

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

Université de 8 Mai 1945 – Guelma -

Faculté des Mathématiques, d'Informatique et des Sciences de la matière

Département d'Informatique



## Mémoire de Fin d'études Master

**Filière :** Informatique

**Option :** Sciences et Technologie de l'Information et de la Communication.

**Thème :**

---

---

### Recommandation des Annonces aux Différentes Communautés d'Intérêt sur le Web

---

---

**Encadré Par :**

Mme. Boughareb Djalila

**Présenté par :**

Bensalah Hazem

**Septembre 2021**

## Remerciements

*Tout d'abord je tiens à remercier Dieu, le tout puissant et miséricordieux,  
qui m'a donné la*

*force, l'intelligence et la patience d'accomplir ce modeste travail.*

*Un grand remerciement à Mme Bougharb Djalila pour son encadrement.*

*Elle était toujours*

*montré à l'écoute et très disponible tout au long de la réalisation de ce  
mémoire, ainsi pour*

*l'inspiration, l'aide et le temps qu'elle a bien voulu me consacrer, je te dis  
merci.*

*Des remerciements à ma famille et mes amis pour leurs encouragements,  
grâce à eux j'ai pu*

*surmonter tous les obstacles.*

*Enfin, mes vifs remerciements vont également aux membres de jury pour  
l'intérêt qu'ils ont*

*porté à mon projet en acceptant d'examiner et juger mon travail.*

## *Dédicaces*

Je tiens c'est avec grand plaisir et honneur que je dédie ce modeste travail :

A l'être la plus chère de ma vie, ma mère Aicha.

A celui qui m'a fait de moi un homme, mon père Salah.

A mon frère Mohammed Et mes sœurs Karima, Madjda, Sana, Nawel Qui m'ont toujours soutenue et encouragé durant toutes ces années d'études.

Et Finalement à mes amis **Zaim**, Hamza, fares, **Salah**, Oussama, iskander, abdou, Dayou, Amar, **Aymen** qui n'ont jamais cessé de me soutenir.

# Résumé

**Mots clés : systèmes de recommandation (SR), SR collaboratifs, réseaux sociaux, communautés, Kmeans, publicité, PMF, KD-TRee.**

Actuellement, l'utilisation des réseaux sociaux fait parti du quotidien de l'internaute algérien. Sur ces réseaux les annonces publicitaires sur différents sujets ne cessent d'apparaître. Dans la plus part des cas ces annonces ne conviennent pas à l'internaute et elles ne correspondent pas à ses goûts et intérêts. Dans ce contexte, nous avons pensé à proposer un système de recommandation qui permet s'assurer la diffusion d'une publicité intéressante aux différentes communautés. Dans ce projet, nous avons collecté des données réelles et spécifiques sur les préférences de l'utilisateur algérien, et ceci afin de donner à notre projet une utilité réelle. Ensuite, nous avons traité ces données afin de créer des communautés selon 3 méthodes différentes : la première c'était Kmeans dans le but de classifier les utilisateurs à partir de leurs vecteurs de préférences, la deuxième approche c'est une structure arborescente KD-Tree et la troisième c'était les PMF.

Nous avons testé la première méthode sur le terrain en visionnant des publicités sous forme de vidéo et nous avons obtenu de bons résultats puisque 92% des participants ont aimé ce que le système a proposé. Dans la deuxième méthode, nous avons testé notre système en divisant les données en deux parties : entraînement et test. La troisième méthode était les PMF. Les résultats obtenus ont été encourageants en terme de précision et temps de réponse.

# Abstract

**Keywords :** recommendation systems (SR), collaborative SR, social networks, communities, Kmeans, advertisement, PMF, KD-Tree.

Currently, the use of social networks is part of the daily life of the Algerian internet user. On these networks advertisements on different subjects cease to appear. In most cases, these advertisements are not suitable for the Internet user and they do not correspond to his tastes and interests. In this context, We thought proposing a recommender system which allows to ensure the diffusion of an interesting advertisement to the various communities. In this project, we collected real and specific data on the preferences of the Algerian user, and this in order to give our project a real utility. Then, we processed this data in order to create communities according to 3 different methods : the first was Kmeans with the aim of classifying users based on their preference vectors, the second approach is a tree structure KD-Tree and the third was the PMF. We tested the first method in the field by viewing advertisements in the form of video and we got good results since 92% of the participants liked the proposed ads. In the second method, we tested our system by dividing the data into two parts : Training and test. The third method was PMF. The obtained results were encouraging in terms of precision and response time.

# Sommaire

<b>Table des figures</b>	1
<b>Liste des tableaux</b>	1
<b>Introduction Générale</b>	1
<b>Problématique</b>	2
<b>Structure du mémoire</b>	3
<b>1 LES SYSTÈMES DE RECOMMANDATION</b>	4
1.1 Introduction	5
1.2 Que sont les systèmes de recommandation ?	5
1.3 Comment fonctionnent les systèmes de recommandation ?	6
1.3.1 Comprendre les relations	6
1.4 Les approches de recommandation	6
1.4.1 Basées sur le contenu	7
1.4.2 Basées sur le filtrage collaboratif	8
1.4.3 Basée sur l'intelligence computationnelle	8
1.4.4 Basée sur les réseaux sociaux (SNA)	9
1.4.5 SR Hybride	9
1.5 Collecte de données	10
1.5.1 Explicite	10
1.5.2 Implicite	11
1.5.3 Hybrides	11
1.6 Domaines des systèmes de recommandation	11
1.7 Recommandation des annonces	12
1.8 Conclusion	12

<b>2 LES RÉSEAUX SOCIAUX</b>	<b>13</b>
2.1 Introduction	14
2.2 Définition	14
2.3 Caractéristiques des Réseaux Sociaux	14
2.3.1 Basés sur l'utilisateur	14
2.3.2 Interactifs	14
2.3.3 La collectivité	15
2.3.4 Les Relations	15
2.4 Fonctionnement des réseaux sociaux	15
2.5 Les principaux réseaux sociaux classés par catégories	15
2.5.1 Les réseaux sociaux généralistes	16
2.5.2 Les réseaux sociaux professionnels	17
2.5.3 Les réseaux sociaux de blogging	18
2.5.4 Les réseaux sociaux de vidéo	20
2.6 La publicité sur les réseaux sociaux	21
2.7 Conclusion	22
<b>3 CONCEPTION</b>	<b>23</b>
3.1 Introduction	24
3.2 Collection de données	24
3.3 Les caractéristiques de notre site web	24
3.4 Architecture du système de recommandation proposé	25
3.4.1 Kmeans + elbow architecture	25
3.4.2 Notre KDTree algorithme	32
3.4.3 Factorisation matricielle probabiliste	38
3.5 Conclusion	39
<b>4 Implémentation et résultats</b>	<b>40</b>
4.1 Introduction	41
4.2 Présentation de notre site	41
4.2.1 Environnement de développement	41
4.2.2 Les interfaces graphiques	42
4.3 Traitement des données collectées	44
4.3.1 Environnement de développement	44
4.3.2 Les bibliothèques utilisées	45
4.3.3 Kmeans + la methode Elbow	45

4.3.4 KDTree VS Factorisation matricielle probabiliste . . . .	46
4.4 Conclusion . . . . .	51
<b>Conclusion générale</b>	<b>54</b>
<b>Bibliographie</b>	<b>58</b>



# Table des figures

1.1 SR Basées sur le contenu [34]. . . . .	7
1.2 SR basées sur le filtrage collaboratif [34]. . . . .	8
1.3 SR Hybride [34] . . . . .	10
2.1 Facebook logo. . . . .	16
2.2 Twitter logo. . . . .	17
2.3 Google+ logo. . . . .	17
2.4 LinkedIn logo. . . . .	18
2.5 viadeo logo. . . . .	19
2.6 Medium logo. . . . .	19
2.7 Tumblr logo. . . . .	20
2.8 Youtube logo. . . . .	20
2.9 Tiktok logo. . . . .	21
3.1 Les etapes de notre algorithme . . . . .	31
3.2 Points et leur position dans un système de coordonnées [10]. . . . .	32
3.3 Dessiner la médiane x de 6 dans le système de coordonnées [10]. . . . .	32
3.4 La séparation spatiale finale [10]. . . . .	33
3.5 l'ensemble de données présenter par le dashboard . . . . .	34
3.6 notre deuxième ensemble de données (dataset enrichie) . . . . .	35
3.7 notre structure KD-tree. . . . .	37
3.8 La matrice R de l'ensemble de données . . . . .	38
4.1 les pages d'inscription et de connexion de notre site. . . . .	43
4.2 la page accueil . . . . .	44
4.3 Le code de notre algorithme KD-Tree . . . . .	47
4.4 les valeurs RMSE et le taux d'exécution de KD-Tree . . . . .	47
4.5 les valeurs RMSE et le taux d'exécution de PMF . . . . .	50
4.6 les préférences de l'utilisateur 350 . . . . .	50

TABLE DES FIGURES

---

4.7 recommandation de publicités à l'utilisateur 350 . . . . .	50
--	----

# Liste des tableaux

1.1 Quelques techniques de recommandation.	10
3.1 Exemple de descripteur de données	26
3.2 deux exemple de la méthode du Elbow appliquée à la catégorie téléphone et alimentation	28
3.3 exemple de calcul de la distance entre le vecteur utilisateur et les centres des clusters	29
3.4 Les caractéristiques du cluster numéro 5.	30
3.5 le pourcentage de différence de goûts pour tous les clusters	31
4.1 quelques exemples d'évaluations par les utilisateurs des publicités recommandées	46
4.2 valeurs observées et prédites pour l'ensemble de test	49
4.3 comparaison entre KDTree algorithme et PMF algorithme	51

# Introduction Générale

Au cours des dernières années, la publicité a largement envahi divers sites web et réseaux sociaux. Parfois l'utilisateur peut fréquenter en naviguant plusieurs annonces, parmi ces annonces un nombre important n'intéressent pas cet utilisateur. Alors, il faut penser à donner plus d'importance aux annonces électroniques diffusées un peu partout sur la toile.

Dans le but d'améliorer la pertinence des annonces affichées à l'utilisateur, on a introduit les systèmes de recommandation (SR) basée sur le filtrage collaboratif qui s'agissent d'un moyen d'aide à trouver des items (films, livres, papiers, etc.) intéressants pour l'utilisateur alors pourquoi ne pas les employer pour nous aider à voir également des publicités qui pourraient nous satisfaire.

Pour cela, nous avons collecté des données réelles sur les préférences de l'utilisateur algérien. Ensuite, nous avons traité ces données pour créer des communautés selon 3 méthodes différentes : Kmeans, KD-Tree et PMF.

Les résultats expérimentaux étaient encourageants.

# Problématique

Avec la grande diversité et richesse de la popularité algérienne et leurs préférences ainsi que le grand nombre de publicités qui leur apparaissent, dont la plupart ne les intéressent pas, il est devenu nécessaire d'utiliser les systèmes de recommandation pour suggérer ce qui est approprié pour chaque communauté.

Les annonceurs se plaignent aussi que leurs annonces ne parviennent pas à des gens qui sont réellement intéressés. Par exemple, une annonce d'emploi peut aller à des retraités, ou une publicité pour un match de football apparaît aux femmes.

Dans ce travail nous proposons une solution pour ce problème en essayant de fournir des annonces appropriées à l'individu algérien, afin d'assurer le profit pour les annonceurs et les utilisateurs à la fois.

