

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique



Université Abderrahmane Mira-Béjaïa
Faculté des Sciences Exactes
Département d'Informatique

Mémoire de fin d'études
En vue de l'obtention du diplôme de Master Académique
en Informatique
Option Intelligence Artificielle

Un système de recommandation de nutrition pour les personnes obèses

Réalisé par Mlle :
Fadia DJAIDANI
Asma BOUCETTA

La présidente du jury
Dr. Nadia BATTAT
Examinatrice
Dr. Hayette KHALED
Encadrante
Dr. Houda EL BOUHISSI

Table des matières

1	Introduction Générale	1
1.1	Introduction	1
1.2	Contexte et Problématique	2
1.3	Motivation	3
1.4	Objectif	3
1.5	Contribution	3
1.6	Organisation du mémoire	4
2	Généralités sur les systèmes de Recommandation	5
2.1	Introduction	5
2.2	Historique	6
2.3	Définition des systèmes de recommandation	6
2.4	Types des systèmes de recommandation	7
2.4.1	Filtrage basé sur le contenu :	7
2.4.2	Filtrage collaboratif :	8
2.4.2.1	Filtrage collaboratif basé sur la mémoire :	9
2.4.2.2	Filtrage collaboratif basé sur un modèle :	10
2.4.3	Approches Hybrides :	11
2.5	Domaines d'utilisation des systèmes de recommandation	12
2.5.1	Recommandation des produits	12
2.5.2	Services de streaming	12
2.5.3	Recommandation d'articles	12
2.5.4	Services	12
2.5.5	Healthcare	13
2.6	Conclusion	13
3	État de l'art	14
3.1	Introduction	14
3.2	Travaux connexes	15
3.2.1	Tableau comparatif	19
3.3	Discussion et comparaison	23
3.4	Conclusion	25
4	Contribution	26
4.1	Introduction	26
4.2	Approche proposée	26
4.2.1	Création du profil utilisateur :	28
4.2.2	Recherche de profils similaires :	28
4.2.3	Apprentissage supervisé	29
4.2.3.1	Classification :	30
4.2.3.2	La similarité Cosinus	31

4.2.4	Recommandation :	32
4.3	Conclusion	35
5	Implementation	36
5.1	Introduction	36
5.2	Jeux de données	36
5.3	Environnement de développement	41
5.3.1	Anaconda	41
5.3.2	Jupyter notebook	42
5.4	Langage de programmation "Python"	42
5.5	Bibliothèques de python	43
5.5.1	Pandas	43
5.5.2	Sckit-Learn	44
5.5.3	Scipy	44
5.5.4	Numpy	44
5.5.5	FuzzyWuzzy	44
5.5.6	Matplotlib	44
5.6	Sublime text3	45
5.6.1	Html	45
5.6.2	CSS	46
5.7	Interface d'inscription :	46
5.8	Interface d'authentification :	47
5.9	Évaluation du système	47
5.9.1	Métriques de classification	47
5.10	Conclusion	51
6	Conclusion Générale	52
	Bibliographie	53

Table des figures

1.1	Classification d'obésité	2
2.1	Filtrage Collaboratif	8
2.2	User-based VS Item-based	9
2.3	FC vs FBC	11
4.1	Approche proposée	27
4.2	K-plus proche voisin	31
4.3	Similarité de cosinus	32
4.4	Diagramme d'activité	34
5.1	Les différentes phases des Jeux de données	37
5.2	Dataset "Epicurious"	38
5.3	Jeux de données "Food and their calories"	38
5.4	"Food and their calories"	39
5.5	"Epicurious"	39
5.6	Recommandation de recettes	40
5.7	Recommandation d'aliments	40
5.8	L'interface du navigateur Anaconda	41
5.9	Jupyter notebook	42
5.10	Bibliothèques de python	45
5.11	Interface d'inscription	46
5.12	Interface d'authentification	47
5.13	L'ensemble de données	49
5.14	Données d'entraînement et données de test	50
5.15	La matrice de confusion	50
5.16	La matrice de confusion.	50
5.17	Les résultats de l'évaluation dans le cas d'une classification multiple	51
5.18	Taux de Accuracy	51

Liste des tableaux

- 3.2 Etat de l'art de quelques contributions relatives aux systèmes de recommandation des soins de la santé. 23
- 4.1 Feedback explicite 29

Liste des Abréviations

ACO	Ant Colony Optimization.
BDD	Base De Donnés
BLS	Bundes Lebensmittel Schluesse.
CSS	Cascading Style Sheets
FBC	Filtrage basé sur le Contenu.
FC	Filtrage Collaboratif.
HFO	Healthy For Obesity
HTML	HyperText Markup Language
IA	Intelligence Artificielle
IOT	Internet of things
IMC	Indice de Masse Corporelle.
KNN	K-Nearest Neighbors.
ML	Machine Learning.
OMS	Organisation Mondiale de la Santé.
SR	Système de Recommandation.
TF-IDF	Term Frequency-Inverse Document Frequency.
TIC	Technologies de l'information et de la Communication.

Remerciements

Notre remerciement va en premier lieu à ALLAH le tout puissant pour ses bénédictions et de nous avoir donné la foi et l'audace pour dépasser toutes les difficultés tout au long de notre recherche ,ainsi ,pour terminer ce modeste travail.

Nous tenons à remercier particulièrement notre encadrant Mme ELBOUHISSI de nous avoir proposé ce travail qui nous semble très intéressant, pour l'intérêt qu'elle a manifesté à notre mémoire, pour son écoute , ses conseils et surtout encouragements.

Nous remercions très sincèrement, les membres de jury de l'intérêt qu'ils ont porté à ce travail en faisant partie de la commission d'examineur.

Enfin, un énorme remerciement à tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail.

Dedicaces

Je dédie ce modeste travail à mes très chers parents. ma mere pour ses conseils, patience et surtout prières qui m'ont toujours accompagné, motivé et encouragé à avancer depuis tout le début. Mon père pour ses sacrifices et son soutien psychologique et matériel durant tout mon parcours.

A mes frère ABY et ADAM .

A ma tante Lynda malgré la distance, elle ne cesse de m'envoyer ses ondes positives et son soutien depuis tous mes debuts.

A mes oncles TAHAR,KAMEL et DJAMEL qui ont toujours été à mes côtés dans le bon et le mauvais .

A mes grands-parents maternels Maman Ouardia ,et Papa AHMED , ainsi que Yemma Zouina . Et paternels Maman Fafa et Papa Mustapha que dieu nous bénisse , ils ont toujours été là aussi à m'encourager, me motiver et surtout prier pour que tout se passe a merveille.

A toute ma chère famille de loin ou de près, mes oncles et tantes paternels qui malgré la distance aussi prennent de mes nouvelles et me soutiennent au long de ce travail .

Enfin a mes chères cousines , cousins , collègues, amis et tous ceux qui me portent dans leurs coeur. Sans oublier ma binôme, pour sa patience, sa compagnie , et surtout son amitié.

Fadia

Dedicaces

« Je crois que le hasard n'existe pas et que chaque personne qu'on croise dans notre vie a une leçon ou quelque chose à nous apporter, il suffit d'être attentif aux détails et toutes les personnes que je vais citer ici y sont pour quelque chose »

Je tiens à dédié ce modeste travail à mes chers parents et je les remercie pour toutes les valeurs qui ont pu nous inculquer, pour leur accompagnement tout au long de ce parcours fastidieux, je les remercie de m'avoir appris que la connaissance et la science été les gisements inépuisables de la vie et que c'était le meilleur investissement qu'on pouvait faire sur soi. Quoi que je dise je ne saurai vous décrire vous êtes mes exemples que dieu vous bénisse.

Je le dédie aussi à mon arrière grand mère Jida Nouara qui m'a donné une éducation exemplaire qui malgré la maladie ne ma jamais laisser tomber, elle ma bercer de tendresse et d'affection et j'ai eu l'immense honneur ma sœur et moi d'être ses 'filles' préférées par excellence. Je remercie tata Djaouida aussi qui m'a toujours soutenue depuis que j'étais enfant, qui m'a aimée pour ce que je suis.

Je dédie cet événement marquant de ma vie à la mémoire de mes grands-parents jedi et jida qui m'ont appris la rigueur et l'organisation et de YAYA disparue trop tôt, elle qui m'a appris que les valeurs été ce qu'on n'a de mieux a montré que la générosité venait du cœur et non de richesse et tant d'autres choses encore je vous garderai dans mon cœur éternellement.

Je remercie aussi tonton Hamid et tata Nacera de m'avoir accompagné tout au long de mon parcours universitaire vous êtes des personnes formidables

A tous mes oncles et tantes je vous remercie pour tous, à ma tendre jumelle et mes frères, à mes chers cousines et cousins qui m'ont soutenue tout au long de cette trajectoire académique. Pour finir je remercie mon binôme pour sa patience et sa compréhension. Je ne peux exprimer ma profonde gratitude envers vous tous votre amour a fait de moi ce que je suis aujourd'hui.

Asma

Chapitre 1

Introduction Générale

1.1 Introduction

En cette dernière décennie, le mode de vie des populations a complètement changé. Nous sommes passés d'une alimentation saine et bio à une alimentation malsaine et rapide, ajoutant à cela la diminution de l'activité physique car avec l'essor de la technologie, les gens sont de plus en plus sédentaires, et ceci engendre de gros dégâts sur notre état de santé. Selon l'OMS¹ d'ici 2030, 8 des 10 principales causes de décès seront liées à des conditions à risque de "Maladies liées au mode de vie", sans distinction de sexe [9]. On compte parmi elles l'obésité.

L'obésité semble avoir atteint des proportions épidémiques. En Algérie seulement, les chiffres ont triplé depuis 2015 [8]. Selon les derniers résultats de l'enquête 'Stepwise' du ministère de la Santé en 2017, le taux de l'obésité et du surpoids est de 63,3 % chez les femmes et de 48,3% chez les hommes, mais aussi, elle gagne du terrain pour concerner même les enfants. Ainsi, dans le monde, le nombre d'adultes en surpoids dans le monde s'élève à plus de 1,9 milliard, dont 650 millions sont obèses [9].

L'obésité est généralement définie comme une augmentation anormale de la teneur en graisse du corps. Cela est principalement dû à une mauvaise éducation nutritionnelle [26], elle demeure parmi les facteurs favorisant l'apparition du diabète, des maladies cardiovasculaires, des cancers du foie, du sein chez les femmes... etc

La mesure de référence internationale actuelle de l'obésité est l'IMC² présenté dans la Figure 1.1 [1] qui est égal au rapport du poids (en kg) sur le carré de la taille (en mètres).

1. Organisation Mondiale de la Santé
2. Indice de masse corporelle

Une personne est dite en [7] :

- Surpoids lorsque son IMC ≥ 25 .
- Obésité si IMC ≥ 30 .
- Obésité grade I si $30 < \text{IMC} < 34,9$.
- Obésité grade II si $35 < \text{IMC} < 39,9$.
- Obésité grade III si IMC ≥ 40 .

Classification obésité

- Poids santé : 18,5-24,9
- Surpoids (non-obèses) : 25-29,9
- Obésité classe 1 : 30,0 -34,9
- Obésité classe 2 : 35,0-39,0
- Obésité classe 3 : 40,0-49,9
- Obésité classe 4 : 50,0-69,9 (super obésité)
- Obésité classe 5 : > 70 (méga obésité)

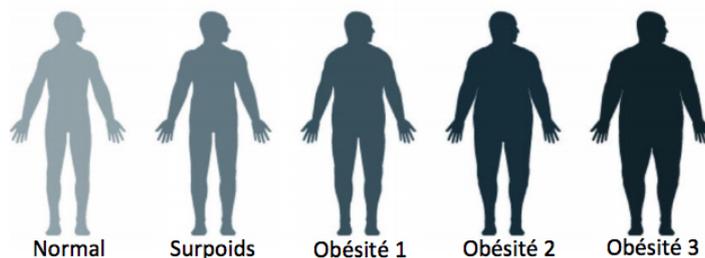
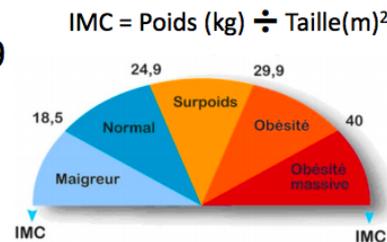


FIGURE 1.1 – Classification d'obésité

1.2 Contexte et Problématique

La prévalence de l'obésité ne cesse d'accroître à l'échelle mondiale. Étant donné qu'elle dégrade grandement la qualité de vie des personnes adultes et peut avoir de sérieuses conséquences sur leur santé, dont on peut citer l'augmentation des risques de maladies cardiovasculaires, la réduction de l'espérance de vie, le risque d'atteindre le diabète également, l'hypertension artérielle, le cancer du côlon...

Ainsi, chez l'enfant et l'adolescent, le surpoids et l'obésité ont pareillement des conséquences néfastes, telles que les problèmes respiratoires ; le risque de fractures ; Le développement d'une

résistance à l'insuline, conduisant au diabète ...

Chez l'adulte comme chez l'enfant, le surpoids et l'obésité peuvent aussi favoriser l'apparition d'une dépression.

Devant l'évolution de cette maladie multifactorielle et pour une meilleure gestion de l'état de santé des personnes souffrantes de cette maladie ; Traiter le surpoids et l'obésité est donc nécessaire pour préserver la santé des personnes concernées. Pour cela, on a pensé à développer un outil permettant de recommander quotidiennement des plats sous forme de recettes avec leur apport calorique adéquates à chaque personne concernée.

1.3 Motivation

Nous assistons aujourd'hui à une véritable épidémie d'obésité qui touche les adultes ainsi que les enfants. Étant donné les conséquences et les complications que cette dernière leur cause, et le statut social précaire de certaines personnes qui ne leur octroie pas l'accès à un suivi chez un diététicien, nous avons étudié la possibilité de mettre en œuvre une application numérique qui pourrait assister les personnes au quotidien afin de choisir la bonne alimentation en fonction de leur état de santé.

1.4 Objectif

L'objectif principal de ce mémoire est de créer une application munie d'un système de recommandation offrant aux utilisateurs souffrants d'obésité ou de surpoids des suggestions de plats ,de recettes et des régimes personnalisés adéquats à chaque métabolisme,accordant ainsi une assistance diététique au quotidien.

1.5 Contribution

Les contributions que va comporter notre mémoire sont :

- Établir un état de l'art des principales contributions de ce domaine.
- Proposer un nouveau modèle de recommandation basé sur le profil de l'utilisateur,son historique et sur le contenu (utilisation de l'approche collaborative).
- Utiliser des connaissances de l'utilisateur pour accélérer le processus de recommandation.

- Proposer un outil offrant un suivi aux personnes souffrant de surpoids et d'obésité.

Nous discutons des détails techniques du cadre, et de la façon dont il permet aux patients de gérer leurs poids.

1.6 Organisation du mémoire

Le reste du mémoire est organisé comme suit :

Chapitre 01 : Aborde une introduction générale sur le healthcare plus précisément l'obésité, suivis des différents types de cette dernière .On a défini l'objectif de notre travail, nos motivations, ainsi que notre problématique.

