding</i>	74
<i>Figure3-43 : Code de l'algorithme KNN</i>	75
<i>Figure3-44 : Matrice de confusion Crp1 -TF-IDF word n-gram-</i>	76
<i>Figure3-45 : Matrice de confusion Crp2 -TF-IDF word n-gram-</i>	77
<i>Figure3-46 : Matrice de confusion Crp3 -TF-IDF char n -gram-</i>	78
<i>Figure3-47 : Courbe rappel et précision Crp 1 -TF-IDF word n-gram-</i>	79
<i>Figure3-48 : Courbe rappel et précision Crp2 -TF-IDF word n-gram-</i>	79
<i>Figure3-49 : Courbe rappel et précision Crp3 -TF-IDF char n-gram-</i>	80
<i>Figure3-50 : Matrice de confusion Crp2 -CNN-</i>	83
<i>Figure3-51 : Matrice de confusion Crp3 -CNN-</i>	83
<i>Figure3-52 : Courbe rappel et précision Crp 2 -CNN-</i>	84
<i>Figure3-53 : Courbe rappel et précision Crp3 -CNN-</i>	85
<i>Figure3-54 : Matrice de confusion Crp2 -RNN-</i>	85
<i>Figure3-55 : Matrice de confusion Crp3 -RNN-</i>	86
<i>Figure3-56 : Courbe rappel et précision Crp2 -RNN-</i>	87
<i>Figure3-57 : Courbe rappel et précision Crp3- RNN-</i>	87
<i>Figure4.1 : Logo du langage de programmation Python.</i>	90
<i>Figure4.2 : Logo de l'IDE Visual Studio Code.</i>	91
<i>Figure4.3 : Logo de Google Colab.</i>	92
<i>Figure4.4 : Logo de Google Flask</i>	92
<i>Figure4.5 : Page d'accueil</i>	94
<i>Figure4.6 : Page d'identification en anglais</i>	94

<i>Figure4.7 : Page d'identification en arabe.....</i>	<i>95</i>
<i>Figure4.8 : Page de l'expression de sentiment en anglais.....</i>	<i>95</i>
<i>Figure4.9 : Page de l'expression de sentiment en arabe</i>	<i>96</i>
<i>Figure4.10 : Page de questionnaire</i>	<i>96</i>
<i>Figure4.11 : Page de diagnostic et des Conseils</i>	<i>97</i>
<i>Figure 4.12 : Exemple d'identification de l'utilisateur en anglais</i>	<i>97</i>
<i>Figure 4.13 : Exemple d'expression de sentiment en anglais.....</i>	<i>98</i>
<i>Figure 4.14 : Exemple de questionnaire pour la maladie "Dépression"</i>	<i>98</i>
<i>Figure 4.15 : Exemple de détection et Conseils pour la maladie "Dépression"</i>	<i>99</i>
<i>Figure 4.16 : Exemple d'identification de l'utilisateur en arabe.....</i>	<i>99</i>
<i>Figure 4.17 : Exemple d'expression de sentiment en arabe</i>	<i>100</i>
<i>Figure 4.18 : Exemple de questionnaire pour la maladie "القلق"</i>	<i>100</i>
<i>Figure 4.19 : Exemple de détection et Conseils pour la maladie "القلق"</i>	<i>101</i>

Liste des tableaux

<i>TABLEAU 1.1 : Les meilleures et les pires fonctionnalités de PsyTests & Mental Health Tests et Wysa & Woebot</i>	25
<i>Tableau 2.1 : Applications et tests en ligne utilisés</i>	32
<i>Tableau 2.2: Tableau résumant les cas du diagramme de cas d'utilisation</i>	42
<i>Tableau 3-1 : Le dataset en chiffres</i>	56
<i>Tableau 3-2: Résultat des Processus de Vectorisation avec et sans POS, NER et lemmatisation pour les Classifieurs de ML sur les quatre corpus</i>	81
<i>Tableau 3-3: Résultat de Accuracy et F1-score pour les classifieurs de DL sur les quatre corpus</i>	88

Liste des acronymes

Ipsos : Institut de Publique Sondage d'Opinion Secteur

NHS : National Health Service.

UML : Langage de Modélisation Unifié.

BD : Basse De Données.

IA : Intelligence Artificielle.

TAL : Traitement Automatique de la Langue.

FAQ: Frequently Asked Questions.

NLP: Natural Language Processing.

UI : Interface Utilisateur.

CL : Ligne de Commande.

GUI : Interface Utilisateur Graphique.

MDI : Interface Pilotée par Menu.

FBI : Interface Basée sur les Formulaires.

NLI : Interface en Langage Naturel.

OMS : Organisation Mondiale de la Santé.

TCC : Thérapie Comportementale et Cognitive.

PTSD : Syndrome de Stress Post Traumatique.

NLTK: Natural Language Toolkit

TF: Term Frequency.

IDF: Inverse Document Frequency

POS : Part-Of-Speech.

NER : Reconnaissance d'Entités Nommées.

ML: Machine Learning.

DL: Deep learning.

KNN: K-Nearest Neighbors Algorithm

NB: Naive Bayes algorithm.

RNN: Réseau Neuronal Récurrent.

CNN : Réseaux de Neurones dits Convolutifs.

Introduction Générale

1.1 Contexte global

L'influence et la diffusion des Chatbots dans la société ne cessent de croître. En 2016, la taille du marché mondial des Chatbots s'élevait à 190,8 millions de dollars américains. En 2025, la valeur de ce marché est évaluée à 1250 millions de dollars américains [1]. Cela indique une augmentation d'environ 600 %. L'expansion de ce secteur peut être une solution à l'insuffisance de la qualité des soins de santé dans certains pays. Selon Ipsos¹, 45 % des adultes dans le monde considèrent que la qualité des soins de santé dans leur pays est bonne. 37 % ont déclaré qu'elle n'était ni bonne ni mauvaise et les 23 % restants ont jugé que les soins de santé auxquels ils avaient accès étaient mauvais.

Les raisons possibles de ce mécontentement peuvent être la longue durée d'obtention d'un rendez-vous chez le médecin généraliste ou à l'hôpital. La pénurie de personnel dans les hôpitaux d'aujourd'hui peut également être une cause possible de ces mauvaises notes. Les habitants du Royaume-Uni ont évalué ces deux aspects à 52 %, ce qui constitue la principale raison de leur insatisfaction à l'égard du National Health Service (NHS). En outre, 22 % des participants à l'étude n'étaient pas satisfaits de la qualité des services fournis par le NHS [2].

En outre, les patients dans le monde critiquent le fait que les médecins n'ont pas assez de temps à consacrer à leurs patients ou qu'ils ne sont pas satisfaits des heures d'ouverture des cabinets médicaux. En outre, ils ont le sentiment de ne pas être pris au sérieux par les médecins spécialistes [3].

L'utilisation de Chatbots pourrait réduire le mécontentement mentionné. Comme les machines n'ont pas besoin de pause comme les êtres humains, elles pourraient fournir un service 24 heures sur 24 heures à toute

¹ Ipsos Institut de Publique Sondage d'Opinion Secteur

personne ayant accès à un smartphone ou à un ordinateur. Les patients n'ont donc pas besoin de se conformer aux horaires d'ouverture des établissements médicaux. En outre, les Chatbots sont capables de répondre instantanément et évitent aux patients de devoir attendre des heures avant de pouvoir rencontrer un médecin spécialiste.

La réponse instantanée d'un Chatbot est également attendue par 55 % des clients [4]. Étant donné que la majorité de notre société a accès à un smartphone, les assistants virtuels peuvent également accroître l'accessibilité du soutien médical ou psychologique.

Les habitants des petites villes n'ont pas besoin de parcourir de longues distances pour consulter un expert humain. Au lieu de cela, ils pourraient simplement chatter avec un assistant virtuel. Un autre avantage de l'utilisation des Chatbots dans les soins de santé est l'évolutivité des Chatbots, car ils peuvent consulter plusieurs patients en même temps et peuvent par conséquent réduire les coûts de personnel.

1.2 Problématique et objectifs

Selon l'Institut de Publique Sondage d'Opinion Secteur, 70% des patients dans le monde ont jugé que les soins de santé mentale auxquels ils avaient accès étaient mauvais.

Les raisons possibles de ce déplaisir peuvent être [5]:

- La longue durée d'obtention d'un rendez-vous chez le médecin généraliste ou à l'hôpital.
- Les heures d'ouverture des cabinets médicaux.
- Le sentiment de ne pas être pris au sérieux par les médecins spécialistes.

Les principaux objectifs de ce travail de recherche sont les suivants :

- Développement d'un système qui peut être utilisé dans les soins de santé mentale, il pourrait s'agir, par exemple, de soutenir les personnes dépressives ou de rendre possible un diagnostic en fonction des symptômes des utilisateurs.

- Ce système doit être adaptable et extensible et contenir les composants suivants : une application web, un back-end et une base de connaissances.
- -L'application web doit être responsable du questionner entre un patient et le système.
- -Le back-end et la base de connaissances peuvent être utilisés pour gérer les données requises.

Toutefois, l'ensemble du système n'est pas conçu dans le but de remplacer les spécialistes. Il doit plutôt aider les professionnels et les patients.

1.3 Organisation du mémoire

Dans le but de faciliter la lecture du présent mémoire, nous allons présenter un bref aperçu des chapitres qui le composent. Ces derniers sont organisés comme suit :

- **Chapitre I - Etat de l'Art :** Ce chapitre présentera brièvement une définition sur les Chatbots, une brève historique du domaine des agents conversationnels. Les genres ou types de Chatbots, le processus général de fonctionnement d'un Chatbot et les méthodologies.
- **Chapitre II – Conception et Modélisation de la solution proposée :** Ce chapitre présentera brièvement le processus général de notre solution, la constitution du dataset.
 - Définition des étapes principales de traitement et
 - La modélisation UML de la solution.
- **Chapitre III – Tests et Validation de la solution :** Dans ce chapitre, nous allons présenter l'environnement de développement et les détails de chaque étape de développement.
- **Chapitre IV – Présentation de l'application finale :** Les résultats de notre projet de Master sont conclus dans ce chapitre

Chapitre I :

Etat de l'art

I.1 Introduction

Un bot est un programme qui interagit automatiquement avec un utilisateur : la machine à café en est un, par exemple. Quand on appuie sur un bouton, elle produit du café mécaniquement. Au cours d'une conversation, sur un chat de messaging avec un amis ou un inconnu, dans une communauté sur les réseaux sociaux, sur un site d'achat en ligne... ... nous puerons, de nos jours voir des bots qui répondent à des humains à la place d'autres humains.

Ce chapitre présentera brièvement une définition sur les Chatbots qui sera donnée dans la section I.2.1 Ensuite, une bref historique du domaine des agents conversationnels sera présentée dans la section I.2.2 Les genres ou types de Chatbots viendront juste après dans la section I.2.3, Puis, le processus général de fonctionnement d'un Chatbot dans la section I.2.4 , les méthodologies dans la section I.2.5, Les Principes des Chatbots psychologiques dans la section I.3.1 et Les Travaux connexes dans la section, I.3.2. Pour terminer avec une conclusion dans la section 1.4.

I.2 les Chatbots

Cette section est consacrée aux Chatbots.

I.2.1 Définition des Chatbots

Selon l'Oxford English Dictionary, un Chatbot est défini comme suit :

Chatbot (n.) : Un programme informatique conçu pour simuler une conversation avec des utilisateurs humains, en particulier sur Internet.

Dans la littérature scientifique, les Chatbots sont plus formellement appelés agents conversationnels. Dans le contexte de ce document, les termes Catbot/Agent Conversationnel seront utilisés de manière interchangeable.

Le principe sous-jacent de chaque Chatbot est d'interagir avec un utilisateur humain (dans la plupart des cas) via des messages texte et de se comporter comme s'il était capable de comprendre la conversation et de répondre à l'utilisateur de manière appropriée.

En outre, une autre hypothèse générale que les concepteurs de Chatbots gardent à l'esprit est que les utilisateurs ont généralement un objectif qu'ils souhaitent atteindre à la fin de la conversation lorsqu'ils initient une interaction avec un Chatbot. Cela influence ensuite le flux et les sujets de la conversation afin d'atteindre l'objectif choisi. Cela peut être exploité par les développeurs car certains modèles de comportement ont tendance à apparaître en conséquence.

Par conséquent, la définition d'un chatbot adoptée pour ce document est la suivante : « Un Chatbot est programme informatique communiquant par texte de manière humaine et qui fournit des services aux utilisateurs humains afin d'accomplir un objectif bien défini ». [6]

I.2.2 Historique des Chatbots

Dans cette section, nous donnerons quelques exemples sur la façon dont les Chatbots ont évolué au cours des 50 dernières années, du bot ELIZA aux actuels assistants personnels tels que Siri d'Apple ou Cortana de Microsoft. [7]

I.2.2.1 Eliza-Bot

Le premier exemple d'agent conversationnel né en 1966, ELIZA était un programme informatique qui simulait un psychiatre et reformulait la saisie de l'utilisateur en utilisant des techniques de base du Traitement du Langage Naturel.

En dépit d'être relativement simple, le programme a réussi à donner l'illusion de comprendre les problèmes de l'utilisateur et a dupé avec succès un grand nombre de personnes. Son créateur, Joseph Weizenbaum, a même

mentionné que sa secrétaire lui demanderait de la quitter pour qu'elle puisse avoir une conversation privée avec ELIZA.

I.2.2.2 SmarterChild

Pendant plusieurs décennies, les Chatbots ont fortement suivi l'approche d'ELIZA avec des ajouts mineurs apportés sur le terrain comme la synthèse vocale et la gestion des émotions. Puis, en 2001, est venu SmarterChild, un agent conversationnel développé par ActiveBuddy, Inc. (maintenant Colloquis) qui fonctionnait sur AOL Instant Messenger et MSN Messenger. Inspiré par l'essor des plates-formes de messagerie instantanée telles que les SMS, SmarterChild a été créé pour fournir un accès rapide aux actualités, aux prévisions météorologiques, aux résultats sportifs, etc.

La principale innovation était que SmarterChild était connecté à une base de connaissances qui contenait des informations utiles pour ses utilisateurs.

Malheureusement, les limitations techniques du Traitement du Langage Naturel ont rattrapé les bots sur ces plateformes à l'époque et ils ont été oubliés par l'histoire.

I.2.2.3 Watson Artificiel Intelligence

La prochaine avancée pour les agents conversationnels a été réalisée par une équipe d'IBM via le projet Watson AI qui était en développement depuis 2006.

L'agent a été conçu dans le seul but de gagner l'émission de télévision américaine Jeopardy. Ce qu'il a fait en 2011 en affrontant deux des anciens champions de l'émission.

C'est intéressant d'un point de vue TAL (Traitement Automatique de la Langue) puisque les questions impliquent beaucoup de jeux de mots et nécessitent une recherche rapide d'informations dans de vastes bases de connaissances. Malheureusement, cette Intelligence Artificielle (IA) sous sa

forme passée ne pouvait répondre qu'à des questions simples et était incapable de mener une conversation appropriée avec quelqu'un d'autre.

I.2.2.4 Une nouvelle génération de Chatbots

Au début des années 2010, des assistants virtuels sont apparus comme Siri d'Apple, Cortana de Microsoft, l'assistant Google de Google, Alexa d'Amazon et plein d'autres. Ces agents ont apporté sur le terrain le concept de conversation ainsi que le dialogue axé sur les objectifs. Un autre événement majeur dans le domaine des Chatbots a été la sortie de la plateforme Messenger pour Facebook Messenger en 2016 et a permis la création d'agents conversationnels pour les entreprises non liées à l'IA.

La figure qui suit, récapitule quelques-uns des Chatbots les plus connus depuis Eliza jusqu'à TAY.

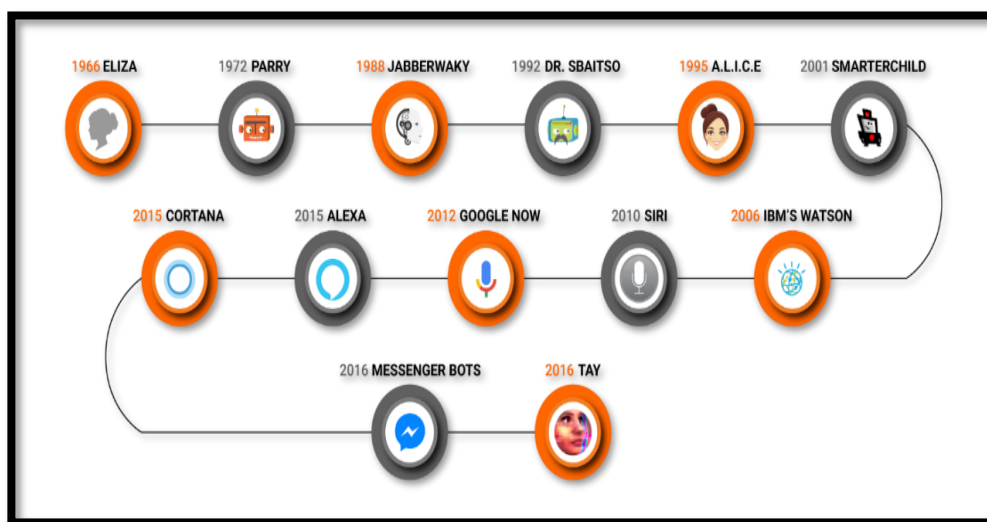


Figure 1-1 : Schéma récapitulatif sur les Chatbots.

I.2.3 Types de Chatbots

Dans cette sous-section, nous essaierons de parler de certains des types de Chatbots les plus courants. [8]

La figure (Figure1-2), ci-dessus résume les différents types de Chatbots que nous allons aborder dans cette section



Figure 1.2 : Les types de chatbots.

I.2.3.1 Chatbots basés sur des règles (linguistiques)

Afin de prédire les types de questions que des clients peuvent poser, un chatbot linguistique pourrait être la solution idéale. Les Chatbots linguistiques ou Chatbots basés sur les règles créent des flux de conversation en utilisant la Logique (Si/ Alors/ Sinon) . Tout d'abord, il est impératif de définir les conditions linguistiques du Chatbot.

Des conditions peuvent être créées pour évaluer les mots, l'ordre des mots, les synonymes, etc. Si la requête entrante correspond aux conditions définies par le Chatbot, les clients (utilisateurs) peuvent recevoir l'aide appropriée en un rien de temps.

Cependant, l'enjeu est de s'assurer que chaque permutation et combinaison de chaque question est définie, sinon le Chatbot ne comprendra pas l'entrée du client. C'est pourquoi un modèle linguistique, bien qu'incroyablement courant, peut être lent à se développer. Ces Chatbots exigent rigidité et spécificité.

I.2.3.2 Chatbots basés sur l'apprentissage automatique

Les Chatbots contextuels sont de loin les plus avancés des trois bots évoqués précédemment. Ces types de Chatbots utilisent l'apprentissage automatique (ML pour Machine learning en Anglais) et l'IA pour mémoriser

les conversations avec des utilisateurs spécifiques afin d'apprendre et de grandir au fil du temps de répondre adéquatement aux besoins en information des utilisateurs. Contrairement aux Chatbots basés sur la reconnaissance de mots clés, les Chatbots contextuels sont suffisamment intelligents pour s'améliorer en fonction de ce que les utilisateurs demandent et de la manière dont ils le demandent.

Par exemple, dans un Chatbot contextuel qui permet aux utilisateurs de commander de la nourriture, le chatbot stockera les données de chaque conversation et apprendra ce que l'utilisateur aime commander. Le résultat est que finalement, lorsqu'un utilisateur discute avec ce chatbot, il se souviendra de sa commande la plus courante, de son adresse de livraison et de ses informations de paiement et lui demandera simplement s'il souhaite répéter cette commande. Au lieu d'avoir à répondre à plusieurs questions, l'utilisateur n'a qu'à répondre par « Oui » et la nourriture est prête !

Bien que cet exemple de commande de nourriture soit élémentaire, il est facile de voir à quel point le contexte de conversation peut être puissant lorsqu'il est exploité avec l'IA et le ML. L'objectif ultime de tout Chatbot devrait être de fournir une expérience utilisateur améliorée par rapport à l'alternative dans la situation prévalant auparavant. Tirer parti du contexte de conversation est l'un des meilleurs moyens de raccourcir de tels processus via un Chatbot.

I.2.3.3 Chatbots basés sur des menus/boutons

Les Chatbots basés sur des menus/boutons sont le type de Chatbots le plus basique actuellement mis en œuvre sur le marché d'aujourd'hui. Dans la plupart des cas, ces Chatbots sont des hiérarchies d'arbres de décision² magnifiés présentées à l'utilisateur sous forme de boutons.

² Un arbre de décision est un moyen de classer visuellement et logiquement les informations et de prendre des décisions. Cette méthode permet grâce à un simple système de calcul de trouver quel est le choix le plus viable.

Semblables aux menus téléphoniques automatisés avec lesquels nous interagissons tous presque quotidiennement, ces Chatbots nécessitent que l'utilisateur fasse plusieurs sélections pour creuser plus profondément vers la réponse ultime.

Alors que ces Chatbots sont suffisants pour répondre aux Frequently Asked Questions (FAQ) qui représentent 80% des requêtes d'assistance ; ils sont bien insuffisants dans des scénarios plus avancés dans lesquels il y a trop de variables ou trop de connaissances en jeu pour prédire comment les utilisateurs devraient obtenir des réponses spécifiques avec confiance. Il convient également de noter que les Chatbots basés sur des menus/boutons sont les plus lents pour amener l'utilisateur à la valeur souhaitée. Aussi, les utilisateurs se trouvent parfois démotivés pour jouer le jeu et sélectionner à chaque fois une réponse parmi plusieurs propositions au lieu de converser de manière libre.

I.2.3.4 Chatbots basés sur la reconnaissance de mots clés

Contrairement aux Chatbots basés sur des menus, les Chatbots basés sur la reconnaissance de mots clés peuvent écouter ce que les utilisateurs tapent et répondre de manière appropriée. Ces Chatbots utilisent des mots-clés personnalisables et une application d'IA -Natural Language Processing (NLP) pour déterminer comment fournir une réponse appropriée à l'utilisateur.

Ces types de Chatbots sont insuffisants lorsqu'ils doivent répondre à de nombreuses questions similaires. Les Chatbots commenceront à faire fausse route lorsqu'il y aura des redondances de mots clés entre plusieurs questions liées.

Il est assez courant de voir des Chatbots hybrides basés sur la reconnaissance de mots clés et sur des menus/boutons. Ces Chatbots offrent aux utilisateurs le choix d'essayer de poser leur question directement ou d'utiliser les boutons de menu du Chatbot si la fonctionnalité de

reconnaissance des mots clés donne de mauvais résultats ou si l'utilisateur a besoin de conseils pour trouver sa réponse.

I.2.3.5 Les Chatbots hybrides

Les entreprises aiment la sophistication des IA-Chatbots, mais n'ont pas toujours les talents ou les gros volumes de données pour les prendre en charge. Alors, ils optent pour le modèle hybride. Le modèle hybride offre le meilleur des deux mots : la simplicité des Chatbots basés sur des règles, avec la complexité des IA-bots.

I.2.3.6 Robots vocaux

Pour rendre les interfaces conversationnelles encore plus vernaculaires, les entreprises commencent maintenant à utiliser des Chatbots vocaux, ou bots vocaux. Les bots vocaux ont augmenté ces dernières années, avec Siri d'Apple, Alexa d'Amazon, et pourquoi ? En raison de la commodité qu'ils apportent.

Il est beaucoup plus facile pour un client de parler plutôt que de taper. Les robots vocaux apportent des expériences sans friction directement au client.

I.2.4 Processus général d'un Chatbot

Il y a sept étapes pour concevoir le processus Chatbot : la portée et les exigences, l'identification des entrées, la compréhension des éléments de l'interface utilisateur, la création d'une première interaction, la création d'une conversation et enfin le test. [9]

La figure du processus de conception de Chatbot est illustrée dans la figure (Figure 1-3) ci-dessous :

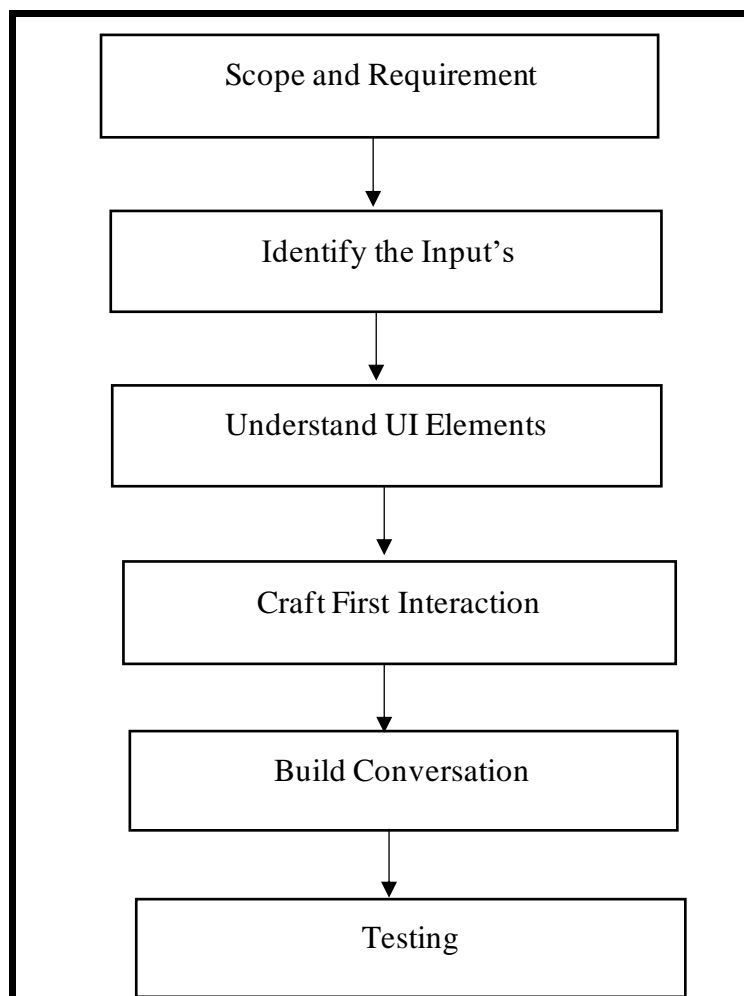


Figure 1.3 : Processus général d'un Chatbot.

La première étape de la conception d'un Chatbot consiste à connaître le champ d'application et les exigences, par exemple pourquoi un Chatbot, la plate-forme de lancement des Chatbots et ses limites.

La deuxième étape consiste à identifier les entrées des utilisateurs sous forme de requêtes par texte, voix ou images, à partir d'appareils et de systèmes d'intelligence.

La troisième étape consiste à comprendre les éléments de l'interface utilisateur (UI) que nous pouvons voir dans nos applications.

Les éléments de l'interface utilisateur sont de cinq types : ligne de commande (CL), interface utilisateur graphique (GUI), interface pilotée par menu (MDI), interface basée sur les formulaires (FBI) et interface en langage naturel (NLI).

Après avoir compris les éléments de l'interface utilisateur, l'étape suivante consiste à créer la première interaction et à établir une conversation. La dernière étape du processus de conception de Chatbot est le test, qui est effectué sur les mobiles et les sites Web pour savoir comment cela fonctionne.

I.2.5 Méthodologies des Chatbots

Dans cette section, nous allons parler des méthodologies utilisées par les Chatbots.

I.2.5.1 Bots basés sur des règles

L'idée principale sous-jacente au RG basé sur des règles est qu'un bot contient une base de connaissances avec des documents, où chaque document contient un <pattern> et un <template>. Lorsque le bot reçoit une entrée qui correspond au <pattern>, il envoie le message stocké dans le <template> en réponse. Le <motif> peut être soit une phrase « Comment vous appelez-vous ? » ou un modèle « Mon nom est * », où étoile est une pression régulière. En règle générale, ces paires <pattern> <template> sont fabriquées à la main. [10]

Le principal défi du programmeur est de choisir un algorithme à utiliser pour la correspondance entre la phrase d'entrée et les documents du corpus de données. Nous pouvons considérer cela comme un problème de plus proche voisin, où notre objectif est de définir la fonction de distance et de récupérer le document le plus proche de la phrase d'entrée. Ci-dessous, nous allons donner un exemple à ce sujet.