# Structure du mémoire

Nous avons proposé d'organiser notre mémoire comme suit :

- Dans le premier chapitre, nous allons présenter un état de l'art sur les systèmes de recommandation.
- Dans le deuxième chapitre, nous parlons des réseaux sociaux les plus populaires.
- Dans le troisième chapitre, nous allons présenter l'approche proposée et la conception du système développé.
- Dans le quatrième chapitre, nous allons présenter le système, nous allons présenter les outils et les technologies utilisées et les détails d'implémentation.

# Chapitre 1

## LES SYSTÈMES DE RECOMMANDATION

## 1.1 Introduction

Les systèmes de recommandations de la dernière décennie ont connu une amélioration importante dans les algorithmes et méthodes de recommandation. Un système de recommandation peut être décrit comme un ensemble de programmes qui tentent de recommander les items les plus appropriés (produits ou services) aux utilisateurs uniques (particuliers ou entreprises) en anticipant l'intérêt de l'utilisateur pour un item donnée en fonction des données relatives aux items, aux utilisateurs et aux interactions entre les items et les utilisateurs [17].

Un système de recommandation est un système d'apprentissage automatique qui aide les utilisateurs à découvrir de nouveaux produits et services. A Chaque fois que vous naviguez en ligne, le système de recommandation vous guidera vers l'article le plus susceptible de vous intéresser.

Les systèmes de recommandations sont des facteurs déterminants de notre monde numérique, car les utilisateurs choisissent souvent d'être dépassés et ont besoin d'aide pour trouver ce qu'ils veulent. Cela apporte des clients plus satisfaits et bien sûr plus de ventes. Le système de recommandation est comme un vendeur connaissant vos préférences en fonction de votre historique.

Dans ce chapitre, nous allons donner une définition plus précise des SR, de leurs types et de leurs fonctionnements.

## 1.2 Que sont les systèmes de recommandation ?

Les SR sont si populaires maintenant que beaucoup d'entre nous les utilisent sans même le savoir. Parce que nous ne pouvons pas regarder tous les produits ou contenus sur un site Web, un système de recommandation joue un rôle important pour nous aider à avoir une meilleure expérience utilisateur, tout en nous exposant à plus d'inventaire que nous ne pourrions pas découvrir autrement.

Le but de son utilisation est de faciliter la recherche et l'accès aux informations en proposant des suggestions d'items (annonce, film, musique... etc.) adaptés aux besoins des différents utilisateurs. Un composant important de l'un de ces systèmes est la fonction de recommandation, qui prend en considération des informations sur l'utilisateur et prédit la valeur que l'utilisateur pourrait attribuer à un item donnée. Prédire les évaluations de l'utilisateur, avant même



qu'il les fourni, fait des systèmes de recommandation un outil puissant [8].

## 1.3 Comment fonctionnent les systèmes de recommandation ?

### 1.3.1 Comprendre les relations

Les relations fournissent aux systèmes de recommandation une excellente compréhension des clients et leurs besoins. Il existe trois principaux types de relations : relation utilisateur-produit, relation produit-produit et relation utilisateur-utilisateur.

#### 1. Relation utilisateur-produit

La relation utilisateur-produit est créée lorsque certains utilisateurs ont une affinité ou une préférence pour des produits spécifiques dont ils ont besoin. Par exemple, un joueur de football pourrait avoir une préférence pour les items liés au foot, ainsi les réseaux sociaux construisent une relation utilisateur-produit de joueur->foot cite [8].

#### 2. Relation produit-produit

Les relations produit-produit se produisent lorsque les items sont de nature similaire, soit par apparence, soit par description. Quelques exemples incluent des livres ou de la musique du même genre, des plats de la même cuisine, ou des items de nouvelles d'un événement particulier cite [8].

#### 3. Relation utilisateur-utilisateur

Les relations utilisateur-utilisateur se produisent lorsque certains clients ont des goûts similaires en ce qui concerne un produit ou un service particulier par exemples : amis communs, préférences similaires, wilaya similaires, etc cite [8].

## 1.4 Les approches de recommandation

Il existe plusieurs approches de recommandation ; nous expliquerons quelques-uns ci-dessous :

### 1.4.1 Basées sur le contenu

Cette approche recommande des éléments similaires aux éléments préférés par un utilisateur spécifique. Deux méthodes ont été utilisées pour générer des recommandations. La première est simple, dans laquelle des mesures de similitude sont utilisées, telles que la distance euclidienne [28], et l'autre méthode génère des recommandations à l'aide de méthodes d'apprentissage évaluatif [54] et d'apprentissage automatique [54], qui sont des modèles largement constructifs capables d'apprendre les intérêts d'un utilisateur en ayant l'historique de ses intérêts passés [37].

Shu et al [45] proposent un algorithme de recommandation basé sur le contenu et sur un réseau neuronal à convolution (CNN) pour permettre aux étudiants de découvrir de nouvelles ressources d'apprentissage qui correspondent à leurs goûts, et permet au système d'apprentissage en ligne de cibler les ressources d'apprentissage vers les bons étudiants. Dans le travail de Suglia [47], les auteurs ont étudié l'efficacité des réseaux neuronaux récurrents (RNN) dans une recommandation basée sur le contenu.

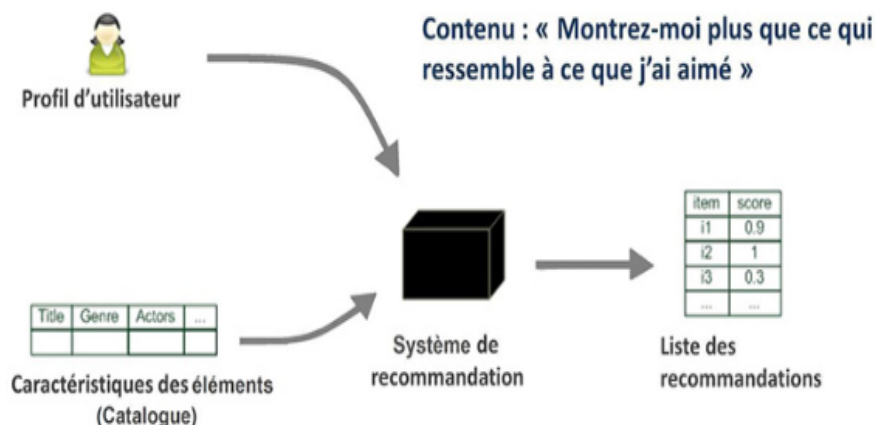


FIGURE 1.1 – SR Basées sur le contenu [34].

#### 1.4.1.1 Avantage et Inconvénients :

L'avantage principal de la recommandation basée contenu est qu'elle fonctionne même lorsqu'un produit n'a pas eu d'avis d'utilisateurs. Cependant, parmi ses inconvénients nous pouvons citer qu'elle nécessite des données descriptives de tout le contenu à recommander, ce qui prend du temps. De plus, elle est également difficile à mettre en œuvre sur de grandes bases de données de produits car les utilisateurs ont des opinions différentes sur chaque produit.

## 1.4.2 Basées sur le filtrage collaboratif

Les approches de recommandation basées sur le filtrage collaboratif (CF) aident les gens à faire des choix en fonction de l'évaluation et des préférences d'autres personnes qui partagent des intérêts similaires comme le pays, le sexe et l'âge ou qui aiment les mêmes éléments. Les SR basés sur la technique de CF travaillent sur les préférences des utilisateurs ayant des préférences similaires pour certains produits afin d'en suggérer un nouveau [16].

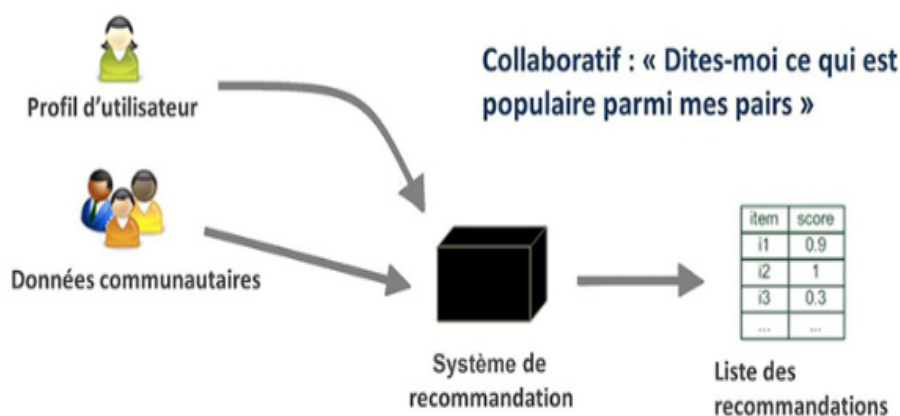


FIGURE 1.2 – SR basées sur le filtrage collaboratif [34].

### 1.4.2.1 Avantage et Inconvénients :

L'avantage principal de la recommandation basée sur le filtrage collaboratif est qu'elle n'exige aucune description de produits. Tandis que, ses inconvénients, consiste dans le fait que nous ne pouvons pas recommander l'item s'il n'y a pas d'avis d'utilisateurs (souffre du problème de démarrage à froid). De plus, il est difficile de recommander de nouveaux utilisateurs et enclin à privilégier les produits populaires ayant beaucoup d'avis. Ainsi, il devient difficile de recommander de nouveaux produits car ils ont moins d'avis.

## 1.4.3 Basée sur l'intelligence computationnelle

Les techniques d'intelligence computationnelle (CI) incluent les techniques bayésiennes [42], les réseaux de neurones artificiels [50], les techniques de clustering [53], les algorithmes génétiques [18] et les techniques d'ensemble flou [49], dans les systèmes de recommandation, ces techniques d'intelligence computationnelle sont largement utilisées pour construire des modèles de recommandation [37].

#### 1.4.4 Basée sur les réseaux sociaux (SNA)

Cette technologie contribue à fournir plus d'informations sur l'utilisateur comme il est apparu fortement avec le grand développement des médias sociaux, ce qui nous permet de connaître les amis des utilisateurs et leurs commentaires sur les différents produits et pages qu'ils aiment.

La logique et la réalité dans lesquelles nous vivons nous disent que la plupart des gens achètent des produits suggérés par leurs proches, et l'impact peut être plus que les publicités pour le produit.

«Trust» est une relation largement discutée dans les études sur les réseaux sociaux et dans un système de recommandation, le mot «Trust» est généralement défini comme «how well does Alice trust Bob concerning the specific product or taste» [21]. Il existe toujours une corrélation positive entre la confiance et la similitude des utilisateurs dans les communautés en ligne [56].

Zhang et al. [55] proposent un modèle appelé Comprehensive Evaluation Trust à utiliser dans un algorithme de recommandation qui fournit des recommandations personnalisées basées sur la confiance dans les réseaux sociaux.

#### 1.4.5 SR Hybride

Pour obtenir des performances plus élevées et surmonter les inconvénients des techniques de recommandation traditionnelles, une technique de recommandation hybride qui combine les meilleures caractéristiques de deux ou plusieurs techniques de recommandation en une seule technique hybride a été proposée [26]. Selon Burke [26], il existe 7 mécanismes d'hybridation de base des combinaisons utilisées dans les systèmes de recommandation pour construire des hybrides : pondéré [39], mixte [46], commutation [23], combinaison de caractéristiques, augmentation de caractéristiques [51, 48], cascade [41] et méta-niveau [20]. La pratique la plus courante dans les techniques de recommandation hybrides existantes est de combiner les techniques de recommandation CF avec les autres techniques de recommandation afin d'éviter les problèmes de démarrage à froid, de parcimonie et / ou d'évolutivité [15, 25].