Chapitre 02 : Ce chapitre donne un aperçu général sur les concepts de base des systèmes de recommandation et leurs types, ainsi que leurs domaines d'application.

Chapitre 03 : Ce chapitre comporte un état de l'art des principales contributions proposées dans ces domaines. Cet état de l'art sera suivi d'une synthèse et étude comparative de ces travaux.

Chapitre 04 : Concerne notre approche, et l'explication de ses différentes étapes.

Chapitre 05 : Ce chapitre se base sur les détails d'implémentation et de mise en œuvre de notre approche , ainsi que les résultats obtenus.

Chapitre 06 : Conclusion et travaux futurs.

.

Chapitre 2

Généralités sur les systèmes de Recommandation

2.1 Introduction

Le concept de la recommandation existe depuis la nuit des temps ; depuis toujours les gens ont eu besoin de se concerter pour prendre les bonnes décisions. À l'ère actuelle d'Internet et avec l'accroissement des informations, les systèmes de recommandation ont vu le jour.

Ces mécanismes ont la capacité de filtrer et d'hierarchiser de grands volumes d'informations générés dynamiquement, afin de permettre aux utilisateurs d'accéder à ce qui les intéresse dans un temps record .

Avec la tendance récente du big data en constante augmentation, l'analyse de grands ensembles de données et la mise en œuvre de systèmes de recommandation sont devenus incontournables. Ces systemes sont utilisés aussi comme stratégie marketing qui aide les commerçants à augmenter leur chiffres d'affaires. Ces derniers compte encore se répandre dans diverses autres domaines.

Ce type de système a pour objectif de réduire aux utilisateurs le temps de recherche et minimise le nombre et le parcours de visites non intéressantes vis à vis de ce dernier.

2.2 Historique

L'origine des systèmes de recommandations remonte à l'étude approfondie dans le domaine de la science cognitive [30], ce dernier a conçu un système bibliothécaire, classait les utilisateurs en "stéréotypes" en se basant sur un court interview, et utilisait ces stéréotypes pour produire des recommandations de livres.

En 1990, avec l'essor du Web, conforté par le rapide développement des nouvelles technologies de l'information et de la communication a conduit à la surcharge d'informations. Pour faire face à ce problème, le filtrage collaboratif a été développé.

Puis en 1992 le centre de recherche de "Xerox" a conçu un nouveau SR appelé « Tapestry » qui a permis aux utilisateurs de créer des requêtes permanentes, basées sur les annotations des utilisateurs.

Deux ans après, le laboratoire de recherche « GroupLens » [28] a été créé. Ce dernier a pour premier objectif d'élaborer des SRs automatiques, il utilise alors l'approche de type "plus proches voisins" à partir de l'historique de l'utilisateur.

En 1995, apparaissent successivement « Ringo » [24], un SR de musique, basé sur les appréciations des utilisateurs, et « Bellcore » [21], un SR de vidéos. Dans la même année, « GroupLens » conçoit la société "Net Perceptions" dont le premier client a été "Amazon".

En 1996 apparait un SR de films [20] puis, la recommandation des pages Web.

A l'heure actuelle, les SR ne cessent d'accroître, et leur utilisation devient de plus en plus indispensable dans divers domaines.

2.3 Définition des systèmes de recommandation

La recommandation consiste en la suggestion de choses ou de solution qui sont susceptible de plaire à des personnes spécifiques.

Les systèmes de recommandation sont des systèmes de filtrage d'informations qui traitent le problème de la surcharge ces dernières en réduisant les fragments d'informations vitales d'une grande quantité d'informations générées dynamiquement en fonction des préférences, des intérêts ou du comportement observé de l'utilisateur à propos de l'élément. Le système de recommandation a la capacité de prédire si un utilisateur particulier préférerait un article ou non en fonction du profil de l'utilisateur [15].

2.4 Types des systèmes de recommandation

Les techniques de recommandation peuvent être classées en trois grandes catégories [6] : le filtrage collaboratif, les recommandations basées sur le contenu et les recommandations hybrides.

2.4.1 Filtrage basé sur le contenu :

Ce filtrage analyse les attributs des éléments (titre, nom..) pour générer des prédictions. Consiste à comparer tout nouveau document avec la base des listes qui plaisent ou déplaisent à cet utilisateur déjà enregistrées dans le but de pouvoir bien le classer dans le cadre de le suggérer ou non à ce dernier . Chaque utilisateur du système doit alors contenir un espace décrivant son centre d'intérêt. En même temps chaque utilisateur est indépendant des autres, ce qui lui permet d'avoir ses propres suggestions même s'il est le seul utilisateur du système [25] . Cela se fait par la technique d'indexation et l'intelligence artificielle. Ce système utilise les feedbacks de chaque utilisateur pour mettre à jour son profil.

On distingue deux types de recommandation par le contenu :

Recommandation basée sur les mots clés : Les mots-clés sont généralement soit extraits sur la base d'une indexation automatique, soit attribués manuellement. Par exemple, si un utilisateur a tendance à consulter des articles portant sur la géologie, il lui sera recommandé des articles ayant comme mots clefs "géo", "terre", "roches"..

Cette approche a comme inconvénient le manque d'intelligence. Par exemple , quand le mot clé est composé , le système lui recommandera des articles comportants un mot de cette combinaison, ce qui risque de sortir du contexte, ou bien l'inverse (des articles loupés car ils ne contiennent pas ce mot malgré qu'ils sont du domaine).

Recommandation basée sur la sémantique : Elle se base sur : Le type de connaissance impliquée (lexique, ontologie...), les techniques adoptées pour l'annotation ou la représentation d'items, le type de contenu inclus dans le profil utilisateur, la stratégie de correspondance entre items et profil.

Avantages :

- Prise en compte de chaque profil qui est la clé, pour trouver des recommandations pertinentes pour chacun.
- La non nécessité des données d'autres utilisateur.
- Pas de démarrage a froid.
- Possibilité de recommander aux utilisateurs ayant un goût rare ou unique.[6]

Inconvénients :

- La nécessité d'analyse du contenu pour pouvoir définir les caractéristiques de l'article.
- L'excellence du produit ne peut être estimée.
- Les images par exemple ne peuvent être représentées par un mot clé .
- Lorsqu'un nouvel utilisateur démarre dans ce système, il est sans historique.

2.4.2 Filtrage collaboratif :

Les systèmes de recommandation collaboratifs, cités dans la Figure 2.1 [32] sont les plus utilisés à l'heure actuelle. Cette approche se base sur les appréciations d'un ensemble d'utilisateurs en comparant leurs opinions(en attribuant une note aux items précédemment visités ou choisis) [15] , et en regroupant ceux qui sont en commun. Cette méthode prend en compte la liste des appréciations enregistrées par cet ensemble ,et tient compte de l'historique de notre utilisateur, afin de suggérer à notre utilisateur ce qui l'intéresse. Cette approche est regroupée en deux classes [23]

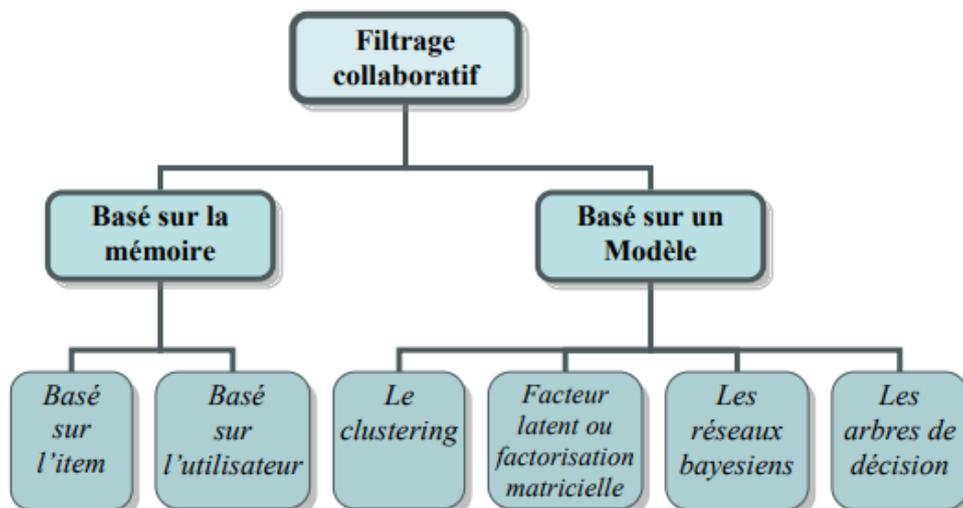


FIGURE 2.1 – Filtrage Collaboratif

2.4.2.1 Filtrage collaboratif basé sur la mémoire :

Ce filtrage utilise une matrice contenant les préférences des utilisateurs sous forme de notes dans le but de prévoir les sujets qui vont intéresser notre utilisateur. Son but est de prédire l'utilité des items pour un l'utilisateur basé sur la base des votes déjà faite. Dans cette méthode, les notes tirées par les utilisateurs sont utilisées pour prédire celles des nouveaux items, cela se fait de deux manières comme nous le montrons dans la Figure 2.2 [11] :

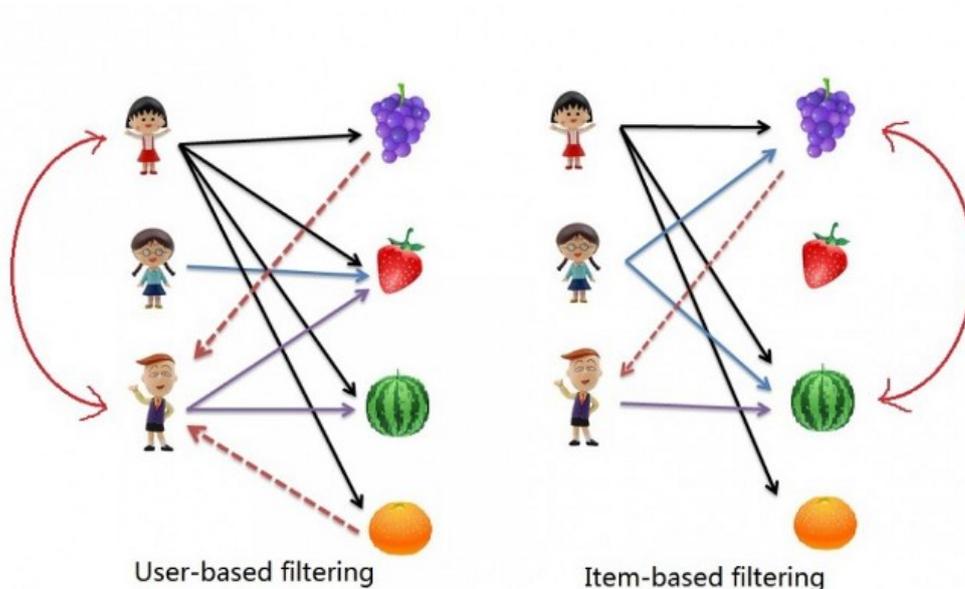


FIGURE 2.2 – User-based VS Item-based

Filtrage collaboratif basé sur les utilisateurs(User-based)

Ce type utilise les utilisateurs voisins , qui sont typiquement les utilisateurs dont les notes sur les items sont les plus proches de celui-ci. Les plus proches voisins sont les utilisateurs les plus similaires dans leur notation. Cette méthode utilise quelques mesures de similarités, dont on peut citer :

- La corrélation de pearson.
- Le cosinus des vecteurs.
- La distance de spearman.

Filtrage collaboratif basé sur les éléments(Item-based) :

Il dépend des éléments en tant qu'utilisateurs. les éléments évalués de manière comparable, sont probablement similaires.

2.4.2.2 Filtrage collaboratif basé sur un modèle :

Ce genre de modèles essayent de réduire la complexité et utilisent l'espérance de l'évaluation pour calculer la prédiction, d'une autre manière ils peuvent se baser sur une classification. Ce filtrage utilise [25] :

- Clustering : pour limiter le nombre d'utilisateurs ayant les mêmes goûts que celui à étudier au lieu de tous les vérifier lorsqu'ils ne nous intéressent pas.
- K-means : elle choisit d'abord k centres dans l'espace de représentation. Ensuite, elle met chaque utilisateur dans le cluster du centre le plus proche. puis, elle calcule la position des centres pour chaque cluster . Ces opérations se répètent jusqu'à ce que les centres ne bougent plus.
- ReckTree : ou arbre de recommandation , son but est de maximiser les similarités entre les membres d'une même clique et à minimiser celles entre les membres de deux cliques différentes..

Avantages :

- L'excellence de l'article peut être estimée par l'utilisateur (réception d'un effet de surprise par l'utilisateur) [17]
- Non nécessité de la connaissance du domaine : Ces méthodes peuvent recommander des items sans avoir besoin de comprendre leurs sens ni disposer de leurs attributs. La recommandation est basée uniquement sur les notes données aux items.
- Plus Il y a d' utilisateurs, plus le nombre de scores augmente. Cela implique un meilleur résultat.
- Le filtrage collaboratif est très utilisé par les systèmes vu ses avantages, parmi ces systèmes on peut citer : Amazon, Netflix, MovieLens, Jester, Citeseer, Phoaks, etc.

Inconvénients :

- Problème de démarrage à froid pour différents utilisateurs et nouveaux produits.
- Difficulté de trouver des groupes d'utilisateurs similaires.

Petite comparaison entre le FC et le FBC : La Figure 2.3 [12] explique globalement la différence entre le filtrage collaboratif et le filtrage basé par le contenu.

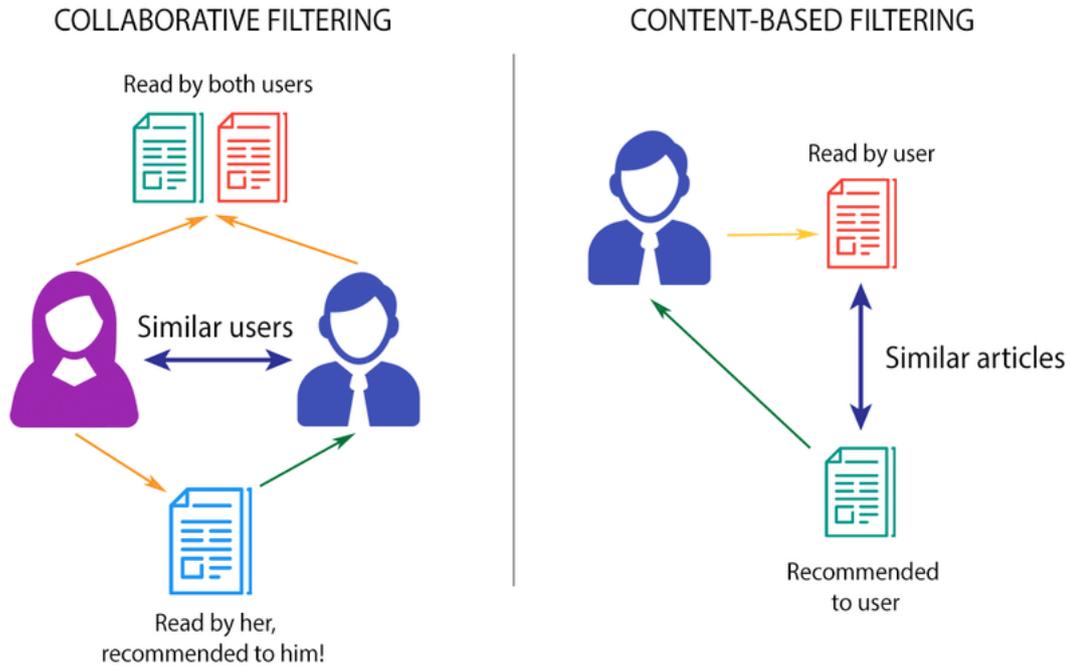


FIGURE 2.3 – FC vs FBC

2.4.3 Approches Hybrides :

Un système de recommandation est dit hybride quand il combine deux ou plusieurs approches de recommandation différentes. En combinant plusieurs systèmes de recommandation dans le but d'éviter les limitations et les soucis des systèmes de recommandation [13], nous pouvons pallier les faiblesses d'une méthode à travers les mérites d'une autre, et construire en conséquence un système plus robuste [17].