ELIZA utilise une analyse incrémentielle ou une approche de « correspondance directe » pour la correspondance de modèle. ELIZA a analysé le texte saisi mot par mot de gauche à droite, recherchant chaque mot dans le dictionnaire, lui donnant un rang en fonction de son importance et le stockant dans une pile de mots clés. Le mot-clé avec le rang d'importance le plus élevé serait traité comme le <pattern> et le modèle serait comparé aux documents du corpus pour trouver le modèle de réponse

approprié. S'il n'existait aucun document correspondant à l'entrée, ELIZA dira « Je vois » ou « Veuillez continuer » dans le script DOCTOR. ELIZA a été créé avec plusieurs « scripts » qui indiquent différentes réponses <template> aux entrées <pattern>.[11]

Les types des Chatbot qui convient à cette méthodologie : Chatbots basés sur des menus/boutons et les Chatbots linguistiques.

I.3 Les Chatbots psychologiques

Cette section est consacrée aux Chatbots psychologiques.

I.3.1 Principes des chatbots psychologiques

Fournir des soins appropriés aux personnes ayant des besoins en matière de santé mentale a toujours été un défi.

Selon l'Organisation Mondiale de la Santé (OMS), la dépression affecte à elle seule plus de 300 millions de personnes par an, induisant un coût économique qui se chiffre à des centaines de milliards de dollars chaque année. Même avec un accès à des services de santé et à des services sociaux compétents, les personnes atteintes de problèmes de santé mentale hésitent à recourir au traitement en raison de la stigmatisation entourant la maladie. En réponse, les entreprises de technologie ont mis au point des applications d'intelligence artificielle qui visent à offrir un soutien de première ligne pour les patients souffrant de troubles mentaux, tout en garantissant la confidentialité et l'anonymat grâce au caractère numérique des échanges.

De plus, contrairement à une application ou une recherche sur le Web, un Chatbot simule une conversation réelle, comme si l'utilisateur parlait avec une autre personne. Son caractère unique réside dans la perception qu'à l'utilisateur de cette interaction.

L'intelligence émotionnelle est une dimension importante, en particulier dans un contexte psychologie / thérapie. Développer des Chatbots

capables de détecter et de répondre de manière appropriée à l'état émotionnel d'un humain grâce à l'IA émotionnellement intelligente et à l'informatique affective ("affective computing") offre des perspectives intéressantes.

Pour l'instant, il apparaît que la plupart des solutions rencontrées sont basées sur des techniques de thérapie cognitive et comportementale ou du "coaching psychologique".[12]

D'un point de vue technique, les Chatbots comprennent divers sous-composants internes. Le composant de reconnaissance des informations des messages de l'interlocuteur est élémentaire. Tout d'abord, les intentions et les entités concernées doivent être identifiées à partir des déclarations de l'interlocuteur.

Cela se fait à l'aide du traitement du langage naturel (NLP), qui est le traitement automatique du langage naturel à l'aide d'algorithmes statistiques. La NLP traduit la saisie de l'utilisateur dans une langue que le Chatbot peut comprendre. Les intentions et les entités identifiées des utilisateurs, à leur tour, servent de base à la réponse appropriée du Chatbot en fonction de ses composants pour la gestion du dialogue et la génération de réponses. La génération de réponse est basée, par exemple, sur des scripts stockés, un dialogue prédéterminé et/ou une base de connaissances intégrée.

La réponse est mise à disposition des utilisateurs via l'interface basée sur le chat, telle qu'une intervention sur Internet ou une application.

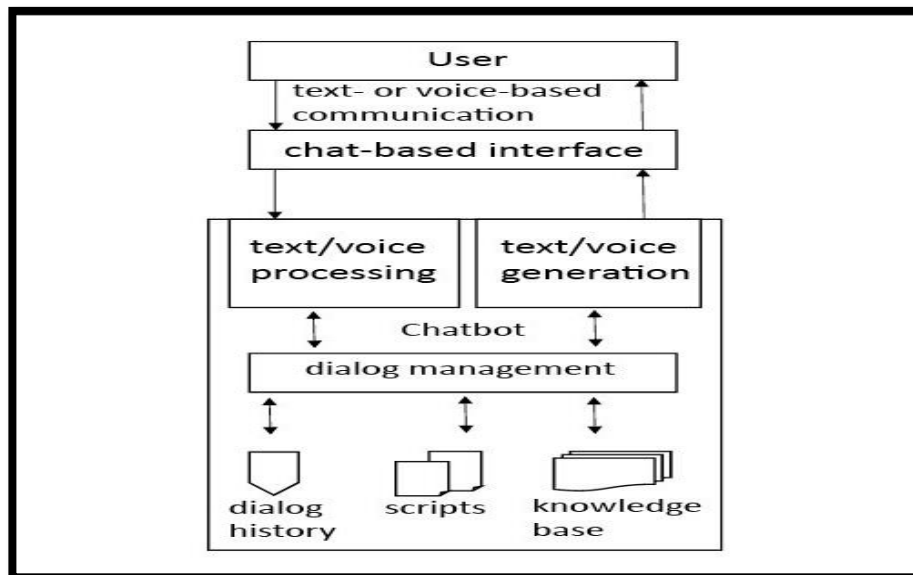


Figure 1.4 : Représentation graphique de la mise en œuvre technique des chatbots[13]

Les Chatbots varient dans leur complexité d'interaction. Les composants de dialogue plus simples sont basés exclusivement sur des séquences basées sur des règles (telles que des règles si-alors). Les conversations sont modélisées comme un réseau d'états possibles.

Les entrées déclenchent les transitions d'état et les réponses associées du bot, de manière analogue aux systèmes stimulus-réponse [14]. Le Chatbot suit en interne un arbre de décision prédéfini [15].

Les Chatbots les plus complexes utilisent également des bases de connaissances ou des modèles d'apprentissage automatique pour construire leurs dialogues et l'IA pour générer des réponses possibles et améliorer la compétence conversationnelle du bot (souvent appelées capacités mentales dans la littérature de recherche) [16].

Il existe des langages de programmation dédiés à la programmation des dialogues de Chatbot. Le langage de balisage artificiel ³ est devenu particulièrement bien établi [17,18]

³ AIML ; <http://alicebot.wikidot.com/learn-aiml>

I.3.2 Travaux connexes

Dans la section suivante nous avons choisi 4 chatbots qui traitent les maladies psychologiques, deux Chatbots basés sur des menus/boutons (psytest, mind Diagnostics) et deux Chatbots basés sur l'apprentissage automatique (Woebot, Wysa).

I.3.2.1 Woebot

Woebot est une solution numérique pour la santé mentale qui s'inspire de la thérapie cognitivo-comportementale et repose sur une plateforme axée sur l'intelligence artificielle pour fournir un contenu personnalisé aux utilisateurs. Lors d'un précédent essai contrôlé randomisé de deux semaines, Woebot a atténué les symptômes dépressifs.[19]

Woebot est utilisé dans une application de messagerie instantanée indépendante de la plate-forme et peut être utilisé sur un ordinateur de bureau ou un appareil mobile. Chaque interaction commence par une enquête générale sur le contexte (par exemple, « Que se passe-t-il dans votre monde en ce moment ? ») et l'humeur (par exemple, « Comment vous sentez-vous ? ») Avec des réponses fournies sous forme de mots ou d'images emoji pour représenter ce moment. Après avoir collecté des données sur l'humeur, les participants se voient présenter les concepts de base liés à la TCC (Thérapie comportementale et cognitive) par un lien vers une courte vidéo ou au moyen de courts « jeux de mots » conçus pour faciliter l'enseignement des distorsions cognitives aux participants. Le premier jour comprenait un processus « d'intégration » qui a introduit le bot, ajoutant que même si le bot peut sembler être une personne, il est plus proche d'un livre d'auto-assistance « choisissez votre propre aventure » et n'est donc pas tout à fait capable de comprendre quels peuvent être les besoins de l'utilisateur.

Le bot a également brièvement expliqué la TCC et a informé l'utilisateur que même si un psychologue « gardait un œil sur les choses » (c'est-à-dire la surveillance), cela ne se produisait pas en temps réel et que le service ne devait donc pas être utilisé en remplacement d'une thérapie. [20]

La figure ci-dessous présente une petite partie d'une conversation avec le Chatbot. L'ensemble de l'échange peut être décrit comme un mélange entre une conversation guidée et une conversation libre. En fonction de la question saisie, l'application mobile fournit des réponses prédéfinies. C'est ce que montre la Figure. En outre, l'utilisateur a la possibilité de saisir des réponses personnalisées.

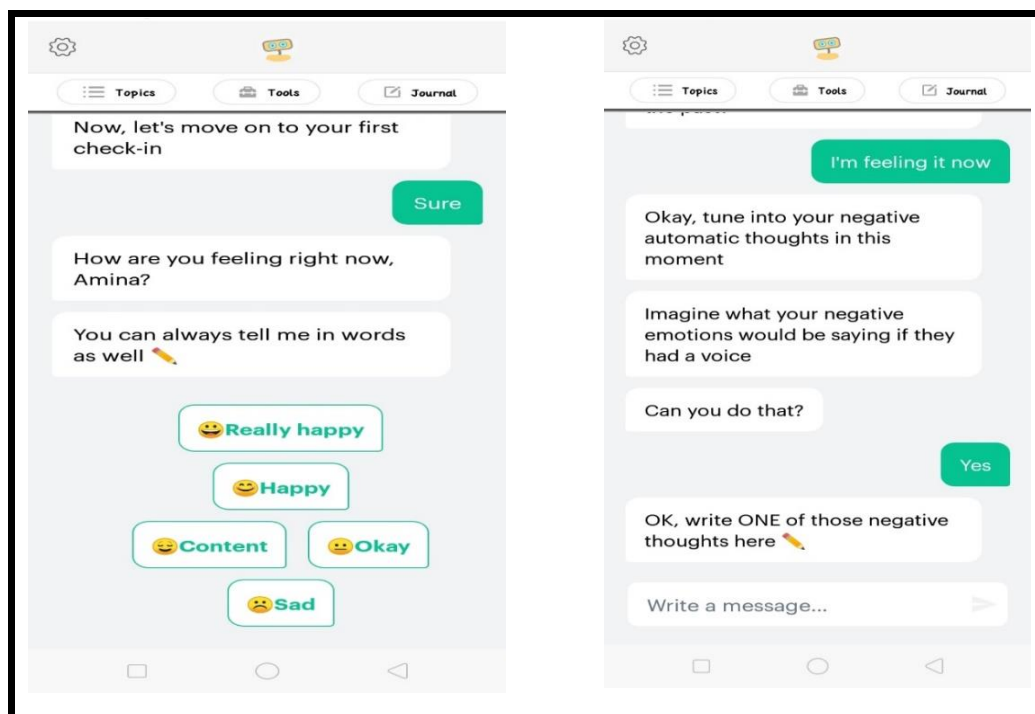


Figure 1.5 : Conversation avec Woebot avec différents types de réponses

I.3.2.2 Wysa

Wysa, développée par Touchkin, est une application chatbot mobile émotionnellement intelligente basée sur l'IA visant à renforcer la résilience mentale et à promouvoir le bien-être mental à l'aide d'une interface conversationnelle textuelle. L'application Wysa aide les utilisateurs à développer une expression positive de soi en utilisant l'IA pour créer un environnement d'autoréflexion externe et réactif. L'utilisation de l'application est gratuite et disponible 24h/24 et 7j/7, mais l'accès à un coach humain via l'application est payant. L'application réagit aux émotions exprimées par l'utilisateur au cours de conversations écrites et dans sa

conversation utilise des pratiques d'auto-assistance fondées sur des données probantes, telles que la TCC, la thérapie comportementale dialectique, l'entretien motivationnel, le soutien comportemental positif, le renforcement comportemental, la pleine conscience, ainsi que des micro-actions et des outils guidés pour encourager les utilisateurs à développer des compétences en matière de résilience émotionnelle.[21]

La figure suivante montre le début d'une conversation en entrant la phrase 'i cant sleep'.

Une réponse peut être choisie en sélectionnant et en confirmant ou en écrivant ce que l'utilisateur veut l'utilisateur peut donc décider s'il veut discuter avec des messages vocaux ou avec texte

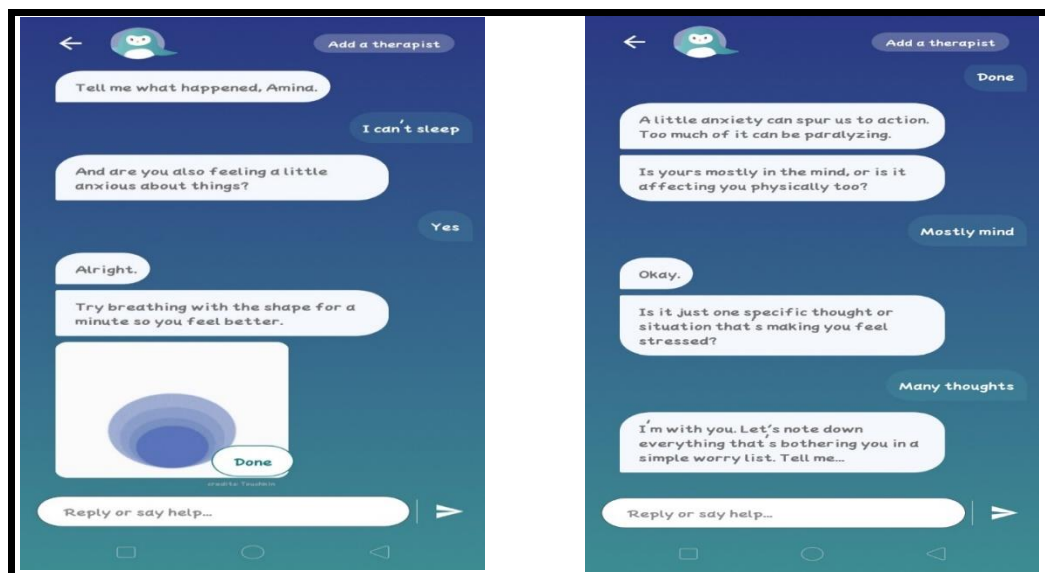


Figure 1.6 : Conversation avec Wysa sur les troubles du sommeil

I.3.2.4 Mind Diagnostics

Mind Diagnostics est un fournisseur d'applications pour la gestion de la santé mentale. Il fournit un test de santé mentale pour déterminer les symptômes d'un problème de santé mentale. Il propose des solutions contre les problèmes de santé mentale, tels que la dépression ou l'anxiété et son processus de récupération.

Les fonctionnalités incluent les installations de suivi, l'envoi des résultats à un ami, un membre de la famille ou un thérapeute et d'autres. Il permet également aux utilisateurs de trouver des thérapeutes dans leur région [22].

La figure suivante montre le début de Dépression Test et les résultats de test et des conseils

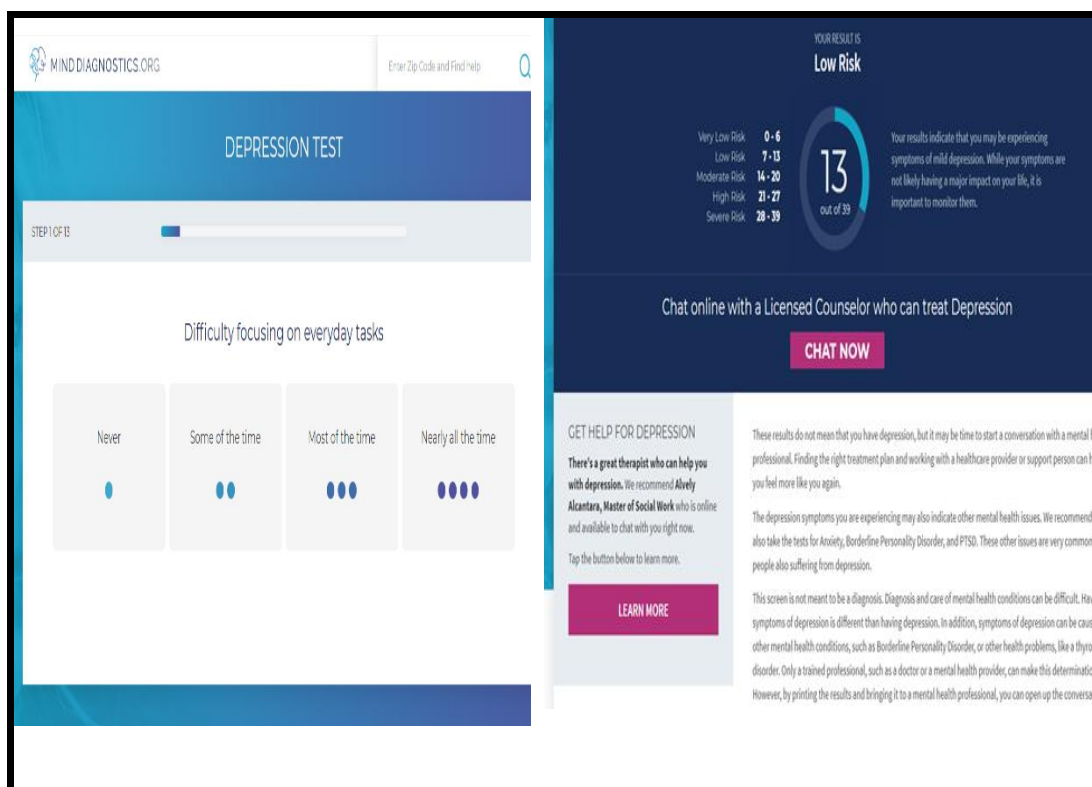


Figure 1.7 : test de dépression avec l'application Mind Diagnostics

I.3.2.5 PsyTests

Cette application contient des tests psychologiques sur plusieurs problèmes que les patients consultent souvent à tous les professionnels de la psychologie. Son contenu est simple à comprendre et très utile. Vous pourrez apprendre, consulter et évaluer de nombreuses caractéristiques de votre personnalité et aussi si vous souffrez de troubles anxieux, de troubles de l'humeur, de troubles de l'alimentation, de problèmes de dépendance, etc. De

plus, il y a des informations utiles sur l'amour, des sujets liés aux relations, sur qui vous êtes et comment fonctionne votre psychisme.

Tous les tests de cette application ont été développés grâce au travail acharné d'un groupe de professionnels de la psychologie.[23]

La figure suivante montre le début de Anxiety Test et les résultats de test :

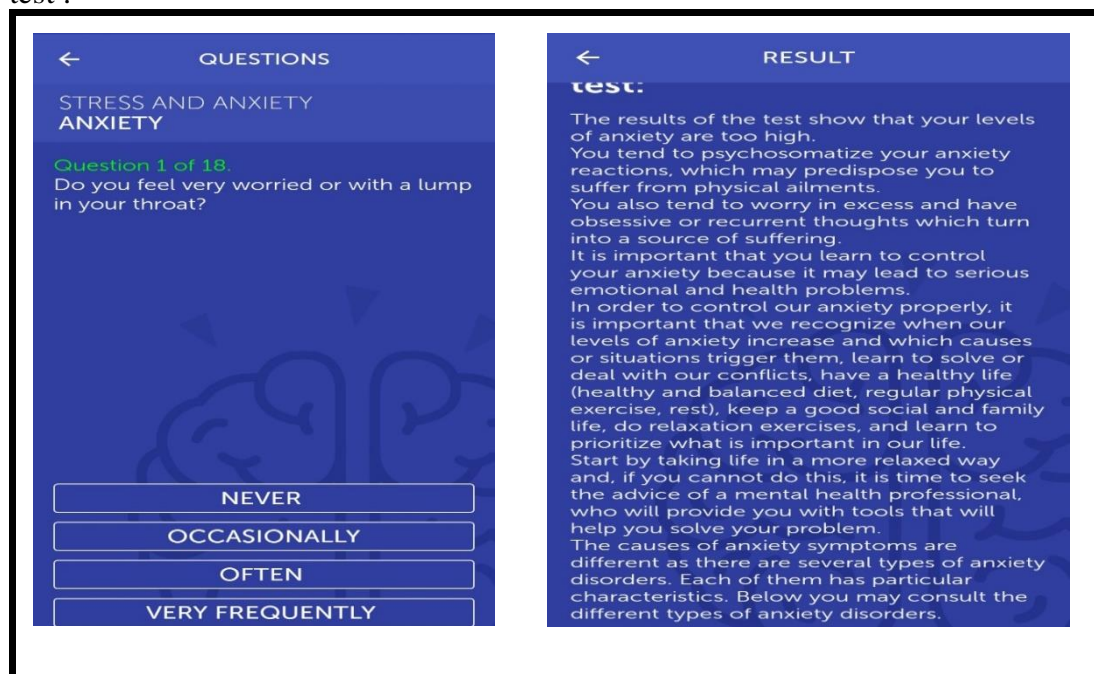


Figure 1.8 : Anxiety Test et les résultats de test .

I.3.2.6 Discussion

Le tableau ci-dessus (TABLEAU 1.1) montre les meilleures et les pires fonctionnalités et les types de chaque chatbots.

Application	Meilleures fonctionnalités	Pires fonctionnalités	Type
Mental Health Tests	<ul style="list-style-type: none"> - Une bonne variété de tests différents à choisir - une possibilité de parler à un conseiller pour toute aide supplémentaire. - Les résultats peuvent être envoyés par e-mail à l'utilisateur. 	<ul style="list-style-type: none"> - Les résultats manquent de profondeur 	
PsyTests	<ul style="list-style-type: none"> - basique, directe et précise. - Tests rapides mais approfondis. - De bons conseils et une très bonne analyse. 	<ul style="list-style-type: none"> - la section des conseils est fondamentalement la même pour chaque test, légèrement différente. - Nombreuses fautes d'orthographe et de grammaire. 	
Wysa	<ul style="list-style-type: none"> -L'accès facile à la boîte à outils aux activités de méditation guidée et de pleine conscience permet de guider les discussions dans des directions utiles. -Possibilité d'« activer coach » (payant) et de discuter de manière anonyme avec un coach qui 	<ul style="list-style-type: none"> -Demandes constantes de plus d'informations sans comprendre pourquoi ni comment cela m'aide. -L'humour était un peu écœurant. 	Chatbots basés sur l'apprentissage automatique

	vous guidera vers les bons outils. -Des réponses pré-remplies et de petites interactions amusantes pour rendre le chat plus amusant.		
Woebot	-Il partage des informations au bon moment et ne traîne pas. -Simple et clair. Woebot est facile à utiliser. Il serait facile d'entrer dans une routine matinale d'enregistrement, puis voir les tendances pourrait m'aider à avoir une perspective plus large sur mes humeurs. -Woebot explique pourquoi le partage d'informations avec lui peut améliorer l'expérience. Le pourquoi est important.	-Il est facile de se désintéresser lorsque vous utilisez un chatbot. Juste quelques faux pas en termes de réponses, et vous perdez de vue pourquoi c'est une bonne chose à faire. -Difficile de se souvenir des invites correctes pour vous amener à la boîte à outils, etc.	Chatbots basés sur l'apprentissage automatique

TABEAU 1.1 : les meilleures et les pires fonctionnalités de PsyTests & Mental Health Tests & Wysa & Woebot.

I.4 Conclusion

Le Chatbot améliore le service client, grâce à cette amélioration, les avantages du Chatbot augmentent de jour en jour. Dans le monde d'aujourd'hui, la messagerie est devenue l'un des moyens de communication les plus populaires, qu'il s'agisse d'un message texte ou via des applications

de messagerie. Les Chatbots sont conçus et développés en fonction des requêtes des clients et de leurs langues.

Les entreprises technologiques du monde entier combinent la puissance de l'intelligence artificielle avec la portabilité des smartphones pour créer des Chatbots conçus pour aider les patients souffrant de problèmes de santé mentale. Ces agents conversationnels accompagnent les patients tout en offrant un haut niveau de confidentialité et d'anonymat.

Dans le chapitre suivant, nous allons présenter la conception de notre solution.

Chapitre II :
Conception et
Modélisation de la
solution proposée

II.1 Introduction

Après avoir présenté l'état de l'art de notre projet, nous nous concentrerons dans ce chapitre sur l'énonciation de la conception de notre Chabot, en décrivant les différents états d'évolution de notre vision du Chabot.

Ce chapitre présentera brièvement : Le processus général de notre solution qui sera donnée dans la section II.2. Ensuite, la constitution du dataset sera présentée dans la section II.3. Définition II.5 Spécification et analyse des besoins. Pour terminer avec une conclusion dans la section II.5.

II.2 Processus global

Le processus général de notre solution comporte quatre étapes : l'identification des entrées, la phase d'apprentissage, la phase de test, et les sorties. La figure ci-dessous (figure 2.1) illustre ce processus.

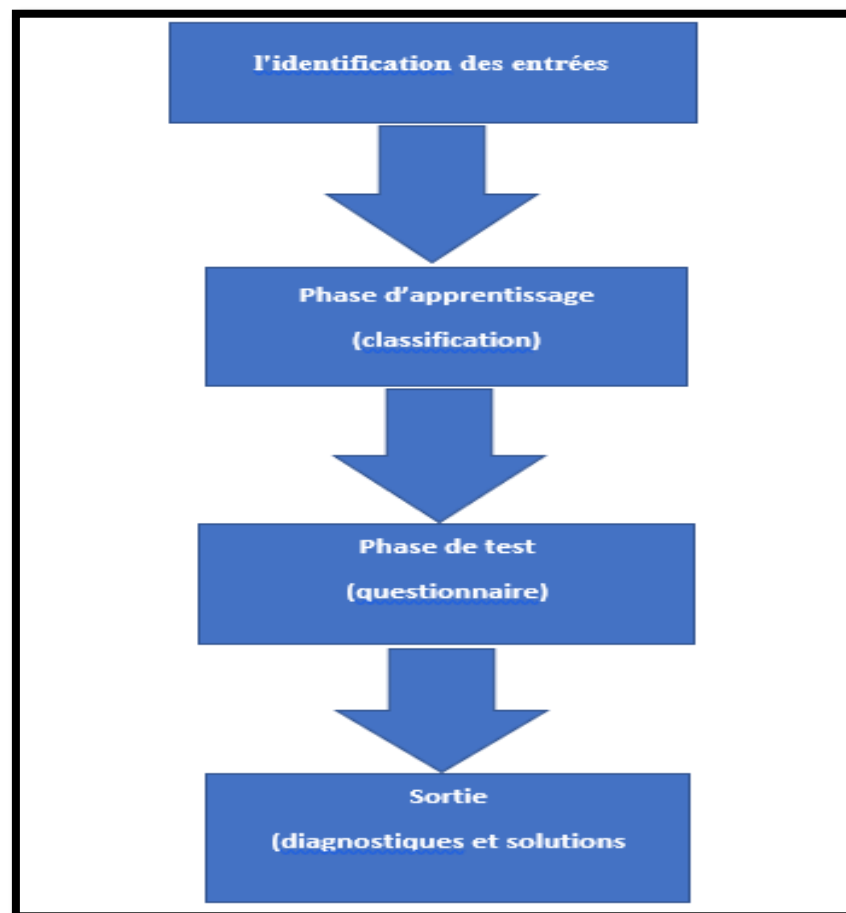


Figure 2.1 : Processus global de notre application

- 1) La première étape consiste à identifier les entrées des utilisateurs sous forme de requêtes libre via texte.
- 2) La deuxième étape consiste à classifier le texte saisi par l'utilisateur afin d'aider à catégoriser les mots importants qui reflètent les symptômes et déduire la maladie de l'utilisateur.
- 3) La troisième étape : la phase de test, 10 questions sur la maladie déduite précédemment sont posées à l'utilisateur qui doit répondre par « yes » ou « no ».
- 4) La quatrième étape consiste à afficher les diagnostics et proposer des Conseils.

Dans le chapitre suivant, nous allons expliquer ces étapes en détail.