Il existe d'autres approches comme les approches de recommandation basées sur les connaissances (KB), les approches de recommandation basées sur la connaissance du contexte (CRS) et les approches de recommandation de groupe (GRS).

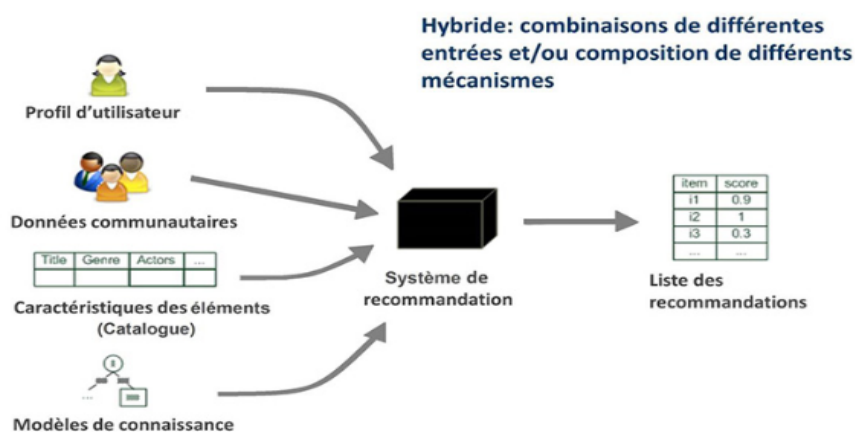


FIGURE 1.3 – SR Hybride [34]

Approach	Techniques de recommandation	
	Data mining technique	References
Content-based	CNN	[45]
	RNN	[47]
	Neuro-Fuzzy Approach	[44]
Collaborative	KMEANS + KNN	[17]
	Clustering	[40]
	Deep learning	[31]
	Bayesian non Negative Matrix Factorization	[24]
	Bio inspired algorithm ant colony + Kmeans	[36]

TABLE 1.1 – Quelques techniques de recommandation.

## 1.5 Collecte de données

La collecte de données est une étape cruciale dans le processus de recommandation. Les données peuvent être fournies de différentes manières. Il existe 3 méthodes particulièrement importantes, collecte explicite, implicite et hybride.

### 1.5.1 Explicite

Le système demande à l'utilisateur de fournir des classes d'éléments afin de construire et d'améliorer son modèle. L'exactitude d'une recommandation dépend du nombre d'avis fournis par l'utilisateur. Le seul inconvénient de cette méthode est qu'elle nécessite des efforts de la part des utilisateurs et aussi que les utilisateurs ne sont pas toujours prêts à fournir suffisamment d'informations. Bien que les commentaires explicites demandent plus d'efforts de la part de

l'utilisateur, ils sont toujours considérés comme fournissant des données plus fiables [33].

### 1.5.2 Implicite

Le système déduit automatiquement les préférences de l'utilisateur en surveillant ses actions diverses telles que l'historique des achats et l'historique de navigation. Les commentaires implicites réduisent le fardeau des utilisateurs en déduisant les préférences des utilisateurs de leur comportement avec le système. Bien que la méthode ne nécessite aucun effort de la part de l'utilisateur, elle est moins précise. En outre, il a été suggéré que les déclarations de préférence implicites peuvent en fait être plus objectives, car il n'y a pas de biais résultant de la réponse des utilisateurs d'une manière socialement souhaitable et pas de problèmes d'image de soi ou de nécessité de préserver l'image pour les autres [33].

### 1.5.3 Hybrides

Les atouts des retours implicites et explicites peuvent être combinés dans un système hybride pour réduire leurs faiblesses et avoir le système le plus performant [33].

## 1.6 Domaines des systèmes de recommandation

Nous regroupons les domaines de recommandation dans huit domaines principaux [37] : e-government, e-business, e-commerce, e-library, e-learning, e-tourism, e-resource services and e-group. Par exemple dans le e-commerce de nombreux systèmes de recommandation pour les achats en ligne qui fournissent des recommandations aux utilisateurs pour des produits adaptés. Sur ces sites, l'évaluation est très courante, par exemple lorsque l'utilisateur évalue des produits avec un nombre d'étoile de 1 à 5, le marquage est une autre façon de connecter les données des éléments de l'utilisateur. Yandi et al. [52] ont proposé un système de recommandation qui utilise des coupons. En outre, différentes méthodes de recommandation de substituts ont été étudiées et sont bien documentées. McAuley et al. [38] ont examiné la relation entre les substituts et

les compléments en fonction des éléments ayant fait l'objet d'examens et leur coût. Ils ont proposé des méthodes pour clarifier la relation entre substituts et compléments pour un sujet donné. Dans les systèmes de recommandation de bibliothèques électroniques (e-library), les approches de recommandation hybrides sont largement utilisées. Une des raisons d'utiliser des approches hybrides est qu'elles tirent parti des avantages de plusieurs techniques de recommandation différentes. Les techniques floues, en particulier la modélisation linguistique floue multi-granulaire, sont utilisées pour représenter et manipuler les informations flexibles des étiquettes linguistiques [37].

## 1.7 Recommandation des annonces

Les publicités apparaissent sur de nombreux sites, elles ont une grande source de profits pour eux, qu'il s'agisse de sites d'actualités ou de sites de médias sociaux et de divertissement. De nombreuses entreprises paient d'énormes sommes d'argent à ces sites pour que leurs annonces apparaissent sur eux et elles paient plus si leurs annonces génèrent plus de clics de la part des utilisateurs. De nombreuses méthodes ne se soucient pas de l'opportunité pour l'utilisateur de bénéficier du contenu affiché, elles ne font que l'afficher. Rutkowski et al. [44] ont concentré leurs recherches sur les bannières publicitaires qui sont généralement placées sur les sites d'information et ont pris en compte la satisfaction de l'utilisateur à l'égard de l'annonce qui lui apparaît, en utilisant des recommandations et une équité raisonnable quant à la fréquence d'apparition de chaque annonce. Le contenu sémantique et sentimental des publicités a été modélisé sur la base de variables extraites, avec un classificateur de réseau neuronal utilisé pour attribuer des publicités à des groupes qui représentent différents canaux médiatiques [44, 24].

## 1.8 Conclusion

Dans ce chapitre, on a fourni des généralités sur les systèmes de recommandation et quelques définitions nécessaires pour comprendre ce qui suit.

Aussi on a mentionné des travaux développés dans ce domaine envisagent d'utiliser une nouvelle génération de SR liés à la technologie d'apprentissage automatique et d'apprentissage profond.

## Chapitre 2

# LES RÉSEAUX SOCIAUX



### 2.1 Introduction

Avec l'énorme développement technologique que nous avons réalisé ces dernières années, l'arrivée d'Internet dans la plupart des pays du monde et la diffusion des sites des réseaux sociaux, la plupart des entreprises ont tendance à afficher leurs publicités sur ces sites et à payer énormément d'argent à cet effet.

Dans ce chapitre, on va parler des réseaux sociaux et de leurs caractéristiques, et nous présenterons notre site social avec ses différentes fonctionnalités.

### 2.2 Définition

Les réseaux sociaux prennent une place de plus en plus importante dans la vie personnelle et professionnelle. Ils connaissent un engouement auprès d'un public très varié.

Nous pouvons définir également les RS comme étant un ensemble d'individus ou d'organisations, reliés entre eux par des interactions sociales, qui permet aux internautes de créer une page personnelle afin de partager et d'échanger des informations et des vidéos avec leur communauté d'amis et leur réseau de connaissances [27].

### 2.3 Caractéristiques des Réseaux Sociaux

#### 2.3.1 Basés sur l'utilisateur

Avant l'apparition des outils Web 2.0 (incluant les réseaux sociaux), le contenu du site web réside uniquement sur la lecture des informations que seul le webmaster a le droit de les ajouter, les modifier et les supprimer. Nous parlons donc d'un flux d'informations unidirectionnel, mais avec l'appréciation des réseaux sociaux les informations sont postulées modifiées et supprimées par les utilisateurs. Donc, dans ce cas on parle d'un flux d'information bidirectionnel [27].

#### 2.3.2 Interactifs

Les réseaux sociaux ne sont pas que des plateformes de chat et discussions car ils nous donnent la chance de tester d'autres outils de divertissement tels que les applications, les jeux et les conversations vidéo [27].

### 2.3.3 La collectivité

Cette caractéristique est un critère de regroupement des utilisateurs. Cela signifie que les réseaux sociaux sont basés sur un esprit communautaire. Autrement dit, que les utilisateurs forment des communautés ou des groupes selon leurs intérêts Communs [27].

### 2.3.4 Les Relations

Les réseaux sociaux s'améliorent sur les relations. La propagation d'informations dépend du nombre de relations qu'un utilisateur effectue. Autrement dit que les informations se propagent à travers des réseaux de contacts [27].

## 2.4 Fonctionnement des réseaux sociaux

Le fonctionnement d'un réseau social est généralement caractérisé par un même et unique processus : création d'un profil, recherche d'autres profils en relation avec mes propres domaines et enfin la mise en relation (directe ou indirecte). Les commentaires, les e-mails, etc. sont des fonctions optionnelles qui peuvent être intégrées ou non selon le réseau social.

Par exemple les réseaux sociaux de partage n'obligent pas leurs utilisateurs à créer des profils, sans qu'il y ait besoin d'avoir une appartenance à la communauté. Cependant, si l'utilisateur souhaite faire partie de cette communauté, en diffusant ses propres vidéos par exemple, il doit s'identifier avec un profil. Ainsi, l'internaute est reconnu en tant que membre de la dite communauté. Une fois est devenue un membre, il est possible de rechercher des profils, diffuser et partager des contenus. Par contre, il n'y a pas une connexion réelle. Les membres du réseau ne peuvent interagir entre eux qu'en réagissant aux publications partagés par des commentaires [27].

## 2.5 Les principaux réseaux sociaux classés par catégories

Ci-dessous on présente quelques réseaux sociaux classés en trois catégories, à savoir les réseaux sociaux généralistes, les réseaux sociaux professionnels, les réseaux sociaux de blogging et les réseaux sociaux de vidéo [14].

### 2.5.1 Les réseaux sociaux généralistes

#### 1. Facebook<sup>[1]</sup>

Récemment Facebook devient le réseau social le plus étendus au monde, Plus que 400 millions personnes ont connectaient chaque jours. À travers ce réseaux on peut discuter, commenter et publier des informations (photographies, liens, textes, etc.) on peut aussi contrôler leur visibilité (public, privé...). Dans ce réseau chaque personne peut exprimer son soutien à une ou plusieurs marques ou organisations et devenir ainsi un fan de cette dernière. Aussi les pages et groupes permettent de faire connaître des institutions, des entreprises, ou des causes et peuvent être consultés par n'importe quel internaute<sup>[14]</sup>.

En 2021 , le fondateur de Facebook Mark Zuckerberg a annoncé que son site a dépassé plus d'un 2,895 milliards de profil actifs. Selon les statistiques de la même année 1,91 milliards d'utilisateurs actifs journaliers<sup>[19]</sup>.



FIGURE 2.1 – Facebook logo.

#### 2. Twitter<sup>[2]</sup>

Cette plateforme est la plus simple des médias sociaux. Twitter permet à ses utilisateurs d'envoyer gratuitement des messages, appelés tweets sur internet, publier des informations (images, vidéos. etc.) et commenter sur n'importe quelle publication. Ces messages sont limités à 140 caractères pas plus <sup>[14]</sup>.