Les deux approches vues précédemment sont souvent considérées comme étant complémentaires, car les approches basées sur le contenu ont l'avantage de pouvoir recommander les nouveaux items non encore évalués par un utilisateur, alors que le filtrage collaboratif ne peut recommander un item que s'il a été noté par un certain nombre d'utilisateurs auparavant, de plus, l'approche basée sur le contenu nécessitent de disposer des attributs des items, en plus d'une étape d'analyse pour pouvoir les extraire et les représenter, alors que le filtrage collaboratif n'utilise pas le contenu des items pour faire de la recommandation.

2.5 Domaines d'utilisation des systèmes de recommandation

Aujourd'hui de nombreux secteurs utilisent les systèmes de recommandation, notamment dans le secteur de l'industries culturelles, et de commerce électronique .On peut donc citer parmi eux ceux-là :

2.5.1 Recommandation des produits

Les algorithmes de recommandation les plus connus sont ceux utilisés sur des sites web de commerce électronique. En effet, chez les commerçants en ligne comme Amazon, l'utilisation des systèmes de recommandation est indispensable. Nous avons remarqué que ces derniers s'efforcent de présenter et suggérer à chaque utilisateur des recommandations de produits qu'il est susceptible d'acheter, ils personnalisent la boutique en ligne, pour chaque client. Le magasin change radicalement en fonction des intérêts de ce dernier. Cette méthode marketing leur permet d'augmenter leur chiffres d'affaires jusqu'à 30% .

2.5.2 Services de streaming

Incluant le cinéma avec des plateformes comme Netflix ,Apple tv qui ont pour objectif de recommander des films et séries qui pourrait plaire à leur clients ,Ces recommandations sont basées sur les évaluations fournies par les utilisateurs.

2.5.3 Recommandation d'articles

Les services d'actualité recommandent des articles aux lecteurs, sur la base des articles qu'ils ont lus dans le passé. La similitude peut être basée sur la similitude des mots importants dans les documents, ou sur les articles qui sont lu par des personnes ayant des goûts de lecture identiques. Les mêmes principes s'appliquent pour recommander des vidéos YouTube ou d'autres sites où le contenu est fourni régulièrement.

2.5.4 Services

Ceci englobe toute sorte de services qui peuvent être recommandés aux clients tel que les services de voyage (Tripadvisor), maisons à louer, etc.

2.5.5 Healthcare

La recommandation dans le domaine de la santé est d'un rapport scientifique, généralement ,sur un thème médical faisant la liste des actes, gestes et même parfois, de traitements médicaux qu'il est recommandé de réaliser dans un contexte donné. la recommandation dans le healthcare , a été étudiée et faite dans plusieurs domaines, la ou nous pouvons citer la recommandation de nutrition pour les personnes diabétiques [18] .

Il y a aussi d'autres domaines ou les systèmes de recommandations sont fortement utiliser comme les réseaux sociaux tel que Instagram, Tik tok, le secteur de la musique comme Spotify.. etc

2.6 Conclusion

Dans ce deuxième chapitre, nous avons introduit la notion des systèmes de recommandation et leur l'historique. Ensuite, nous avons défini les différents types de ces systèmes ,ainsi que leurs avantages et inconvénients. Enfin, nous avons cité les différents domaines d'application suivis des exemples.

Dans le chapitre suivant, nous aborderons l'état de l'art pour présenter et analyser les travaux existants liés aux SR basés sur les approches utilisées.

Chapitre 3

État de l'art

3.1 Introduction

Dans l'univers digital actuel, la santé demeure l'un des domaines qui est en pleine émergence dans l'industrie du numérique. Selon les organisations internationales, pour améliorer la qualité des systèmes de santé, particulièrement ceux des pays en développement, il faudrait utiliser les outils développés pour affronter la complexité de l'environnement et les défis qui en découlent. Parmi ces outils, l'avancée des technologies de l'information et de la communication (TIC) permet la mise en place, dans une approche systémique, d'un système d'information sanitaire intégré.

La e-santé apparaît de plus en plus comme une solution pertinente pour répondre aux défis que doivent relever les systèmes de santé. Ces derniers doivent analyser un vaste nombre de données sur les patients fournissant des informations et des prédictions sur de diverses maladies.

Ces systèmes doivent être intelligents et capables de prédire l'état des patients en fonction de leurs modes de vie, les dossiers de santé physique et activités sociales.

Sachant que les gens ont tendance à utiliser les TIC pour en savoir plus sur leur santé, les systèmes de santé intelligents sont importants pour mettre en œuvre des recommandations diagnostiques, assurance maladie, et la médecine alternative basée sur l'état de santé des malades.

Dans cette partie du mémoire , nous allons présenter un état de l'art des principales

contributions des soins de santé, plus spécifiquement l'obésité, liés à la nutrition. Vers la fin du chapitre, nous allons faire une étude comparative de ces contributions permettant de cibler les principales différences entre les approches proposées.

3.2 Travaux connexes

Jung et al [22] ont présenté une étude des recommandations nutritionnelles basées sur les connaissances pour la prise en charge de l'obésité. Ces recommandations comprennent non seulement des données nutritionnelles diététiques statiques, mais également des menus diététiques individualisés en utilisant des données contextuelles basées sur les connaissances à travers une méthode du FC. Cette méthode utilise les informations de base des jeunes obèses, forme un regroupement de similarités avec une forte corrélation, applique le poids de similarité sur la matrice user-menu dans le regroupement de similarités et utilise le filtrage collaboratif basé sur les connaissances pour recommander le menu nutritionnel diététique. En utilisant également la modélisation contextuelle basée sur les connaissances, l'étude constitue une matrice de fusion menu utilisateur afin de résoudre le problème de la rareté. De cette façon, les utilisateurs recevront des menus personnalisés et préférés de manière omniprésente (quelque soit le temps et quelque soit le lieu) via des appareils mobiles.

Le chercheur avait développé une opération de service mobile de la gestion diététique pour le contrôle de l'obésité chez les adolescents en utilisant la méthode proposée ici. L'application développée a été vérifiée pour ses performances de recommandations nutritionnelles diététiques basées sur les connaissances des jeunes obèses.

Une excellente précision supérieure à la moyenne de 80 % a été trouvée dans la recommandation de menu pour les utilisateurs.

Cette application s'est également avérée avoir un niveau de satisfaction positif de ses utilisateurs sur la qualité du service.

Le problème clairsemé du filtrage collaboratif a été résolu avec succès avec cette étude pour augmenter la précision des recommandations de menu.

Cependant, elle devrait être efficace pour prévenir l'obésité et éviter également les pertes socioéconomiques. Donc il a été jugé nécessaire de disposer d'une plateforme ouverte de médecine préventive clinique pour une approche thérapeutique de la prévention de l'obésité juvénile.

El-Dosuky et al [19] ont proposé un système de recommandation avec un cadre de personnalisation alimentaire en utilisant l'approche des hypermédias adaptatifs. Ils ont étendu le framework Hermès avec une fonctionnalité de recommandation alimentaire. Ils combinent la méthode d'extraction de termes TF-IDF avec la mesure de similarité cosinus. Pour développer un biais en faveur d'une alimentation saine, des heuristiques nutritionnelles examinées sont collectées et une base de données standard sur les aliments sont intégrées à la base de connaissances, La base de connaissances est une ontologie de domaine composée de classes, de relations et d'instances de classes. Sur la base de l'évaluation effectuée, nous concluons que les systèmes de recommandation sémantique surpassent en général les systèmes de recommandation traditionnels en termes d'exactitude, de précision et de rappel, et que la recommandation proposée à une meilleure mesure F que les recommandations sémantiques existantes, et que l'algorithme proposé obtient le meilleur score de précision car la plupart des aliments recommandés sont pertinents.

Leipold et al [27] ont conçu un système de recommandations nutritionnel Nutrilize qui combine des recommandations de recettes personnalisées, un retour visuel et d'autres mesures persuasives, en prenant en compte les caractéristiques personnelles et l'état nutritionnel de 26 macro et micronutriments.

Dans cet article, ils ont présenté les caractéristiques du système Nutrilize ainsi que son étude pilote. Ils ont analysé l'interaction et la perception de leur système sur une période de 21 jours en se servant des données de 14 participants.

Dans leurs système de recommandation nutritionnel Ils ont étudié les aspects nutritionnels personnalisés et basé sur les connaissances.

Ils ont utilisé la base de données BLS (Bundes Lebensmittel Schluessel) pour enregistrer l'apport de l'utilisateur ainsi que pour calculer le profil nutritionnel des recettes. Au cours de l'étude pilote, 26 micro et macronutriments différents ont été dérivés du BLS pour l'apport de l'utilisateur et le profil des recettes. Ils ont fixé une moyenne de 3 jours pour déterminer le profil nutritionnel des utilisateurs. Les recettes utilisées sont obtenues à partir de Koch Wiki1 der , ils ont combiné la base de données de cette dernière avec les informations nutritionnelles pour chaque aliment dans la base de données BLS .

Dans le but de collecter les informations régulières sur le comportement nutritionnel de l'utilisateur, ils se sont servi d'un suivi via le journal alimentaire personnel intégré fourni par les informations nutritionnelles de la base de données BLS.

Ils ont utilisé un jeu de trois couleurs qui servira d'avertissement, attention, continuité ou autre

après les 3 jours de surveillance de consommation, des flèches indiquantes le comportement recommandé, boutons circulaires pour un accès facile au journal afin d'ajouter de nouveaux repas.

Les recommandations de recettes sont sous forme de liste pour chaque utilisateur, si ce dernier veut savoir pourquoi exactement cette recette lui a été recommandée, il n'a qu'à cliquer sur plus d'informations afin d'avoir des explications. En général, ils se sont concentrés sur le groupe d'étude, l'interaction du système, le système perception et comportement alimentaire déclaré. Leur procédure d'étude consistait à faire participer une centaine de personnes en leur attribuant un questionnaire à remplir sur les antécédents. Ensuite, un lien leur sera envoyé pour refaire le début qui servira de confirmation ou de mise à jour sur le comportement alimentaire. En troisième lieu, l'application Nutrilize sera envoyée à tous les participants. Au final, après 3 semaines d'utilisation, les participants vont recevoir l'enquête finale et demander des commentaires sur le système.

L'un des principaux défis qu'ils ont pu tirer des résultats est la convivialité, ce qui était évident lorsqu'il s'agit d'un système prototype. La recommandation devrait être améliorée dans certains cas, les utilisateurs sont confrontés à de nombreuses contraintes qui n'ont pas été modélisées par faute de disponibilité de certains aliments dans leurs bases utilisées. De plus, les utilisateurs ont signalé le manque de certaines fonctionnalités. Par exemple, un suivi d'activité physique manuellement au lieu d'un questionnaire, ainsi que la présentation de l'interface de leur application semble décourageante, ce qui les a poussé à envisager d'améliorer cette perception. Malgré ces lacunes, 43 % des utilisateurs ont déclaré qu'ils utiliseraient fréquemment le système, ce qui montre que l'idée générale et l'objectif du système sont pertinents pour le groupe d'étude. Enfin, certains défis majeurs restent ouverts, tels que l'intégration des informations contextuelles et sociales ainsi que l'exactitude des données d'entrée reçues.

Brink et al [31] ont proposé dans leur article, une perspective personnalisée, dynamique et systémique sur les interactions des déterminants clés et des conseils de coaching sur l'obésité. Ils ont présenté une description du développement d'un programme de coaching intégratif personnalisé en obésité à long terme avec un caractère hautement adaptatif, ciblant les adultes, qui est basé sur une approche de modélisation pour un groupe de patients. Au final, le programme adaptatif sera évalué.

Le but de cette étude était d'utiliser une vision systémique sur le surpoids, pour développer un programme de recherche et de coaching intégratif personnalisé sur cinq ans afin d'ajouter la notion du temps contrairement aux autres travaux qui n'étaient pas assez durables.

Un modèle de santé systémique de l'obésité a été utilisé afin d'identifier les relations causales des variables ayant le plus d'influence sur l'obésité. Le modèle a aidé à concevoir un programme personnalisé de coaching intégratif en obésité et à identifier les variables clés pour suivre les progrès et ajuster le programme personnalisé, en fonction des objectifs et des besoins du participant. Il a été décidé d'utiliser le sous-typage des participants par un biologiste des systèmes, qui se base sur les symptômes de la médecine traditionnelle chinoise, comme nouvelle méthode pour personnaliser l'intervention. L'approche utilisée dans cette étude est l'approche collaborative basée sur une vision systémique de l'obésité. Elle leur a permis de développer un programme personnalisé et adaptatif de cinq ans de coaching et de recherche sur l'obésité.

Rehman et al [29] visent la sélection d'un régime alimentaire approprié qui doit répondre aux besoins nutritionnels des patients .

Ils ont présenté un système de recommandation alimentaire basé sur le cloud, appelé Diet-Right , pour des recommandations diététiques basées sur les rapports pathologiques des utilisateurs. Donc l'utilisateur doit indiquer les résultats de ses tests pathologiques déjà faits, ensuite, le système les compare aux plages normales déjà existantes dans la base de données . Le modèle utilise un algorithme de colonie de fourmis (Ant Colony Optimization) pour générer une liste d'aliments optimale et recommande des aliments appropriés en fonction des valeurs des rapports pathologiques saisies par l'utilisateur. Dans leur application, ils se sont servis de l'approche "méta heuristique ACO" , car elle a été jugée comme l'approche la plus puissante pour la résolution de problèmes d'optimisation combinatoire . Diet-Right gère et met à jour les informations heuristiques de manière à maximiser la diversité des aliments.

Ils ont utilisé une base de données de 345 rapports de tests pathologiques pour catégoriser diverses maladies qui surviennent en raison de l'écart par rapport aux plages normales de composés/paramètres. De plus, à travers ces dernières par rapport aux plages normales, le système génère un plan de régime qui vise à couvrir ces anomalies.