II.3 Constitution du dataset

Notre application se voudra être fonctionnelle en langue anglaise et arabe et dans ce qui suit nous allons parler des deux corpus.

II.3.1 Dataset anglais

La première version de notre dataset a été composée de cinq maladies (panic, Syndrome de stress post-traumatique (ptsd), depression ,anxiety ,narcissim) et pour chaque maladie nous avons associé quatre questions dont la réponse doit être soit par 'YES' ou 'NO'. Ces dernières ont été collectées à partir d'une application de test psychologique « *Mental Health Tests* ».

A	B
maladie	symptomes
panic	In the past 7 days, how many activities have you avoided because they might cause panic attack-like symptoms such as shortness of breath?:
panic	How many panic attacks have you experienced in the past 7 days?:
panic	Over the past 7 days, how concerned have you been that another panic attack may occur?:
panic	Over the past 7 days, how often did panic attacks, or concern about having another attack, prevent you from attending social events?:
panic	If you had one or more panic attacks, in how much distress did you find yourself while they occurred?:
ptsd	part of the traumatic event are hazy or difficult to remember
ptsd	I've been avoiding things like places, people or event that make me think about the trauma
ptsd	Any reminder brought back feelings about the event
ptsd	I avoided letting myself get upset when I thought about it or was reminded of it
ptsd	Images and pictures of the event pop into my mind
depression	difficulty focusing on everyday tasks
depression	finding no joy in life, or feeling like you never will again
depression	being lethargic or having no energy
depression	eating too much, or too little
depression	Feeling down, depressed, or hopeless
anxiety	1 I've been unable to stay in one place, like sitting down, for any length of time (
anxiety	2 I've been very concerned about several things at one time
anxiety	3 I've been agitated, tense, or skittish
anxiety	4 I've been unable to feel calm or relaxed
anxiety	Do you worry about things working out in the future?
narcissism	it's nice when people notice that I've made an effort to look good

Figure 2.2 : capture de la première version du corpus

La deuxième version de notre dataset est une version évoluée de la première, car nous avons augmenté le nombre des maladies à dix (maladies choisies pour leur popularité dans les réseaux sociaux, sitesweb, et applications psychologiques) :

- Dépression : La dépression appelée dans le langage courant "dépression nerveuse", elle se caractérise par des perturbations de l'humeur (tristesse, perte de plaisir).
- Anxiété : C'est un trouble émotionnel qui se manifeste par un sentiment d'insécurité, elle est caractérisée par un sentiment d'appréhension, de tension, de malaise, de terreur face à un péril de nature indéterminée.
- Narcissisme: Le narcissisme est un trouble mental et la poursuite de la gratification de la vanité ou de l'admiration égoïste de son image de soi et de ses attributs idéalisés .
- Schizophrénie : La schizophrénie est une maladie psychiatrique complexe qui se définit comme une perte de contact avec la réalité.

- Panique : est une forme d'anxiété généralement très impressionnante, tant pour la personne qui la subit que pour son entourage.
- Anorexie : est un trouble de l'alimentation caractérisé par un poids corporel anormalement bas, une peur intense de prendre du poids et une perception déformée du poids.
- Bipolaire : est un trouble de l'humeur caractérisé par des périodes de dépression et des périodes d'humeur anormalement élevée qui durent de quelques jours à plusieurs semaines chacune.
- PTSD : Le PTSD (Post Traumatic Stress Disorder, ou Syndrome de Stress Post Traumatique) décrit une réaction psychologique à une situation traumatisante.
- Sleep disorder: des conditions qui perturbent les habitudes de sommeil normales.
- ocd: Le trouble obsessionnel-compulsif présente un schéma de pensées et de peurs indésirables (obsessions) qui vous amènent à adopter des comportements répétitifs (compulsions).

Par la suite, nous avons retenu pour chaque maladie, cent (100) questions. Ce qui nous a permis de constituer un corpus de mille (1000) données. Ceci était possible grâce à une collecte effectuée à partir d'applications et des tests en ligne résumés dans le tableau suivant (tableau 2.1) :

Maladies	Application	Test en ligne
anxiety	-PsyTest - 40+ Psychological Test	-MHA : une collection d'outils de dépistage en ligne, gratuits, confidentiels, anonymes et scientifiquement

	<p>-Mental Health Tests</p>	<p>validés pour aider les individus à comprendre et à en savoir plus sur leur santé mentale.⁴</p> <p>-PSYCOM : Les tests proposés sur ce site sont destinés à aider les patients à déterminer s'ils peuvent bénéficier d'un traitement supplémentaire.⁵</p> <p>-depression.org.nz : Ce site Web regorge d'idées et d'histoires pour aider les personnes souffrant de dépression et d'anxiété à se sentir mieux.⁶</p> <p>-Anxiety.org.⁷</p> <p>- anxietycentre8 : une organisation dédiée à aider les gens à surmonter le trouble anxieux, ainsi qu'un certain nombre d'autres problèmes de santé mentale, y compris le trouble de stress, les problèmes de sommeil (insomnie) et le trouble dépressif, ainsi que les dépendances, le développement personnel et défis des relations interpersonnelles.</p> <p>- theanxietytest9: Ce test est basé sur la classification officielle des troubles anxieux utilisé par les professionnels de la santé mentale .</p> <p>- idrlabs¹⁰ : Test d'anxiété de 3 minutes Basé sur les travaux du Dr. Robert L. Spitzer, M.D</p> <p>-QuizExpo¹¹:est un site des divertissement et d'éducation basé sur des quiz.</p>
<p>narcissistic</p>	<p>- PsyTest -40+ Psychological Test</p>	<p>- PSYCHCENTRAL : vous mettre en relation avec les conseils et le soutien dont vous avez besoin pour</p>

4 <https://screening.mhanational.org/screening-tools/anxiety>

5 <https://www.psycom.net/anxiety-test>

6 https://depression.org.nz/is-it-depression-anxiety/self-test/anxiety-test/?fbclid=IwAR1_4nefHgRDT1wfN1AE8EFq31wBBfGU9h2_cI3n2yix8612k_2PSjQao4

7 <https://www.anxiety.org/do-i-have-anxiety>

8 <https://www.anxietycentre.com/tests/anxiety-test/>

9 <http://www.theanxietytest.com/>

10 <https://www.idrlabs.com/fr/anxiete-stress-depression/test.php>

11 <https://www.quizexpo.com/accurate-anxiety-test/>

	-Mental Health Tests	effectuer des changements significatifs qui peuvent vous être bénéfiques dans votre vie réelle ¹² . -PSYCOM -idrlabs
schizophrenia	-PsyTest - 40+ Psychological Test -Mental Health Tests	-psychcentral -pscho-tests 13 -schizophrene.info14 -Psymed.info¹⁵ -myrealpersonality¹⁶
panic	- PsyTest - 40+ Psychological Test - Mental Health Tests	-anxietycentre -Healthyplace :est le plus grand site de santé mentale grand public sur le net.IL fournie des informations et un soutien faisant autorité aux personnes ayant des problèmes de santé. 17. -le meilleur avis ¹⁸ -Anxiety Network ¹⁹ -verywellhealth²⁰
anorexia	-PsyTest - 40+ Psychological Test -Mental Health Tests	- idrlabs -Psymed.info

12 <https://psychcentral.com/quizzes/narcissistic-personality-quiz>

13 <https://psycho-tests.com/test/schizophrenia-test>

14 <http://www.schizophrene.info/>

15 <https://psymed.info/schizophrenia-test>

16 https://myrealpersonality.com/fr/tests/test-de-schizophr-nie-auto-test-de-d-pistage-de-la-schizophr-nie#google_vignette

17 <https://www.healthyplace.com/psychological-tests/panic-attack-test>

18 <https://www.lemeilleuravis.com/manic-panic-test/>

19 <https://anxietynetwork.com/content/test-panic-disorder/>

20 <https://www.verywellhealth.com/panic-disorder-diagnosis-5114973>

		<p>-Recovery Village : offre des services de traitement complets pour la toxicomanie et les troubles de santé mentale concomitants²¹</p> <p>-Le Journal des Femmes : le premier magazine féminin français en ligne²²</p> <p>- Addictions-Auvergne : est destiné aux personnes en difficulté avec un produit ou un comportement pour aider à se repérer face aux addictions et trouver où et comment se faire aider²³</p> <p>-CLINICAL PARTNERS : un des principaux prestataires de services de santé mentale du Royaume-Uni, qui travaille à la fois dans le secteur privé et avec le NHS pour aider les personnes et les familles rencontrant des difficultés mentales et émotionnelles à aller mieux²⁴</p>
Depression	-depression test	<p>-VIDAL : l'intelligence médicale au service du soin²⁵ .</p> <p>-PSYMONTREAL: Test d'auto-évaluation ²⁶ .</p> <p>-MHA</p> <p>-PSYCOM</p> <p>-CLINICAL PARTNERS</p>
Bipolar	<p>bipolar test</p> <p>-mental health test</p> <p>-bipolar disorder test</p>	<p>-MHA</p> <p>-PSYCOM</p>

21 <https://www.therecoveryvillage.com/mental-health/anorexia/do-i-have-anorexia/>

22 <https://sante-medecine.journaldesfemmes.fr/faq/4849-detecter-l-anorexie-test>

23 <http://addictions-auvergne.fr/site/content/testez-vous-0>

24 <https://www.clinical-partners.co.uk/for-adults/eating-disorders-clinic/a-test-for-anorexia/test>

25 <https://www.vidal.fr/maladies/psychisme/depression-adulte/test.html>

26 <https://psymontreal.com/test-de-depression/>

		<p>-BLACK DOG le seul institut de recherche médicale en Australie à étudier la santé mentale 27 .</p> <p>-Families for Depression Awareness aide les familles à reconnaître et à gérer la dépression et les troubles bipolaires afin d'aider les gens à se rétablir et de prévenir les suicides.²⁸</p> <p>-PSYCHCENTRAL vous mettre en relation avec les conseils et le soutien dont vous avez besoin pour effectuer des changements significatifs qui peuvent vous être bénéfiques dans votre vie réelle. ²⁹</p>
ptsd	-ptsd test	<p>-CLINICAL PARTNERS</p> <p>-MIND DIAGNOSTICS : Il propose un test de santé mentale pour déterminer les symptômes d'un problème de santé mentale 30 .</p> <p>-TALKSPACE : un réseau qui compte des milliers de thérapeutes agréés et expérimentés dans le traitement du Ptsd 31 .</p> <p>-ROGERS : Un groupe de professionnels de premier plan assure des soins de qualité 32 .</p> <p>-NHS : est le plus grand site web de santé du Royaume-Uni, avec plus de 50 millions de visites par mois 33.</p>
Sleep disorder	-mental health test	<p>-Polysomnography : est un test complet utilisé pour diagnostiquer les troubles du sommeil 34.</p>

27 <https://www.blackdoginstitute.org.au/resources-support/digital-tools-apps/bipolar-self-test/>

28 <https://www.familyaware.org/help-someone/take-the-depression-and-bipolar-test/>

29 <https://psychcentral.com/quizzes/bipolar-quiz#1>

30 <https://www.mind-diagnostics.org/ptsd-test>

31 <https://www.talkspace.com/assessments/ptsd-test>

32 <https://rogersbh.org/ptsdquiz>

33 <https://www.nhs.uk/mental-health/conditions/post-traumatic-stress-disorder-ptsd/overview/>

34 <https://www.mayoclinic.org/tests-procedures/polysomnography/about/pac-20394877#:~:text=Polysomnography%2C%20also%20called%20a%20sleep,leg%20movements%20during%20the%20study>

		<p>-Clayton sleep institute : est composé de spécialistes du sommeil et de scientifiques agréés, de technologues du sommeil et de thérapeutes respiratoires agréés ayant des années d'expérience dans le domaine des troubles du sommeil 35</p> <p>-Columbus sleep consultants : il offre un environnement confortable pour votre diagnostic il s'agit d'une équipe multidisciplinaire de médecins hautement qualifiés, spécialisés dans le diagnostic et le traitement des troubles du sommeil 36</p> <p>-WebMd : organisation qui, selon elle, remplit la promesse d'informations sur la santé sur Internet. Elle fournit des informations crédibles, des communautés de soutien et des documents de référence approfondis sur des sujets liés à la santé 37</p> <p>-Sharp il améliore la santé de tous ceux que nous servons. Et c'est le principal fournisseur de soins de santé de San Diego. 38</p>
ocd	-ocd test	<p>-Psycom</p> <p>-Patient : Des informations cliniques fiables, rédigées et examinées par un vaste réseau de médecins et de professionnels de la santé, aident les gens à se sentir mieux et à vivre plus longtemps 39 .</p> <p>-CLINICAL PARTNERS</p> <p>-PSYMONTREAL</p> <p>-Nocd : Les thérapeutes agréés sont spécialisés dans la thérapie d'exposition et de prévention de la réponse, le traitement le plus efficace des TOC. Ils sont formés par</p>

35 <https://claytonsleep.com/adult-sleep-quiz-sleep-disorder/>

36 <https://columbussleep.com/quizzes/>

37 <https://www.webmd.com/sleep-disorders/understanding-sleep-problems-treatment>

38 <https://www.sharp.com/services/sleep-disorders/quiz.cfm>

39 <https://patient.info/news-and-features/quiz-do-i-have-ocd>

		certains des meilleurs experts et chercheurs du monde en TOC. 40
--	--	--

Tableau 2.1 : applications et tests en ligne utilisés

Dans une troisième étape et afin de constituer la version finale du dataset. Un dataset de deux mille (2000) données a pu être construit à partir de la deuxième version en utilisant la reformulation. Pour cela nous avons utilisé les deux sites de reformulation suivants : prepostseo⁴¹ et paraphraz.it⁴².

Il s'agit de présenter un texte sous une autre forme, sans que l'information soit modifiée, en d'autres termes, aucune donnée ne doit être supprimée ou ajoutée à la source.

Pour cela, nous pouvons recourir à :

- synonymes : c'est la ressource la plus commune que nous utilisons pour reformuler une phrase, bien qu'il soit parfois plus difficile de trouver le mot juste qu'il n'y paraît.

- hyperonymes : ce sont des termes dont le sens inclut celui d'autres.

- pronoms : grosso modo, ce sont ces mots qui peuvent désigner une chose sans la nommer directement.

- la paraphrase : quand on ne trouve pas un synonyme, un hyperonyme ou un pronom qui nous permet de faire la reformulation, on peut toujours construire une phrase qui a le même sens que l'originale.

II.3.2 Dataset arabe

⁴⁰ <https://www.treatmyocd.com/blog/where-can-i-find-a-good-online-ocd-test>

⁴¹ <https://www.prepostseo.com/about>

⁴² <https://paraphraz.it/fr/>.

Afin de constituer la version arabe du dataset. Un dataset de deux mille (2000) données a pu être construit à partir de la version finale du dataset anglais en utilisant la traduction. Pour cela nous avons utilisé la fonction GOOGLETRANSLATE⁴³ de l'application Web Google Sheets⁴⁴.

B	A
	المرض
هل أزعجت بسبب الشعور بالانقباض أو الاكتئاب أو العصبية أو اليأس خلال الأسبوعين الأخيرين؟	الانقباض
هل أزعجت أن لديك القليل من الاهتمام أو المتعة في فعل الأشياء خلال الأسبوعين الماضيين؟	الانقباض
هل تم إزعاجك عشاء المتاعب الناعمة أو النوم أو النوم أكثر من اللازم خلال الأسبوعين الأخيرين؟	الانقباض
هل أزعجت أن لديك شهية سيئة وفقدان الوزن أو الإفراط في تناول الطعام على مدار الأسبوعين الماضيين؟	الانقباض
هل أزعجت بسبب الشعور بالتعب، أو وجود طاقة صغيرة على مدى الأسبوعين الأخيرين؟	الانقباض
هل تم إزعاجك من الشعور بالسوء عن نفسك - أو الشعور بأنك فشل، أو أنك تسمح لنفسك أو عائلتك على مدى الأسبوعين الأخيرين؟	الانقباض
هل أزعجت أن لديك مشكلة في التركيز على أشياء مثل العمل المدرسي أو القراءة أو مشاهدة التلفزيون خلال الأسبوعين الماضيين؟	الانقباض
هل تم إزعاجك عن طريق التحرك أو التحدث ببطء شديد يمكن أن يلاحظ الآخرون؟ أو العكس - أن تكون ملموسا جدا أو لا يبدأ أنك كنت تتحرك حوله أكثر من المعتاد خلال الأسبوعين الماضيين؟	الانقباض
كم مرة أزعجت الأفكار التي ستكون أفضل حالا من الأموات، أو إيذاء نفسك بطريقة ما على مدار الأسبوعين الماضيين؟	الانقباض
القليل من الاهتمام أو المتعة في فعل الأشياء	الانقباض
الشعور بالانقباض أو الاكتئاب أو اليأس	الانقباض
مشكلة في السقوط أو البقاء ناعما، أو تنام كثيرا	الانقباض
الشعور بالتعب أو وجود طاقة صغيرة	الانقباض
شهية سيئة أو الإفراط في تناول الطعام	الانقباض
الشعور بالسوء عن نفسك - أو أنك فشل أو السماح لنفسك أو عائلتك	الانقباض
مشكلة في التركيز على الأشياء، مثل قراءة الصحيفة أو مشاهدة التلفزيون	الانقباض
تتحرك أو يتحدث ببطء شديد أن الأشخاص الآخرين قد لاحظوا	الانقباض
الأفكار التي ستكون أفضل قبالة الموت، أو إيذاء نفسك	الانقباض
أفعل الأشياء ببطء	الانقباض
مستقبلي يبدو ميؤوس منها	الانقباض
من الصعب بالنسبة لي التركيز على القراءة	الانقباض
لقد خرج السرور والفرح من حياتي	الانقباض

Figure 2.3 : capture de la première version du corpus d'arabe

Après la vérification manuelle de la traduction nous avons remarques que les phrases reformulées à partir de la deuxième version donnent la même traduction que celles de la deuxième version, Pour plus de clarté, un exemple est donné ci-dessous :

⁴³ **GOOGLETRANSLATE** une fonction permet de traduire un texte en mode automatique ou en fonction des langues spécifiées, Utilisation : =GOOGLETRANSLATE(texte ; langue_source; langue_cible)

⁴⁴ **Google Sheets** est une application Web qui permet aux utilisateurs de créer, mettre à jour et modifier des feuilles de calcul et de partager les données en ligne en temps réel.

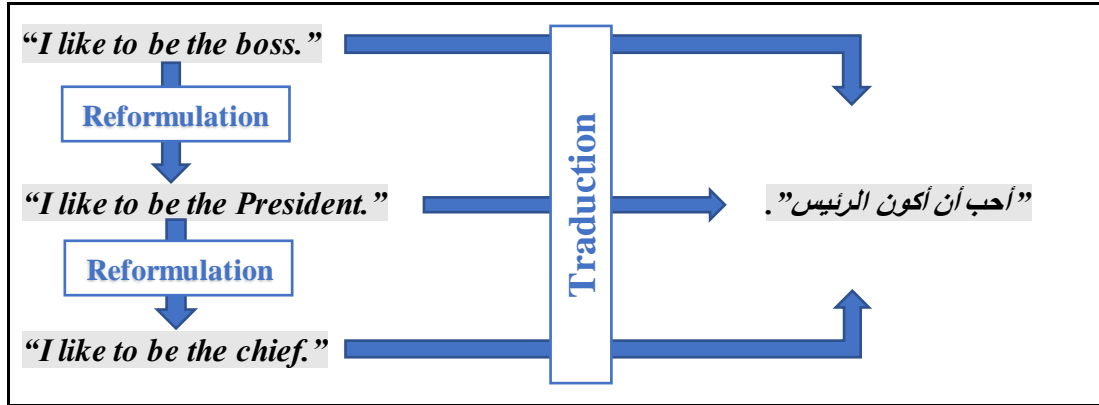


Figure 2.4 : Exemple de traduction de la version finale du dataset anglais.

Par la suite, nous avons constitué un corpus de mille (1000) données. Ceci était possible grâce à une traduction effectuée sur la deuxième version du dataset anglais à partir de la fonction GOOGLETRANSLATE de l'application Web Google Sheets.

972	اضطراب الوسواس القهري	هل تعاني من صور ذهنية مزعجة كالملوث أو الدمار أو الأحداث غير السارة الأخرى
973	اضطراب الوسواس القهري	هل تقلق بشأن التصرف بناءً على رغبة لا معنى لها ، مثل دفع شخص غريب أمام حافلة أو طعن أحد أفراد أسرته بسكين
974	اضطراب الوسواس القهري	هل تشغل بالتأكد من أن الأشياء في ترتيبها الصحيح؟ (فهرسة الكتب ، إعادة ترتيب الأثاث ، ترتيب الأواني الفضية ، إلخ.)
975	اضطراب الوسواس القهري	هل تقلق بشكل مفرط بشأن أشياء مثل الحرائق أو حوادث السيارات أو ميثاقك
976	اضطراب الوسواس القهري	هل تحقق من الأشياء مثل المقايض القرن وأقفال الأبواب وفرامل السيارات مرارًا وتكرارًا في غضون فترة زمنية قصيرة
977	اضطراب الوسواس القهري	هل تقلق بشأن انتشار المرض (حتى لو لم يتم تشخيصك بالمرض من قبل)
978	اضطراب الوسواس القهري	هل لديك أفكار غير مقبولة شخصيًا (غالبًا ما تكون ذات طبيعة دينية أو جنسية) وتشعر بالتطفل وخروجك عن سيطرتك
979	اضطراب الوسواس القهري	هل تكرر الإجراءات الروتينية مثل فتح الباب ، أو ارتداء الحذاء ، أو النوم مرارًا وتكرارًا حتى "تشعر أنك على ما يرام؟"
980	اضطراب الوسواس القهري	هل تمارس عادات الغسل والتنظيف والعناية الشخصية (غسل يديك خمس مرات متتالية ، على سبيل المثال)
981	اضطراب الوسواس القهري	هل تتجنب الأوثان أو أرقامًا معينة لأنك تنظر إليها على أنها غير محظوظة أو شريرة
982	اضطراب الوسواس القهري	هل تقوم بإعادة قراءة الرسائل أو رسائل البريد الإلكتروني أو الرسائل النصية قبل وبعد إرسالها عدة مرات
983	اضطراب الوسواس القهري	هل كنت قادرًا على ممارسة السيطرة على تلك الأفكار الوسواسية
984	اضطراب الوسواس القهري	هل كنت قادرًا على ممارسة السيطرة على السلوكيات القهرية
985	اضطراب الوسواس القهري	هل قضيت الاخرط في سلوكيات الإدمان
986	اضطراب الوسواس القهري	هل تجعلك هذه الأفكار تشعر بالأسى أو بالضيق
987	اضطراب الوسواس القهري	هل تمكنت من منع نفسك من التفكير في الهوس؟
988	اضطراب الوسواس القهري	هل أعاقت هذه السلوكيات القهرية طريقك أو مدرستك أو عملك أو حياتك الاجتماعية
989	اضطراب الوسواس القهري	هل شعرت بالحاجة إلى التفكير في بعض الأفكار أو العبارات بصمت في عقلك؟ يمكن أن تشمل الأمثلة العد أو تكرار الصلاة أو كلمات أخرى مرارًا وتكرارًا أو إعادة قراءة عملك عدة مرات
990	اضطراب الوسواس القهري	هل نسيت للحصول على تطبيقات من العائلة أو الأصدقاء أو زملاء العمل فيما يتعلق بالأشياء التي تقلقك بما في ذلك ما قمت به أو لم تفعله
991	اضطراب الوسواس القهري	هل تحفظ الأشياء التي قد يومية الآخرون ويصعب عليهم التخلص منها
992	اضطراب الوسواس القهري	هل تسحب الشعر من رأسك أو من مناطق أخرى من الجسم
993	اضطراب الوسواس القهري	هل تترجع جلدك أو تجرح عن عيوبك
994	اضطراب الوسواس القهري	هل تشعر أن هناك شيئًا خاطئًا في مظهرك الجسدي مثل شكل أو حجم أنفك أو أذنك أو يديك أو أجزاء أخرى من جسمك عندما لا يشعر الآخرون بذلك
995	اضطراب الوسواس القهري	لقد قمت بحفظ الكثير من الأشياء التي تعرض طريقهم
996	اضطراب الوسواس القهري	أشعر بالضيق إذا لم يتم ترتيب الأشياء بشكل صحيح
997	اضطراب الوسواس القهري	أجد صعوبة في لمس شيء ما عندما أعلم أنه قد لمس غيابه أو أشخاص معينون
998	اضطراب الوسواس القهري	أضطر أحيانًا إلى غسل نفسي أو تنظيف نفسي لمجرد أنني أشعر بالتلوث
999	اضطراب الوسواس القهري	أحجب رمي الأشياء بعيدًا لأنني خائف قد أحتاج إليهم لاحقًا
1000	اضطراب الوسواس القهري	أشعر أن هناك أرقام جيدة وسيئة

Figure 2.5: La version finale du dataset arabe.

II.5 Spécification et analyse des besoins

Dans cette section, nous allons présenter notre modélisation du système avec une approche plus directe axée principalement sur les différents scénarios pouvant survenir lors des interactions entre l'utilisateur et le système.

Nous allons présenter les différents scénarios sous forme de diagrammes de séquence représentant chacun un cas de notre système, mais nous allons d'abord présenter le modèle général des cas possibles qui peuvent se produire comme indiqué dans le diagramme de cas d'utilisation suivant

II.5.1 Diagramme de cas d'utilisation

La figure ci-après (Figure 2.24), représente notre cas d'utilisation général. Il détaille toutes les actions que les utilisateurs peuvent effectuer sur la plateforme.

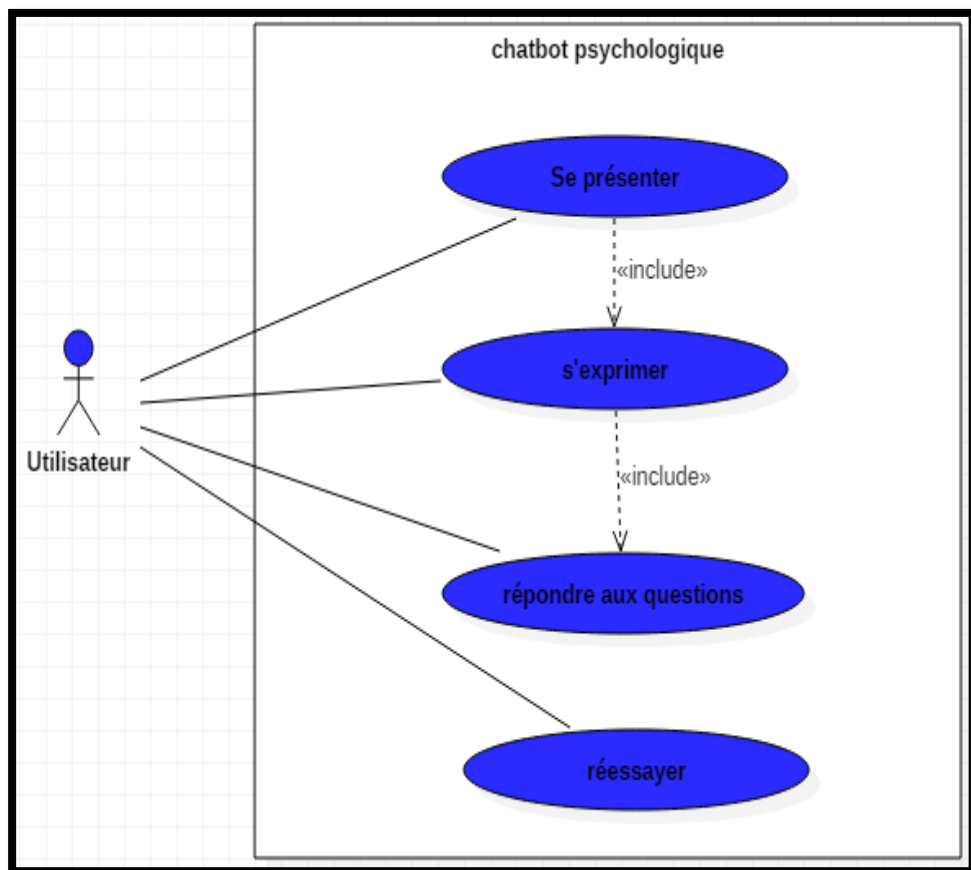


Figure2.6 : diagramme de cas d'utilisation

Comme le montre le schéma précédent, nous pouvons distinguer différentes entités composant notre schéma qui sont : Acteurs et Cas.