#### 3. Google+<sup>[3]</sup>

- 
1. <https://www.facebook.com>
  2. <https://www.twitter.com>
  3. <https://www.google.com>



FIGURE 2.2 – Twitter logo.

Les principes des réseaux mentionnés ci-dessus sont combinés dans une seule plateforme. Google a fournis aux internautes un réseau social qui a un petit quelques choses de tout le monde. D'après ce réseau on peut ajouter du nouveau contenu, distinguant des sujets avec des hashtags et même séparer des contacts en cercles. La configuration du profil G+ est optimisée et non couteuse, elle ne prend que quelques minutes pour être configuré. Note : en 2019 Google + est définitivement fermé[14].



FIGURE 2.3 – Google+ logo.

## 2.5.2 Les réseaux sociaux professionnels

### 1. LinkedIn<sup>4</sup>

Dans le domaine de business, LinkedIn est le meilleur réseau électronique

---

4. <https://www.linkedin.com>

orienté vers ce dernier. Ce réseau est un cyberspace ce que les groupes de réseautage ont été pour les communautés d'affaires locales. La fidélité de LinkedIn est préférable pour rester en contact avec ses anciens collègues. Il nous offre un recrutement de nouveaux employés et se tenir au courant des dernières nouvelles de l'industrie et rencontrer des clients et des fournisseurs<sup>[14]</sup>.



FIGURE 2.4 – LinkedIn logo.

### 2. Viadeo<sup>5</sup>

Viadeo est un site de réseautage social en ligne français. Concurrent de LinkedIn, il nous offre une construction de notre propre réseau professionnel et de favoriser le dialogue entre les membres. Viadeo est caractérisé par des méthodes majeures. Ces derniers lui offrent un réassemblage de professionnelles, petites et moyennes entreprises, car les profils des entreprises de moins de 50 salariés représentent près de la moitié du nombre enregistré. Cependant, on peut regretter que la plateforme soit clairement en perte de vitesse<sup>[14]</sup>.

### 2.5.3 Les réseaux sociaux de blogging

#### 1. Medium<sup>6</sup>

Medium est une plateforme de blogs Internet minimaliste conçue pour les longs textes. Le but de ce réseau est de permettre à la plateforme d'offrir une visibilité maximale au texte quel que soit l'auteur, afin de surmonter la difficulté de promotion de son blog. En conséquence, il existe des

---

5. <https://www.viadeo.com>

6. <https://www.medium.com>



FIGURE 2.5 – viadeo logo.

milliers d'articles très utiles et très intéressants, et en même temps très agréables à lire [14].



FIGURE 2.6 – Medium logo.

## 2. **Tumblr**<sup>7</sup>

Cette plateforme héberge essentiellement des micros blogs pour ses utilisateurs. C'est pour ça, il est différent par rapport aux autres réseaux. Elle offre à leurs internautes et entreprises la possibilité de construire de manière très simple et intuitive leurs blogs avec du multimédia (comme des images et des clips vidéo courts). Tumblr est en force dans ce type de réseaux à cause de son principe qui réside principalement dans la réticulation augmentant ainsi la virilité du contenu [14].

---

7. <https://www.tumblr.com>



FIGURE 2.7 – Tumblr logo.

## 2.5.4 Les réseaux sociaux de vidéo

### 1. Youtube<sup>8</sup>

À cause de ses avantages, YouTube est le meilleur moteur de recherche de vidéos. Les utilisateurs de YouTube peuvent partager, évaluer et commenter toutes les vidéos qu'ils visionnent sur la plateforme.

YouTube est le leader incontestable des réseaux sociaux de vidéos. C'est pour cela il a été déjà racheté par Google à cause de ses bénéfices et sa capacité de stocker des milliards de vidéos<sup>[14]</sup>.



FIGURE 2.8 – Youtube logo.

### 2. TikTok<sup>9</sup>

En 2016, les chinois ont lancé une application avec le nom de TikTok. Ce dernier est un réseau social dans le service de partage de vidéos, qui

---

8. <https://www.youtube.com>

9. <https://www.tiktok.com>

permet aux utilisateurs de regarder, créer, éditer et partager des clips musicaux sous format d'une vidéo avec une taille limitée (de quelques secondes) [14].



FIGURE 2.9 – Tiktok logo.

### 2.6 La publicité sur les réseaux sociaux

Aujourd'hui la présence sur les réseaux sociaux est indispensable pour toute entreprise, organisation, institution, etc. Le marketing électronique est vu comme étant un outil de communication idéal. Il offre aux entreprises la possibilité de créer et engager une communauté, acquérir de nouveaux clients et augmenter le trafic sur leurs sites internet.

Pour rendre vos campagnes plus performantes, la publicité sur les réseaux sociaux ou la publication de contenus sponsorisés vous permet de mettre en place des actions de diffusion publicitaires de contenus sur les différents réseaux.

Les réseaux sociaux permettent de travailler sur votre visibilité dans une démarche prospective. La publicité peut vous aider dans vos objectifs de vente ! Quand tu fais une publicité, les campagnes publicitaires seront un véritable tremplin concernant vos ventes, car elles génèrent du trafic sur le site de votre entreprise. Ce trafic attire plus de visiteurs, et donc augmente vos chances de conversion en nouveaux clients. La publicité sur les réseaux sociaux permet de cibler les personnes que vous souhaitez toucher, selon certains critères. Chaque réseau social dispose de sa propre régie publicitaire et offre des services adaptés à sa plateforme [13].



## 2.7 Conclusion

Presque tout le monde a un ou plusieurs comptes sur l'un des sites de réseaux sociaux, et c'est ce qui en fait un bon marché pour les propriétaires de publicité. Mais, il est nécessaire de faire attention à la façon dont ces publicités sont publiées et bien promues auprès des utilisateurs.

Dans ce chapitre, nous avons passé en revue brièvement les sites de réseautage social et leurs principes.

# Chapitre 3

## CONCEPTION

### 3.1 Introduction

L'objectif de ce chapitre est de présenter les étapes de conception de notre système de recommandation. En fait, nous avons proposé deux approches, la première est basée sur l'apprentissage automatique (K-means) et la seconde est basée sur l'architecture arborescente (KD-Tree).

### 3.2 Collection de données

L'un des problèmes auxquels nous avons confronté est l'incapacité d'obtenir les données des utilisateurs (leurs commentaires, les pages qu'ils préfèrent, etc.) à partir de sites comme Facebook et Twitter en raison de la politique de confidentialité et si nous utilisons la méthode de web scraping, nous violons les lois de ces sites et la loi sur la confidentialité, nous avons donc développé Hazmit .

Hazmit est un réseau social que nous avons créé afin de collecter des données sur les intérêts de personnes de différentes sociétés, compte tenu de la difficulté d'accéder aux bases de données des grands réseaux tels que Facebook et Twitter, et l'absence d'une base de données adaptée à nos recherches.

Nous avons distribué notre site Web à des personnes de 3 wilayas : Guelma, Annaba et Skikda et nous leur avons proposé de s'inscrire avec la garantie que leurs données ne seraient exposées à aucune forme d'exploitation non académique de notre part. Grâce à leur coopération et leur interaction avec les publications et les publicités que nous avons affichées afin de découvrir leur intérêt nous avons créé des vecteurs de préférences pour chacun.

### 3.3 Les caractéristiques de notre site web

Notre site Web se caractérise par une interface simple et une facilité d'utilisation car il contient les fonctionnalités de base de tout autre site de médias sociaux où l'utilisateur crée sa page personnelle en saisissant quelques informations régulières. Après son inscription, l'utilisateur peut bénéficier des services du site comme l'ajout d'amis et la discussion instantanée. Les utilisateurs peuvent également publier sur le site, parcourir les publications d'autres personnes, les aimer ou pas, et les commenter également.

Léger, fluide et sûr à utiliser, vous pouvez également l'utiliser en anglais, arabe il garantit que ses utilisateurs ne compromettent pas leurs données privées

à des fins autres que l'étude.

## 3.4 Architecture du système de recommandation proposé

### 3.4.1 Kmeans + elbow architecture

Le clustering K-means est de simples algorithmes d'apprentissage automatique non supervisés.

les algorithmes non supervisés font des inférences à partir d'ensembles de données en utilisant uniquement des vecteurs d'entrée sans se référer à des résultats connus ou étiquetés.

l'algorithme K-means identifie k nombre de centroïdes, puis alloue chaque point de données au cluster le plus proche, tout en gardant les centroïdes aussi petits que possible.

Les « moyennes » dans les K-moyennes font référence à la moyenne des données, c'est-à-dire à la recherche du centroïde [11].

#### 3.4.1.1 DATASET

L'ensemble de données est collecté à partir de Hazmit<sup>1</sup>, chaque vecteur de données comprend 20 criteres, chacune est évaluée explicitement par plus de 200 utilisateurs. Chaque utilisateur inscrit détermine ses préférences en exprimant son appréciation ou son aversion aux publications proposées. la table 3.1 montre un exemple de descripteur de données.

---

1. <http://servicedzone.com/fr/Hazmit/>

id	24	identificateur
pseudo	DJA	pseudo
c	23	Ville(code wilaya d'utilisateur)
s	-1	sexe (-1 femme/1 homme)
sp	-1	sport annonce rating
pol	-1	politic annonce rating
f	1	fish annonce rating
sw	1	Sandwitch
pz	1	pizza
v	1	végétarien
sam	1	Samsung
op	1	Oppo
App	1	Apple
x	1	Xioami
an	1	Anime
act	-1	Action films
rom	1	Romantic
dr	-1	Drama
com	-1	Comedy
si	1	Science imagination
doc	1	Documentary

TABLE 3.1 – Exemple de descripteur de données

Le descripteur illustré à la table (3.1) comprend plus de 20 caractéristiques dans la plage de -1 à 1 où les valeurs sont normalisées comme suit :

- la valeur 1 correspond à une caractéristique aimée par l'utilisateur.
- la valeur -1 correspond à une caractéristique non aimée par l'utilisateur.

### 3.4.1.2 Détection des communautés

Il y a trois étapes pour détecter et diviser notre ensemble de données en plusieurs communautés nous commençons par la première étape qui s'agit du prétraitement du jeu de données (supprimer les utilisateurs qui n'ont aucune réaction avec nos annonces).

La seconde a pour objectif de trouver le nombre de communautés à créer à partir de l'ensemble de données.

Le but de la dernière étape est de classifier notre jeu de données à l'aide d'une méthode de clustering (Kmeans).

Nous avons appliqué l'algorithme k-Means pour créer différents communautés d'intérêts. Afin de définir le nombre de clusters, nous avons utilisé la méthode Elbow [38].

### Méthode Elbow pour la valeur optimale de k en K-means

Elbow est utilisée avec erreur quadratique (sse) ou avec cluster somme (wcss). Nous avons utilisé wcss pour choisir le nombre de clusters.

**WCSS** peut être défini comme une fonction d'objectif implicite qui permet de donner le bon nombre de centres de gravité ou de clusters à inclure dans l'ensemble de données [1].

En pratique, il peut ne pas y avoir de coup de coude aigu, et en tant que méthode heuristique, un tel "coude" ne peut pas toujours être identifié sans ambiguïté [33].

l'ensemble de données est divisé en 4 catégories d'annonces qui sont : les films, les téléphones, les actualités et les aliments. Par exemple, dans les publicités sur les téléphones, 4 marques célèbres sont choisies : Samsung, oppo, apple et xiaomi.