Il a été observé qu'un nombre croissant de fourmis fait converger la solution vers son coût minimum, mais, il n'est pas possible d'utiliser un nombre élevé de fourmis. De plus, l'utilisation d'un nombre élevé de fourmis pour contracter une solution augmente la complexité temporelle. Les résultats expérimentaux montrent que par rapport à l'exécution en nœud unique, le temps de convergence de l'exécution parallèle sur le cloud est environ 12 fois plus faible. De plus, une précision adéquate peut être obtenue en augmentant le nombre de fourmis.

3.2.1 Tableau comparatif

Nous avons établi un tableau comparatif des différentes méthodes utilisées dans chacun des travaux lus. Ce tableau résume les résultats obtenus des approches proposées, ainsi que les principaux travaux menant à l'implémentation des composants des systèmes de recommandations.

La Table 3.2 figurée ci-dessous est constituée de 7 colonnes comportant les caractéristiques suivantes :

- 1ere colonne "catégorie" représente le type de recommandation pour laquelle appartiennent ces approches (recommandation de recettes, de médicaments, d'activités physiques...).
- 2eme colonne : "Approche" représente l'approche elle même (l'auteur ou les auteurs avec l'année d'édition).
- 3eme colonne : "Titre", c'est le titre de l'article en question.
- 4ème colonne "Source de données" indique les données en entrée.
- 5eme colonne : "Sorties" indique les/le résultat de l'approche.
- 6eme colonne "Technique utilisées" : explicite toutes les techniques utilisées pour aboutir au résultat (les types de filtrages, les outils de calculs de similarités...).
- 7eme colonne "Outils utilisés" : indique le nom de l'outil logiciel utilisé dans le cas où l'approche a été implémentée.

Catégorie de l'approche	Approche	Titre	Source de données (données en entrées)	Sorties (le résultat de l'approche)	Techniques utilisées	Outils utilisés
recommandation alimentaire	[El-Dosuky, M. A., et al,2012]	Food Recommendation using Ontology and Heuristics	*Heuristiques nutritionnelles *Une base de données sur les aliments .	recommandation des produits alimentaires sains et pertinents	*Approche des hypermédias adaptatifs *Combinaison de la méthode d'extraction de termes TF-IDF avec la mesure de similarité cosinus	extension du framework Hermès avec une fonctionnalité de recommandation alimentaire

recommandation de recettes	[Leipold, Nadja, et al, 2018]	Nutrilize a Personalized Nutrition Recommender System : an Enable Study.	<p>*Koch Wiki1 der dérivés de la base de données BLS (Bundes Lebensmittel Schluessel)</p> <p>*site Koch Wiki1 der(pour les recettes)</p> <p>*les caractéristiques personnelles et l'état nutritionnel de 26 macro et micronutriments</p>	<p>*classement des recettes par repas en fonction des notes.</p> <p>*recommander les recettes aux utilisateurs</p> <p>*une explication sur le nutriment dédié.</p>	<p>*Matomo, 4 anciennement Piwik.</p> <p>*échelle d'utilisabilité du système (SUS)</p> <p>*la méthode de saisie FFQ</p> <p>* filtrage collaboratif basé sur la mémoire</p>
----------------------------	-------------------------------	--	--	--	--

recommandation de menus	[Jung, Hoill, and Kyungyong Chung, 2016]	Knowledge-based dietary nutrition recommendation for obese management.	Données statiques et sur la base de données contextuelles	recommandation nutritionnelles diététiques et horaire spécifiques à l'utilisateur	filtrage collaboratif et clustering	Application mobile Health Mate
	[Brink, Sander M., et al.,2022]	Developing a Personalized Integrative Obesity-Coaching Program : A Systems Health Perspective.	*les symptômes de la médecine traditionnelle chinoise *médecins spécialistes, des psychologues cliniciens, des diététiciens, des physiothérapeutes et psychomotriciens et des coachs de style de vie	*perspective personnalisée, dynamique et systémique sur les interactions des déterminants clés. * conseils de coaching sur l'obésité *un programme de recherche et de coaching personnalisé sur cinq ans	*Approche collaborative *diagramme de boucle causale (CLD)	-logiciel MARVEL

recommandation alimentaire	[Rehman, Faisal, et al. 2017]	Diet-right : A smart food recommendation system.	*les résultats des tests pathologiques des utilisateurs (Une matrice de 345 entrées) *Ensemble de données	*Recommander une liste d'aliments optimaux. *suggérer des régimes selon les besoins des utilisateurs.	Algorithme basé sur l'optimisation des colonies de fourmis (ACO)
----------------------------	-------------------------------	--	--	--	--

TABLE 3.2 – Etat de l’art de quelques contributions relatives aux systèmes de recommandation des soins de la santé.

3.3 Discussion et comparaison

Les travaux étudiés nous ont permis d’analyser les performances de différentes techniques et méthodes de conception de SR, à travers les résultats de l’expérience en termes de précision, rappel et d’exactitude.

Nous avons remarqué que la majorité des travaux basés sur la recommandation nutritionnelle utilisaient généralement le filtrage collaboratif en premier lieu, les deux autres approches les plus courantes sont celles basées sur le contenu et hybride.

Cependant, l’utilisation des techniques de filtrage collaboratif peut entraîner le problème du démarrage à froid qui peut concerner les nouveaux utilisateurs (user cold start), ainsi que les nouveaux items car celui-ci ne peut pas être recommandée sans qu’il ne soit noté par certains utilisateurs.

Concernant les différentes méthodes utilisées dans le cadre des systèmes de recommandation, on cite :

- La mesure de similarité entre utilisateurs (user-user), entre éléments (item-item), ou entre

utilisateurs et éléments (user-item). En se basant sur le calcul du cosinus, ou bien en utilisant le coefficient de corrélation de Pearson.

- La classification des utilisateurs ou des éléments (Classification supervisée et non-supervisée).
- Via d'autres techniques (TF-IDF).

Enfin, dans chacun des systèmes, un ou plusieurs jeux de données ont été utilisés en fonction de l'approche étudiée.

Nous avons étudié diverses approches à partir des articles lus et cités auparavant. Ainsi que les articles lus et non cités. Le tableau présenté précédemment montre une comparaison partielle de ces derniers.

On note qu'en 2012 [19], ils ont utilisé l'approche des hypermédias adaptatifs dans le but de recommander des aliments sains et pertinents, la technique utilisée par les chercheurs en question était la combinaison entre l'extraction de termes TF-IDF et le calcul de similarité cosinus. Cette méthode offre une bonne précision en termes de recommandation.

En 2016, d'autres chercheurs ont trouvé que l'utilité du filtrage collaboratif était plus pratique dans le domaine de la recommandation, plus spécifiquement la recommandation dans le domaine des soins de la santé, ce qui les a incité à l'adapter et l'améliorer en utilisant le clustering. Cette méthode les a aidé à rajouter la notion du temps, cela dit leur recommandation ne s'arrêtait pas qu'à la nutrition diététique, mais aussi, elle consistait à l'adapter à des fuseaux horaires spécifiques à l'utilisateur.

Un an après, une recommandation d'une liste d'aliments optimaux a vu le jour. L'approche utilisée a comme particularité l'optimisation qui a été développée en se servant de l'algorithme méta-heuristique des colonies de fourmis (ACO). Diet-Right gère et met à jour les informations heuristiques de manière à maximiser la diversité des aliments.

En 2018, on a vu l'apparition de l'application Nutrilize, qui consistait à recommander des recettes en les classant par repas. Dans leurs cas, ils ont utilisé le feedback explicite du filtrage collaboratif en se servant des notes déjà attribuées. Cette application souffrait de quelques lacunes, par exemple, le manque de disponibilité d'activités physiques. Ceci a été l'une des raisons qui a exhorté d'autres chercheurs [31] à développer une perspective personnalisée, dynamique et systémique pour un programme de recherche et de coaching personnalisé sur 5 ans en utilisant encore une fois le filtrage collaboratif en l'adaptant avec le diagramme de boucles causales.

Après avoir discuté quelques travaux, on a remarqué que le filtrage le plus utilisé est le filtrage collaboratif, car c'est le plus adapté dans le domaine des recommandations du healthcare. Pour

cela, nous avons décidé de reprendre ce dernier en l'améliorant afin de l'assimiler à notre propre approche.

Dans ce cas, on peut tirer qu'à partir du nombre de techniques d'évaluation existantes, tout système de recommandation peut être amélioré et peut gagner en utilité en améliorant le support de l'IA et du ML, ainsi, de l'évaluer en utilisant différentes métriques afin d'avoir les meilleures résultats.

3.4 Conclusion

Dans ce chapitre, nous avons établi un état de l'art, qui consistait dans notre cas , à résumer quelques articles déjà lus sur les systèmes de recommandation du domaine des soins de la santé, et nous avons spécifié et analysé l'utilité de ces systèmes dans chaque application en fonction de leurs approches.

Ensuite, nous avons établi un tableau comparatif entre les différentes techniques et approches utilisées dans chaque article résumé.

Enfin,nous avons terminé ce chapitre, avec une discussion générale autour de notre état de l'art, ce qui nous a permis de constater les elements essentiels d'un SR plus pertinent .

Nous allons présenter dans le chapitre suivant la contribution bien détaillée de notre système de recommandation.

Chapitre 4

Contribution

4.1 Introduction

L'obésité est une maladie chronique qui touche actuellement 60% des adultes et 25 % des adolescents et des enfants. Ce chiffre ne cesse d'être en constante augmentation , comme il pourrait toucher selon les chercheurs jusqu'à 80% des adultes.

Cette surcharge pondérale augmente le risque de maladies cardiovasculaires, de diabète, d'hypertension et de certaines formes de cancer. Cela étant dit, elle est mortelle.

Dans notre mémoire, nous soulignons la question de la sélection d'un régime alimentaire approprié qui doit répondre aux besoins nutritionnels des patients. Pour résoudre ce problème, nous présentons un système de recommandation alimentaire, pour des recommandations diététiques basées sur les rapports pathologiques des utilisateurs.

Dans ce chapitre, nous présentons et expliquons l'approche proposée en détails ainsi que son architecture .

4.2 Approche proposée

Dans cette partie nous présentons notre approche, de création d'un système de recommandation personnalisé dans le healthcare , illustrée dans la figure 4.1, pour assister les personnes en surpoids et/ou obèses à mieux gérer leur nutrition au quotidien en utilisant le filtrage collaboratif basé sur l'utilisateur.

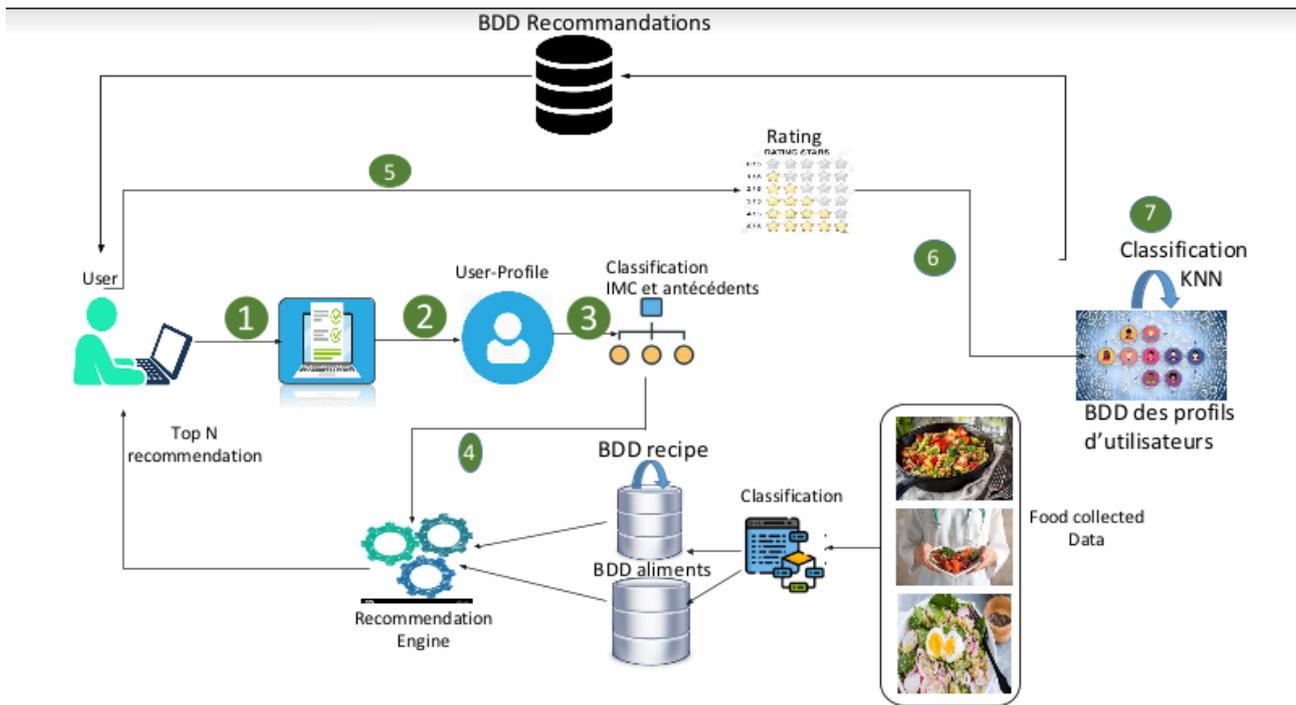


FIGURE 4.1 – Approche proposée

L'architecture de notre système de recommandation consiste en les étapes suivantes :

La première étape consiste en la création du profil utilisateur, qui se fait en remplissant un formulaire d'inscription. Cette étape nous permet de récupérer les informations de l'utilisateur afin de les utiliser dans les prochaines étapes.

La deuxième étape classe l'utilisateur en fonction de son imc et ses antécédents par types d'obésité .

Dans la troisième étape ,le système recommande à l'utilisateur des recettes ou aliments qui lui correspondent le mieux en utilisant la classification précédente .

La quatrième étape se résume en ce que l'utilisateur attribue une note à la recette qui lui a été recommandée.

La cinquième étape fait la recherche de profils similaires, en utilisant de la mesure de similarité cosinus pour trouver les profils qui sont similaires a celui qui vient d'être créé en le comparant (par rapport aux ratings et aux informations déjà collectées) avec les utilisateurs déjà existants.

Ensuite la classification en utilisant l'algorithme kNN pour regrouper les profils les plus similaires , et les recettes nutritionnelles en fonction de leurs calories.

Enfin, le processus de recommandation qui se refait en boucle, à l'aide des étapes citées précédemment.

Dans ce qui suit nous allons présenter les étapes de notre approche en détails :

4.2.1 Création du profil utilisateur :

La création du profil utilisateur se fait à l'aide des informations collectées a partir du formulaire d'inscription contenant le nom, prénom, âge ,poids , taille et antécédents pathologiques comme données explicites , ainsi que l'IMC comme donnée implicite qui se calcule directement en utilisant le poids et la taille de l'utilisateur : $IMC = \text{poids (kg)}/\text{taille}^2(\text{m}) \dots (1)$

Ensuite,après avoir créé le profil, il sera directement stocké dans la BDD utilisateurs, et on lui recommandera des recettes qui lui corespondent par rapport aux criteres collectés précédemment. Par la suite,ce dernier attribuera une note a ces recommandations .