Concernant les acteurs, nous n'avons qu'un seul acteur qui est l'Utilisateur puisque les différentes interactions sont liées aux besoins de l'utilisateur.

En revanche, il existe différents cas qui peuvent se produire en fonction de certaines conditions, nous pouvons les résumer dans le tableau suivant :

Cas	description
Se présenter	Pour commencer la discussion Chaque nouvel utilisateur doit insérer un nom donc la plateforme pourrait l'appeler avec ce nom
S'exprimer	On demande à l'utilisateur de s'exprimer pour le laisser formuler sa pensée par la parole afin que la machine puisse classer son sentiment
Répondre aux questions	Après la classification la machine pose quelques questions à l'utilisateur donc il doit répondre par vrai ou faux
Réessayer	Ce cas se produit uniquement lorsque la discussion est terminée, l'utilisateur peut refaire le test

Tableau 2.2 : Tableau résumant les cas du diagramme de cas d'utilisation

II.5.2 Diagramme de séquence

Dans cette sous-section, nous allons présenter le diagramme de séquence pour montrer le cours des événements qui s'y rapportent.

Le scénario global de l'interaction entre le système et l'utilisateur se présente comme suit :

- L'utilisateur accède au système
- Le système affiche la page d'accueil
- L'utilisateur saisit son nom
- Le système invite l'utilisateur à s'exprimer
- L'utilisateur s'exprime via un texte.
- La classification du texte saisi par l'utilisateur afin d'aider à catégoriser les mots importants qui reflètent les symptômes et déduire la maladie de l'utilisateur.
- Dix questions de test sont apportées de la base de données de test.

- Le système affiche les dix questions de test.
- L'utilisateur répondre aux questions par « yes » ou « no ».
- Le système fait les diagnostique.
- Le système affiche les diagnostique.
- Le système affiche les Conseils.
- Le système invite l'utilisateur à réessayer le test.

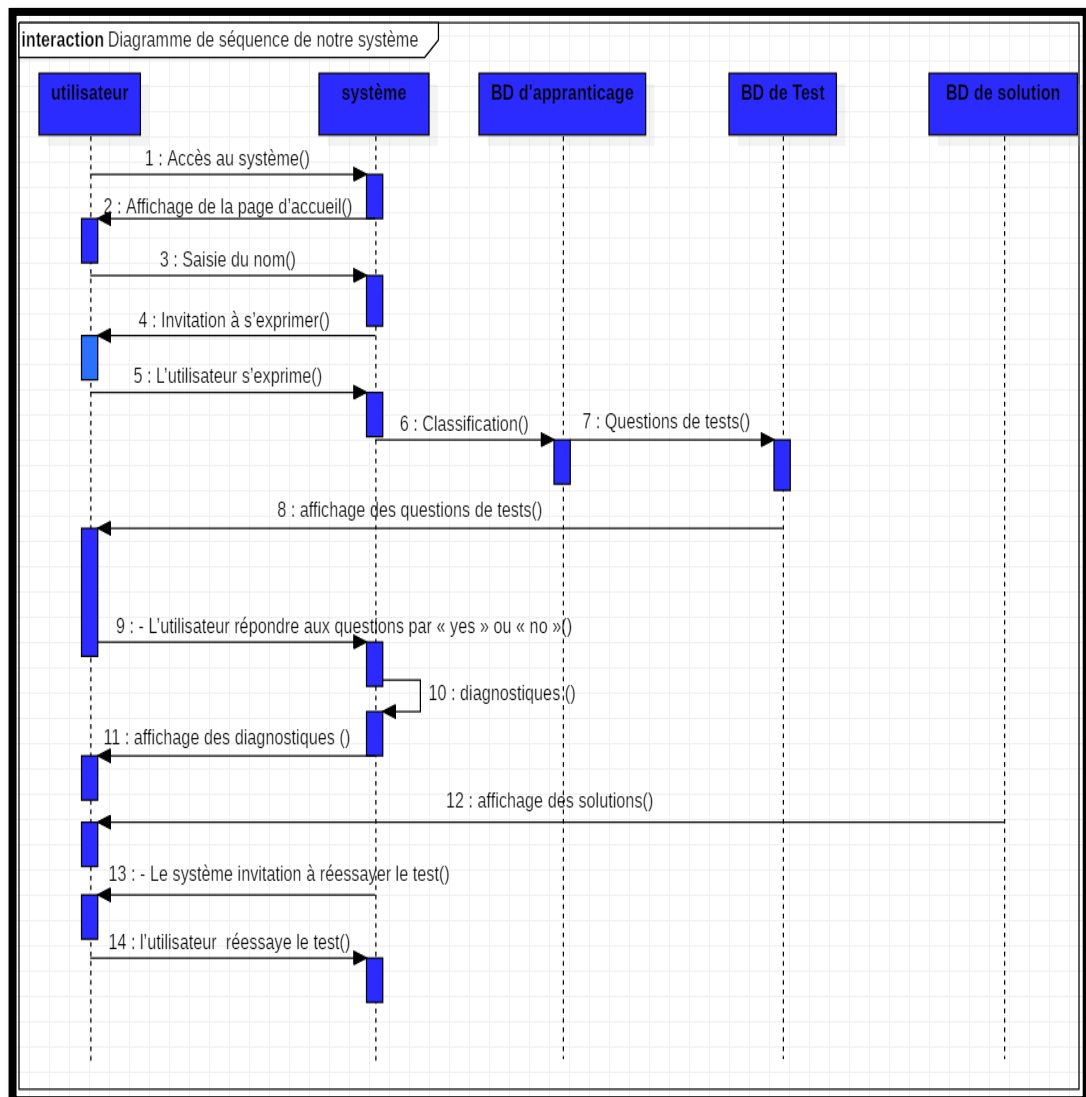


Figure 2.7 : Diagramme de séquence de notre système

Comme montre (la figure 2.25), dès que l'utilisateur accède au système il doit saisir son nom et s'exprimer via un texte.

Après la classification du texte saisi par l'utilisateur afin d'aider à catégoriser les mots importants qui reflètent les symptômes et déduire la maladie de l'utilisateur, 10 questions sur la maladie déduite précédemment sont posées à l'utilisateur qui doit répondre par « yes » ou « no ».

Un affichage des diagnostics et des Conseils qui sont tirés de la base de données de Conseils et une proposition à l'utilisateur de réessayer le test.

Nous allons brièvement parler des différentes BDs dans la prochaine section.

II.5.3 Diagramme de classes

Dans cette sous-section, nous allons présenter le diagramme de classes pour montrer les différentes BDs.

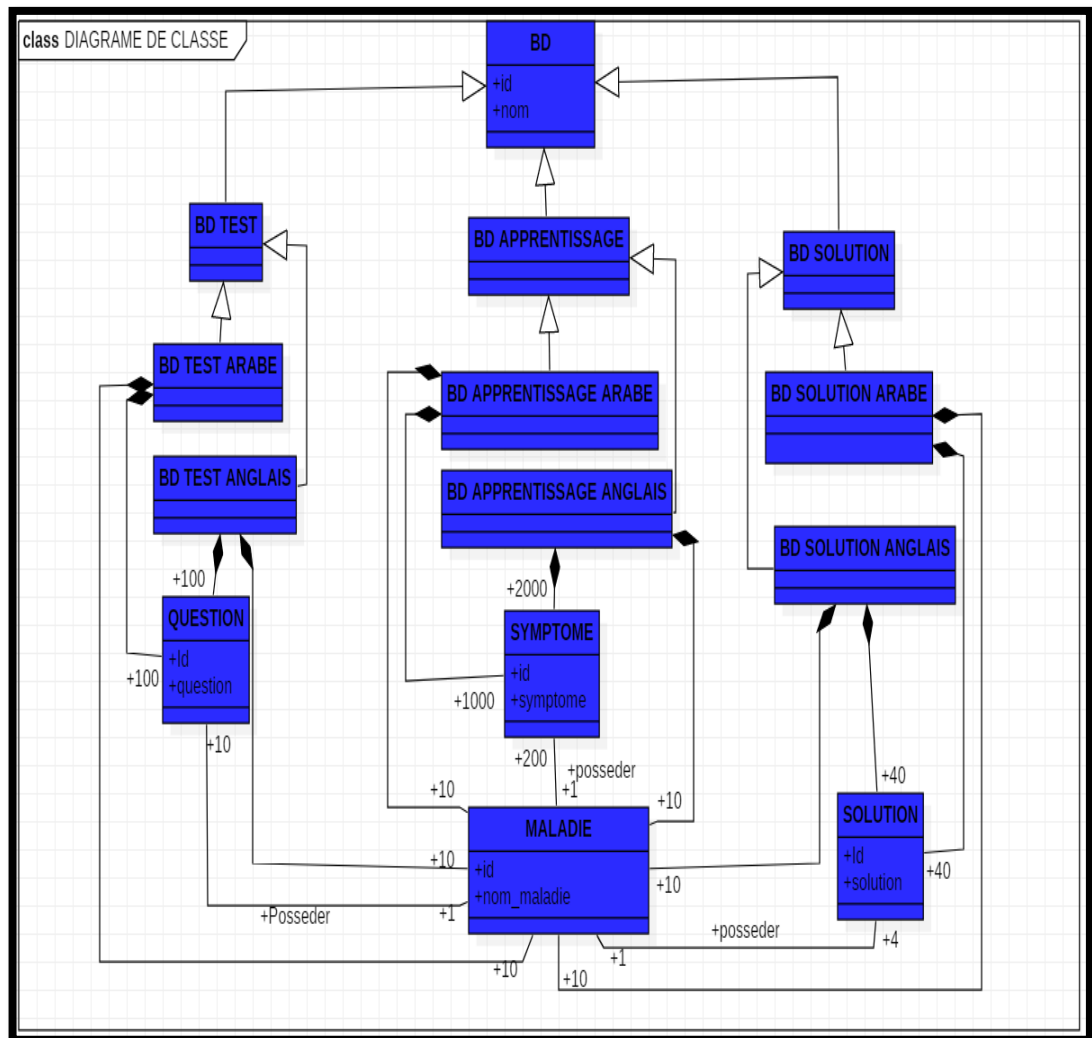


Figure 2.8 : Bases de données du système.

Comme montre (la figure 2.26), la BD est divisée en trois :

- BD d'apprentissage divisée en deux :
 - BD d'apprentissage anglaise, elle est composée de dix maladies et deux milles symptômes, chaque maladie possède deux cent symptômes.
 - BD d'apprentissage arabe, elle est composée de dix maladies et mille symptômes, chaque maladie possède cent symptômes.
- BD de test divisée en deux :
 - BD de test anglaise, elle est composée de dix maladies et cent questions, chaque maladie possède dix questions.
 - BD de test arabe, elle est composée de dix maladies et cent questions, chaque maladie possède dix questions.

- BD des Conseils divisée en deux :
 - BD des Conseils anglaise, elle est composée de dix maladies et pour chaque maladie nous avons associé quatre Conseils. Ces dernières sont le fruit d'une collaboration entre des médecins généraliste, des psychiatres et d'autres professionnels de santé.
 - BD des Conseils arabe, construit à partir de la BD des Conseils anglaise en utilisant la traduction. Elle est composée de dix maladies et quarante Conseils, chaque maladie possède quatre Conseils.

II.6 Conclusion

Après avoir présenté le processus global, expliqué la construction de notre dataset et cité les notions théoriques nécessaires à la compréhension des étapes principales de notre traitement. Dans la le chapitre suivant, nous allons présenter leur fonctionnement exactes dans le cadre de notre système, l'environnement de développement et les détails de chaque étape de développement.

Chapitre III :

Tests et Validation de la solution

III.1 Introduction

Dans les chapitres précédents, nous avons surtout parlé théoriquement des Chatbots en général, et du background théorique de notre solution en particulier. Dans ce chapitre, nous allons présenter l'environnement de développement et les détails de chaque étape de développement.

III.2 Bibliothèques utilisées

Les sections suivantes seront consacrées à la présentation des bibliothèques utilisées pour le développement de notre système.

III.2.1 NumPy

NumPy est la bibliothèque open-source fondamentale du calcul scientifique en Python pour le traitement de tableaux à usage général (voir le logo dans la figure cidessous). Elle fournit des tableaux multidimensionnels de haute performance et des outils pour les manipuler.



Figure3-1 : Logo de la bibliothèque NumPy.

Au-delà de ses nombreuses utilisations dans le cadre scientifique, Numpy peut aussi être utilisée en tant que conteneur multidimensionnel performant pour des données générales. Ce qui rend NumPy accessible et productif est sa syntaxe de haut niveau, sa large variété de fonctions mathématiques complètes et de générateurs de nombres aléatoires ainsi que sa capacité de prise en charge d'un large éventail de plateformes matérielles et informatiques.

III.2.2 Pandas

Pandas est un package Python qui fournit des structures de données rapides, flexibles et expressives conçues pour rendre le travail avec des

données « relationnelles » ou « étiquetées » à la fois facile et intuitif. Il vise à être le bloc de construction fondamental de haut niveau pour effectuer une analyse de données pratique et réelle en Python. De plus, il a pour objectif plus large de devenir l'outil d'analyse/manipulation de données open source le plus puissant et le plus flexible disponible dans n'importe quelle langue. Elle est déjà bien avancée vers cet objectif.



Figure 3-2 : Logo de Pandas.

Voici Caractéristiques principales de pandas :

- Gestion facile des données manquantes
- Mutabilité de taille
- Alignement automatique et explicite des données
- Fonctionnalité de regroupement puissante et flexible pour effectuer des opérations de fractionnement-application-combinaison sur des ensembles de données.

Pandas a été utilisé pour ouvrir, lire et organiser notre data set.

III.2.3 TensorFlow

TensorFlow⁴⁵ est une plateforme d'apprentissage automatique open source de bout en bout (voir le logo dans la figure ci-dessous). Il est doté d'un écosystème exhaustif et flexible de bibliothèques, de ressources et d'outils communautaires permettant aux chercheurs de promouvoir les avancées de l'apprentissage automatique, et aux développeurs de créer et de déployer aisément des applications d'apprentissage automatique.

⁴⁵ <https://www.tensorflow.org/>



Figure 3-3 : Logo de la bibliothèque TensorFlow.

TensorFlow a vu le jour aux mains d'ingénieurs et de chercheurs membres de l'équipe Google Brain au sein de l'organisation de recherche sur l'intelligence artificielle de Google. TensorFlow fournit des API Python et C++ statistiques, ainsi qu'une API rétro compatible non cautionnée pour d'autres langages.

III.2.4 Keras

Keras est une bibliothèque open source pour la création d'applications de deep learning. Keras est écrit en Python et fournit une interface commune pour différents backends de deep learning tels que TensorFlow et Theano. Le Deep Learning fait partie du machine learning et se base sur les réseaux de neurones artificiels. Keras vise à simplifier l'entrée dans le deep learning comme le montre la figure suivante :

```
from keras.preprocessing.text import Tokenizer
from keras.preprocessing import sequence
from keras.models import Sequential
from keras.layers import Dense, Dropout, Activation
from keras.layers import Embedding
from keras.layers import Conv1D, GlobalMaxPooling1D
from keras.utils import np_utils
```

Figure3-4 : Code d'importation de la bibliothèque Kears.

Nous l'utilisons pour certaines opérations de prétraitement comme le montre la figure suivante :

```
from keras.preprocessing.text import Tokenizer
from keras.preprocessing import sequence
```

Figure3-5 : Quelques opération de prétraitement avec Kears.

III.2.5 Scikit Learn

Scikit Learn est un outil simple et efficace pour l'analyse des données de prédiction.



Figure3-6 : Logo de Scikit Lean.

Cette bibliothèque a été utilisée pour les classificateurs, l'extraction de caractéristiques et pour les mesures d'évaluation, comme le montrent les figures et qui présentent le code.

```
from sklearn import model_selection, preprocessing, naive_bayes, metrics, svm, linear_model, neighbors
```

Figure3-7 : classificateurs de Sklearn.

```
from sklearn.feature_extraction.text import TfidfVectorizer  
from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer  
from sklearn import decomposition, ensemble  
from sklearn.pipeline import Pipeline
```

Figure3-8 : Sklearn Feature Code.

```
from sklearn.metrics import accuracy_score  
  
from sklearn.metrics import classification_report  
from sklearn.metrics import confusion_matrix  
from sklearn.metrics import cohen_kappa_score  
from sklearn.metrics import precision_score
```

Figure3-9 : Sklearn Metrics Code.

III.2.6 FastText

FastText est une bibliothèque pour un apprentissage efficace des représentations de mots et de la classification des phrases.



Figure3-10 : Logo de FastText.

La bibliothèque FastText fournit les capacités suivantes via ses outils :

- **Formation de classificateur :** Formation de classificateur supervisé pour la classification de texte.
- **Entraînement du modèle SkipGram :** Apprentissage des représentations de mots/des vecteurs de mots à l'aide de la technique de skipgram.
- **Quantification :** La quantification est un processus appliqué sur un modèle afin de réduire l'utilisation de la mémoire lors de la prédiction.
- **Prédictions :** Prédiction d'étiquettes pour un texte donné : Classification de texte.
- **Prédictions avec Probabilités :** Prédiction de probabilités en plus des étiquettes pour un texte donné , Classification de Texte.
- **Entraînement du modèle CBOW :** Apprentissage des représentations de mots/vecteurs de mots à l'aide de la technique CBOW (Continuous Bag Of Words).
- **Impression des vecteurs de mots :** Impression de vecteurs de mots pour un modèle entraîné avec chaque ligne représentant un vecteur de mot.
- **Impression des vecteurs de phrases** [print-sentence-vectors] Impression de vecteurs de phrases pour un modèle entraîné avec chaque ligne représentant un vecteur pour un paragraphe.
- **Interrogation des voisins les plus proches.**
- **Recherche d'analogies.**

III.2.7 TextBlob

TextBlob est une bibliothèque Python (2 et 3) pour le traitement de données textuelles.

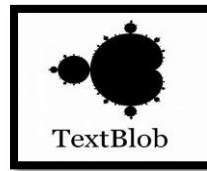


Figure3-11: Logo de TextBlob.

Il fournit une API simple pour plonger dans les tâches courantes de traitement du langage naturel (NLP) telles que le balisage de parties de discours, l'extraction de phrases nominales, l'analyse des sentiments, la classification, la traduction.

III.2.8 Seaborn

Seaborn est une bibliothèque permettant de créer des graphiques statistiques en Python. Elle est basée sur Matplotlib, et s'intègre avec les structures Pandas.



Figure3-12 : Logo de Seaborn.

Cette bibliothèque est aussi performante que Matplotlib, mais apporte une simplicité et des fonctionnalités inédites. Elle permet d'explorer et de comprendre rapidement les données.

III.2.9 Stanza

Stanza est un package d'analyse du langage naturel Python. Il contient des outils, qui peuvent être utilisés dans un pipeline, pour convertir une chaîne contenant du texte en langage humain en listes de phrases et de mots, pour générer des formes de base de ces mots, leurs parties du discours et leurs caractéristiques morphologiques, pour donner une analyse syntaxique des dépendances de la structure, et de reconnaître les entités nommées. La boîte à outils est conçue pour être parallèle à plus de 70 langues, en utilisant le formalisme des dépendances universelles .



Figure 3-13: Logo de Stanza.

Pour résumer, Stanza propose :

- Implémentation native de Python nécessitant un minimum d'efforts de configuration.
- Pipeline de réseau neuronal complet pour une analyse de texte robuste, y compris la tokenisation, l'expansion de jetons multi-mots (MWT), la lemmatisation, le balisage des parties du discours (POS) et des caractéristiques morphologiques, l'analyse des dépendances et la reconnaissance des entités nommées.
 - Modèles neuronaux pré-entraînés prenant en charge 66 langues (humaines) .
 - Une interface Python stable.

III.2.10 Googletrans

Googletrans est une bibliothèque Python gratuite et illimitée qui implémente l'API Google Translate. Cela utilise l'API Google Translate Ajax pour faire des appels à des méthodes telles que la détection et la traduction.

III.3 Chargement des données

Cette étape représente l'ouverture, la lecture et l'organisation de notre dataset.

Pour cela nous avons utilisé Pandas, la figure suivante présente ces opérations :

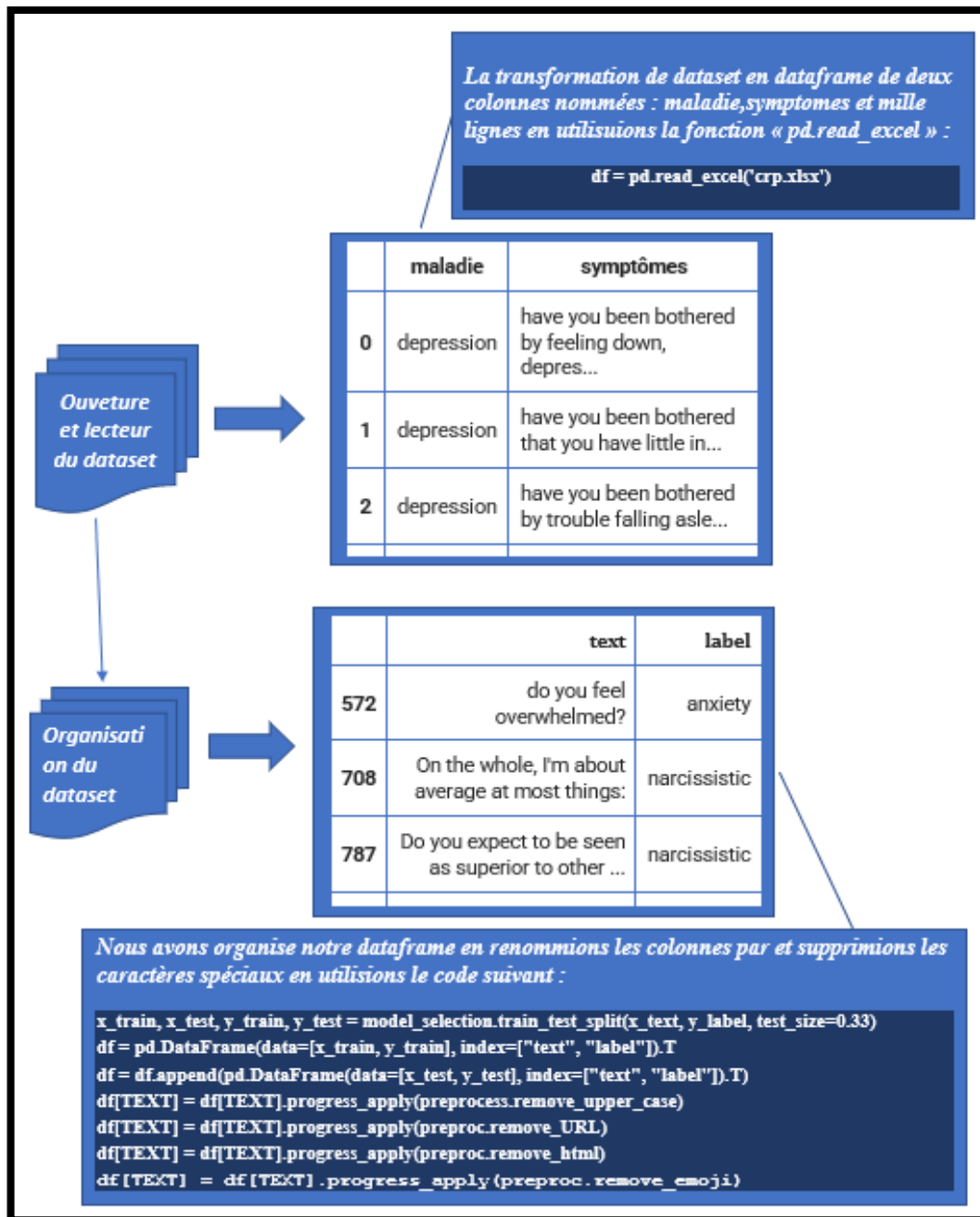


Figure 3-14 : organigramme de chargement des données.

Comme c'est indiqué dans le chapitre II, sous-section « Construction du dataset », nous avons trois versions de dataset. Nous avons divisé la version finale en deux parties en tant qu'ensemble de données d'entraînement et ensemble de données de test.

Le tableau qui suit (Tableau 3-1) résume le nombre de phrases, le nombre de classes et le nombre de phrases retenues par classe pour chacune des versions du dataset.

	La première Version	La deuxième version	Version finale (Ensemble des données d'entraînement)	Version finale (Ensemble des données de test)
Nombre de phrases	25	1000	2000	100
Nombre de classes	5	100	1000	10
Nombre de phrases par classe	5	10	10	10

Tableau 3-1 : le dataset en chiffres.

III.4 Prétraitement

Dans cette section, nous allons présenter des parties de codes de l'étape de prétraitement.

III.4.1 Prétraitement surfacique

III.4.1.1 Suppression des caractères spéciaux

Dans cette étape, nous avons procédé à (voir figure récapitulative 3-15):

- Une suppression de la capitalisation
- Une suppression des URL
- Une suppression des tag html
- Et enfin une suppression des différents emojis.

```
df[TEXT] = df[TEXT].progress_apply(preprocess.remove_upper_case)
df[TEXT] = df[TEXT].progress_apply(preproc.remove_URL)
df[TEXT] = df[TEXT].progress_apply(preproc.remove_html)
df[TEXT] = df[TEXT].progress_apply(preproc.remove_emoji)
```

Figure 3-15 : Fonction de suppression des caractères spéciaux.

III.4.1.2 Suppression des mots vides

Dans cette étape, nous avons procédé à une suppression des mots vide comme montré dans la figure 3-16.

```
df.loc[:,TEXT+"_sw"] = df.loc[:,TEXT].progress_apply(lambda x : preproc.remove_stop_words(x, stop_word))
100%|██████████| 1001/1001 [00:00<00:00, 13321.42it/s]

if df[TEXT+"_sw"].isnull().sum()>0:
    print("Empty text")
    df[TEXT+"_sw"][df[TEXT+"_sw"].isnull()] = "empty_text"
```

Figure 3-16 : Fonction de suppression des mots vides.

III.4.2 Prétraitement morphologique

Dans cette sous-section, nous allons présenter la fonction de lemmatisation, POS Tagger et la fonction de NER.

```

def pos_ner_lemma(df, column, lang="en"):
    """
    Function to do part-of-speech with stanza library. The pipeline is configured to
    tokenize, mwt, pos, lemma, ner parameters, it will does part-of-speech, lemmatized words
    and name entity recognition.
    @param df: (pandas.dataframe) dataframe containing the data
    @param column: (str) name of the column containing text
    @param lang: (str) language of the text
    """
    print("Building an French pipeline...")
    nlp = stanza.Pipeline(lang, processors="tokenize,mwt,pos,lemma,ner") # creation of the pipeline
    text_lem = []
    text_pos = []
    text_ner = []
    for i in tqdm(df[column]): # loop on each text
        doc = nlp(i) # fit the text with the pipeline
        try:
            text_lem.append(' '.join([word.lemma for sent in doc.sentences for word in sent.words])) # lemmatize
        except: # if
            text_lem.append('')
        try:
            text_pos.append(' '.join([f'word: {word.text} upos: {word.upos} xpos: {word.xpos} feats: \
{word.feats if word.feats else "_"}' for sent in doc.sentences for word in sent.words])) # pos
        except:
            text_pos.append('')
        try:
            text_ner.append(' '.join([f'token: {token.text} ner: {token.ner}' for sent in doc.sentences for token in sent.tokens])) # NER
        except:
            text_ner.append('')

    df.loc[:,column+"_lemma"] = text_lem
    df.loc[:,column+"_pos"] = text_pos
    df.loc[:,column+"_ner"] = text_ner

    return df

```

Figure 3-17 : Fonction de lemmatisation, POS et NER.