Après avoir utilisé la méthode Elbow, les nombres de clusters sont choisis et leurs centres sont calculés.

#### 3.4.1.3 Recommandation

Chaque nouvel utilisateur devra introduire la ville et le sexe puis ces informations seront utilisées pour détecter la communauté à laquelle il appartient en calculant la distance euclidienne entre ces informations et les centres des clusters.

Après cela, des publicités vidéo sont créées et chaque publicité est recommandée à une communauté cible. Ensuite, les utilisateurs sont invités à exprimer leur satisfaction sur ces annonces.

Categorie	les téléphones	les aliments
nombre de clusters choisis	5	4

Table 3.2 continued from previous page

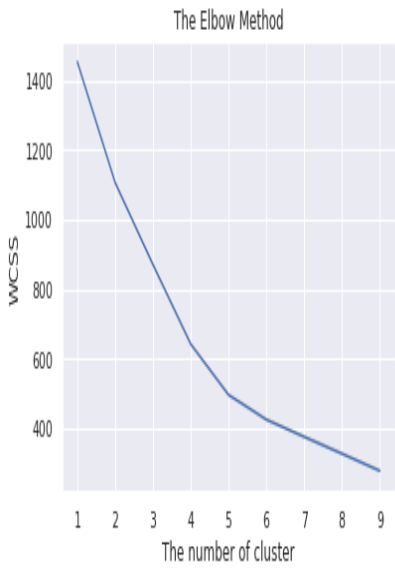
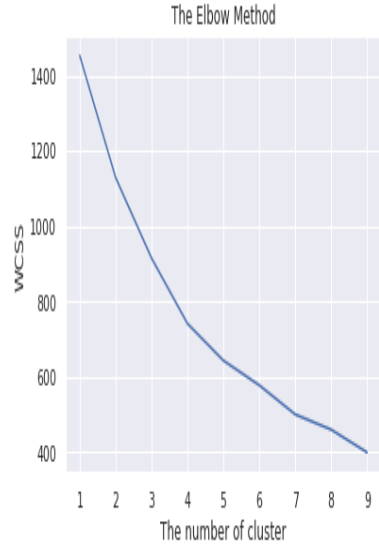
Categorie	les téléphones	les aliments
Résultat de la méthode Elbow		

TABLE 3.2: deux exemple de la méthode du Elbow appliquée à la catégorie téléphone et alimentation

Notez que tous les utilisateurs d’une même communauté ont des goûts communs. Donc lorsqu’un nouvel utilisateur arrive, il suffit d’identifier sa communauté puis de lui proposer ce que cette communauté aime.

Pour identifier sa communauté nous pouvons simplement calculer les distances avec chaque centre de cluster et l’utilisateur et détecter la distance min, par exemple s’il s’agit d’une distance min avec le cluster 1 donc nous recommanderons les annonces les plus appréciées par le cluster 1.

Prenons l’exemple avec la catégorie téléphones :

Nous avons appliqué Kmeans avec  $K=5$ (tableau 3.2) puis nous avons obtenu 5 clusters avec 5 centres, et par exemple lorsqu’un nouvel utilisateur homme de skikda se presente alors( sexe=1 et ville=-1). Ensuite, nous calculons la Distance Euclidienne entre les centres et les préférences du nouvel utilisateur en utilisant la distance euclidienne.

Laissez le point  $p$  avoir des coordonnées cartésiennes  $(p_1, p_2)$ .

Et laissez le point q ont des coordonnées  $(q_1, q_2)$ .

Alors la distance entre p et q est donnée par [43] :

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (3.1)$$

N° centre	cardinalité des centres	utilisateur	Distance euclidienne
1	( 0.80311781, 0.67545023)	(-1,1)	1.83
2	( 0.00422124, 0.07932574)	(-1,1)	1.36
3	( -0.01280443,-0.77705977)	(-1,1)	2.03
4	( -0.32308249, -0.20408812)	(-1,1)	1.38
5	( -1.05562639,-0.2137995 )	(-1,1)	1,21 (valeur minimale )

TABLE 3.3: exemple de calcul de la distance entre le vecteur utilisateur et les centres des clusters

Comme le tableau [3.3] le montre, distance minimale est avec le centre 5, nous suggérerons donc les annonces aimées par la communauté 5 (cluster5).

Le cluster 5 est représenté dans ce tableau [3.4] ci-dessous ci-dessous :

id	wilaya	sexe	Samsung	OPPO	apple	Xiaomi
58	guelma	femme	1	-1	-1	1
53	guelma	femme	1	-1	-1	1
25	guelma	femme	1	-1	-1	1
194	guelma	femme	1	-1	-1	1
89	guelma	femme	1	-1	-1	1
224	annaba	femme	1	-1	-1	-1
68	guelma	femme	1	-1	-1	1
139	guelma	homme	1	-1	-1	-1
34	guelma	femme	1	-1	-1	1
43	guelma	femme	1	-1	-1	1
258	annaba	femme	1	-1	-1	1
189	guelma	femme	1	-1	-1	1
92	guelma	femme	1	-1	-1	1
184	guelma	femme	1	-1	-1	1
190	guelma	femme	1	-1	-1	1
95	guelma	femme	1	-1	-1	1
128	guelma	femme	1	-1	-1	1
87	guelma	homme	1	-1	-1	-1
129	guelma	femme	1	-1	-1	1
261	annaba	femme	1	-1	-1	1
157	annaba	femme	1	-1	-1	-1
63	guelma	femme	1	-1	-1	1
48	guelma	femme	1	-1	-1	1
209	annaba	homme	1	-1	-1	1
85	guelma	femme	1	-1	-1	1
202	guelma	femme	1	-1	-1	1
81	guelma	femme	1	-1	-1	1
90	guelma	femme	1	-1	-1	1
213	guelma	femme	1	-1	-1	-1
153	guelma	femme	1	-1	-1	-1
26	guelma	femme	1	-1	-1	1
33	guelma	femme	1	-1	-1	1



id	wilaya	sexe	Samsung	OPPO	apple	Xiaomi
181	guelma	femme	1	-1	-1	1
50	guelma	femme	1	-1	-1	1
45	guelma	femme	1	-1	-1	1
204	guelma	femme	1	-1	-1	1
269	annaba	femme	1	-1	-1	1
123	annaba	femme	1	-1	-1	-1
183	guelma	femme	1	-1	-1	1
244	annaba	femme	1	-1	-1	-1
56	guelma	femme	1	-1	-1	1
174	guelma	femme	1	-1	-1	1
Total	/	/	100% likes	0% likes	0% likes	80.95% likes

TABLE 3.4: Les caractéristiques du cluster numéro 5.

Le pourcentage de différence de likes pour le cluster 5 :4,76%.

On le calcule par cet algorithme :

Si le pourcentage des likes d'une annonce  $\geq 50\%$  alors

erreur = 100 - Pourcentage des J'aime de l'annonces.

Sinon si le pourcentage des likes  $< 50\%$  alors

erreur = Pourcentage des j'aimes.

Ensuite, nous calculons la somme des erreurs de toutes les annonces et divisons par le nombre d'annonces existantes qui est 4 (Samsung, oppo..).

Par exemple dans le cluster 5 :

Les publicités sur Samsung ont 100% likes ( $\geq 50\%$ )  $100-100 = 0$  donc nous avons 0% d'erreur.

Les publicités sur Oppo ont 100% likes ( $< 50\%$ ) donc nous prenons 0 et ici nous aurons 0% d'erreur.

Les publicités sur Apple ont 100% likes ( $< 50\%$ ), nous prenons donc 0 et on aura à nouveau 0% d'erreur.

Les publicités sur Xiaomi ont 80,95% likes ( $\geq 50\%$ )  $100-80,95 = 19,05$  donc nous aurons 19,05% d'erreur.

Ensuite, nous divisons la somme d'eux sur 4 :  $(0+0+0+19.05)/4 = 4,76\%$  et c'est le pourcentage de différence de likes pour le cluster 5.

Nous pouvons également obtenir le pourcentage de différence de goûts pour tous les clusters (voir table 3.5) en calculant leur somme pour tous et en divisant le resultat par 5 (le nombre total de clusters existants).

le pourcentage de différence de likes pour tous les clusters est égal à 7,86% et c'est un bon rapport de similarité en général.

Pour la recommandation, nous proposons à cet utilisateur les publicités des téléphones Samsung (100% de la communauté aime ça) ou Xiaomi (80.95% de la communauté aime ça) mais nous prenons le maximum entre cas deux valeurs

cluster	pourcentage de différence (%)	Note
1	0.5	Similitude parfaite
2	0	Similitude parfaite
3	25	Pas trop de similitude
4	8.14	Bonne similitude
5	4.76	très bonne similitude
Total / 5	$(0.5+0+25+8.14+4.76)/5 = 7.86$	bonne similitude

TABLE 3.5 – le pourcentage de différence de goûts pour tous les clusters

et nous suggérons des publicités sur samsung, à noter que nous recommandons pas des publicités sur Apple ou Oppo.

La même technique sera appliquée pour les autres catégories (movies, news, et food).

La figure 3.1 represent les etapes de notre algorithme.

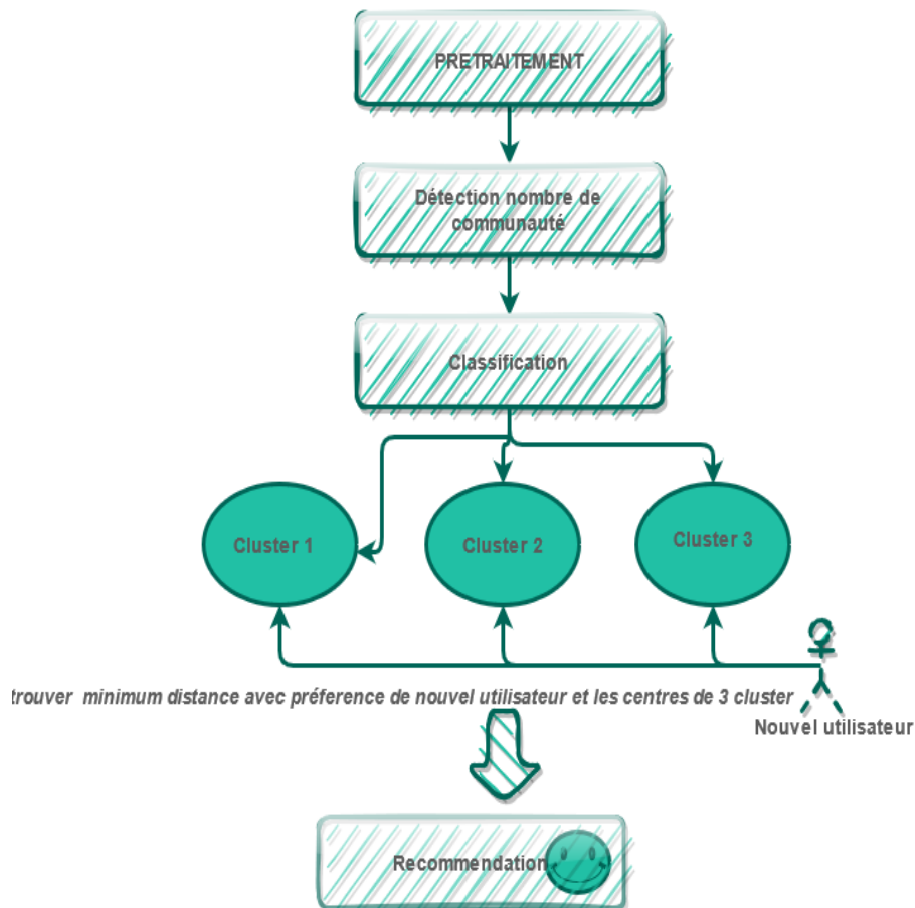


FIGURE 3.1 – Les etapes de notre algorithme

### 3.4.2 Notre KDTree algorithme

L'algorithme d'arbre KD est l'un des algorithmes de voisin le plus couramment utilisés. Les points de données sont divisés à chaque nœud en deux ensembles. L'arbre KD est également un algorithme d'arbre binaire se terminant toujours par un maximum de deux nœuds. Les critères de partage choisis sont souvent la médiane [10]. la figure 3.2 représente la position exacte des points et leur position spatiale