Après avoir stocké les infos dans la bdd, on procède à la vectorisation des informations afin qu'elles puissent être reconnue et utilisées dans le Machine learning.

4.2.2 Recherche de profils similaires :

La deuxième étape consiste à trouver un profil similaire à celui créé dans l'étape précédente et ceci en le comparant a tous les profils voisins existants dans la BDD en se basant sur les calculs de similarité. Nous allons l'appliquer à l'aide des notes déjà attribuées par chaque utilisateur pour les recettes précédemment consultés , cela etant dit , les appreciations des utilisateurs en tant que feedback explicite, l'évaluation implicite en se servant du timestamp qui represnte le temps de consulation d'une recette ou aliment par l'utilisateur ; Ainsi que leurs IMC et leurs pathologies.

User/Recipe	Recipe1	Recipe2	Recipe3	Recipe4
User1	5	3	5	4
User2	5	2	2	1
User3	3	2	5	2
User4	2	4	0	1
User5	3	5	4	3

TABLE 4.1 – Feedback explicite

Il existe plusieurs métriques qui nous permettent de calculer la similarité entre les profils. Dans notre cas ,nous avons choisis celle de cosine similarity qui est le produit scalaire des vecteurs divisé par leur amplitude (l'angle entre les vecteurs) en utilisant l' IMC et les antécédents de chaque utilisateur . Plus la distance entre deux profils ou deux objets est petite, plus la similarité est grande.

- Norme d'un vecteur T de dimension n :

$$|T| = \sqrt{T_1 * T_1 + T_2 * T_2 + \dots + T_n * T_n} \dots (2)$$

- Produit scalaire de deux vecteurs T et U de dimension n :

$$T \cdot U = T_1 * U_1 + T_2 * U_2 + \dots + T_n * U_n \dots (3)$$

-Similarité du cosinus de T et U....(4)

$$\text{simCos}(T,U) = (T \cdot U) / (|T| * |U|) \dots (5)$$

Nous avons choisi le cosine similarity pour sa faible complexité et,sa simplicité. Donc en utilisant cette métrique , ça nous permet d'avoir un gain de temps et rend la recommandation plus rapide en terme de recherche.

4.2.3 Apprentissage supervisé

L'apprentissage supervisé est le paradigme d'apprentissage le plus populaire en Machine Learning et en Deep Learning. Comme son nom l'indique, il consiste à superviser l'apprentissage

de la machine en entraînant un modèle, basé sur des données préalablement labellisées. Ainsi, résoudre la majorité des problèmes de ML. Les domaines d'applications de ce mécanisme sont nombreux, dont la classifications ,la régressions, la vision par ordinateur, la reconnaissance vocale. . .

4.2.3.1 Classification :

La classification est l'une des techniques de l'apprentissage supervisé , qui permet de prédire la catégorie des points discrets déjà labellisées dans le but de les classer dans un certain nombre de classes ou de groupes.

Dans notre cas, nous avons choisi la methode KNN comme le montre la Figure 4.2, car nos données sont labellisées.Cet algorithme est le plus adapté à notre modèle en termes de taille de dataset,et en termes de réponse à nos besoins.Il est jugé comme étant l'un des plus simples et robustes algorithmes de ML supervisé.

Le KNN développe une fonction qui prédit avec précision la sortie à partir des variables d'entrée (celles déjà saisies par les utilisateurs) , afin de classer les utilisateurs en fonction de leurs IMC, antécédents pathologiques, et leurs préférences . Cette étape permet de réduire l'espace des profils en les rassemblants.

Cette méthode utilise principalement deux paramètres : une fonction de similarité pour comparer les individus dans l'espace de caractéristiques, et le nombre k de voisins qui influencent la classification .

Le fonctionnement de KNN peut être expliqué sur la base de l'algorithme ci-dessous [10] :

- La sélection du nombre k de voisins.
- Le calcul de la distance.
- La sélection des k voisins les plus proches selon la distance calculée.
- Compter le nombre de points appartenant a chaque catégorie parmi ses k voisins.
- Attribuer le nouveau point à la catégorie la plus présente ou fréquente parmi ses voisins .
- Le modèle est prêt.

Cette méthode nous permet de gagner du temps et d'accélérer le processus de recommandation , ceci en rassemblant les utilisateurs ayant un intérêt similaire dans les mêmes groupes afin d'éviter de comparer chaque nouvel utilisateur arrivant avec tous les utilisateurs du système. En effet,il existe plusieurs méthodes pour calculer la similarité entre deux vecteurs, celle que nous avons choisis est la similarité cosinus .

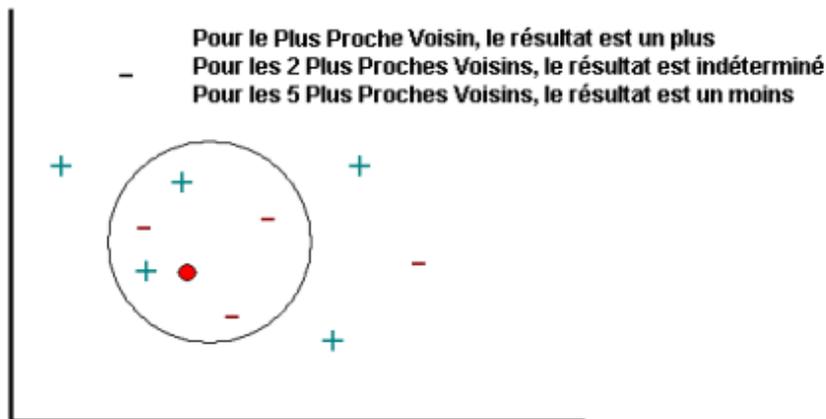


FIGURE 4.2 – K-plus proche voisin

4.2.3.2 La similarité Cosinus

La fonction Cosinus est utilisée pour calculer la Similitude ou la Distance des observations dans un espace de grande dimension, comme le montre Figure 4.3 .

Si deux ensembles d'attributs sont identiques, notre fonction de similarité renvoi nsembles d'attributs sont complètement différents, la fonction renvoie (0). S'il y a un chevauchement significatif entre les deux ensembles, elle renvoie un nombre réel plus proche de 1 (ex : 0,95) et vice versa, les deux ensembles d'éléments les plus dissemblables se rapprochent de 0 (ex : 0,05). Heureusement , les valeurs positives du cosinus présentent ce comportement :

Cosinus (0) = 1,0 -angle de 0 degrés

Cosinus($\pi/4$) = 0,7 -angle de 45 degrés

Cosinus($\pi/2$) = 0.0 -angle de 90 degrés

Si nous voulons comparer la similarité des points, nous représentons d'abord chaque entité comme un vecteur dans un espace à N dimensions, puis nous calculons la valeur du cosinus de l'angle entre ces deux vecteurs. Ainsi, cette valeur de cosinus peut être facilement calculée comme le produit scalaire des deux vecteurs divisé par le produit de leurs grandeurs.

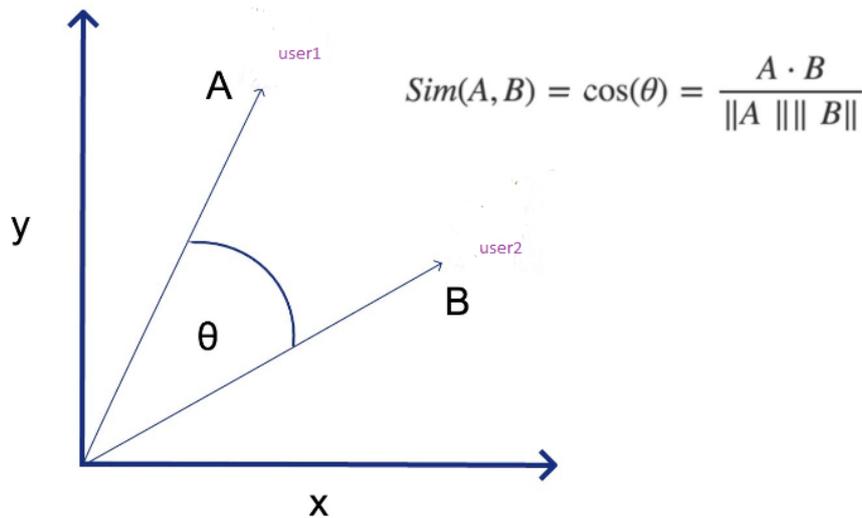


FIGURE 4.3 – Similarité de cosinus

Avec cosine similarity, même si deux utilisateurs sont éloignés par rapport à la distance euclidienne en raison de leur taille, ils peuvent toujours avoir un angle plus petit entre eux. Plus l'angle est petit, plus la similitude est grande.

4.2.4 Recommandation :

En premier lieu, le moteur de recommandation reçoit en entrée les informations collectées du profil de l'utilisateur y compris l'IMC et les pathologies afin de le classifier par types d'obésité, et lui renvoie par la suite les recommandations correspondantes à ces critères. L'utilisateur attribue une note à la recette qui lui a été proposée. Cette étape est ensuite enregistrée dans la bdd des profils utilisateurs, puis elle est utilisée pour calculer la similarité entre cet individu et les individus déjà classés afin d'effectuer une classification.

Après avoir comparé le nouveau profil arrivant avec les profils déjà enregistrés, l'utilisateur sera associé à un autre profil ou une catégorie dont il a été jugé plus proche voisin, en utilisant le KNN et la mesure de similarité cosinus. La bdd des recommandations va ensuite intervenir en affectant à l'utilisateur les mêmes recommandations déjà attribuées à ses profils similaires.

Ajoutant à ceci, que nous avons utilisé deux bdd afin de pouvoir recommander des recettes complètes ainsi que des aliments diététiques.

Pour résumer, L'algorithme suivant illustre les étapes clés de conception d'un système de recommandation :

Algorithme de recommandation

——- Algo : Conception du SR ; ——-

Entrée :

— Données de l'utilisateur.

Sortie :

— Recommandation de Recettes.

Debut

1- Si formulaire <>vide alors :

enregistrement des données et création du profil de l'utilisateur

Sinon écrire('veuillez remplir le formulaire pour votre inscription');

2-Comparer le profil avec les profils existants dans la BDD

3-Si Cosine similarity<>0 alors

4- Procédure de Recommandation

Sinon recommandation personnalisé

5-Afficher les résultats

Fin.

Le diagramme illustré dans la Figure 4.4 , représente une vue globale des activités des autres systèmes :

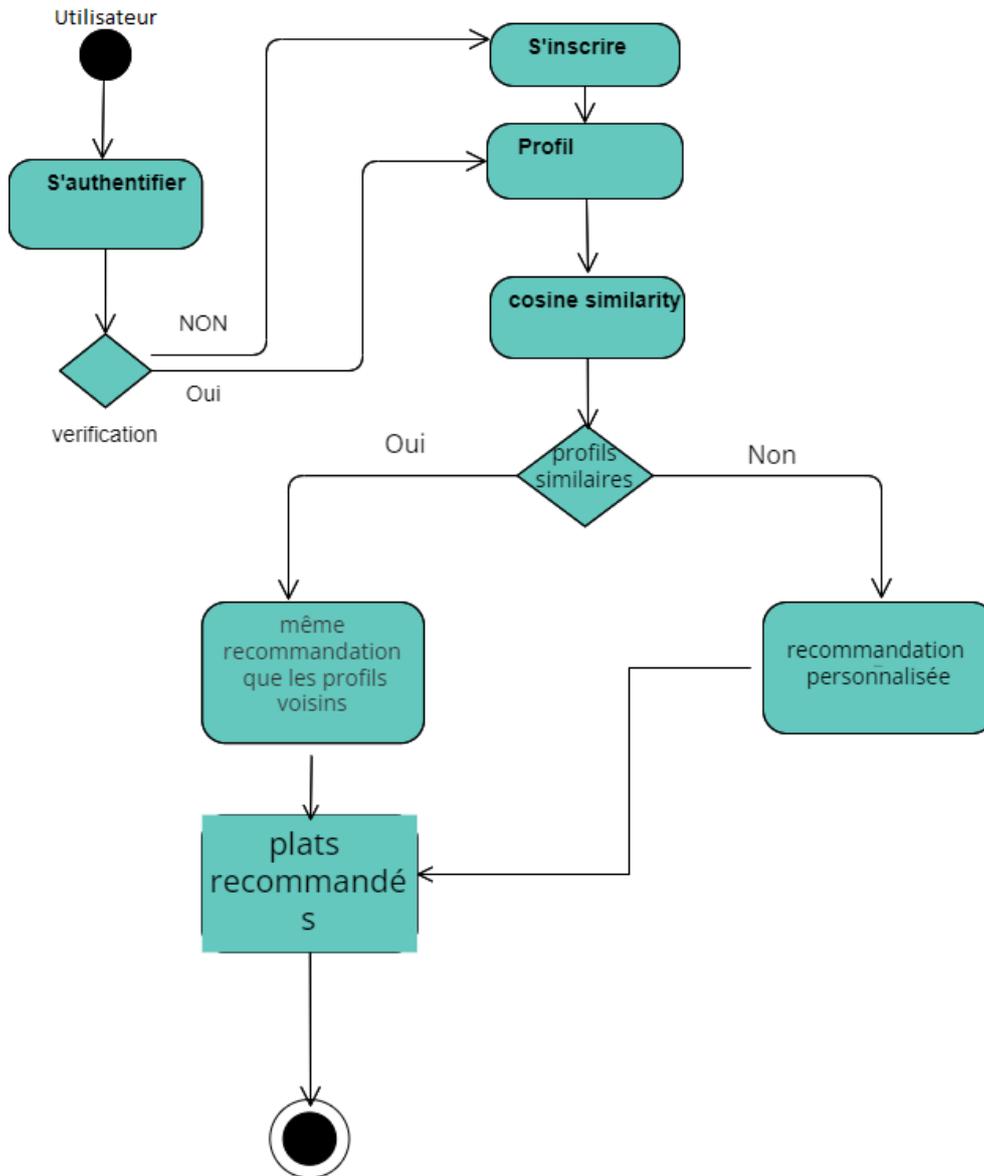


FIGURE 4.4 – Diagramme d'activité

- Inscription : l'utilisateur doit tout d'abord s'inscrire pour que le système crée son profil si cette étape est déjà faite il pourra s'authentifier directement.
- Authentification : l'utilisateur doit obligatoirement se connecter à son compte s' il en possède un sinon il devra en créer un en s'inscrivant.
- Profil : Après inscription, le système crée le profil de l'utilisateur à partir des informations récupérées pendant l'inscription. Si cette dernière a déjà été faite il accédera directement à son profil.
- Similarité de cosinus : cette étape est crucial pour la recommandation . La similarité cosinus est la mesure du calcul de la différence d'angle entre deux vecteurs. Nous utilisons

cette métrique pour calculer les plus proches voisins pour qu'on puisse trouver les profils similaires au nouvel utilisateur. Ceci peut mener a deux cas :le premier est de trouver un profil similaire ainsi il lui recommandera la meme chose , le deuxième est de ne pas en trouver un.