Et voici un petit extrait du résultat de l'étape de lemmatisation, POS et NER

	text	label	text	lemmatize	POS
832	In the course of recent days, how concerned ha...	panic	In the course of recent day , how concern have...	[[In, IN), (the, DT), (course, NN), (of, IN), ...	
594	Even when things are going well you thinks it ...	anxiety	Even when thing be go well you think it wo n't...	[(Even, RB), (when, WRB), (things, NNS), (are,...	
408	Do you feel afraid that you might get contamin...	ocd	Do you feel afraid that you might get contamin...	[(Do, VBP), (you, PRP), (feel, VB), (afraid, I...	
184	I've had moments when I can't stop crying and ...	bipolar	I 've have moment when I can 't stop cry and...	[(I, PRP), (' , VBP), (ve, NNS), (had, VBD), (m...	
606	do you have strong feelings or beliefs about ...	schizophrenia	do you have strong feeling or belief about be ...	[(do, VBP), (you, PRP), (have, VB), (strong, J...	
...
216	I have tried to remove the situation from my m...	ptsd	I have try to remove the situation from my memory	[(I, PRP), (have, VBP), (tried, VBN), (to, TO)...	
473	o you get caught up making sure things are in ...	ocd	o you get catch up make sure thing be in their...	[(o, NN), (you, PRP), (get, VBP), (caught, VBN...	

Figure3-18 : implémentation de la fonction de lemmatisation, POS et NER.

- **Lemmatisation :**

Une analyse morphologique est appliquée aux mots, visant à supprimer uniquement les terminaisons flexionnelles et à renvoyer la forme de base ou dictionnaire d'un mot, connue sous le nom de lemme.

Pour plus de clarté, un exemple de lemmatisation de phrase est donné ci-dessous :

```
def pos_ner_lemma(df, column, lang="en"):
    nlp = stanza.Pipeline(lang, processors="tokenize,mwt,pos,lemma,ner") # creation of the pipeline
    text_lem = []
    for i in tqdm(s): # loop on each text
        doc = nlp(i) # fit the text with the pipeline
        try:
            text_lem.append(' '.join([word.lemma for sent in doc.sentences for word in sent.words])) # lemmatize
        except: # if
            text_lem.append('')
    return text_lem

['Being', 'the', 'focus', 'of', 'others', 'attention', 'makes', 'me', 'feel', 'good']
```

Figure 3-19 : Fonction de lemmatisation de phrase.

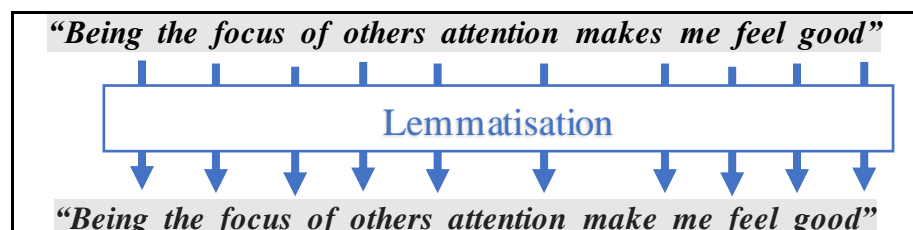


Figure 3-20 : Exemple de lemmatisation de phrase.

- **POS tagger :**

Après la lemmatisation, des mots sont restés les mêmes. C'est parce que ces mots sont traités comme un nom dans la phrase donnée plutôt que comme un verbe.

Pour contourner cela, nous utilisons des balises POS (Part of Speech) pour la catégorisation les classes de mots, telles que les noms, les verbes, les adjectifs, etc.

Pour plus de clarté, un exemple de POS de phrase est donné ci-dessous :

```
def pos_ner_lemma(df, column, lang="en"):
    text_pos = []
    s="Being the focus of others attention makes me feel good"
    for i in tqdm(s): # loop on each text
        doc = nlp(i) # fit the text with the pipeline
        try:
            text_pos.append(' '.join([word.pos for sent in doc.sentences for word in sent.words]))
        except: # if
            text_pos.append('')
    return (text_pos)
[('Being', 'VBG'), ('the', 'DT'), ('focus', 'NN'), ('of', 'IN'), ('others', 'NNS'), ('attention', 'NN'), ('makes', 'VBZ'), ('me', 'PRP'), ('feel', 'VB'), ('good', 'JJ')]
```

Figure3-21 : Fonction de POS de phrase.

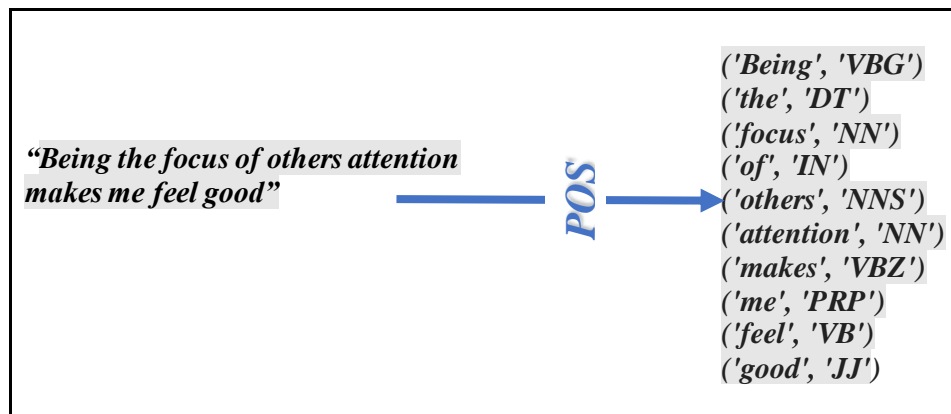


Figure3-22 : Exemple de POS de phrase.

- **NER**

La reconnaissance d'entité nommée (NER) est appliquée aux mots qui consiste à repérer des entités nommées (personnes, lieux, organisations, etc.) et à les classer dans un ensemble prédéfini de catégories.

Pour plus de clarté, un exemple de NER de phrase est donné ci-dessous :

```
def pos_ner_lemma(s, lang="en"):
    nlp = stanza.Pipeline(lang, processors="tokenize,mwt,pos,lemma,ner") # creation of the pipeline
    text_ner = []
    s="Apple is looking at buying U.K. startup for $1 billion"
    for i in tqdm(s): # loop on each text
        doc = nlp(i) # fit the text with the pipeline
        try:
            text_ner.append(' '.join([f'token: {token.text} ner: {token.ner}' for sent in doc.sentences for token in sent.tokens]))
        except:
            text_ner.append('')
    return text_ner
```

Apple 0 5 ORG
 U.K. 27 31 GPE
 \$1 billion 44 54 MONEY

Figure 3-23 : Fonction de NER de phrase.



Figure 3-24 : Exemple de NER de phrase.

Dans la sortie, la première colonne spécifie l'entité, les deux colonnes suivantes les caractères de début et de fin dans la phrase/le document et la dernière colonne spécifie la catégorie.

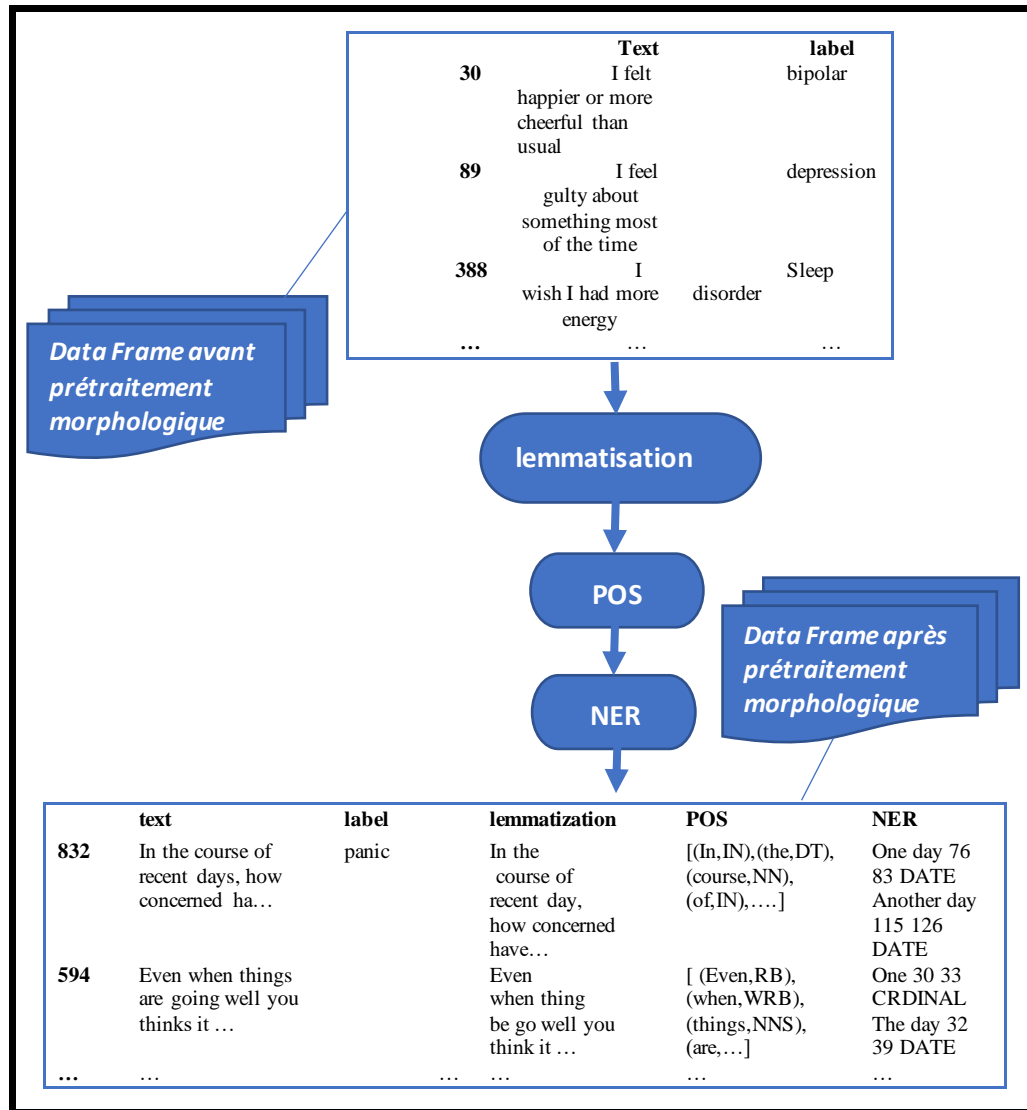


Figure 3-25 : Organigramme du Prétraitement Morphologique.

III.5 processus de vectorisation

III.5.1 Count vectoriser

Le code ci-dessous montre l'utilisation du Count vectoriser qui permet de calculer la fréquence des termes.

Cette méthode est importée depuis la bibliothèque **Scikit-learn**

```
# create a One-Hot encoding object
count_vect = CountVectorizer(analyzer='word', token_pattern=r'\w{1,}')
count_vect.fit(df[TEXT]+"_sw")

# transform the training and validation data
xtrain_count = count_vect.transform(train_x_sw)
xvalid_count = count_vect.transform(valid_x_sw)
```

Figure 3-26 : code de count vectoriser

La figure suivante montre la fréquence de quelques termes :

```
{'fearful': 903, 'that': 2317, 'something': 2137, 'terrible': 2310, 'will': 2572,
```

Figure 3-27 : fréquence de quelques termes

III.5.2 N-grammes

Pour bien explorer notre corpus nous avons utilisé la bibliothèque preproc qui nous donne les différents bi-grammes en utilisant un texte avec et/ou sans les mots vides

Dans notre étude nous avons utilisé jusqu'à 5 combinaisons, la figure ci-dessus montre le code de bi-gramme sans suppression des mots vides

```
number = 20
common_words = preproc.get_top_n_bigram(df[TEXT], number)
df3 = pd.DataFrame(common_words, columns = ['words' , 'count'])
plt.figure(figsize=(14,8))
plt.barh(df3["words"][::-1], df3["count"][::-1])
plt.ylabel("Words")
plt.xlabel("Frequency")
plt.title("Top %d words in text before removing stop words" %(number))
plt.grid(True)
```

Figure 3-28 : le code de bi-gramme avant la suppression des mots vides

Dans le cas où nous avons utilisé notre corpus sans supprimer les mots vides, nous pouvons remarquer que les bi-grammes les plus répétés (courants) sont des mots vides , le résultat est montré dans la figure suivante

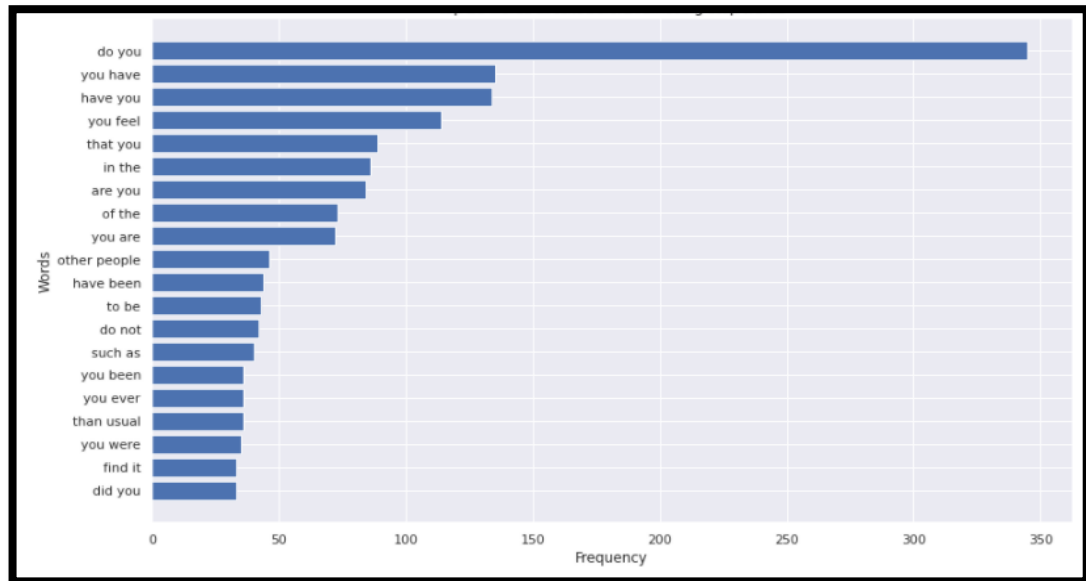


Figure3-29 : La distribution des premiers bi-gramme avant la suppression des mots vides

Dans l'autre cas (avec la suppression des mots vides), nous remarquons que les bi-grammes les plus fréquents sont des mots qui ont une relation avec notre étude donc cette opération est très importante pour la compréhension du corpus.

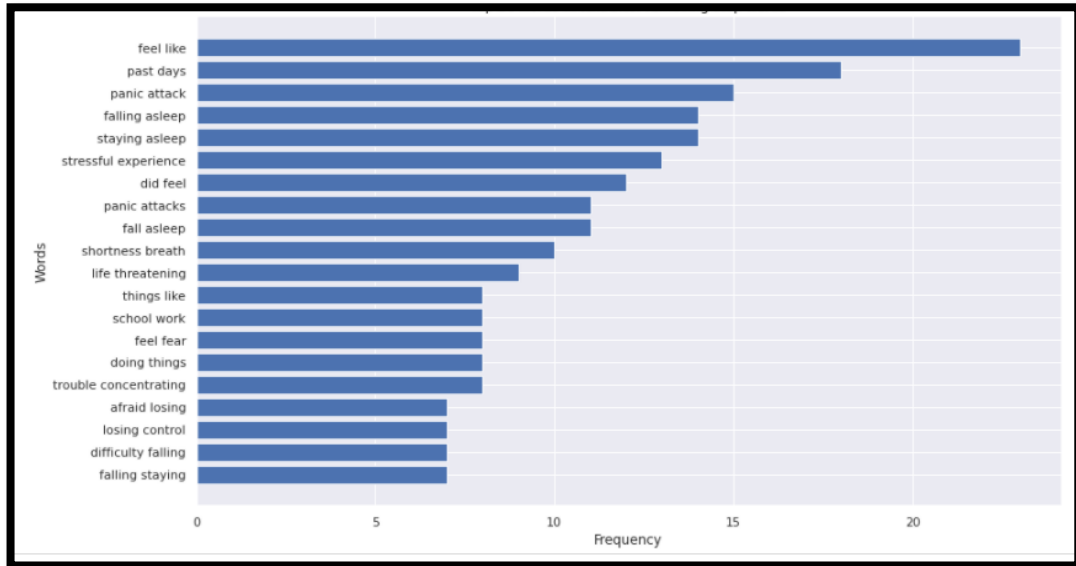


Figure3-30 : La distribution des premiers bi-gramme après la suppression des mots vide

III.5.3 Terme Fréquence-Inverse Document Fréquence (TF-IDF)

Tant que la fonctionnalité doit être constituée de n-grammes de mots et de caractères nous avons utilisé

- un analyseur {'word', 'char', 'char_wb'} ,
- token_pattern qui supprime les caractères alphanumériques et la ponctuation(on utilise cet paramètre seulement dans le cas ou analyzer=word) ,
- ngram_range pour choisir la limite inférieure et supérieure des valeurs n pour différents n-grammes
- max_features pour construire un vocabulaire qui ne prend en compte que les principaux max_features classés par fréquence de termes à travers le corpus.

La figure suivante représente le code expliqué ci-dessus :

```
# word level tf-idf
tfidf_vect_ngram = TfidfVectorizer(analyzer='word', token_pattern=r'\w{1,}', ngram_range=(1,3), max_features=5000)
tfidf_vect_ngram.fit(df[TEXT])
xtrain_tfidf_ngram = tfidf_vect_ngram.transform(train_x_sw)
xvalid_tfidf_ngram = tfidf_vect_ngram.transform(valid_x_sw)
print("word level tf-idf done")
# characters level tf-idf
tfidf_vect_ngram_chars = TfidfVectorizer(analyzer='char', ngram_range=(1,3), max_features=5000)
tfidf_vect_ngram_chars.fit(df[TEXT])
xtrain_tfidf_ngram_chars = tfidf_vect_ngram_chars.transform(train_x_sw)
xvalid_tfidf_ngram_chars = tfidf_vect_ngram_chars.transform(valid_x_sw)
print("characters level tf-idf done")
# characters_wb level tf-idf
tfidf_vect_ngram_chars_wb = TfidfVectorizer(analyzer='char_wb', ngram_range=(1,3), max_features=5000)
tfidf_vect_ngram_chars_wb.fit(df[TEXT])
xtrain_tfidf_ngram_chars_wb = tfidf_vect_ngram_chars_wb.transform(train_x_sw)
xvalid_tfidf_ngram_chars_wb = tfidf_vect_ngram_chars_wb.transform(valid_x_sw)
print("characters_wb level tf-idf done")
```

Figure3-31 : le code de tf-idf

Pour bien comprendre le fonctionnement de l'opération TF-IDF, nous représentons ci-dessous le code en schéma (figure 3-32).

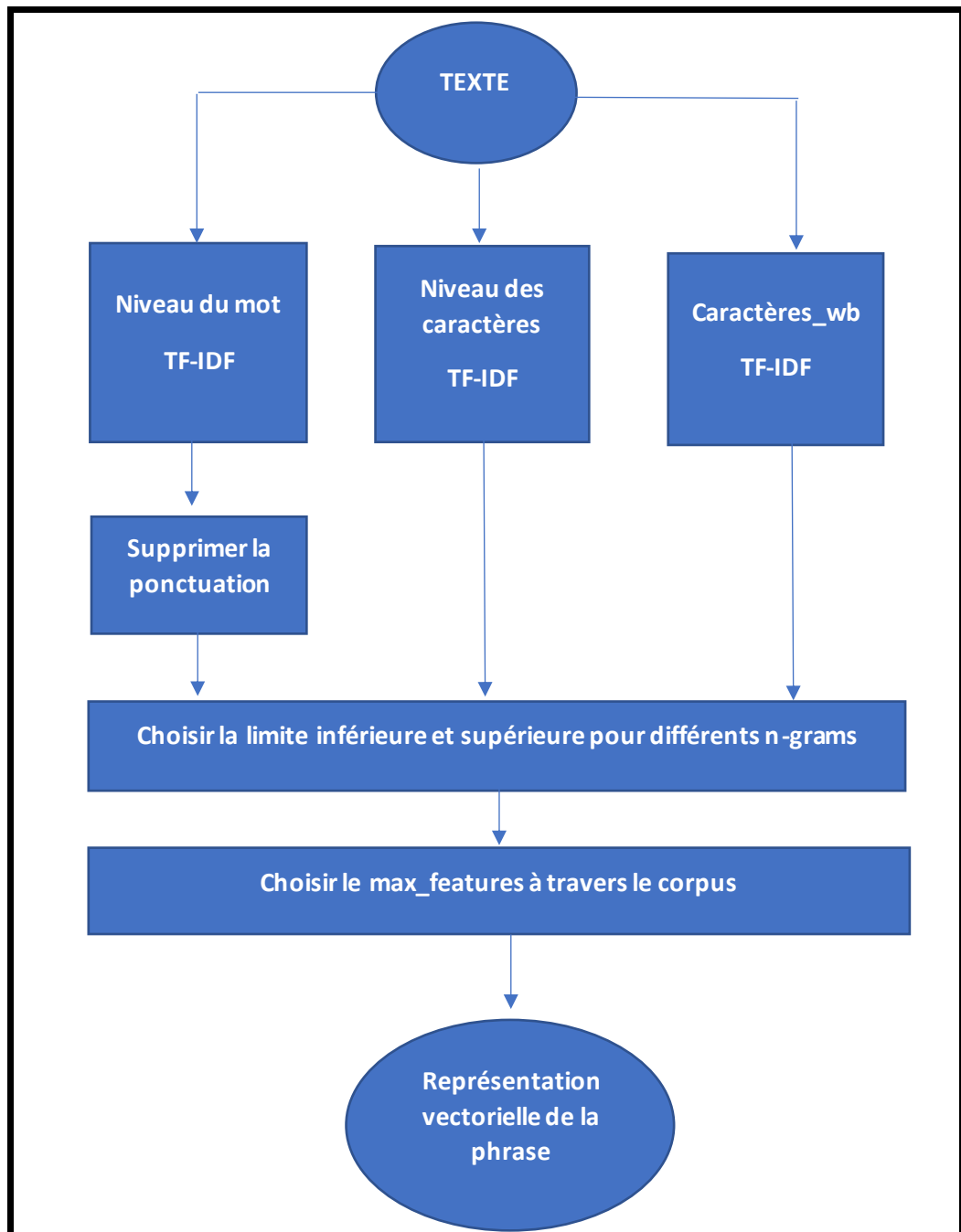


Figure 3-32 : schéma du fonctionnement de l'opération TF-IDF

III.5.4 word embedding

De la bibliothèque Keras, nous importons tokenizer pour effectuer la tokenisation comme suit :

```
# create a tokenizer
token = Tokenizer()
token.fit_on_texts(df[TEXT])
word_index = token.word_index
```

Figure 3-33 : Code de la tokenisation

Ensuite, dans le code, nous préparons un Embedding Matrix

```
# create token-embedding mapping
embedding_matrix = np.zeros((len(word_index) + 1, 300))
for word, i in tqdm(word_index.items()):
    embedding_vector = pretrained.get_word_vector(word) #embeddings_index.get(word)
    if embedding_vector is not None:
        embedding_matrix[i] = embedding_vector
```

Figure3-34 : Code de Embedding matrix

Le schéma suivant illustre la phase de la création du Embedding matrix

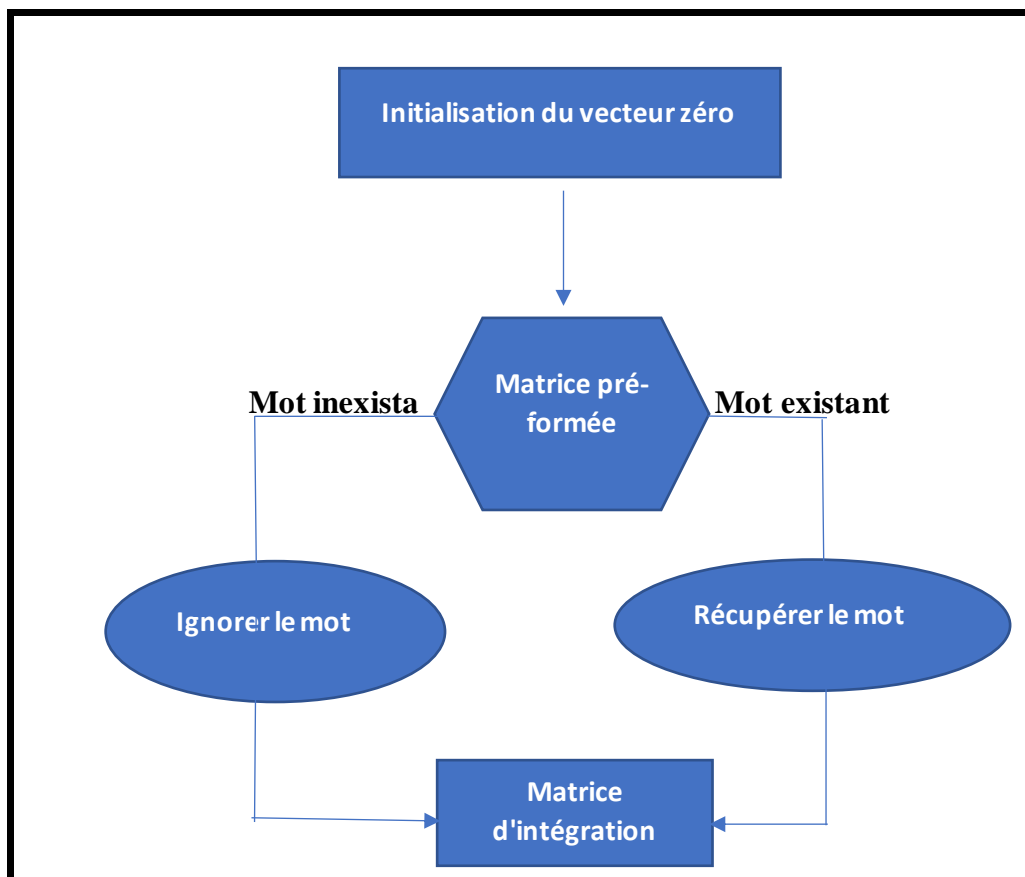


Figure3-35 : Schéma de la création de la matrice d'intégration

III.6 Classification

Nous avons appliqué ces algorithmes pour les trois corpus préalablement décrits.

III.6.1 Apprentissage Automatique

La bibliothèque Scikit-learn de Python destinée à l'apprentissage automatique fourni les deux modules de classification de Machine Learning.

Pour chaque classifieur nous avons utilisé quatre processus de vectorisation : One-Hot encoding, TF-IDF word n-grams, for TF-IDF char n-grams, TF-IDF char_wb n-grams en parallèle.

III.6.1.1 Algorithme de Naive Bayes

D'abord nous avons défini une fonction `metrics_ML` pour calculer le modèle de classification. Elle nous donne les différents résultats dont nous avons besoin dans notre étude : temps d'exécution, score, matrice de confusion, précision, recall et kappa puis nous allons l'utiliser en précisant le classifieur MNB (Naive Bayes)et la matrice après la vectorisation comme suit :

```
if multinomial_naive_bayes:
    print("\nMultinomial Naive Bayes for One-Hot encoding\n")
    metrics_ML(naive_bayes.MultinomialNB(), xtrain_count, train_y_sw, xvalid_count, valid_y_sw, gb=True)

if multinomial_naive_bayes:
    print("\nMultinomial Naive Bayes for TF-IDF word ngrams\n")
    metrics_ML(naive_bayes.MultinomialNB(), xtrain_tfidf_ngram, train_y, xvalid_tfidf_ngram, valid_y, gb=True)

if multinomial_naive_bayes:
    print("\nMultinomial Naive Bayes for TF-IDF char n-grams\n")
    metrics_ML(naive_bayes.MultinomialNB(), xtrain_tfidf_ngram_chars, train_y, xvalid_tfidf_ngram_chars, valid_y, gb=True)

if multinomial_naive_bayes:
    print("\nMultinomial Naive Bayes for TF-IDF char_wb n-grams\n")
    metrics_ML(naive_bayes.MultinomialNB(), xtrain_tfidf_ngram_chars_wb, train_y, xvalid_tfidf_ngram_chars_wb, valid_y, gb=True)
```

Figure3-36 : code de L'algorithme MNB

Dans la section qui suit, nous n'allons montrer que le meilleur résultat des quatre processus de vectorisation pour chaque classifieur en utilisant les trois corpus.