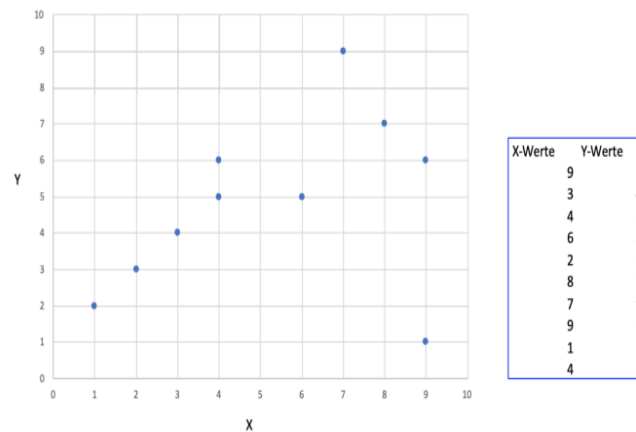


FIGURE 3.2 – Points et leur position dans un système de coordonnées [10].

L'algorithme KD-Tree utilise d'abord la médiane du premier axe puis, dans la deuxième couche, la médiane du deuxième axe. Nous allons commencer par l'axe X. Les valeurs x triées par ordre croissant sont : 1,2,3,4,4,6,7,8,9,9. Ensuite, la médiane est de 6. Le dessin de la médiane de 6 dans le système de coordonnées nous montre les deux clusters visualisés (voir figure 3.3).

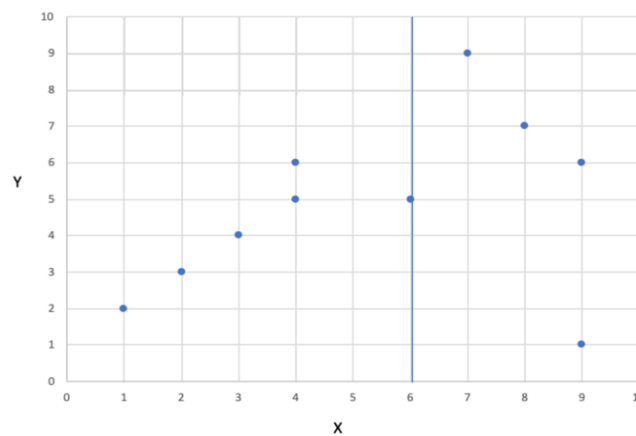


FIGURE 3.3 – Dessiner la médiane x de 6 dans le système de coordonnées [10].

Nous allons maintenant utiliser l'axe Y. Nous avons déjà deux clusters, nous devons donc les examiner séparément.

Sur le côté gauche, nous avons les valeurs y triées : 2,3,4,5,6. La médiane est alors de 4. Cela conduit à une ligne de séparation à la valeur 4.

Sur le côté droit, nous avons les valeurs y triées : 1,5,6,7,9. La médiane est alors de 6. Cela conduit à une ligne de séparation à la valeur 6.

Le système de coordonnées final résultant peut être vu sur la figure 3.4. Les points de données ont été divisés en 4 clusters avec une profondeur de 2 axes.

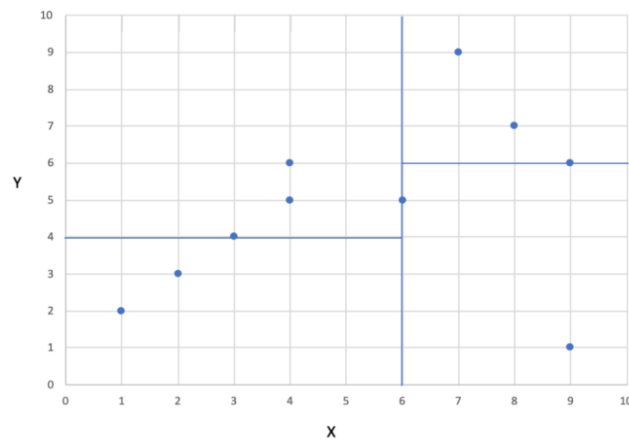


FIGURE 3.4 – La séparation spatiale finale [10].

### 3.4.2.1 DATASET

En raison du nombre modeste des utilisateurs de différentes wilayas, nous avons procédé à collecter un deuxième ensemble de données qui contient plus d'informations telles que l'âge et nous avons essayé également d'enrichir la base par des utilisateurs qui habitent d'autres villes algériennes.

L'ensemble de données est collecté à partir de Hazmit, chaque descripteur comprend 5 caractéristiques, chacune est évaluée explicitement par 171 utilisateurs 122 sont des hommes et 51 femmes de différentes wilayas (nord, sud, est et ouest) et âgé entre 18 et 62.

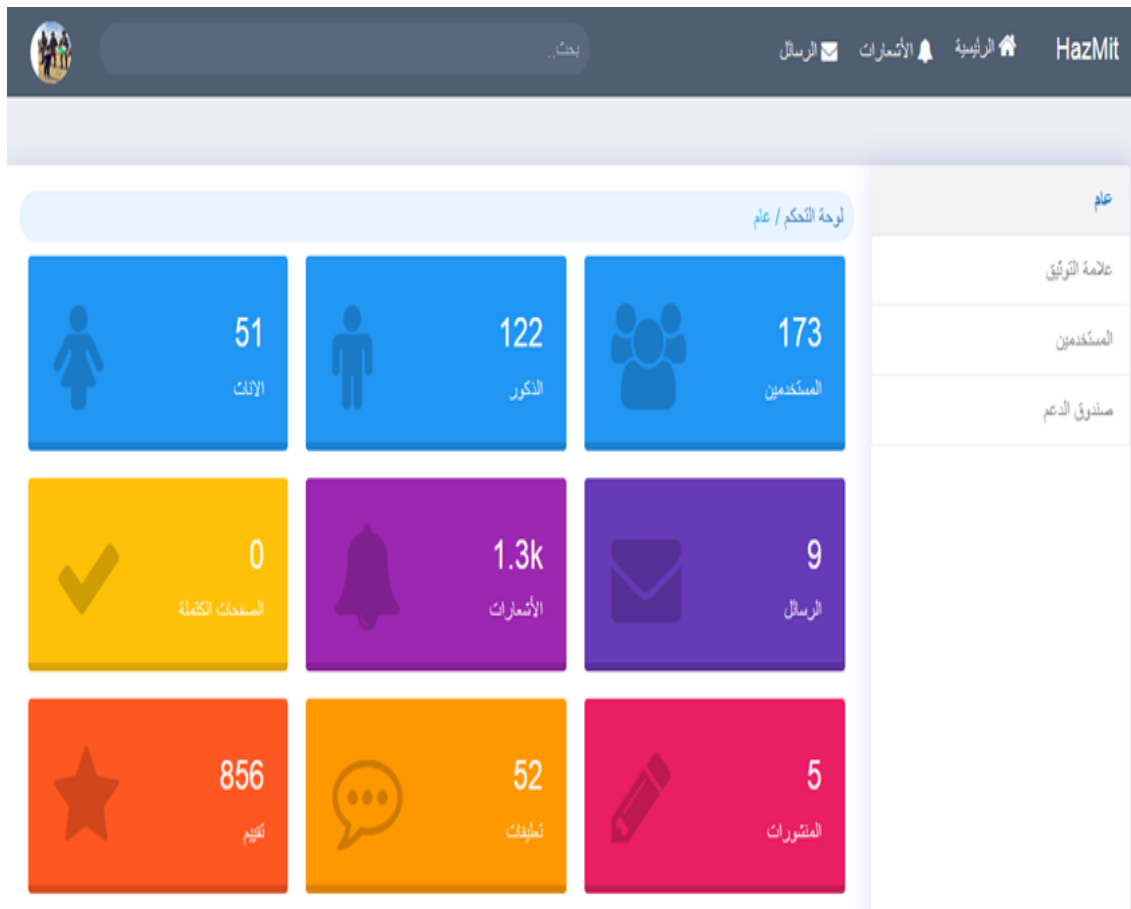


FIGURE 3.5 – l'ensemble de données présenter par le dashboard

iduser	age	sexe	east	north	coast	tech ad	food ad	job offres	clothes ads	news ads
1	22	1	1	1	0	4	3	2	5	1
2	21	1	1	1	0	4	3	2	5	1
3	25	1	1	1	0	4	3	2	5	1
4	25	1	1	1	0	4	3	2	5	1
5	26	1	1	1	0	4	3	2	5	1
...	...	...	...	...	...	...	...	...	...	...
167	41	1	1	1	1	2	3	0	3	4
168	48	0	1	1	1	2	3	1	5	5
169	45	0	1	1	1	3	3	1	5	3
170	51	0	1	1	1	2	3	0	5	4
171	51	0	1	1	1	2	3	1	5	5

FIGURE 3.6 – notre deuxième ensemble de données (dataset enrichie)

Dans l’objectif de partitionner les utilisateurs du système en plusieurs communautés d’intérêts, nous avons utilisé une approche de décomposition spatiale basée sur un arbre KD. Un arbre à  $k$  dimensions (KD) [22] est un arbre binaire dans lequel chaque nœud est un point à  $k$  dimensions. Il peut être utilisé pour partitionner un ensemble d’objets dans un plan 2D et un espace 3D.

Une application des arbres KD dans le domaine des systèmes de recommandation peut être trouvée dans la Réf [29], où les auteurs ont proposé un algorithme de recommandation et de localisation qui partitionne les données des utilisateurs sur la base d’une structure de données KD-tree.

Le travail proposé par Das et al. en 2019 [30] consiste à utiliser les KD-tree et les Quadtree pour partitionner l’ensemble d’utilisateurs en clusters plus petits sur la base de l’emplacement et appliquer des algorithmes de recommandation basés sur CF séparément pour chaque clusters.

### 3.4.2.2 Algorithme de construction de KDTree

Dans cette section nous avons utilisé la deuxième dataset (DS2).