- Recommandation : le système de recommandation recommandera à l'utilisateur des plats et recettes saines et équilibrées en fonction de ce qui lui conviendra le mieux.

4.3 Conclusion

Au cours de ce chapitre, nous avons défini les étapes suivies pour la modélisation de notre système . Nous avons également discuté en détail sur notre approche en l'illustrant avec des figures et diagrammes . Ainsi , un algorithme représentant l'enchaînement du processus de notre système de recommandation.

D'un point de vue global, notre système consiste à créer un profil pour chaque nouvel utilisateur arrivant, pour ensuite lui fournir des recommandations personnalisées propres à ses besoins. Pour s'y faire, nous nous sommes servis du filtrage collaboratif basé sur l'utilisateur.

Le chapitre suivant, va présenter les outils d'implémentation utilisés et la mise en œuvre de notre système.

Chapitre 5

Implementation

5.1 Introduction

Afin de recommander des recettes ou des aliments sains et diététiques aux utilisateurs, nous avons adopté l'apprentissage supervisé. Le processus de classification sera effectué sur l'ensemble d'utilisateurs arrivés afin qu'ils puissent être classés sous leurs types d'obésité .

Dans ce chapitre, nous décrirons les jeux de données utilisés pour former notre système , présenterons les outils matériels et logiciels utilisés lors du développement de notre application, tels que le langage de programmation et la métrique de classification servie.

Enfin, nous allons discuter des résultats de notre expérience pour évaluer le système créé ainsi que sa capacité.

5.2 Jeux de données

Un jeu de données ou “Dataset” en Machine Learning, est une collection de données cohérentes qui peuvent se présenter sous différents formats : vidéos,images,statistiques.. Le terme “jeux de données” a été créé par IBM qui l'utilisait dans un sens proche d'un fichier .

Le dataset en ML se divise en trois catégories comme l'illustre la Figure 5.1 [16] :

- Un dataset d'entraînement : En apprentissage supervisé, le dataset contient une variable d'entrée et une variable de sortie, ou ce dernier devrait apprendre au modèle de faire la corrélation entre les deux. Le but de cette étape est d' apprendre à un modèle de

machine learning à réaliser une prédiction ou effectuer une tâche. Son exploitation consiste à extraire et modifier les fonctionnalités nécessaires pour s'adapter aux modèles de ML avant qu'il soit déployé.

- Un dataset de validation : après l'étape de formation, vient la phase de validation. le dataset vérifie les paramètres du modèle et apporte des modifications si nécessaire pour avoir une meilleure configuration. et enfin, valider l'architecture du modèle.
- Le dataset de test : dans cette phase, le dataset évalue la précision et la performance finale du modèle avant le déploiement , en utilisant de nouvelles valeurs de fonctionnalités et de fonctions prédictives. Les prédictions obtenues sont comparées à celles attendues.

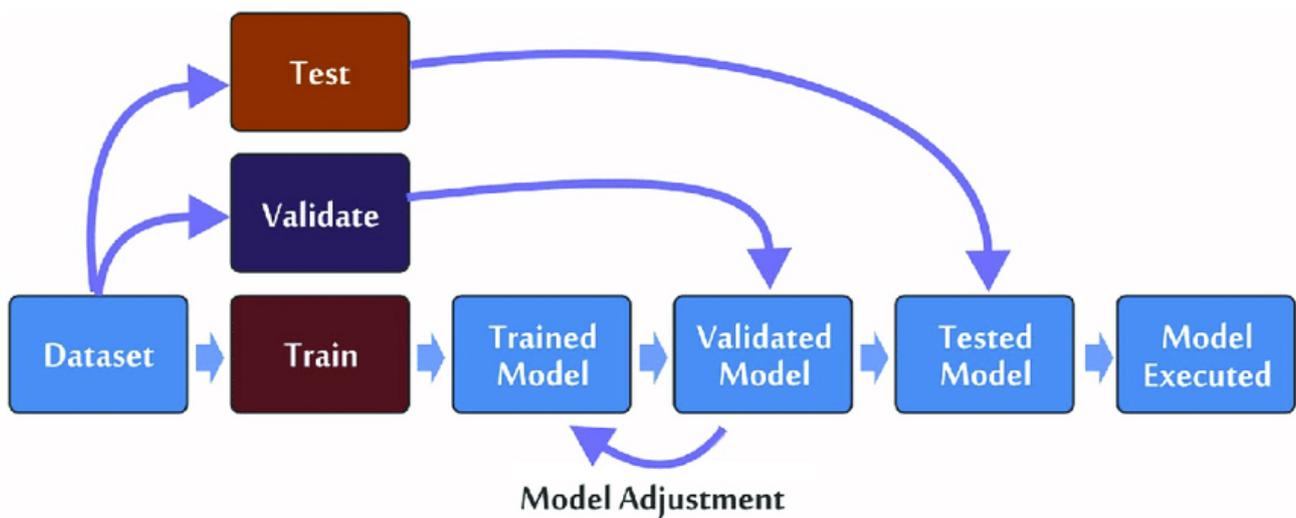


FIGURE 5.1 – Les différentes phases des Jeux de données

Dans notre cas, nous avons utilisé les datasets "Epicurious" [2], et "Food and their calories" [3], du site officiel 'kaggle' en format csv pour faciliter l'exploitation des données en python.

Le premier est un dataset de recettes contenant plus de 20000 recettes saines, avec leurs informations nutritionnelles, leurs apports caloriques, catégorie attribuée, ainsi que leurs ingrédients comme peut le montrer la Figure 5.2 .

	title	rating	calories	protein	fat	sodium	#cakeweek	#wasteless	22-minute meals	3-ingredient recipes	...	yellow squash	yogurt
0	Lentil, Apple, and Turkey Wrap	2.500	426.0	30.0	7.0	559.0	0.0	0.0	0.0	0.0	...	0.0	0.0
1	Boudin Blanc Terrine with Red Onion Confit	4.375	403.0	18.0	23.0	1439.0	0.0	0.0	0.0	0.0	...	0.0	0.0
2	Potato and Fennel Soup Hodge	3.750	165.0	6.0	7.0	165.0	0.0	0.0	0.0	0.0	...	0.0	0.0

3 rows × 680 columns

FIGURE 5.2 – Dataset "Epicurious"

Le deuxième, est d'une variété d'aliments (plus de 560 environs) et leurs calories en fonction de la portion pour laquelle les calories sont calculées.

	FoodId	Food
0	1	Artichoke
1	2	Arugula
2	3	Asparagus
3	4	Aubergine
4	5	Bell Pepper

	FoodId,Food,Serving,Calories
0	1,Artichoke,1 artichoke (128 g),60 cal
1	2,Arugula,1 leaf (2 g),1 cal
2	3,Asparagus,1 spear (12 g),2 cal
3	4,Aubergine,1 aubergine (458 g),115 cal
4	5,Bell Pepper,1 pepper (73 g),15 cal

FIGURE 5.3 – Jeux de données "Food and their calories"

De plus, nous avons rajouté les colonnes "FoodId" dans le dataset "Food and calories",comme vous pouvez le voir sur la Figure 5.3, ainsi que "RecipeId" pour le dataset

"Epicurious" comme attributs clés afin de les associer aux ratings.

Les jeux de données "ratingsN" et "ratingsrecipe" ont été créés par nous même dans le but d'appliquer le feedback explicite du filtrage collaboratif. Ces derniers contiennent chacun des "userId" , "title/food" , "rating" . ou chaque utilisateur ayant déjà consulté ou essayé une recette donnée, va noter cette dernière . Ce qui va faciliter et enrichir le processus de recommandation. Les Figures 5.4 et 5.5 représentent des captures de ces dernières.

	UserId	FoodId	rating
0	50	1	4.0
1	51	2	4.0
2	52	3	4.0
3	53	4	5.0
4	54	5	5.0

FIGURE 5.4 – "Food and their calories"

	UserId	recipeld	rating
0	40	1	2.500
1	41	2	4.375
2	42	3	3.750
3	43	4	5.000
4	44	5	3.125

FIGURE 5.5 – "Epicurious"

5.3 Environnement de développement

Dans cette partie, nous allons spécifier les outils utilisés au cours de notre développement d'application ainsi que pour la création de notre système.

5.3.1 Anaconda

Anaconda est une plateforme open-source et un logiciel gratuit fournissant une boîte à outils adaptée au développement ML et à la science des données. L'installation de ce dernier, donne accès à différents environnements nous permettant de coder en Python ou R. Ces environnements, également appelés environnements de développement intégrés (IDE), sont des plateformes ou des applications qui facilitent grandement le développement du code.

Anaconda a plus de 250 packages et plus de 7500 packages qui peuvent être facilement installés à partir de PyPI et gérés par le système de gestion de paquets appelé conda.

anaconda est l'une des meilleures façons pour installer et utiliser python. cela se fait à travers GUI (interface utilisateur graphique) qui est simple et bien prise en charge comportant les bibliothèques les plus importantes et IDE dans son installation. Elle simplifie également le processus de mise à jour de toutes ces bibliothèques. Ainsi, plutôt que d'installer Python séparément avec différents IDE, bibliothèques et fonctionnalités, Anaconda le fait dans une seule installation. La Figure 5.8 illustre l'interface du navigateur anaconda.

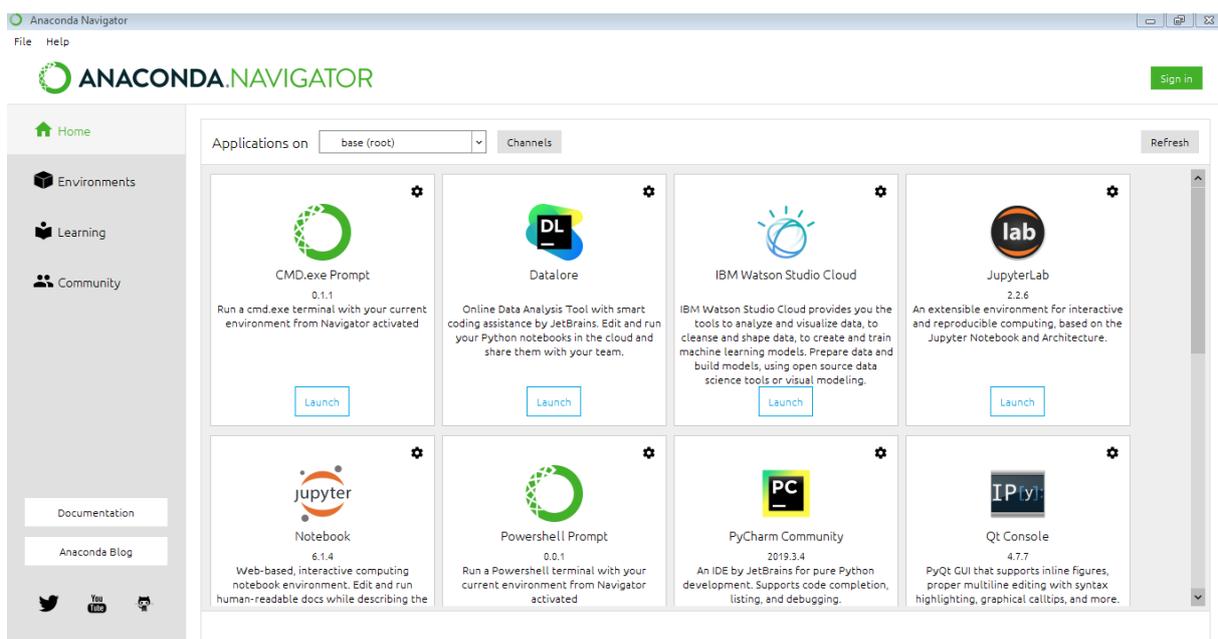


FIGURE 5.8 – L'interface du navigateur Anaconda

5.3.2 Jupyter notebook

Jupyter Notebook, est une application web open-source interactive .La première fois qu'il s'est déployé, son nom était IPython ,et puis, Jupyter en 2014 qui vient des principaux langages de programmation pris en charge : Julia, Python et R [5]. Malgré cela , le format ipynb de ses fichiers lorsqu'ils sont enregistrés a été conservé. Les notebooks Jupyter peuvent être convertis au format HTML, PDF et LaTeX en les téléchargeant a partir du menu de l'interface.

Jupyter notebook permet d'exploiter des ensembles de données dans des expériences et des modélisations ML.On peut l'utiliser également pour créer et partager des documents contenant du code en direct, des équations, des visualisations et du texte.

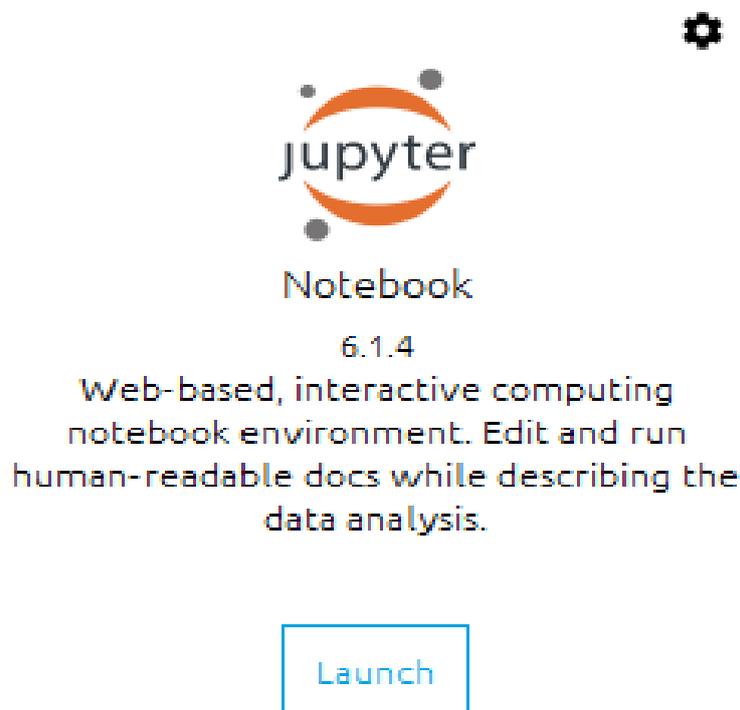


FIGURE 5.9 – Jupyter notebook

5.4 Langage de programmation "Python"

Python est un langage de programmation orienté objet open source de haut niveau, connu pour sa puissance, sa rapidité, et surtout sa simplicité ,ce qui l'a rendu très populaire. C'est un langage interprété, qui ne nécessite donc pas d'être compilé pour fonctionner .

Ce langage a été créé par le programmeur Guido van Rossum en 1991 [14] . Il tire son nom de l'émission Monty Python's Flying Circus. Les principales utilisations de python sont :

- La programmation d'applications
- La création de services web
- La génération de code
- La métaprogrammation
- Le scripting
- L'automatisation.