III.6.1.1.1 Matrice de confusion

- La taille de la matrice dépend du nombre de classes (5 maladies dans le 1^{er} corpus et 10 dans le 2^{ème} et 40 le 3^{ème})
- Pour connaître le nombre de phrases de test pour chaque maladie en additionne la première ligne de matrice (une phrase pour le 1^{er} corpus 20 pour le 2^{ème} et 40 pour le 3^{ème})
- Pour chaque matrice, nous nous basons sur la diagonale principale qui correspond aux fois où la machine connaît correctement la maladie et le reste du matrice correspond aux fois où la machine se trompe dans la détection de la maladie.
- Le meilleur résultat est obtenu lorsque les valeurs de la diagonale principale sont grandes (proche du nombre des phrases de test).

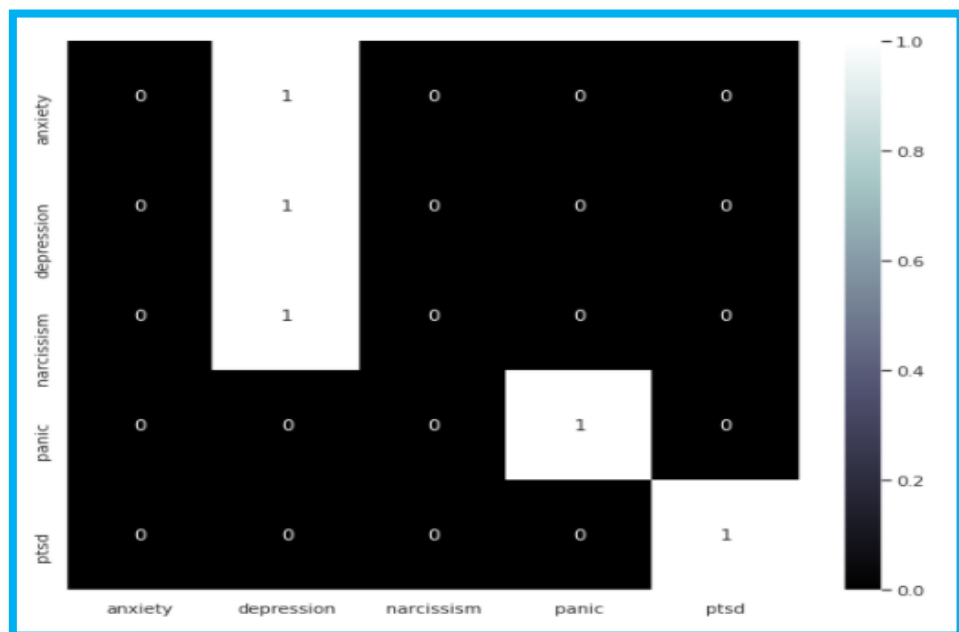


Figure3-37 : Matrice de confusion Crp1- one hot encoding-

Le meilleur résultat de processus de vectorisation one hot encoding pour le classifieur MNB en utilisant le premier corpus est obtenu lorsque les valeurs de la diagonale principale sont égaux à 1 (proche du nombre des phrases de test qui est 1), la machine connaît correctement les maladies :panic, ptsd, depression et elle se trompe dans la détection de les maladies : anxiety et narcissism.

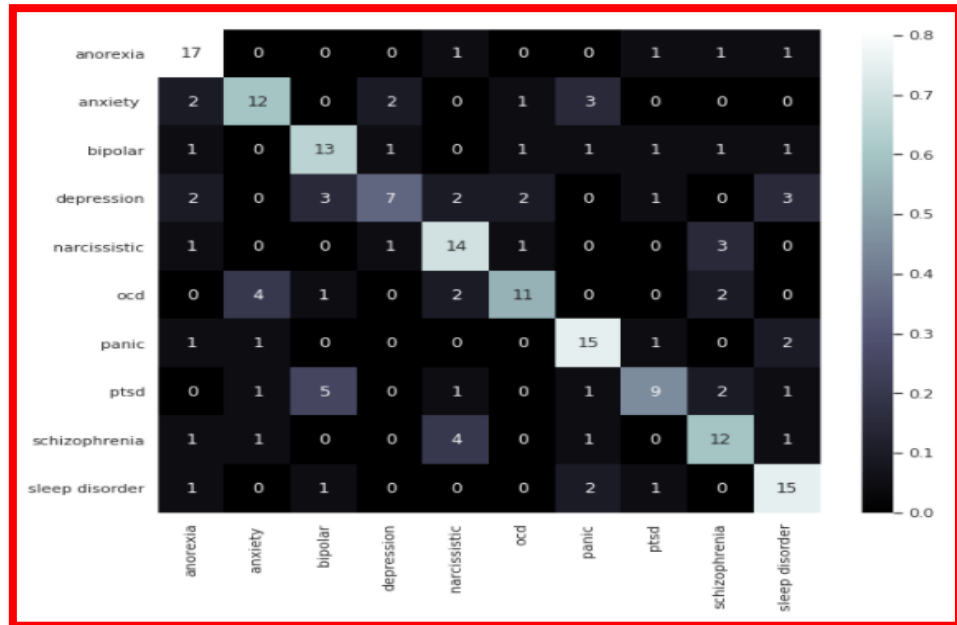


Figure3-38 : Matrice de confusion Crp2 -TF-IDF word n-gram-

Le meilleur résultat de processus de vectorisation TF-IDF word n-gram pour le classifieur MNB en utilisant le deuxième corpus est obtenu lorsque les valeurs de la diagonale principale sont grandes (proche du nombre des phrases de test qui est 20), la machine connaît correctement les maladies : anorexia,anxiety,bipolar,narcissism,ocd,panic, sleep disorder,schizophrenia et elle se trompe dans la détection de les maladies : ptsd et depression.

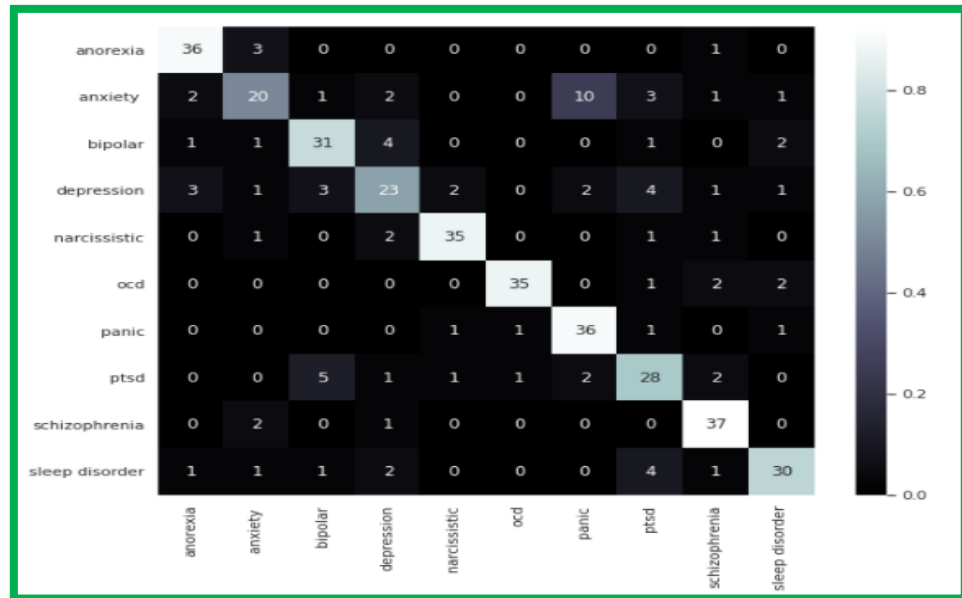


Figure3-39: Matrice de confusion Crp3 -one hot encoding-

Le meilleur résultat de processus de vectorisation one hot encoding pour le classifieur MNB en utilisant le troisième corpus est obtenu lorsque les valeurs de la diagonale principale sont grandes (proche du nombre des phrases de test qui est 40), la machine connaît correctement les maladies : anorexia ,bipolar , narcissism , ocd ,panic, sleep disorder ,schizophrenia et elle se trompe dans la détection de les maladies : anxiety et depression.

III.6.1.1.2 Rappel et précision

Notre travail vise à mieux cerner les différences de classification selon les classes grâce à une lecture graphique de la courbe rappel-précision. Nous considérons pour cela disposer de la courbe rappel-précision fournie par chaque classifieur, et nous déduisons des informations de cette courbe pour des constructions graphiques.

Cette courbe se concentre principalement sur la performance de la classe positive qui est cruciale lorsqu'il s'agit de classes déséquilibrées. Dans l'espace PR, l'objectif est d'être dans le coin supérieur droit - le coin supérieur droit (1, 1) signifie que nous avons classé tous les positifs comme positifs (rappel = 1) et que tout ce que nous classons comme positif est vrai positif (précision = 1) - ce dernier se traduit par zéro faux positif. Dans ce

cas, la « courbe » pour un classifieur sans compétence² est une ligne horizontale Parce que les courbes rappel-précision sont tellement en zigzag dans ces graphiques, il est difficile de comparer deux modèles différents et de juger lequel est le meilleur.

Les figures ci-dessus illustrent une mise en correspondance des valeurs de rappel et de précision obtenues pour nos différentes classes.

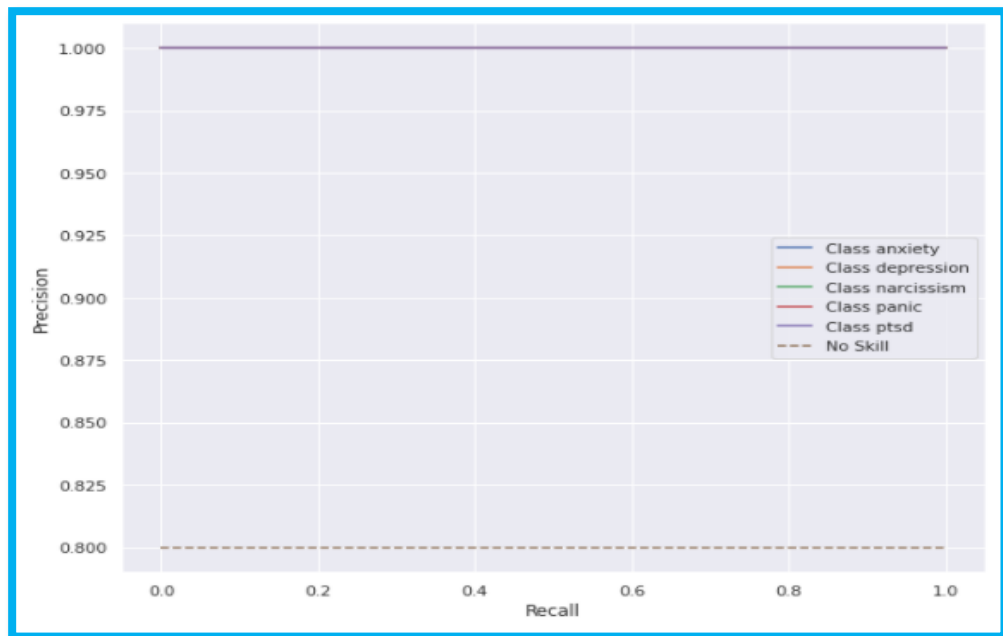


Figure3-40 : Courbe rappel et précision Crp1 -one hot encoding-

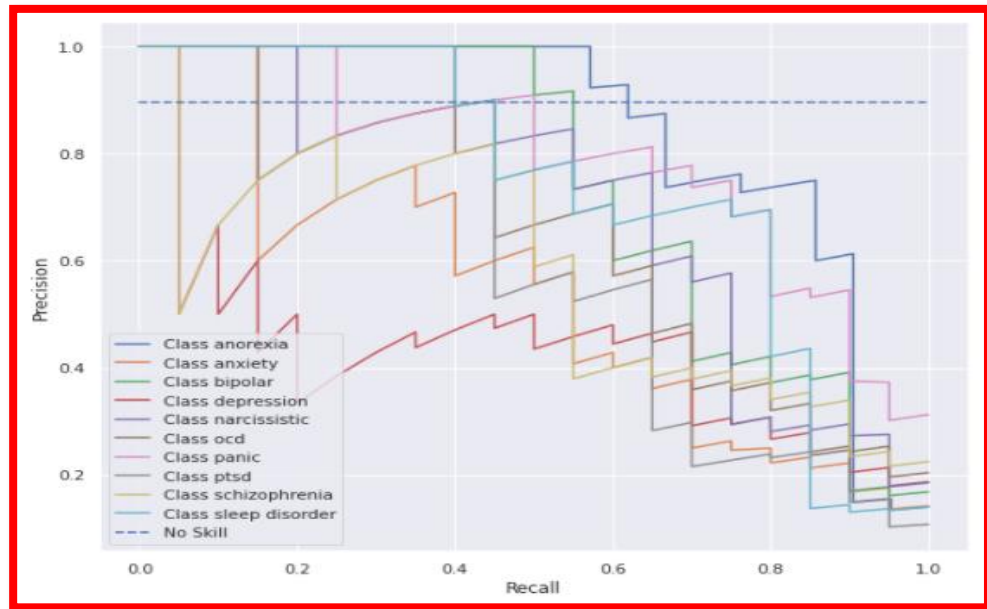


Figure3-41 : Courbe rappel et précision Crp2 -TF-IDF word n-gram-

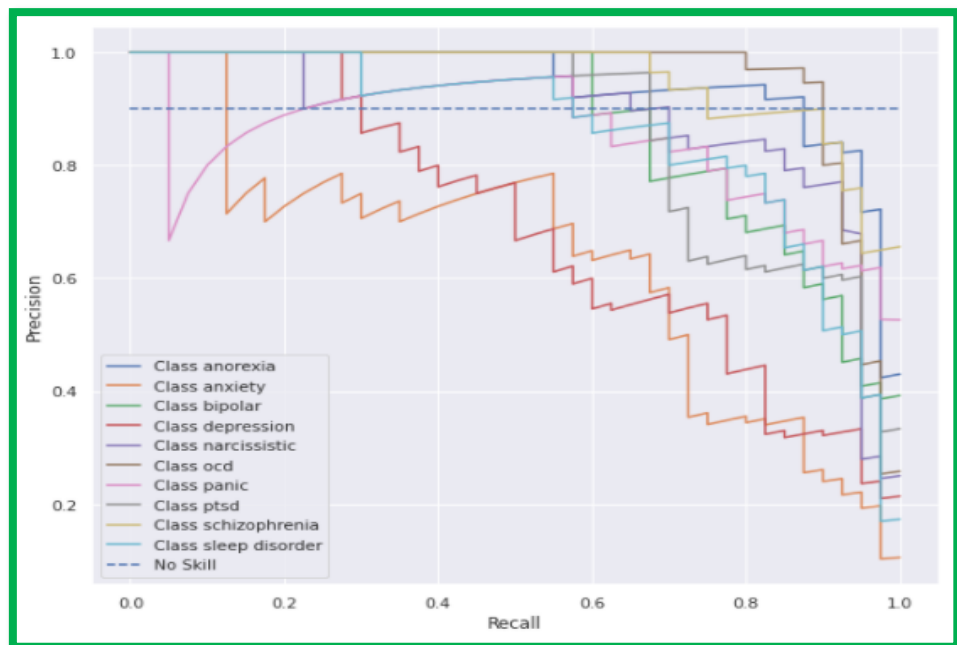


Figure3-42 : Courbe rappel et précision Crp3 one hot encoding

III.6.1.2 Algorithme du K plus proche Voisin

On utilise la même fonction metrics_Ml en précisant cette fois le classifieur KNN qui a comme paramètre

- n_neighbors : c'est le nombre de voisins à utiliser.

•Weights : c'est une fonction de poids utilisée dans la prédiction et la valeur distance. Ça correspond aux poids des points par l'inverse de leur distance. Dans ce cas, les voisins les plus proches d'un point de requête auront une plus grande influence que les voisins les plus éloignés.

•n_jobs : c'est le nombre de travaux en parallèles à exécuter pour la recherche de voisins. Tel que -1 signifie l'utilisation de tous les processeurs.

```
if k_nn_model:
    print("\nk-NN for One-Hot encoding\n")
    metrics_ML(neighbors.KNeighborsClassifier(n_neighbors=20, weights='distance', n_jobs=-1),
               xtrain_count, train_y_sw, , valid_y_sw, gb=True)

if k_nn_model:
    print("\nk-NN for TF_IDF word n-grams\n")
    metrics_ML(neighbors.KNeighborsClassifier(n_neighbors=20, weights='distance', n_jobs=-1),
               xtrain_tfidf_ngram, train_y, xvalid_tfidf_ngram, valid_y, gb=True)

if k_nn_model:
    print("\nk-NN for TF_IDF char n-grams\n")
    metrics_ML(neighbors.KNeighborsClassifier(n_neighbors=20, weights='distance', n_jobs=-1),
               xtrain_tfidf_ngram_chars, train_y, xvalid_tfidf_ngram_chars, valid_y, gb=True)

if k_nn_model:
    print("\nk-NN for TF_IDF char_wb n-grams\n")
    metrics_ML(neighbors.KNeighborsClassifier(n_neighbors=20, weights='distance', n_jobs=-1),
               xtrain_tfidf_ngram_chars_wb, train_y, xvalid_tfidf_ngram_chars_wb, valid_y, gb=True)
```

Figure3-43 : Code de l'algorithme KNN

Avec la même méthode utilisée dans MNB on va choisir les résultats suivants

III.6.1.2.1 Matrice de confusion

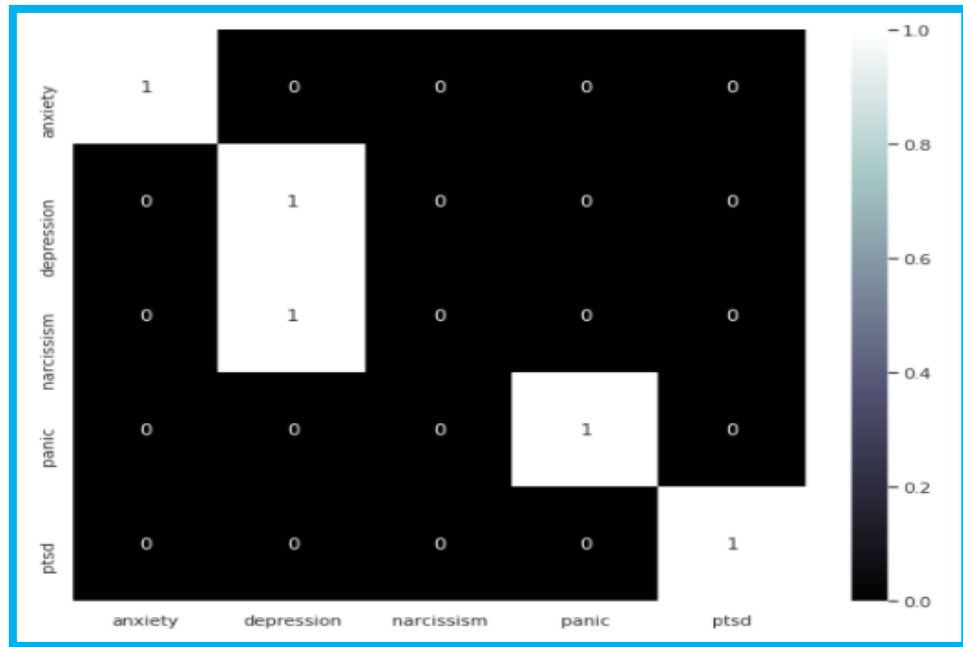


Figure3-44 : Matrice de confusion Crp1 -TF-IDF word n-gram-

Le meilleur résultat de processus de vectorisation TF-IDF word n-gram pour le classifieur KNN en utilisant le premier corpus est obtenu lorsque les valeurs de la diagonale principale sont égaux à 1 (proche du nombre des phrases de test qui est 1), la machine connaît correctement les maladies : anxiety , depression, ptsd , panic et elle se trompe dans la détection de la maladie : narcissism.

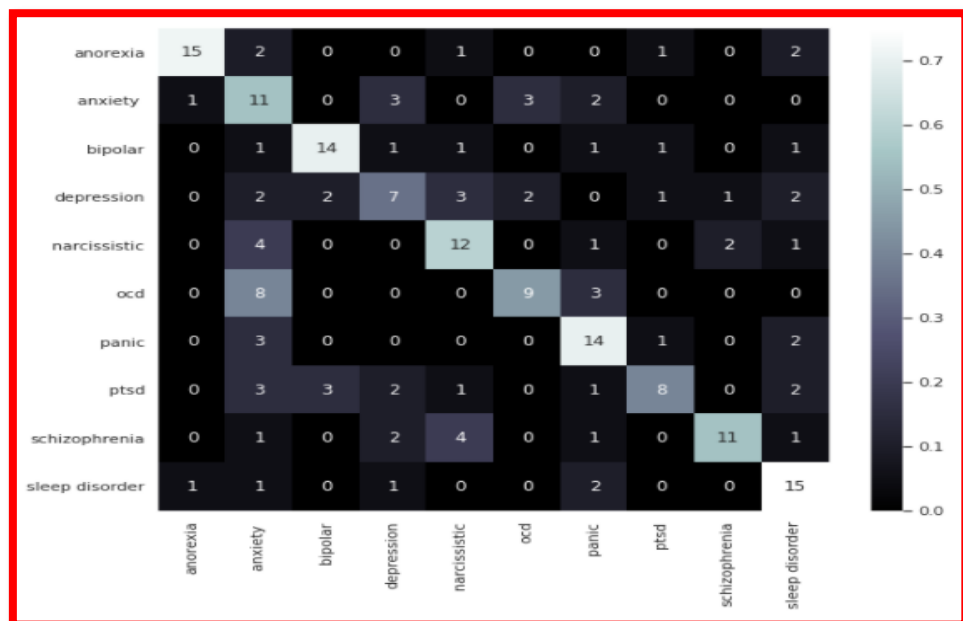


Figure3-45 : Matrice de confusion Crp2 -TF-IDF word n-gram-

Le meilleur résultat de processus de vectorisation TF-IDF word n-gram pour le classifieur KNN en utilisant le deuxième corpus est obtenu lorsque les valeurs de la diagonale principale sont grandes (proche du nombre des phrases de test qui est 20), la machine connaît correctement les maladies : anxiety ,anorexia , bipolar, narcissism,panic,schizophrenia,sleep disorder et elle se trompe dans la détection de les maladies :ptsd,ocd,depression

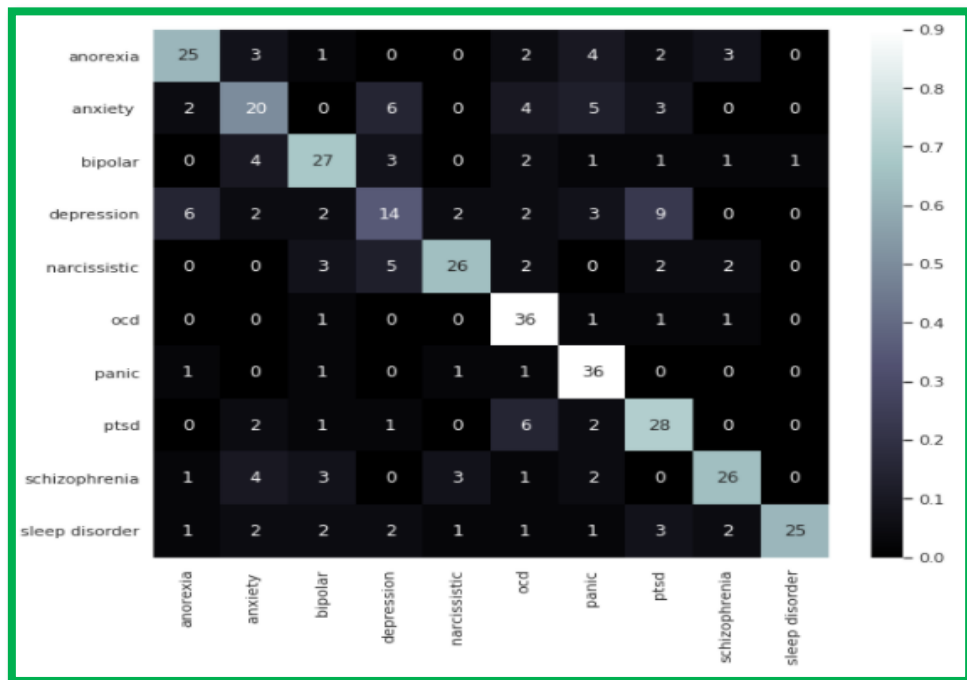


Figure3-46 : Matrice de confusion Crp3 -TF-IDF char n -gram-

Le meilleur résultat de processus de vectorisation TF-IDF word n-gram pour le classifieur KNN en utilisant le troisième corpus est obtenu lorsque les valeurs de la diagonale principale sont grandes (proche du nombre des phrases de test qui est 40), la machine connaît correctement les maladies : ocd, anxiety ,anorexia , bipolar, ptsd , narcissism, panic, schizophrenia, sleep disorder et elle se trompe dans la détection de la maladie :depression.

III.6.1.2.2 Rappel e précision

Les figures ci-dessus illustrent une mise en correspondance des valeurs de rappel et de précision obtenues pour nos différentes classes.