Le DS2 a été conçu à 80% pour la construction KD-Tree et à 20% pour le test, et après avoir appliqué différentes étapes de construction, une communauté différente est créée dans chaque nœud, par exemple au niveau 1, deux communautés se distinguent par leur âge : la plus âgée que la moyenne et la plus jeune que la moyenne. Pour le niveau 2, quatre communautés peuvent être distinguées selon le sexe et l'âge qui sont : les hommes plus jeunes que la moyenne, les femmes plus âgées que la moyenne, les hommes plus âgés que la moyenne et les femmes plus jeunes que la moyenne. Après cela, au niveau 3, chaque communauté est conçue par l'emplacement nord-est, le sexe et l'âge en 8 communautés différentes. Par la suite, l'efficacité de cette structure dans la recherche de la communauté adéquate à laquelle appartiennent les utilisateurs est testée. Le KD-Tree a une complexité de recherche de  $O(\log n)$  dans le cas moyen ce qui représente un gros gain de temps. Pour la classification, l'âge, le sexe et l'emplacement de l'utilisateur sont introduits avec le nombre  $K$ . Par exemple, si  $k=5$ , le système renverra les 5 utilisateurs les plus proches de l'utilisateur actuel. Le principe de recherche consiste à considérer que la moyenne d'âge est de 33 ans. Ainsi, étant donné que l'âge de l'utilisateur est de 25 ans, l'algorithme recherche à gauche de l'arbre et si l'utilisateur est une femme l'algorithme recherche à droite du niveau 2.

notre structure KD-tree est représenté sur la figure [3.7](#)

Entrée : wilaya, âge, sexe des utilisateurs.

Sortie : KD-Tree.

1. Construisez l'arbre complet jusqu'à un niveau spécifié (4 niveaux).
2. Divisez la région entière (niveau 0) en deux parties (cellule gauche et cellule droite) avec la valeur moyenne de l'âge.

2.2 Au niveau suivant (niveau 1), divisez à nouveau les deux cellules individuellement en deux parties selon le sexe des utilisateurs.

2.3 Au niveau suivant (niveau 2), divisez à nouveau chacune des quatre cellules individuellement en deux parties selon la position de la wilaya est à gauche et ouest à droite.

2.4 Au niveau suivant (niveau-3), divisez à nouveau chacune des huit cellules individuellement en deux parties selon la position de la wilaya nord à gauche et sud à droite.

2.5 Au niveau suivant (niveau-4), divisez à nouveau chacune des seize cellules individuellement en deux parties selon la position de la wilaya cote à gauche et intérieure droite.

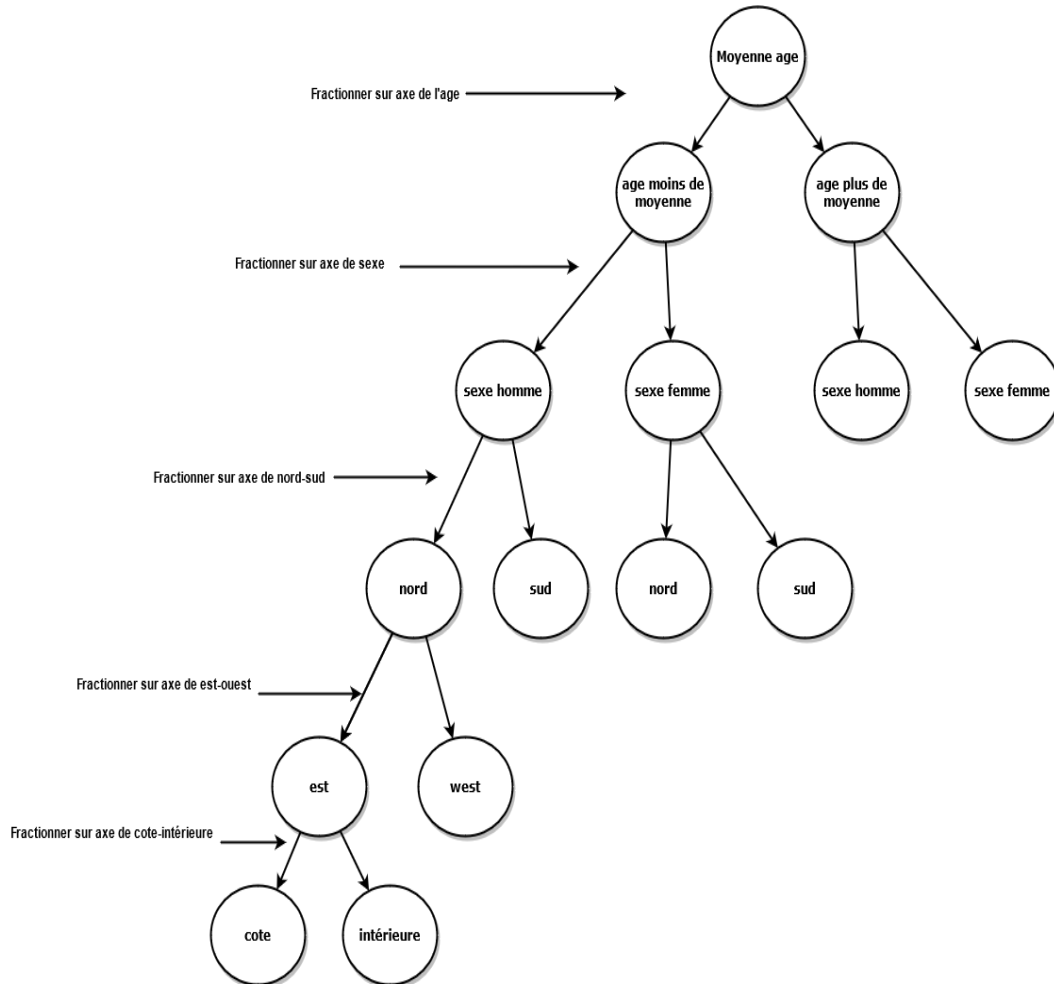


FIGURE 3.7 – notre structure KD-tree.

### 3.4.2.3 Algorithme de recommandation

Maintenant, en utilisant la structure Kd-tree, les résultats sont présentés sur la base d'un exemple concret qui est l'utilisateur 9. Donc le système, recherche dans l'arbre les 3 les plus proches de l'utilisateur 9 dans la même communauté puis la somme des notes données à chaque catégorie d'annonces est calculé et conçu par 3, par exemple si les utilisateurs les plus proches attribuent les notes 3, 5 et 2 aux annonces technologiques, la note prévue des annonces technologiques pour l'utilisateur actuel sera  $(3 + 5 + 2) / 3 = 3.33$ .

Entrée : wilaya, âge et le sexe de l'utilisateur.

Sortie : recommandation.



1. Recherchez dans l'arbre les 15 proches à cet utilisateur.
2. Calculez la moyenne de chaque annonce notée par les 15 plus proches utilisateurs.
3. Affichez les notes recommandées.

### 3.4.3 Factorisation matricielle probabiliste

#### La matrice R

Supposons que nous avons un ensemble d'utilisateurs  $u_1, u_2, u_3 \dots u_N$  qui évaluent un ensemble d'éléments  $v_1, v_2, v_3 \dots v_M$ . Nous pouvons alors structurer les notations comme une matrice  $R$  de  $N$  lignes et  $M$  colonnes, où  $N$  est le nombre d'utilisateurs et  $M$  est le nombre d'éléments à évaluer. La matrice  $R$  est creuse. Seules certaines de ses cellules auront une valeur d'évaluation non vide, tandis que d'autres ne l'auront pas. Pour un utilisateur  $A$  donné, le système doit pouvoir proposer des recommandations d'items en fonction de ses préférences ainsi que des choix effectués par des utilisateurs similaires. L'utilisateur  $A$  n'a pas besoin d'avoir explicitement évalué un élément particulier pour qu'il soit cibler par des recommandations. D'autres utilisateurs ayant des préférences similaires compenseront les données manquantes sur l'utilisateur  $A$ . Par conséquent, la factorisation matricielle probabiliste entre dans la catégorie des systèmes de recommandation basée sur le filtrage collaboratif. La matrice  $R$  de notre ensemble de données est représentée sur la figure [3.8](#).

[4	2	5	1	3]
[4	2	0	1	3]
[0	2	5	1	3]
[4	2	0	1	3]
[5	2	5	0	3]
[4	0	4	1	0]
[0	2	4	0	3]
[0	1	5	0	0]

FIGURE 3.8 – La matrice  $R$  de l'ensemble de données

$$R = U^T V \tag{3.2}$$

$$U^T = D * M \tag{3.3}$$

Tels que  $M$  est le nombre d'utilisateurs et  $D$  est l'intervalle des valeurs comprises entre 1 et 171 il correspond aux utilisateurs du système.

$$V = D * N \tag{3.4}$$

Telle que  $N$  est le nombre d'items et  $D$  est un intervalle des valeurs qui sont comprises entre 1 et 5 types d'annonces (news, clothes, food, tech et job).

### Initialisation et mise à jour des paramètres

Afin d'initialiser  $V$  nous tirons des nombres aléatoires d'une gaussienne de moyenne nulle avec un écart-type de  $1/\lambda V$ .

Ensuite, pour mettre à jour  $U$  et  $V$ , nous utilisons les équations suivantes :

$$\begin{aligned} U_i &= \left[ (V_j V_j^\top)_{j \in \Omega_{V_i}} + \lambda_U I \right]^{-1} (R_{ij} V_j^\top)_{j \in \Omega_{U_i}} \\ V_j &= \left[ (U_i U_i^\top)_{i \in \Omega_{V_j}} + \lambda_V I \right]^{-1} (R_{ij} U_i^\top)_{i \in \Omega_{V_j}} \end{aligned} \tag{3.5}$$

## 3.5 Conclusion

La phase conceptuelle est une étape fondamentale pour la réalisation de n'importe quel projet. Elle permet de faciliter la réalisation et l'implémentation du projet. Cela nous a permis d'une part de répondre aux questions se rapportant à la réalisation du système à développer et d'autre part à rechercher les moyens et les outils pour la concrétisation de notre système et le traitements du dataset collectée que nous présenterons dans le chapitre suivant.

# Chapitre 4

## Implémentation et résultats

## 4.1 Introduction

L'objectif de ce chapitre est de présenter les étapes de l'implémentation de l'approche proposée dans le cadre d'un système de recommandation des annonces aux différentes communautés d'intérêts. Nous commençons par la présentation de notre site Hazmit. Les ressources utilisées pour créer le site, et les interfaces graphiques. Ensuite le traitement des données collectées, et les résultats obtenus. Alors, ce chapitre est composé de trois parties à savoir : la présentation de notre site, le traitement des données collectées, la recommandation et les résultats expérimentaux.

## 4.2 Présentation de notre site

### 4.2.1 Environnement de développement

Nous avons opté pour le langage PHP manipulé via Sublime Text et Easy-PHP 17, pour tout ce qui est modélisation UML nous avons utilisé Bouml, pour la base de données nous avons utilisé phpMyAdmin.

#### Langage PHP [7] (Hypertext Preprocessor)

nous avons utilisé un langage de programmation libre, principalement utilisé pour produire des pages Web dynamiques via un serveur HTTP, mais pouvant également fonctionner comme n'importe quel langage interprété de façon locale. PHP est un langage impératif orienté objet.

#### HTML [12] (Hypertext Markup Language), CSS [12] (Cascade Style Sheet)

Nous avons utilisé le HTML, CSS pour le design de l'interface graphique et la bibliothèque BOOTSTRAP pour réduire l'écriture des codes CSS et aussi pour rendre le site adaptée à tous les appareils et machines.

#### JavaScript [3]

Pour rendre nos pages interactives nous avons utilisé le langage JavaScript et son bibliothèque JQuery.

#### MySQL [4]

Est un système de gestion de bases de données relationnelles (SGBDR). Il fait partie des logiciels de gestion de base de données les plus utilisés au monde vu de sa simplicité.

### 4.2.2 Les interfaces graphiques

Comme nous l'avons mentionné précédemment, les interfaces de notre site sont simples et empruntées à l'ancien design de Facebook comme il est présenté par les figures 19 et 20. L'utilisateur peut s'inscrire sur le site en saisissant son nom d'utilisateur, nom complet, email, mot de passe et wilaya, et il peut également ajouter quelques informations telles que le lieu de travail ou d'étude lors de la création de sa page personnelle.