Il existe 2 versions de Python : 2.x et 3.x. Il y a beaucoup de concurrence entre ces deux dernières et les deux semblent avoir un certain nombre de fans différents .

Python 3.x n'est pas une simple amélioration ou extension de Python 2.x. Cela étant-dit ,tant que les auteurs de bibliothèques n'auront pas effectué la migration, les deux versions devront coexister.

5.5 Bibliothèques de python

Les bibliothèques sont des collections de modules avec du code pré-écrit qui peuvent être facilement importées et utilisées par les développeurs pour implémenter n'importe quelle fonctionnalité en réduisant le temps du codage. Le Python se fait avec un traitement de données régressif et continu en accédant aux données, en les manipulant et en les transformant à travers ces bibliothèques.

Python contient de nombreuses bibliothèques qui offrent un large éventail d'outils lui permettant d'effectuer des tâches comme les mathématiques, l'exploration des données, et la visualisation. Parmi ces dernières, on cite :

5.5.1 Pandas

Utilisée pour les structures et analyses de données de haut niveau. Elle simplifie la lecture de fichiers de tout type : csv, xlsx, txt, json ou d'un dictionnaire python. Il permet de fusionner et de filtrer des données, ainsi que de les rassembler à partir d'autres sources externes comme Excel. Les données lues sont stockées dans des dataframes.

5.5.2 Scikit-Learn

Scikit-learn est la bibliothèque la plus robuste et utile pour l'apprentissage automatique en Python. Elle fournit des outils efficaces pour le ML et la modélisation statistique, notamment la classification, la régression, le regroupement et la réduction de la dimensionnalité via une interface de cohérence.

5.5.3 Scipy

Bibliothèque Python a usage scientifique notamment mathématiques, elle dépend de NumPy, qui fournit une manipulation des tableaux à N dimensions pratique et rapide. SciPy est organisé en sous-packages couvrant différents domaines de calcul scientifique. Dans notre cas , nous avons utilisé **Scipy.spatial** :Structures et algorithmes de données spatiales.

5.5.4 Numpy

Numerical Python ou Numpy est une bibliothèque essentielle dédiée aux calculs numériques avec python .Cette dernière a été la clé de l'expansion des capacités de Python avec des fonctions mathématiques, sur la base desquelles des solutions de machine learning seraient construites. Numpy permet d'effectuer des opérations de base et avancées sur des tableaux multidimensionnels tels que l'algèbre linéaire,transformée de fourier et et les nombres aléatoires.

5.5.5 FuzzyWuzzy

FuzzyWuzzy est une bibliothèque de Python qui a comme méthodes la comparaison de chaînes en leur faisant une correspondance où nous pouvons avoir un score sur 100, indiquant si deux chaînes sont égales en donnant un indice de similarité [4] .

En résumé :

- NumPy et SciPy pour le calcul scientifique
- Pandas, Matplotlib pour l'exploration et la visualisation des données.
- Scikit learn pour le machine learning.

5.5.6 Matplotlib

Matplotlib est une bibliothèque python dédiée au traçage graphique et à la visualisation de données. La bibliothèque Matplotlib possède une sous-bibliothèque appelée « Pyplot » qui

nous permet de créer des interfaces similaires au logiciel MATLAB. En fait, les développeurs préfèrent utiliser la bibliothèque matplotlib car elle est open source et offre plus de fonctionnalités que la langage statistique MATLAB. Parmi les fonctionnalités qu'il offre, on peut désigner la fonctionnalité de dessiner tous types de graphiques différents par exemple : histogrammes, nuages de points également des diagrammes 2D et 3D, par conséquent, ces graphiques nous permettent de visualiser toutes les données et d'établir une analyse détaillée complète qui contribue à construire des modèles ML robustes.

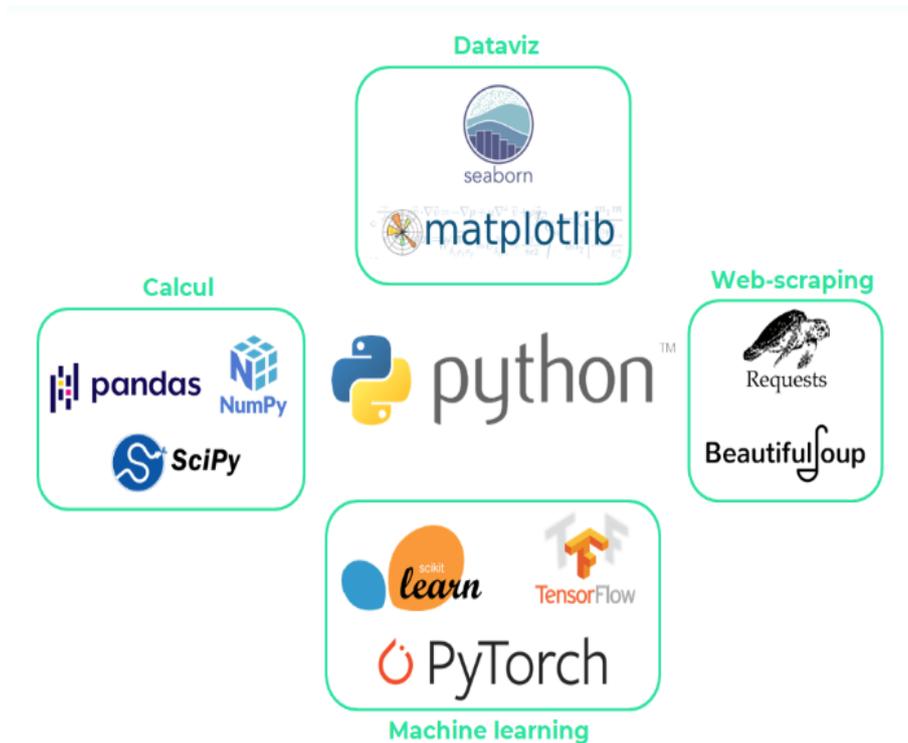


FIGURE 5.10 – Bibliothèques de python

5.6 Sublime text3

C'est un éditeur de texte générique gratuit, codé en C++ et Python, disponible sur Windows, Mac et Linux. Ce logiciel est riche en fonctionnalités et prend en charge plusieurs langages de programmation, dont CSS, HTML, XML, JavaScript, etc.

5.6.1 Html

Acronyme de HyperText Markup Language, qui est un type de langage informatique descriptif. Il s'agit plus précisément d'un format de données utilisé dans l'univers d'Internet pour la mise en forme des pages Web. Il permet, entre autres, d'écrire de l'hypertexte, mais aussi

d'introduire des ressources multimédias dans un contenu. Il utilise des balises pour modifier la structure ou la façon dont le texte et les objets figurent sur une page web.

5.6.2 CSS

Acronyme de Cascading Style Sheets, ou feuilles de style en cascade. Complémentaire avec le Html. Il donne des indications qui concernent le design et l'apparence des éléments de la page. Il indique par exemple la police et la couleur du texte. Le code CSS indique au logiciel quels devraient être la mise en page, la palette de couleurs et le formatage à utiliser pour le contenu HTML de la page web.

5.7 Interface d'inscription :

La figure ci-dessous représente l'interface d'inscription qui contient en en-tête un menu ayant un bouton Accueil , Se connecter et S'inscrire . l'interface contient un formulaire avec les champs a renseigner comme le nom, prenom , age ,poids, tailles ainsi qu'une listes de maladies chroniques que l'utilisateur doit selectionner et enfin le bouton envoyer permettant d'envoyer le formulaire.

The image shows a registration form for a website titled "Healthy for obesity". The form is overlaid on a background image of various healthy dishes. The form includes the following fields and elements:

- Header: "Healthy for obesity" on the left, and navigation links "Accueil", "se connecter", and "s'inscrire" on the right.
- Form fields: "Entrez Votre nom", "Entrez Votre prenom", "Entrez Votre pseudo", "Entrez votre mot de passe", "Enter Votre poids", and "Enter Votre taille".
- Section: "Maladies Chroniques:" with a dropdown menu.
- Dropdown menu options: "Diabete type1", "Diabete type1", "Diabete type2", "Hypertension artérielle", "Maladie du foie gras", "Allergie au gluten", and "Insuffisance renale".

FIGURE 5.11 – Interface d'inscription

5.8 Interface d'authentification :

La figure ci-dessous représente l'interface d'authentification qui contient des champs à remplir qui sont le 'username' et le mot de passe sans oublier le bouton se connecter qui permet d'envoyer la requête, l'authentification se fait seulement si l'utilisateur possède déjà un compte sinon il devra en créer un sur l'interface d'inscription.



FIGURE 5.12 – Interface d'authentification

5.9 Évaluation du système

Dans cette section, nous évaluerons notre système et discuterons des résultats en utilisant des métriques de classification dédiées au système de recommandation de nutrition pour les personnes obèses.

5.9.1 Métriques de classification

Dans le but d'établir la qualité de notre modèle, et de définir la mesure pour laquelle il a réussi à prédire la recommandation la plus pertinente, basée sur la similarité cosinus et la classification KNN, nous évaluons les performances, en générant quelques utilisateurs cibles aléatoires X et prédire leur classe correspondante qui est ensuite comparée à l'étiquette réelle.

Notre évaluation se base sur l'une des manières les plus utilisées pour mesurer la performance d'un modèle de classification, qui est **“la matrice de confusion”**. Cette méthode

correspond à un résumé tabulaire du nombre de prédictions correctes et aléatoires, faites par le modèle, dont chaque ligne correspond à une classe réelle et chaque colonne correspond à une classe estimée.

- **Vrai positif** : TP : C'est une instance qui a été correctement prédite par le modèle ML comme un modèle positif, autrement dit, le vrai positif est lorsque la classe réelle et la classe estimée sont toutes les deux positives. Dans notre cas le nombre de recettes/aliments pertinents recommandés.
- **Vrai négatif** : TF : C'est une instance correctement prédite par le modèle ML comme négatif; dans notre cas le nombre de recettes/aliments non pertinents a non recommandés.
- **Faux positif** : FP : C'est une instance mal classée par le modèle; Au lieu de la classer comme négative, l'instance est prédite comme positive. Dans notre cas, le nombre d'aliments non pertinents recommandés.
- **Faux négatif** : FN : La classe positive est prédite comme négative. Le modèle ne recommande pas une recette qui devrait être recommandée.

Différentes métriques peuvent être calculées à partir de la matrice de confusion afin de faciliter l'interprétation. C'est par exemple le cas du taux d'erreur, de la précision, du rappel et du F1 score. Ces indicateurs permettent de mieux apprécier la qualité de précision du modèle.

- **Accuracy** : Est une métrique dédiée à l'évaluation de la classification des modèles; le résultat obtenu correspond au nombre de prédictions correctes du modèle (TP, TN), il est calculé en divisant le nombre de prédictions correctes par le nombre total de prédictions. Ceci peut être calculé en utilisant la formule suivante en se basant sur les valeurs de la matrice de confusion :

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \dots (6)$$

Cette mesure est utilisée lorsque le nombre de True Positive et de True Negative sont les plus importants.

- **Precision** : Est un indicateur de ML qui évalue les performances d'un modèle. Elle mesure l'exactitude d'un classifieur. Il est calculé en divisant les éléments pertinents (TP) par le nombre total d'éléments positifs.

$$\text{Precision} = \frac{TP}{(TP+FP)} \dots (7)$$

Elle reflète le point de vue de l'utilisateur : si la précision est faible, l'utilisateur sera insatisfait, car il perd du temps à lire des informations qui ne l'intéressent pas.

- **Rappel** : Le rappel mesure la sensibilité et la capacité d'un système de classification à détecter les recettes correctement classées par rapport à toutes les autres recettes.

Cependant, un système de classification qui considérerait toutes ces dernières comme pertinentes ,obtiendrait un rappel de 100% et vice versa.Si le rappel est faible, l'utilisateur n'aura pas accès à une information qu'il souhaitait avoir.

Un rappel fort ou faible n'est pas suffisant pour évaluer les performances d'un système.

$$\text{Rappel} = \frac{TP}{(TP+FN)} \dots (8)$$

Un rappel fort ou faible n'est pas suffisant pour évaluer les performances d'un système.

- **Score F1** : est une combinaison des deux métriques : précision et rappel du test pour calculer le score. c'est une mesure de la précision d'un test. Généralement, nous considérons le score F1 plus que la précision comme une évaluation métrique en particulier lorsque le jeu de données est déséquilibré (les classes ne sont pas égales).

$$\text{F1-mesure} = \frac{2 * (\text{precision} * \text{rappel})}{(\text{precision} + \text{rappel})} \dots (9)$$

Pour appliquer le processus d'évaluation, nous effectuons un test de recommandation manuel, en sélectionnant des Id d'aliments ainsi que les ratings qui leurs ont été attribués.

Nous avons commencé par diviser l'ensemble de nos données en 80% comme données d'entraînement , et 20% comme données de test.