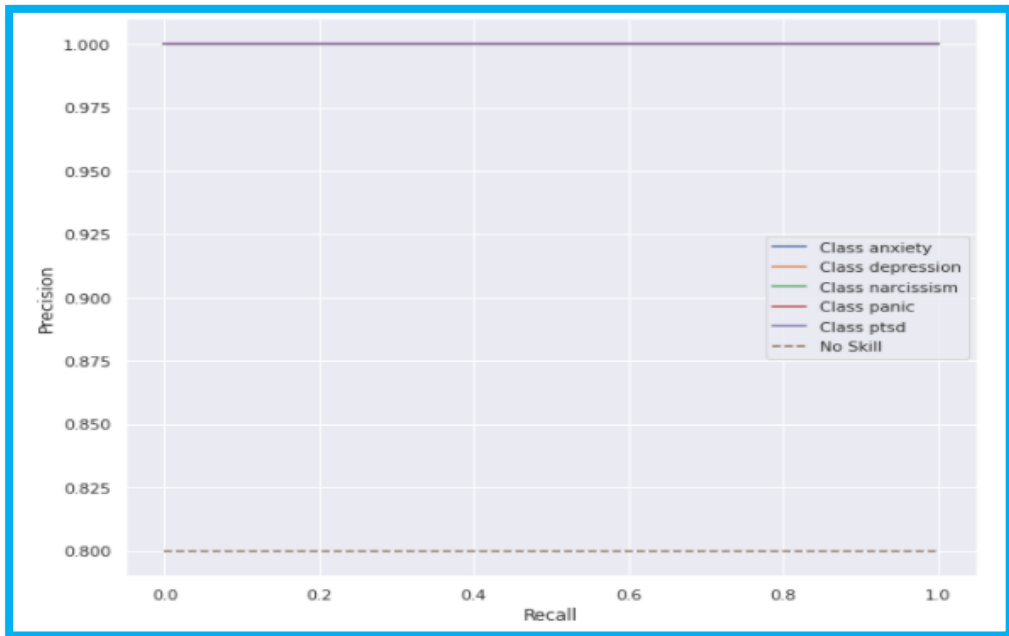


Figure3-47 : Courbe rappel et précision Crp 1 -TF-IDF word n-gram-

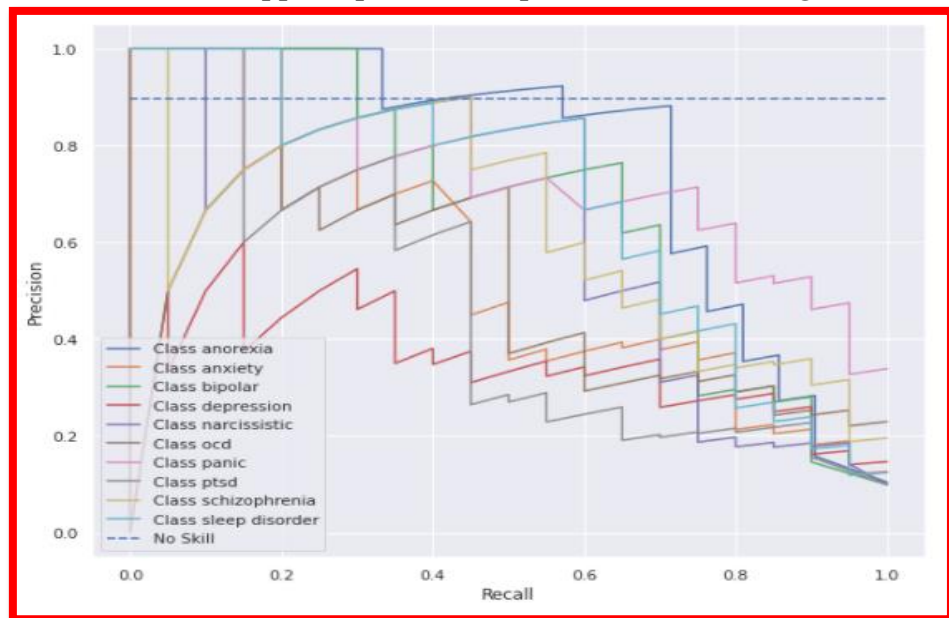


Figure3-48 : Courbe rappel et précision Crp2 -TF-IDF word n-gram-

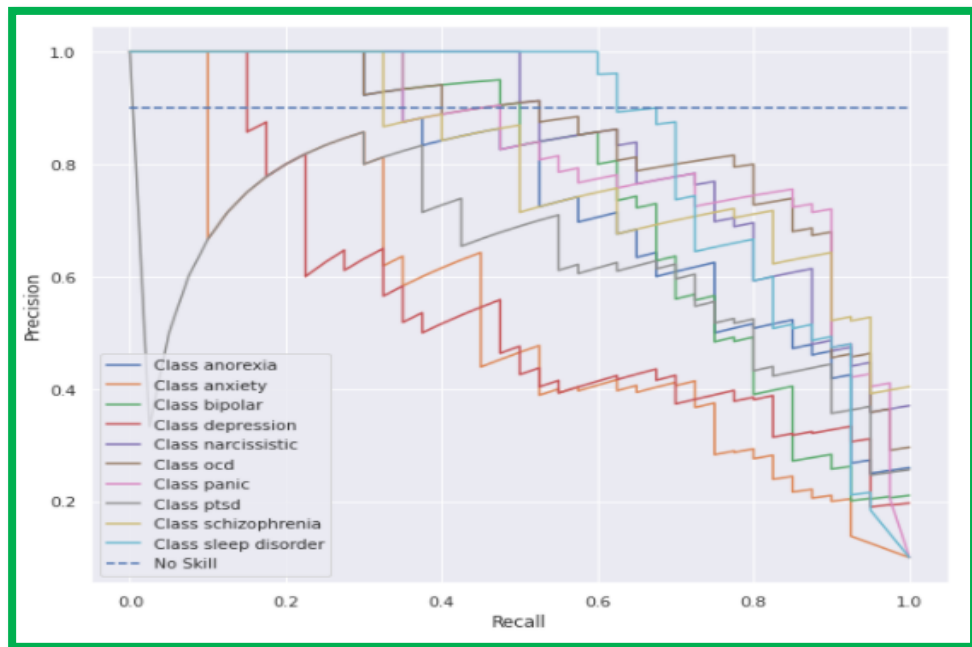


Figure3-49 : Courbe rappel et précision Crp3 -TF-IDF char n-gram-

III.6.1.3 Performance des différents algorithmes

Dans les tableaux suivants, nous aurons :

Acc : accuracy

F1 : f1-score

			Corpus 1		Corpus 2		Corpus 3		Corpus Arabe	
			Type de vectorisation		Acc	F1	Acc	F1	Acc	F1
MNB	Pos Lem ner	One-Hot encoding	80	73.33	68.16	67.18	67.75	67.71	59	58.28
		TF-IDF word n-grams	80	73.33	71.14	69.93	70.5	70.18	63.5	62.93
		TF-IDF char n-grams	40	23.33	56.72	56.75	62.25	62.42	61	60.23
		TF-IDF char_wb n-grams	40	23.33	59.2	59.43	62.75	62.77	62	60.65
	Aucun	One-Hot encoding	60	50	66.67	65.89	73.25	72.88	61.5	60.35
		TF-IDF word n-grams	60	50	66.67	65.88	76	75.59	66	65.07
		TF-IDF char n-grams	60	53.33	58.71	57.74	63.75	63.94	64.5	63.37
		TF-IDF char_wb n-grams	60	53.33	61.19	60.65	63.5	63.64	64.5	63.02
KNN	Pos Lem ner	One-Hot encoding	40	28	18.41	14.96	28.25	26.56	32	27.53
		TF-IDF word n-grams	80	73.33	54.23	54.66	52.5	53.58	51	51.42
		TF-IDF char n-grams	40	23.33	47.76	48.07	62.5	62.27	55.5	54.88
		TF-IDF char_wb n-grams	40	28	47.26	47.66	63	62.97	53	52.41
	Aucun	One-Hot encoding	50	60	20.4	18.77	32	31.63	27	22.85
		TF-IDF word n-grams	40	30	55.53	56.22	60.75	61.8	52.5	52.85
		TF-IDF char n-grams	60	53.33	60.2	60.3	67.5	67.43	58.5	65.84
		TF-IDF char_wb n-grams	40	30	56.2	57.96	67.25	67.32	61	59.66

Tableau 3-2 : Résultat des Processus de Vectorisation avec et sans POS, NER et lemmatisation pour les Classifieurs de ML sur les quatre corpus.

D'après les résultats montrés dans le tableau ci-dessus (tableau 3-3), le meilleur processus de vectorisation était le processus de vectorisation TF-IDF word n-grams sans POS, NER et lemmatisation pour le classifieur MNB en utilisant le troisième corpus (Acc=76, F1=75.59) et le meilleur processus de vectorisation était le processus de vectorisation TF-IDF word n-grams sans POS, NER et lemmatisation pour le classifieur MNB en utilisant le corpus arabe (Acc=66, F1=65.07).

III.6.2 Deep learning

Nous avons implémenté les deux classifieurs de deep learning en utilisant les bibliothèques keras, numpy, pandas, matplotlib, et sklearn.

III.6.2.1 Algorithme des Réseaux de Neurones Convolutionnel

Dans un premier lieu, nous avons créé une fonction « *create_conv_model* » qui génère un réseau de neurones récurrent pour une classification binaire ou multi-classes qui a les paramètres suivants :

- `word_index` : pour chaque token du corpus.
- `Label` : liste d'étiquettes pour déterminer s'il s'agit d'un binaire ou multi-classes.
- `embedding_matrix` : matrice d'entiers pour chaque mot dans le corpus.
- `pre_trained` : (booléen) déterminant si la classification utilisera un modèle pré-entraîné.

Et une autre fonction « *Metric.metrics_deep_learning* » a été utilisée pour tracer les différentes métriques pour les algorithmes d'apprentissage profond qui ont les paramètres suivants :

- `Model` :
(`tensorflow.python.keras.engine.sequential.Sequential`) modèle d'apprentissage profond.
- `history` : (`tensorflow.python.keras.callbacks.History`) historique du modèle d'apprentissage.
- `X` : données.
- `Y` : données cibles.
- `Labels` : liste contenant les étiquettes.

III.6.2.1.1 matrice de confusion

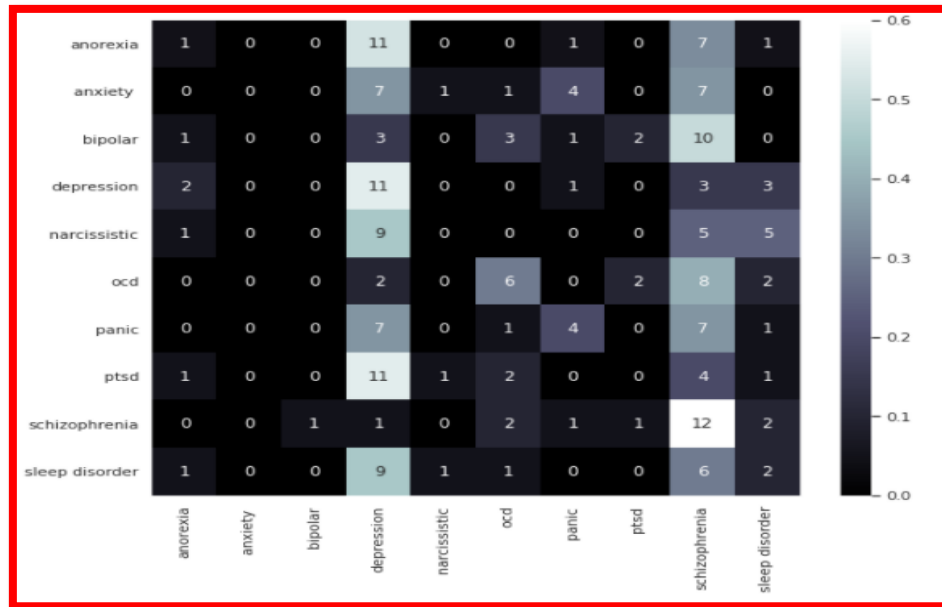


Figure3-50 : Matrice de confusion Crp2 -CNN-

Le meilleur résultat de classifieur CNN en utilisant le deuxième corpus est obtenu lorsque les valeurs de la diagonale principale sont grandes (proche du nombre des phrases de test qui est 20), la machine connaît correctement les maladies : depression, schizophrenia et elle se trompe dans la détection des maladies : ocd, anxiety ,anorexia , bipolar, ptsd , narcissism, panic, sleep disorder.

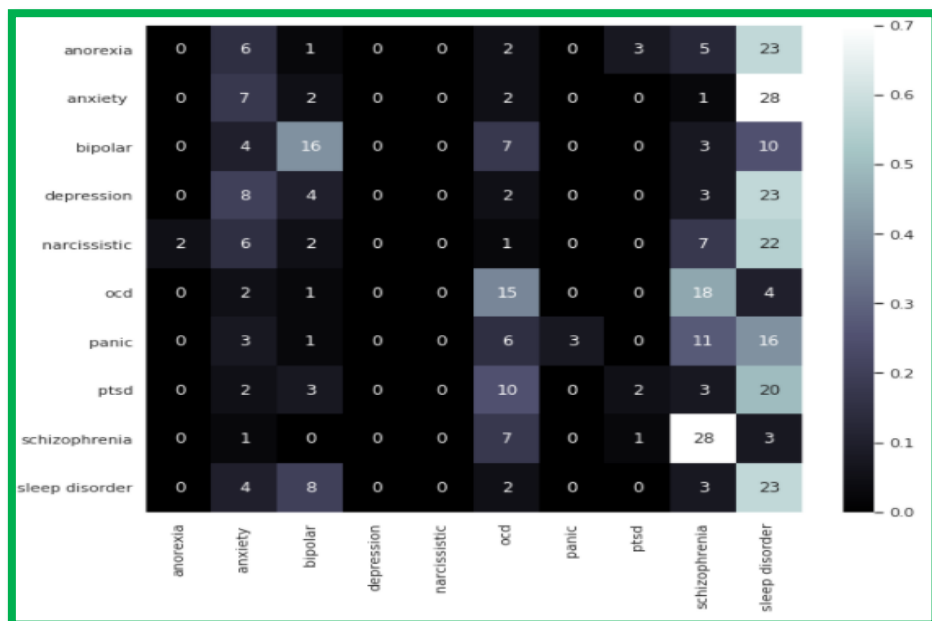


Figure3-51 : Matrice de confusion Crp3 -CNN-

Le meilleur résultat de classifieur CNN en utilisant le troisième corpus est obtenu lorsque les valeurs de la diagonale principale sont grandes (proche du nombre des phrases de test qui est 40), la machine connaît correctement les maladies : schizophrenia , sleep disorder et elle se trompe dans la détection des maladies : ocd, anxiety ,anorexia , bipolar, ptsd , narcissism, panic,depression.

III.6.2.1.2 Rappel et précision

Les figures ci-dessus illustrent une mise en correspondance des valeurs de rappel et de précision obtenues pour nos différentes classes.

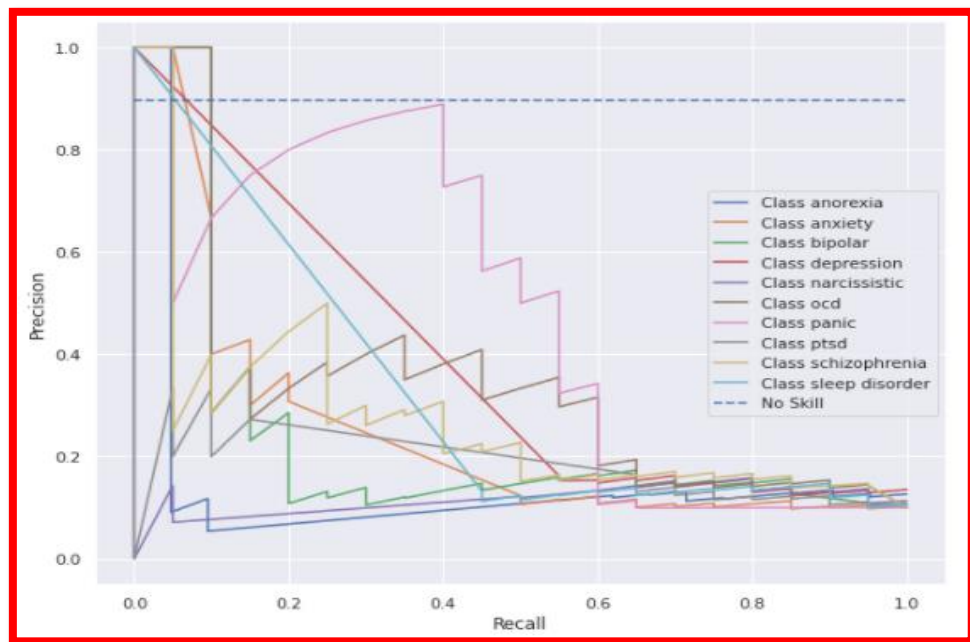


Figure3-52 : Courbe rappel et précision Crp 2 -CNN-

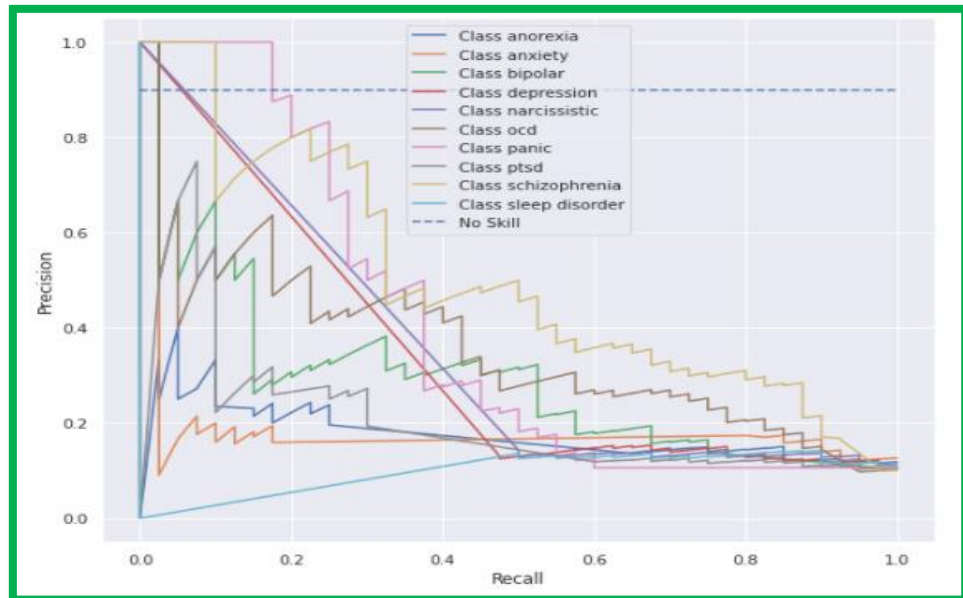


Figure3-53 : Courbe rappel et précision Crp3 -CNN-

III.6.2.2 RNN algorithm

Premièrement on crée une fonction `create_rnn_model` pour générer un réseau de neurones convolutifs pour une classification binaire ou multi-classe qui a les mêmes paramètres que `create_conv_model`, on utilise aussi la fonction `Metric.metrics_deep_learning`

III.6.2.2.1 Matrice de confusion

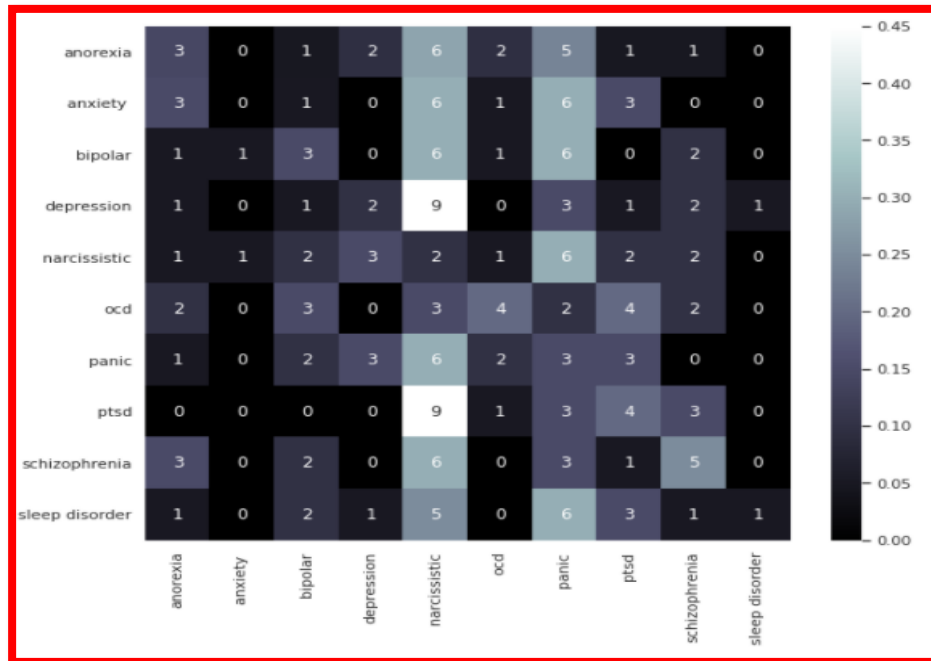


Figure3-54 : Matrice de confusion Crp2 -RNN-

Le meilleur résultat de classifieur RNN en utilisant le deuxième corpus est obtenu lorsque les valeurs de la diagonale principale sont grandes (proche du nombre des phrases de test qui est 20), la machine n'a reconnu aucune maladie et elle se trompe dans la détection des maladies : ocd, anxiety ,anorexia , bipolar, ptsd , narcissism, panic,depression , schizophrenia , sleep disorder.

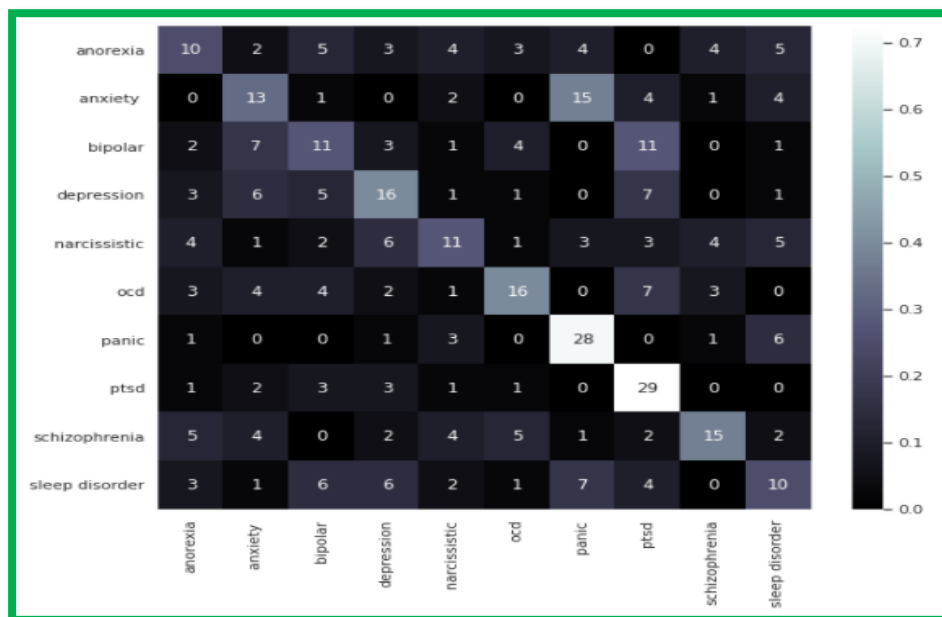


Figure3-55 : Matrice de confusion Crp3 -RNN-

Le meilleur résultat de classifieur RNN en utilisant le troisième corpus est obtenu lorsque les valeurs de la diagonale principale sont grandes (proche du nombre des phrases de test qui est 40), la machine connaît correctement les maladies :panic,ptsd et elle se trompe dans la détection des maladies : ocd, anxiety ,anorexia , bipolar, narcissism,,depression , schizophrenia , sleep disorder.

6.2.2.2 Rappel et précision

Les figures ci-dessus illustrent une mise en correspondance des valeurs de rappel et de précision obtenues pour nos différentes classes.

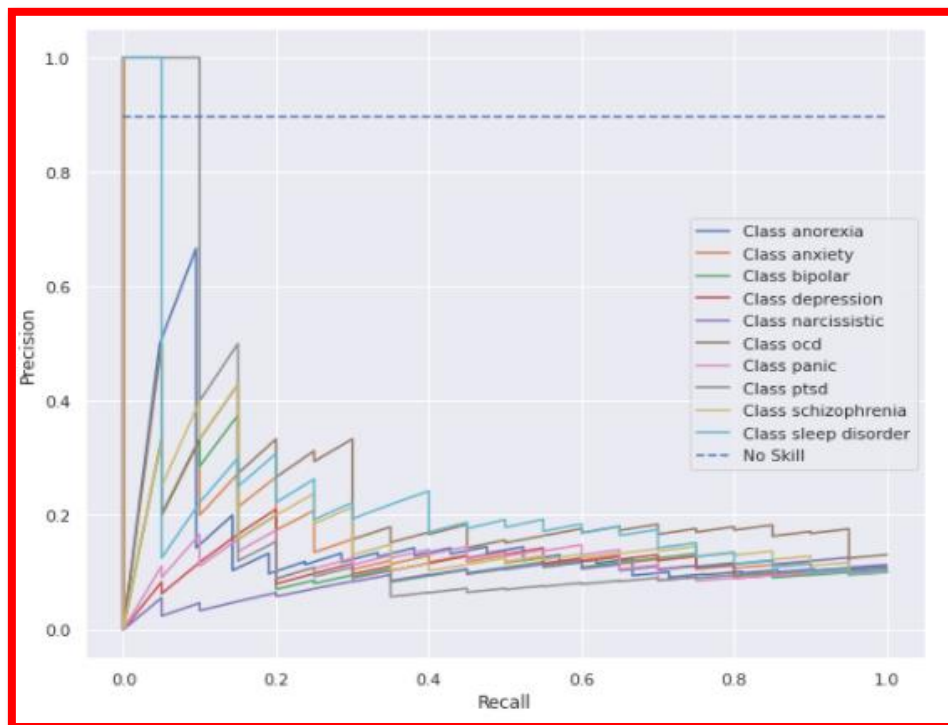


Figure3-56 : Courbe rappel et précision Crp2 -RNN-

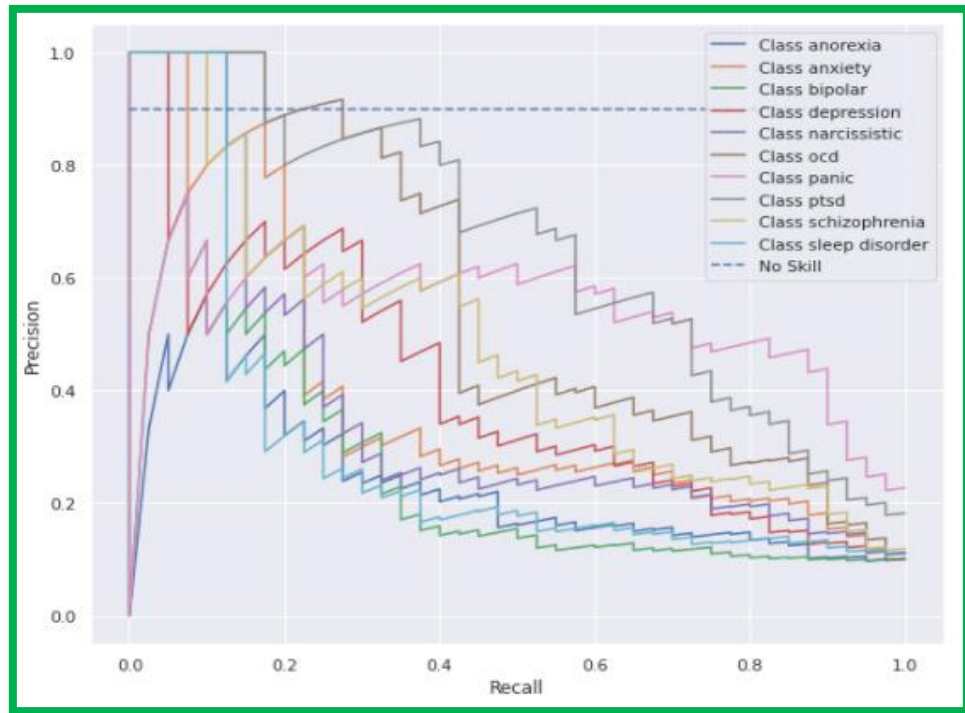


Figure3-57 : Courbe rappel et précision Crp3- RNN-

III.6.2.3 Des informations sur chaque algorithme

Le tableau ci-dessous (Tableau 3-4), représente une récapitulation des résultats de l'accuracy et F1 score pour les classifieurs de DL sur les quatre coprus.

		Corpus 1		Corpus 2		Corpus 3	
		Acc	F1	Acc	F1	Acc	F1
CNN	FastText=true	12	20	20	14	25	22
	Aucun	16	18	30	34	20	14
RNN	FastText=true	20	13	25	24	48	47
	Aucun	20	19	26	25	35	35

Tableau3-3 : Résultat de Accuracy et F1-score pour les classifieur de DL sur les quatre corpus

D'après les résultats montrés dans le tableau ci-dessus (tableau 3-3), les scores était mauvaises donc le meilleur processus de vectorisation était le processus de vectorisation TF-IDF word n-grams sans POS, NER et lemmatisation pour le classifieur MNB en utilisant le troisième corpus (Acc=76, F1=75.59) et le meilleur processus de vectorisation était le processus de vectorisation TF-IDF word n-grams sans POS, NER et

lemmatisation pour le classifieur MNB en utilisant le corpus arabe (Acc=66, F1=65.07).

III.6 Conclusion

À travers ce chapitre, nous avons rappelé la problématique initiale et les objectifs à atteindre de notre système. Pour ce faire, nous avons présenté l'environnement de développement, énuméré les différentes étapes de prétraitement par lesquelles passent les corpus utilisés pour l'apprentissage de notre système. Ensuite, nous avons présenté les détails de chaque étape de développement. Le chapitre suivant sera dédié à l'implémentation et la mise en œuvre de notre système.

Chapitre IV :

Présentation de

l'application finale

IV.1 Introduction

Pour mettre en œuvre un système ou une application informatique, l'environnement et le domaine d'application de ces derniers doivent constituer les principaux critères de sélection des approches et des outils utilisés pour la conception et l'implémentation.

Le présent chapitre sera donc consacré à la présentation de l'application finale. Nous commençons par présenter le langage de programmation, les environnements de développement et les outils que nous avons utilisés pour construire notre système.

IV.2 langages

Pour l'implémentation de notre système, nous avons utilisé le célèbre langage de programmation multi-usage et multiparadigme Python (voir le logo dans la figure ci-dessous). Ce dernier est un langage de script structuré et open source de haut niveau. Initialement développé en 1969 par Guide Van Rossum, il est comme la majorité des applications et outils open source, maintenu par une communauté de développeurs éparpillés à travers le monde.

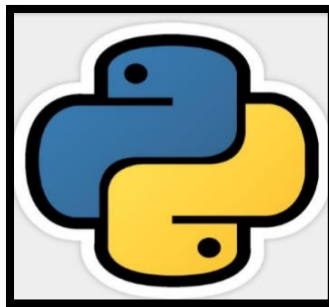


Figure4.1 : Logo du langage de programmation Python.

Grâce à ses bibliothèques spécialisées, ce langage peut être utilisé dans divers contextes et s'accommoder à tout type d'application. Néanmoins, il est considérablement utilisé pour l'automatisation des tâches élémentaires mais tant fastidieuses, notamment la mise en œuvre de modèles d'apprentissage automatique. Python, tout comme d'autres langages de

programmation, dispose de plusieurs versions. Les deux principales versions sont la version 2, publiée en 2000 et la version 3 publiée à son tour en 2006. Dans notre travail nous avons à l'évidence, utilisé la version 3 qui est plus récente, puisque le support de la deuxième version est fini depuis le 1er janvier 2020.

IV.3 Environnements

- **Visual Studio Code**

Pour développer notre système, nous avons utilisé l'environnement de développement pour Python, Visual Studio (voir le logo dans la figure ci-dessous).

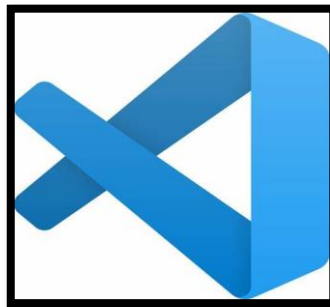


Figure4.2 : Logo de l'IDE Visual Studio Code.

L'environnement de développement intégré de Visual Studio est une plateforme de lancement créative pour Python (et d'autres langages) qui permet de modifier, déboguer et tester du code, puis de publier une application. Un environnement de développement intégré (IDE) est un programme riche en fonctionnalités qui peut être utilisé pour de nombreux aspects du développement de logiciels. En plus de l'éditeur et du débogueur standard fournis par la plupart des IDE, Visual Studio offre des outils de complétion du code, des environnements REPL interactifs et d'autres fonctionnalités qui facilitent le processus de développement logiciel.

- **Google Colab**

Google Colab ou Colaboratory est un service cloud, offert par Google (voir le logo dans la figure ci-dessous), basé sur Jupyter

Notebook ⁴⁶ et destiné à la formation et à la recherche dans l'apprentissage automatique. Cette plateforme permet d'entraîner des modèles de Machine Learning directement dans le cloud.



Figure4.3 : Logo de Google Colab.

IV.4 Outils

Pour construire nos corpus et notre application, nous utilisons l'outil suivant :

- **Flask**

Flask est un Framework web⁴⁷ Python construit avec un petit noyau et une approche Facile à étendre. Nous l'avons utilisé pour développer notre application.



Figure4.4 : Logo de Google Flask.

⁴⁶ **Jupyter Notebook** est une application Web Open Source permettant de créer et de partager des documents contenant du code (exécutable directement dans le document), des équations, des images et du texte. Avec cette application il est possible de faire du traitement de données, de la modélisation statistique, de la visualisation de données, du Machine Learning, etc. Elle est disponible par défaut dans la distribution Anaconda .

⁴⁷ En programmation informatique, un framework (appelé aussi infrastructure logicielle, infrastructure de développement, environnement de développement, socle d'applications, cadre d'applications ou cadriciel) désigne un ensemble cohérent de composants logiciels structurels, qui sert à créer les fondations ainsi que les grandes lignes de tout ou d'une partie d'un logiciel (architecture).

IV.5 Présentations de l'application

Nous avons voulu que notre interface soit conviviale et simplifiée et cette dernière est composée de cinq pages :

- Page d'accueil.
- Page d'identification.
- Page de l'expression de sentiment.
- Page de questionnaire.
- Page de diagnostic et de Conseils.

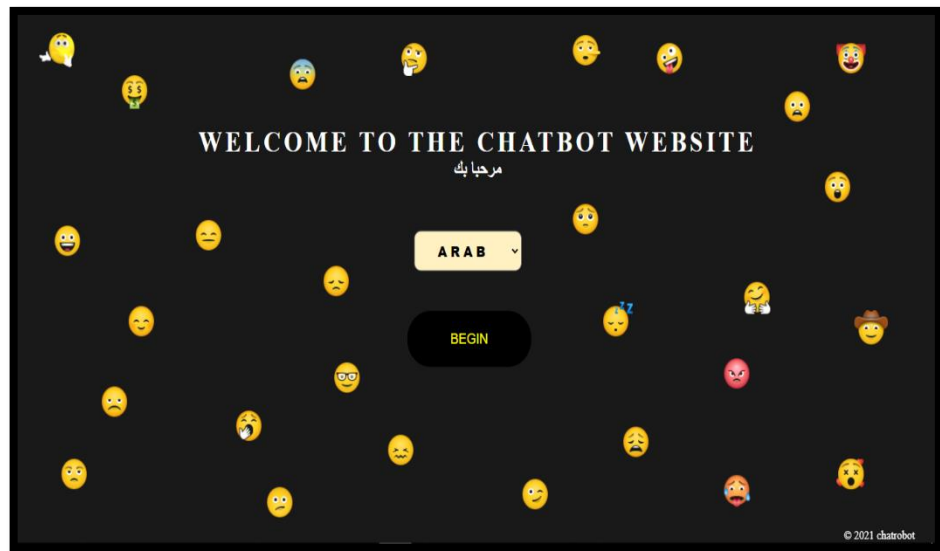


Figure4.5 : Page d'accueil.

Comme montre (Figure 4.5), Lorsque l'utilisateur se connecte, il aura la possibilité de :

- Choisir la langue (anglaise ou arabe).
- Passer à la page suivante en cliquant sur le bouton BEGIN.

Les figures suivantes présentent la page d'identification après la sélection de la langue anglaise et la page d'identification après la sélection de la langue arabe.

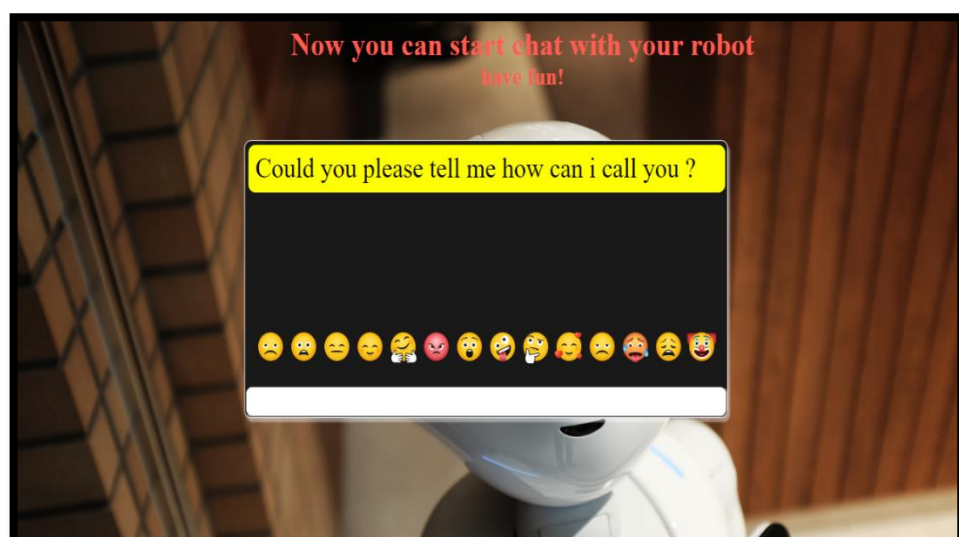


Figure4.6 : Page d'identification en anglais.



Figure4.7 : Page d'identification en arabe.

Comme montrent (Figure 4.6 et Figure 4.7), l'utilisateur aura la possibilité de saisir son vrai nom ou un pseudonyme.

Les figures suivantes présentent les pages de l'expression de sentiment en anglais et en arabe.

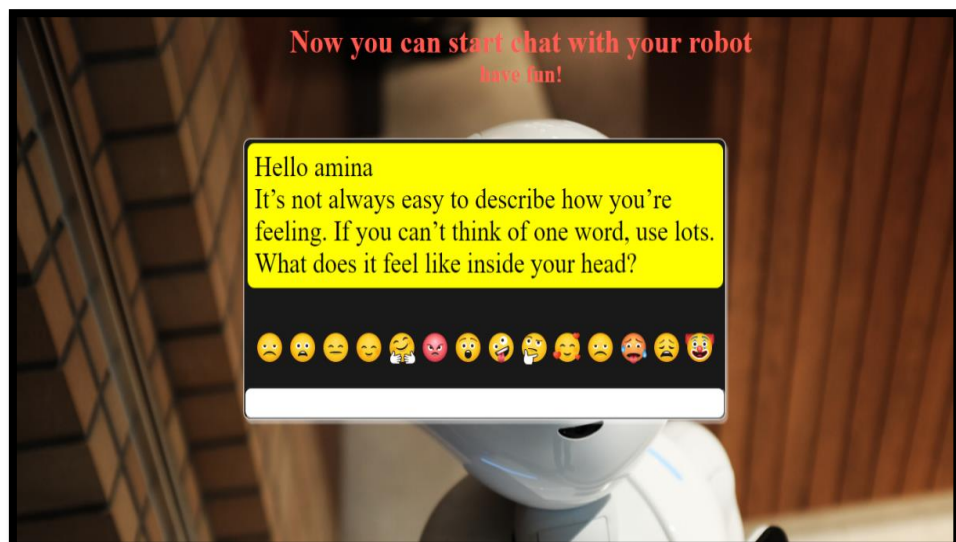


Figure4.8 : Page de l'expression de sentiment en anglais.



Figure4.9 : Page de l'expression de sentiment en arabe.

Comme montrent (Figure 4.8 et Figure 4.9), l'utilisateur aura la possibilité de s'exprimer.

La figure suivante présente la page de questionnaire.

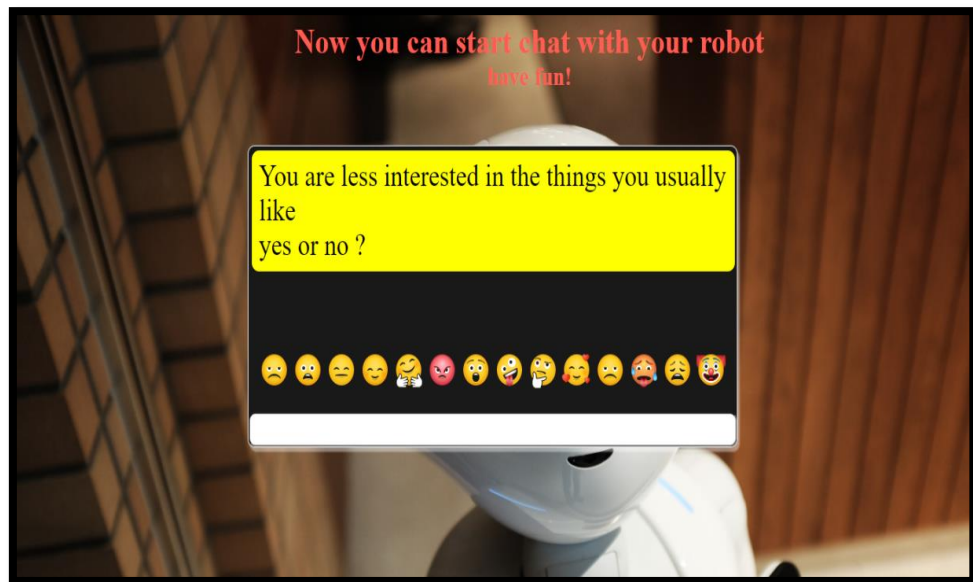


Figure4.10 : Page de questionnaire.

Comme montre (Figure 4.10), l'utilisateur aura la possibilité de répondre aux dix questions de test par yes ou no.

La figure suivante présente la Page de diagnostic et de solution.

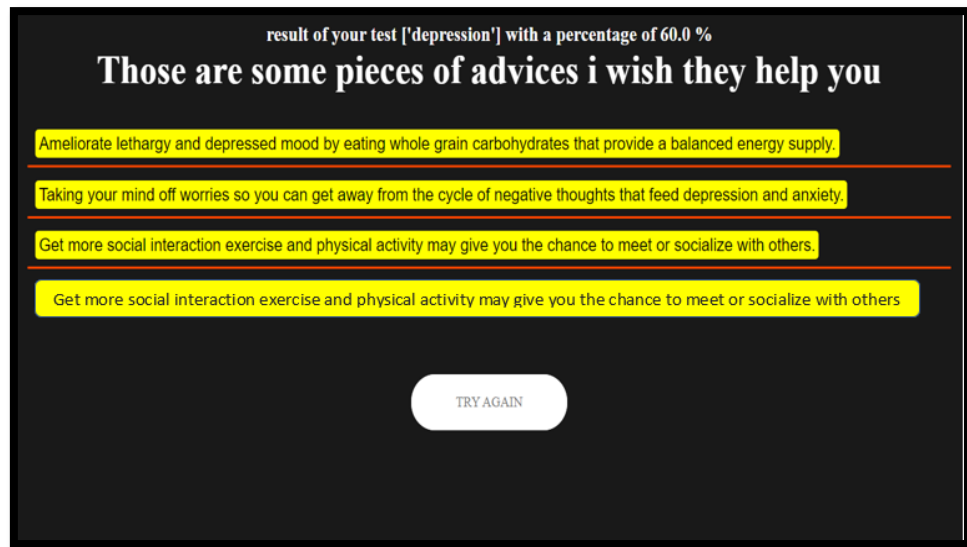


Figure4.11 : Page de diagnostic et des Conseils.

Comme montre (Figure 4.11), l'utilisateur aura la possibilité de :

- Consulter son diagnostic et les Conseils proposées.
- Retourner à la page d'accueil en cliquant sur le bouton TRY AGAIN.

Les figures suivantes présentent le déroulement d'un exemple pour la maladie "Dépression" après la sélection de la langue anglaise.

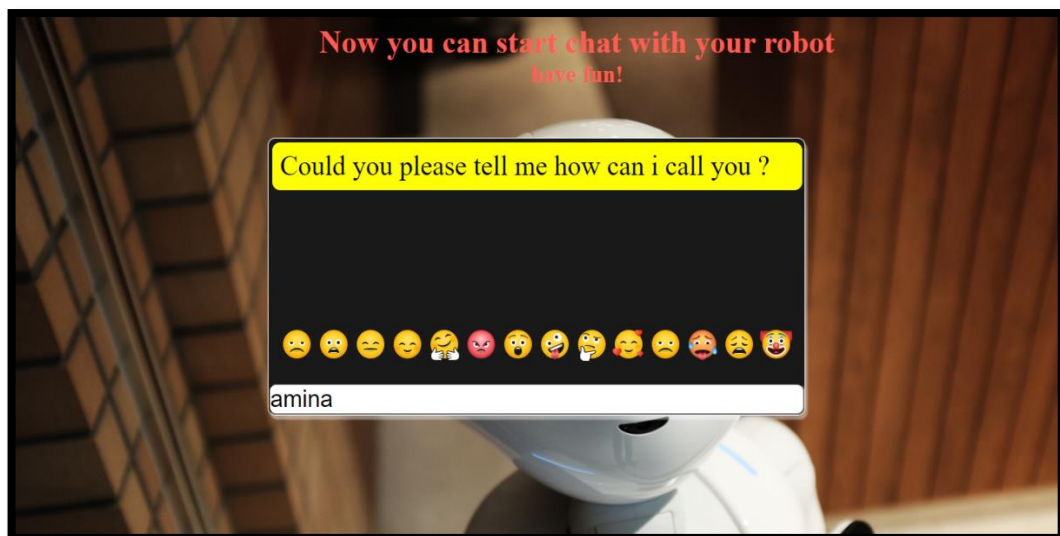


Figure 4.12 : Exemple d'identification de l'utilisateur en anglais.

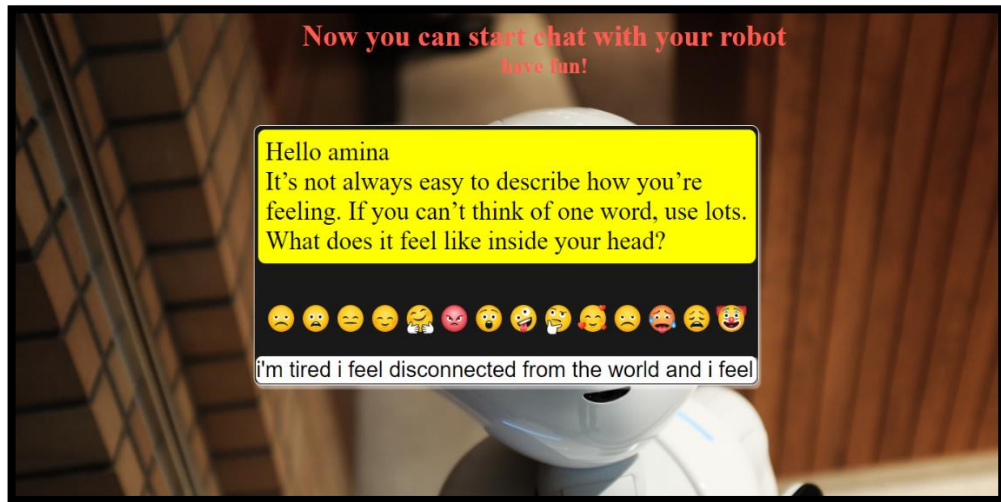


Figure 4.13 : Exemple d'expression de sentiment en anglais.

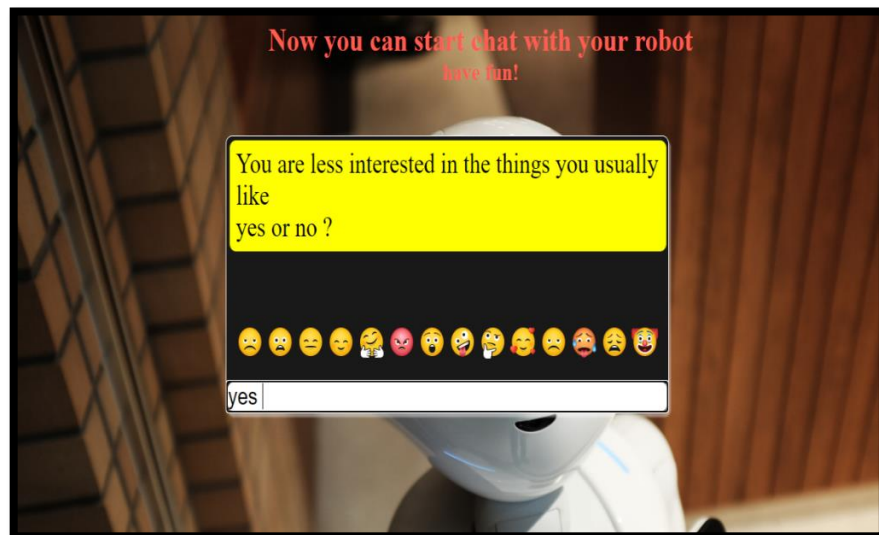


Figure 4.14 : Exemple de questionnaire pour la maladie "Dépression".



Figure 4.15 : Exemple de détection et Conseils pour la maladie "Dépression".

Les figures suivantes présentent le déroulement d'un exemple pour la maladie "القلق" après la sélection de la langue arabe.



Figure 4.16 : Exemple d'identification de l'utilisateur en arabe.



Figure 4.17 : Exemple d'expression de sentiment en arabe.



Figure 4.18 : Exemple de questionnaire pour la maladie "القلق".



Figure 4.19 : Exemple de détection et Conseils pour la maladie "القلق"

IV.6 Conclusion

Ce dernier chapitre a été consacré à la présentation de l'implémentation de notre système, en commençant par les langages utilisés. Nous avons par la suite décrit les environnements de développements et les Outils. Pour conclure, nous avons exposé la présentation de l'application finale.

Conclusion générale

Synthèse

Au cours des dernières années, le besoin de Chatbots a considérablement augmenté, et il augmentera encore plus dans les années à venir. Et c'est parce que les Chatbots offrent un moyen de communication rapide, automatisé et peu coûteux entre les entreprises et les utilisateurs, ce qui est bénéfique pour les deux.

Dans ce projet de Master, l'objectif était de créer un système générique qui aide les patients du secteur médical psychologique. Pour cela, nous avons dû voir de nombreux exemples de Chatbots et tests en ligne disponibles, et essayer de comprendre comment ils fonctionnent, et comment nous pouvons implémenter notre propre version à partir de zéro pour notre domaine spécifique sans utiliser de plateforme de création pour ce genre de système, car cela ne nous aiderait pas à comprendre comment fonctionne le cœur du système.

Au cours de nos recherches, nous avons découvert que contrairement à ce que nous pensions initialement et à ce que la plupart des gens pensent, le domaine des Chatbots et du Traitement Automatique du Langage Naturel en général souffre de nombreuses carences, trop de travail reste encore à faire même pour les pionniers comme Google ou Amazon.

Les objectifs que nous nous étions fixés initialement ont été partiellement atteints, nous avons réussi à créer un système fonctionnel pour soutenir psychiquement les utilisateurs ou de rendre possible un diagnostic en fonction des symptômes des utilisateurs.

Nous avons rencontré de nombreuses difficultés et nous avons réussi à en surmonter certaines, mais d'autres nous ont obligés à changer notre façon de faire et à nous réadapter, encore et encore. Le principal problème que nous avons rencontré était le manque de données, les choses auraient été beaucoup mieux avec plus de données à portée de main, "Les données sont le nouveau pétrole" comme l'a dit Clive Humby , un mathématicien

et entrepreneur britannique dans le domaine de la science des données et des stratégies commerciales centrées sur le client.

Perspectives

Au cours de ce travail, nous avons appris beaucoup de choses, à la fois théoriques et pratiques, c'était une introduction pour nous au domaine du Machine Learning, et cela nous ouvrira certainement des portes à explorer. Et de manière générale nous pouvons dire que nous avons atteint les objectifs fixés initialement dans le projet.

Toutefois, il reste quelques perspectives d'amélioration de l'application qui pourrait être résumées en :

- L'ajout d'un espace chat.
- Tester d'autres algorithmes de classification.
- Comparer les résultats des tests TAL arabe sans passer par la traduction.

Bibliographie

[1]: Chatbot Market Size To Reach \$1.25 Billion By 2025. August 2017. URL: <https://www.grandviewresearch.com/press-release/global-chatbotmarket> (visited on 01/03/2019).

[2] :Ruth Robertson, John Appleby, and Harry Evans. Public satisfaction with the NHS and social care in 2017. February 2018. URL: <https://www.kingsfund.org.uk/publications/public-satisfaction-nhs-2017> (visited on 01/03/2019).

[3]: Healthcare-Barometer.March2018.
URL:<https://www.pwc.de/de/gesundheitswesenund-pharma/pwc-ergebnisse-healthcare-barometer-2018-final.pdf> (visited on 01/03/2019).

[4]: The 2018 State of Chatbots Report. January 2018. URL: <https://www.drift.com/wp-content/uploads/2018/01/2018-state-of-chatbotsreport.pdf> (visited on 01/03/2019).

[5]: What Consumers Really Think About AI: A Global Study. April 2017.
URL: <https://www.pega.com/sites/pega.com/files/docs/2017/Apr/whatconsumers-really-think-about-ai.pdf> (visited on 01/03/2019).

[6] : Mémoire de Master : Conception et mise en place d'un chatbot dans le cadre du support client /Auteur : Peters, Florian Promoteur(s) : Wehenkel, Louis Faculté : Faculté des Sciences Appliquées Diplôme : Master en ingénieur civil en informatique, à finalité spécialisée en "systèmes intelligents" Année académique : 2017-2018

[7] : Mémoire de Master : Conception et mise en place d'un chatbot dans le cadre du support client /Auteur : Peters, Florian Promoteur(s) : Wehenkel, Louis Faculté : Faculté des Sciences Appliquées Diplôme : Master en ingénieur civil en informatique, à finalité spécialisée en "systèmes intelligents" Année académique : 2017-2018 URI/URL : <http://hdl.handle.net/2268.2/4625>

[8]: Rosruen, Nudtaporn, and Taweesak Samanchuen. "Chatbot utilization for medical consultant system." *2018 3rd technology innovation management and engineering science international conference (TIMES-iCON)*. IEEE, 2018.

[9]: Chatbot:Architecture,Application and Design Process Step
.URL:<https://www.elprocus.com/chatbot-design-process-and-its-architecture/>.(visité le 16/07/2021).

[10] : Wallace.RS, "L'anatomie d'ALICE," 2007. [En ligne]. URL :
<https://alicebot.org/> (visté le 16/07/2021)

[11] : Lokman.Abbas.S et Zain.Jasni.M, "Catégories à une correspondance et à toutes les correspondances pour la correspondance des mots-clés dans Chatbot", *American Journal of Applied Sciences*, 2010.(visté le 16/07/2021)

[12] : Les Chatbots, état des lieux et perspectives en santé, Joséphine Arrighi de Casanova et Jacques Durand, Note de synthèse, Février 2019 (MAJ Mars 2019)(visté le 20/07/2012)

[13] : Gregori, E. (2017). Évaluation des outils modernes pour un chatbot conseiller omcs.

[14] : Storp, Michaela. "Chatbots. Possibilités et limites du traitement du langage naturel par les machines". *Networx* n° 25 25 (2002).

[15] : Beitzel, S. M., Jensen, E. C., Chowdhury, A. et Frieder, O. (2008, juin). Analyse des différentes approches de la classification des requêtes web topiques. Dans les actes de la 3e conférence internationale sur les systèmes d'information évolutifs (pp. 1-5).

[16]: Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2019). La prochaine génération : les chatbots en psychologie clinique et en psychothérapie pour favoriser la santé mentale-une revue de cadrage. *Verhaltenstherapie*, 1-13.

[17] : Klopfenstein, L. C., Delpriori, S., Malatini, S. et Bogliolo, A. (2017, juin). The rise of bots : Une enquête sur les interfaces, les modèles et les paradigmes

conversationnels. In Proceedings of the 2017 conference on designing interactive systems (pp. 555-565).

[18] : Bendig E, Erb B, Schulze-Thuesing L, Baumeister H (2019) Prochaine génération : les chatbots en psychologie clinique et psychothérapie pour favoriser la santé mentale — un examen de portée. *Verhaltenstherapie* : 1-15. <https://doi.org/10.1159/000499492>

[19] : Prochaska, J. J., Vogel, E. A., Chieng, A., Kendra, M., Baiocchi, M., Pajarito, S., & Robinson, A. (2021). A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study. *Journal of Medical Internet Research*, 23(3), e24850.

[20]: Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health*, 4(2), e7785.

[21]: Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*, 6(11), e12106.

[22]: Mind Diagnostics. URL: <https://tracxn.com/d/companies/minddiagnostics.org> (visité le 23/07/2021)

[23] : Tests Psy.org URL: <https://apps.apple.com/us/app/psytests/> (visité le 23/07/2021)

[24] : Digital guide URL : <https://www.ionos.fr/digitalguide/web-marketing/search-engine-marketing/deep-learning-vs-machine-learning/> (visité le 25/08/2021)