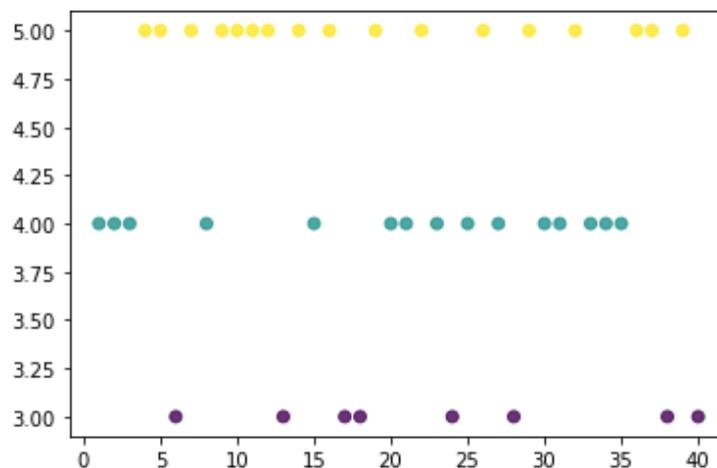


FIGURE 5.13 – L'ensemble de données

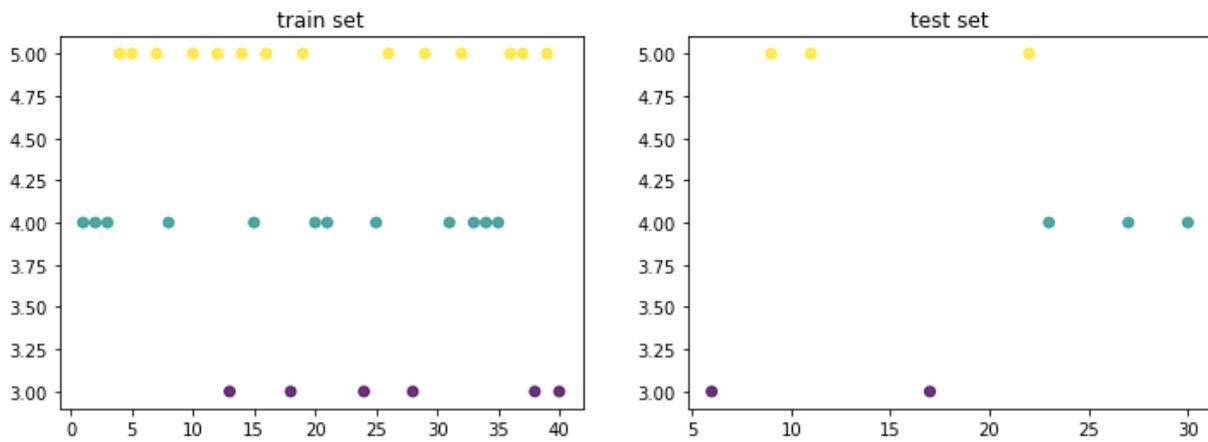


FIGURE 5.14 – Données d’entraînement et données de test

Puis, nous avons créé notre matrice de confusion avec seulement quelques données de celles déjà divisées, pour évaluer notre système .

```
confusion_matrix(y_true, y_pred)
array([[0, 0, 0, 0, 0],
       [1, 2, 0, 0, 0],
       [0, 0, 2, 0, 0],
       [0, 0, 0, 1, 0],
       [0, 0, 1, 0, 1]], dtype=int64)
```

FIGURE 5.15 – La matrice de confusion

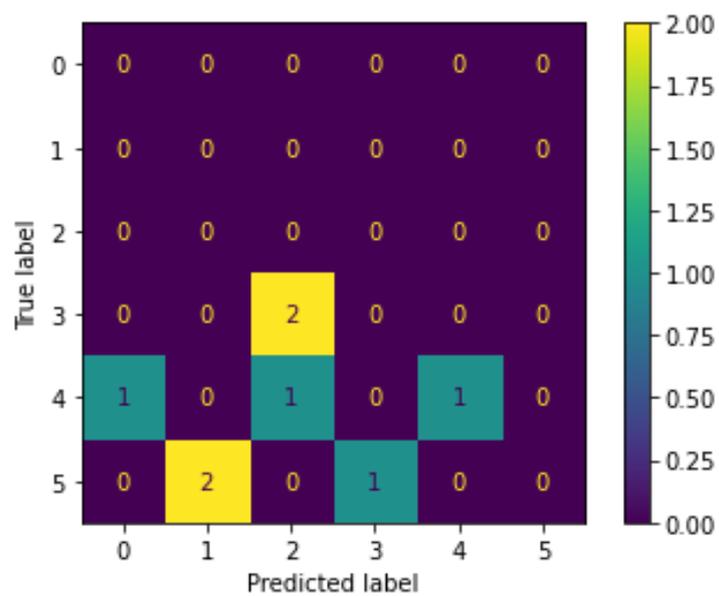


FIGURE 5.16 – La matrice de confusion.

Nous choisissons la mesure F1-score car elle tient compte a la fois la précision et le rappel dans un seul résultat plus exact et précis.

	precision	recall	f1-score
1	1.00	0.67	0.80
2	0.67	1.00	0.80
3	1.00	1.00	1.00
micro avg	0.83	0.83	0.83
macro avg	0.89	0.89	0.87
weighted avg	0.89	0.83	0.83

FIGURE 5.17 – Les résultats de l'évaluation dans le cas d'une classification multiple

Dans le cas d'une classification multiple, les performances en termes de précision sont de 89% et de rappel de 83%. En fonction de ces derniers, nous avons obtenu 83% en termes de la mesure choisie qui est le F1-score. Ainsi, de 75% accuracy. C'est un résultat acceptable, et satisfaisant pour un modèle a peu de données.

```
print("Accuracy du modele est :", accuracy)
Accuracy du modele est : 0.75
```

FIGURE 5.18 – Taux de Accuracy

En conséquent, ce dernier est le plus adapté a notre modèle en donnant un résultat exacte, précis, et surtout assez bon.

5.10 Conclusion

Dans ce chapitre, nous avons évoqué les principales étapes de notre implémentation. En premier lieu, nous avons défini les jeux de données en général, et présenté les datasets que nous avons utilisés pour former notre système, illustrés des captures d'écran de quelques affichages obtenus. Ensuite, nous avons donné l'environnement et les outils utilisés pour construire un système solide permettant de recommander des recettes et aliments aux personnes obèses. Enfin, nous avons expliqué l'utilité et l'objectif des métriques utilisées pour les calculs et l'évaluation des performances de notre système.

Dans le dernier chapitre, nous clôturons notre mémoire avec une conclusion générale comportant une idée générale du plan suivi lors de chaque chapitre, en discutant vers la fin sur nos travaux futurs et leurs objectifs.

Chapitre 6

Conclusion Générale

En raison d'hygiène de vie nuisible et malade, ajoutant à cela le stress et l'alimentation malsaine, le problème d'obésité/surpoids ne cesse d'augmenter à l'échelle mondiale. Ainsi, avec l'essor de la technologie et sa démocratisation nous avons opté pour une solution numérique automatisée, car il est nécessaire d'offrir un suivi diététique aux patients qui n'auront plus besoin de se déplacer ou de dépenser de l'argent. L'approche proposée se base sur la recommandation des recettes et aliments sains suivis de leurs apports caloriques. Et ceci, en utilisant les notions des systèmes de recommandation et les mesures de similitude à l'aide des datasets tirés du site officiel "kaggle".

notre thèse est constituée de six chapitres organisés comme suit :

Le premier chapitre est sous forme d'une introduction générale qui définit le contexte et la problématique, nos objectifs, ainsi que la méthodologie suivie au long du mémoire.

Le deuxième chapitre contient la notion du domaine de notre étude qui est les systèmes de recommandation, les principaux types de ce dernier, suivi de quelques avantages et inconvénients. Enfin, nous avons cité quelques domaines d'utilisation des SR.

Le troisième chapitre élabore l'état de l'art, où nous avons synthétisé quelques travaux existants catégorisés par les méthodes utilisées. Nous avons résumé ce travail dans un tableau explicatif, puis nous avons procédé à une analyse comparative de ces approches. Cette dernière, nous a inspiré à concevoir notre système et à réaliser notre propre approche "Healthy for obesity" dans le domaine de santé choisi "l'obésité".

Le quatrième chapitre fait une présentation de notre approche proposée dans laquelle on explique les étapes détaillées la façon de récupérer les informations de chaque utilisateur et le type de recommandation choisi ainsi que la méthode de classification et la mesure de similarité suivie, illustrées d'algorithmes, diagrammes, tableaux, et figures explicatives.

Le cinquième chapitre aborde les différents outils de mise en œuvre de notre approche, le langage de programmation, les datasets utilisés, ainsi que les divers aspects liés à l'implémentation de notre système.

Enfin, nous avons conclu avec une discussion des résultats obtenus, ainsi que les métriques utilisées pour réaliser ce processus.

Les résultats de l'accuracy est de 0.88, soit 88 % dans la classification binaire (88 prédictions correctes sur 100 exemples au total). Cela signifie que notre classificateur d'utilisateurs obèses fait un excellent travail d'identification. Nous avons réussi à avoir des recommandations personnalisées nutritionnelles suivies d'informations pour toute personne obèse ou en surpoids utilisant notre système, en se basant sur les informations de chaque utilisateur comme l'IMC et les antécédents, ainsi que leurs préférences à l'aide des ratings attribués par les utilisateurs.

La réalisation de ce projet, nous a permis d'acquérir plus d'expertise dans le domaine de la recommandation. Ceci nous a motivé à fixer notre objectif pour les travaux futurs qui comportera une combinaison des deux datasets utilisés et réaliser un nouveau à double recommandation en une seule approche. Pour chaque recette, nous allons afficher les calories et les informations de chaque ingrédient. De plus, nous comptons étendre notre application afin qu'elle puisse assurer un suivi complet en utilisant des capteurs pour calculer le taux de calories brûlées. Ainsi des recherches ont montré que le sommeil est un facteur majeur pour une bonne santé. Pour cela, nous comptons ajouter un IoT tel qu'une smart watch pour contrôler le sommeil pour un meilleur mode de vie. Et enfin renforcer notre système afin d'obtenir des résultats encore plus solides et intéressants.

Bibliographie

- [1] Classification obésité sur la base de l'imc ou bmi dr salsano. <https://chirurgie-digestive-montpellier.fr/centre-obesite/prise-en-charge-obesite/1ere-cs-chirurgien-obesite/classification-obesite-sur-la-base-de-limc-ou-bmi-dr-salsano/>. Page consultée le 01 septembre 2022.
- [2] Epicurious - recipes with rating and nutrition. <https://www.kaggle.com/datasets/hugodarwood/epirecipes>. page consultée le 23 avril 2022.
- [3] Food and their calories. <https://www.kaggle.com/datasets/vaishnavivenkatesan/food-and-their-calories?select=Food+and+Calories+-+Sheet1.csv&fbclid=IwAR0cJiZ8WTfmOGRIRvE-ICk7709P5eiGH1W0kd6gG5QW9qRs4Qj-5MkzM78>. page consultée le 24 avril 2022.
- [4] Fuzzywuzzy python library. <https://www.geeksforgeeks.org/fuzzywuzzy-python-library/>. Page consultée le 19 aout 2022.
- [5] Jupyter notebook : An introduction. <https://realpython.com/jupyter-notebook-introduction/>. Page consultée le 25 aout 2022.
- [6] Les systèmes de recommandation : une catégorisation. <https://interstices.info/les-systemes-de-recommandation-categorisation>. Page consultée le 06 mars 2022.
- [7] L'obésité chez l'adulte : ligne directrice de pratique clinique. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7721377/>. Page consultée le 19 mars 2022.
- [8] Obésité : Une maladie qui pèse sur le système de santé. <http://www.santemaghreb.com/actus.asp?id=29300>. Page consultée le 27 fevrier 2022.
- [9] Obésité et surpoids. <https://www.who.int/fr/news-room/fact-sheets/detail/obesity-and-overweight>. Page consultée le 15 mars 2022.
- [10] Qu'est ce que l'algorithme knn ? <https://datascientest.com/T1\guilsinglrightknn>. page consultée le 13 juillet 2022.
- [11] Recommender systems — user-based and item-based collaborative filtering. <https://medium.com/@cfpinela/recommender-systems-user-based-and-item-based-collaborative-filtering-5d5f375a127f>. Page consultée le 01 septembre 2022.
- [12] Systèmes de recommandation. <https://www.olcya.com/2018/09/22/2018-09-22-systemes-de-recommandation/>. Page consultée le 15 juillet 2022.
- [13] Systèmes de recommandation : ce que vous devez savoir. <https://www.affde.com/fr/what-is-recommendation-system.html>. Page consultée le 25 janvier 2022.

- [14] Un peu de l'histoire. <http://leaderswithouttitle.github.io/PythonLWT/History.html>. Page consultée le 16 mai 2022.
- [15] Benouaret.I. *Un système de recommandation contextuel et composite pour la visite personnalisée de sites culturels*. mémoire d'une thèse de doctorat dans la technologies de l'information et des systèmes, Université de Technologie de Compiègne, 2017.
- [16] Crochet-Damais.A. Dataset en machine learning : définition et techniques. 2022.
- [17] Debashis.D, Laxman.S, and Sujoy.D. A survey on recommendation system., international journal of computer applications. vol160.7 :page 5, 2017.
- [18] El-Bouhissi.H, Kettam.A, and Adel.M. Towards an efficient knowledge-based recommendation system. 2021.
- [19] El-Dosuky.M.A, Rashad.M, Hamza.T, and El-Bassiouny.A. Food recommendation using ontology and heuristics. 2012.
- [20] Fisk.D. An application of social filtering to movie recommendation. vol1198, 1997.
- [21] Hill.W, Stead.L, Rosenstein.M, and George.F. Recommending and evaluating choices in a virtual community of use. 1995.
- [22] Hoill.J and Kyungyong.C. Knowledge-based dietary nutrition recommendation for obese management. vol17.1, 2016.
- [23] John.S.B, David.H, and Carl.K. Empirical analysis of predictive algorithms for collaborative filtering. In *Appears in Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI1998)*, page 10, 2013.
- [24] Joseph.A.K. Introduction to recommender systems : Algorithms and evaluation. vol22, 2004.
- [25] KARAOUZENE.M. *Système de recommandation des services web sémantiques*. mémoire d'un master en informatique, Université Abou Bakr Belkaid– Tlemcen, 2014.
- [26] Mckensy-Sambola.D, Rodríguez-García.M.A, García-Sánchez.F, and Valencia-García.R. Ontology-based nutritional recommender system. vol12.1, 2022.
- [27] Nadja.L, Mira.M, Hanna.S, Martin.L, Nada.T, Georg.G, Markus.B, Kurt.G, and Helmut.K. Nutrilize a personalized nutrition recommender system : an enable study. 2018.
- [28] Paul.R, Neophytos.I, Mitesh.S, Peter.B, and John.R. GroupLens : An open architecture for collaborative filtering of netnews. 1994.
- [29] Rehman.F, Osman.K, Nuhman.U.H, Atta.U.R.K, Kashif.B, and Sajjad.A.M. Diet-right : A smart food recommendation system. vol11.6, 2017.
- [30] Rich.E. User modeling via stereotypes. vol3, 1979.
- [31] Sander.M.B, T. Heleen.M.W, Cornelis.H.E, and H. A.V.W. Developing a personalized integrative obesity-coaching program : A systems health perspective. vol19.2, 2022.
- [32] TAOULI.S and BENACHENHOU.W. *Utilisation de factorisation matricielle dans les Systèmes de recommandation sensible au contexte*. Mémoire master, université Abou Bakr Belkaid –Tlemcen, 2017.

Résumé

Les systèmes de recommandation sont utilisés pour fournir des items (par exemple : les films, la musique, les articles, les recettes..) adaptées aux préférences des utilisateurs. Ils sont capables d'estimer l'intérêt d'un utilisateur pour une ressource donnée en se basant sur certaines de ses informations et de ce qui a été suivi par d'autres utilisateurs similaires .

Dans le cadre de ce mémoire, nous nous intéressons à élaborer un système de recommandation basé sur le filtrage collaboratif qui consiste à trouver l'information qui satisfait l'utilisateur en utilisant les évaluations des autres personnes. Notre objectif est de recommander des recettes ou aliments censés être appréciés par l'utilisateur, qui sont aussi jugés diététiques aux personnes souffrants d'obésité ou de surpoids, et ceci en tenant compte leurs IMC ainsi que les antécédents.

En outre, nous proposons "Healthy For Obesity", un nouveau SR alimentaire basé sur la connaissance et les informations de l'utilisateur, en intégrant des techniques d'apprentissage automatique tel que KNN, qui peuvent être utilisées pour classifier les utilisateurs et les aider à trouver des aliments pertinents suivis de leur apport calorique. En particulier, nous mettons en œuvre HFO dont nous présentons l'idée générale dans ce travail en se servant de deux datasets tirés du site kaggle, les analyser , et enfin obtenir des résultats précis et pertinents . l'évaluation de notre système a montré d' assez bons résultats avec un accuracy de 75%.

Mots-clés

Item ,IMC , Apprentissage automatique, KNN, SR, Filtrage collaboratif, HFO, kaggle.