La page d'accueil du site contient la barre de navigation qui à son tour permet à l'utilisateur de naviguer entre les différentes interfaces des sites (sa page et les pages personnelles des utilisateurs). L'utilisateur peut également visualiser ses messages, afficher les notifications qu'il reçoit ou rechercher une page ou un ami. Il contient également un formulaire dans lequel il peut ajouter ses propres publications.

En bas de page, les différentes publications sont affichées en fonction de l'heure de leur publication, et l'utilisateur peut interagir facilement avec ces publications : il peut poser un like ou dislike, la partager ou la commenter.

Sur le côté gauche, il y a les pages populaires et les articles les plus interactifs. De l'autre côté, on trouve les moyens de paramétrage du site telles que la modification du profil, de la langue, etc.

The image displays two side-by-side screenshots of a website's user interface, specifically the registration and login pages.

**Left Panel (Login Page):**

- Title: تسجيل الدخول الآن (Login Now)
- Input fields: البريد الإلكتروني أو اسم المستخدم (Email or Username) and كلمة السر (Password).
- Text: نيت كلمة السر? (Forgot Password?)
- Button: تسجيل دخول (Login)
- Footer: لا تمتلك حساب؟ إنشاء حساب مجاناً. (Don't have an account? Create a free account.) العربية • English

**Right Panel (Registration Page):**

- Title: إنشاء حساب جديد (Create New Account)
- Input fields: الاسم الكامل (Full Name), إسم المستخدم (باللغة الإنجليزية) (Username in English), البريد الإلكتروني (Email), كلمة السر (Password), تأكيد كلمة السر (Confirm Password), and جنس (Gender).
- Dropdown menu: دولة الولاية (State) with 'Guelma' selected.
- Text: بالضغط على إنشاء الحساب، أنت توافق على الشروط سياسة الخصوصية و استخدام ملفات تعريف الارتباط (By clicking on Create Account, you agree to the Terms of Service, Privacy Policy, and Cookies Policy).
- Button: إنشاء الحساب (Create Account)
- Footer: بالفعل تمتلك حساب؟ تسجيل الدخول الآن. (Already have an account? Login now.) العربية • English

FIGURE 4.1 – les pages d'inscription et de connexion de notre site.

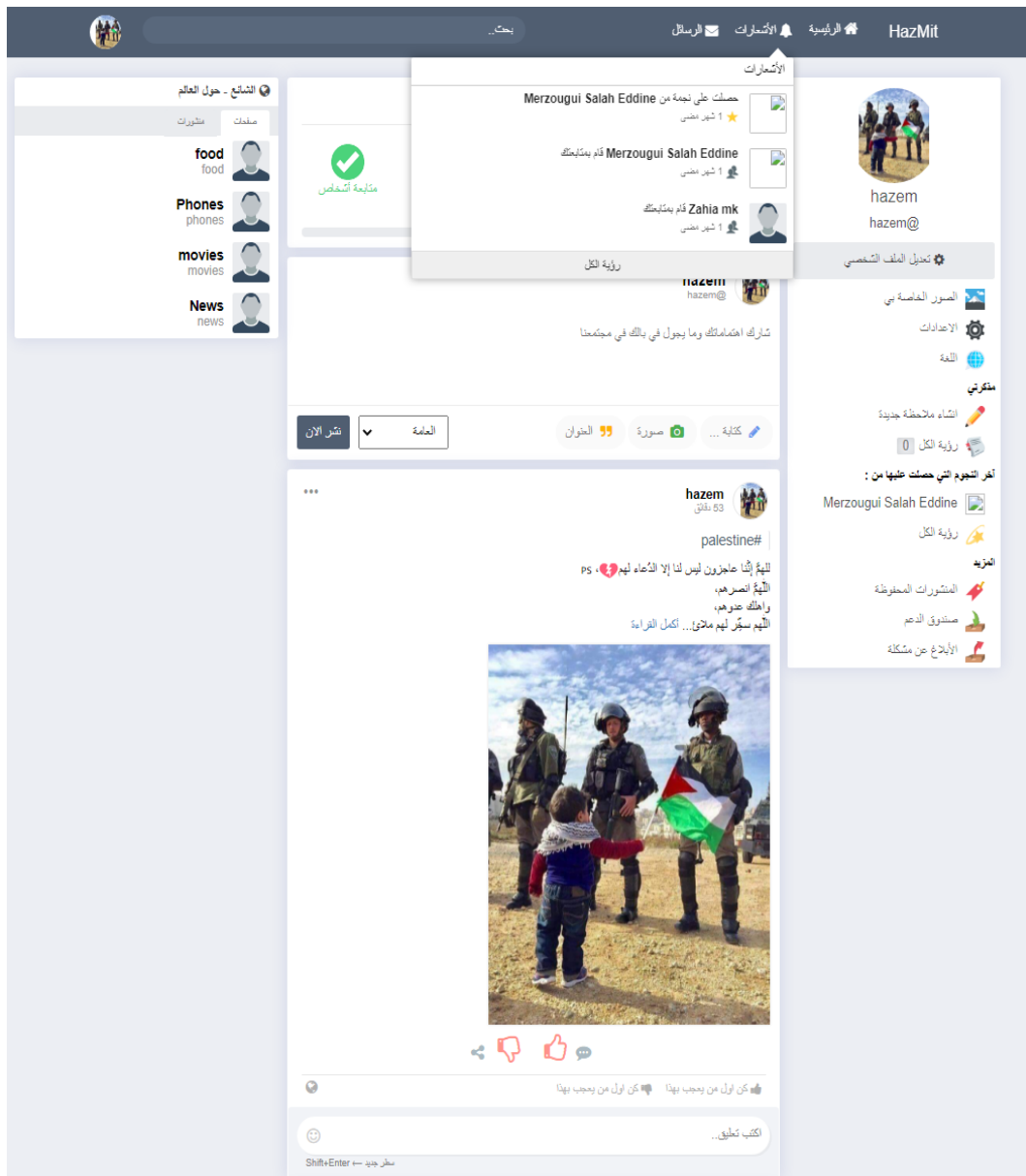


FIGURE 4.2 – la page accueil

## 4.3 Traitement des données collectées

### 4.3.1 Environnement de développement

#### GOOGLE COLAB

C'est un produit de Google Research<sup>1</sup>. Colab permet à n'importe qui d'écrire et d'exécuter le code Python de son choix par le biais du navigateur. C'est un environnement particulièrement adapté à l'apprentissage automatique (machine

1. <https://colab.research.google.com>

learning), à l'analyse de données (data analysis) et à l'éducation (e-learning). En termes plus techniques, Colab est un service hébergé de notebooks Jupyter qui ne nécessite aucune configuration et il permet d'accéder gratuitement à des ressources informatiques, dont des GPU (Graphics Processing Unit) [2].

### 4.3.2 Les bibliothèques utilisées

#### **Pandas**

Pandas est une bibliothèque écrite pour le langage de programmation Python permettant la manipulation et l'analyse des données. Elle propose en particulier des structures de données et des opérations de manipulation de tableaux numériques et de séries temporelles [6].

#### **Scikit-learn**

Scikit-learn (anciennement scikits.learn et également connu sous le nom de sklearn) est une bibliothèque logicielle gratuite d'apprentissage automatique en Python. comporte divers algorithmes de classification, de régression et de clustering, notamment les machines vectorielles de support, les forêts aléatoires, l'amplification de gradient, k-means et DBSCAN, et est conçu pour interagir avec les bibliothèques numériques et scientifiques Python NumPy et SciPy [9].

#### **Numpy**

NumPy est une bibliothèque pour le langage de programmation Python, ajoutant la prise en charge de grands tableaux et matrices multidimensionnels, ainsi qu'une grande collection de fonctions mathématiques de haut niveau pour opérer sur ces tableaux [5].

### 4.3.3 Kmeans + la methode Elbow

#### **Evaluation et Résultats expérimentaux de l'approche de k-means**

Nous créons des publicités vidéo, chaque publicité est recommandée à une communauté cible. Ensuite, les utilisateurs sont invités à exprimer leur satisfaction sur ces annonces. Les résultats présentés dans le tableau 4.1 montrent quelques exemples. Nous avons obtenu 4 dislikes sur 55 évaluations.



userid	Sexe	wilaya	Annonce id	Type annonce	rating
295	Male	guelma	1622819917	fish restaurant	like
295	Male	guelma	1622813557	romantic comedy movie	like
295	Male	guelma	1622813257	apple	like
296	Female	skikda	1622813017	samsung	like
296	Female	skikda	1622813557	romantic comedy movie	like
296	Female	skikda	1622813667	fast food	like
296	Female	skikda	1622813567	political	like
297	Female	skikda	1622813667	fast food	like
297	Female	skikda	1622813017	samsung	like
297	Female	skikda	1622813557	romantic comedy movie	Doesn't like
297	Female	skikda	1622813567	political	like
298	Female	skikda	1622813667	fast food	like
298	Female	skikda	1622813567	political	like
298	Female	skikda	1622813017	samsung	Doesn't like

TABLE 4.1 – quelques exemples d'évaluations par les utilisateurs des publicités recommandées

### 4.3.4 KDTree VS Factorisation matricielle probabiliste

#### 4.3.4.1 KDTree algorithme

##### Construction de l'arbre et la recommandation

Pour la construction d'arbre KD nous avons utilisé la bibliothèque Scipy et pour la recommandation nous avons appliqué le principe de l'algorithme écrit sur le chapitre de conception (voir la section 3.4.2.2).

L'algorithme est exécuté dans 0.78 seconde et il a atteint 1.5 comme une valeur de RMSE (voir figure [4.4](#)).

```

from scipy import spatial
tree = spatial.KDTree(train_set)#construction arbre KD
for index, row in test_set1.iterrows(): # boucle sur les information des
                                         #utilisateurs dans 20% test dataset
    age = row.age
    sexe = row.sexe
    east = row.east
    north = row.north
    coast = row.coast
    X=[(age,sexe,east,north,coast)]
    idxs = tree.query(X[0], k)# recherche les 15 plus proches dans arbre KD
    for x in idxs[1]:
        techAdsPredction += int(datax.loc[[x]]['tech ad'])
        foodAdsPredction += int(datax.loc[[x]]['food ad'])
        jobAdsPredction += int(datax.loc[[x]]['job offres'])
        clothesAdsPredction += int(datax.loc[[x]]['clothes ads'])
        newsAdsPredction += int(datax.loc[[x]]['news ads'])
    #calculer les resultats de prediction
    techAdsPredction = techAdsPredction/k
    foodAdsPredction = foodAdsPredction/k
    jobAdsPredction = jobAdsPredction/k
    clothesAdsPredction = clothesAdsPredction/k
    newsAdsPredction = newsAdsPredction/k

```

FIGURE 4.3 – Le code de notre algorithme KD-Tree

```

kdtree executed in 0.7811784049999915
RMSE value for KDTree algorithm 1.5070393493042253

```

FIGURE 4.4 – les valeurs RMSE et le taux d’exécution de KD-Tree

Le tableau [4.2](#) montre les résultats de la prédiction.

id	valeurs observées					valeurs prédites				
	Tech ads	Food ads	Job ads	Clothes ads	News ads	Tech ads	Food ads	Job ads	Clothes ads	News ads
9	3	2	1	5	0	3.266667	3.466667	1.600000	3.666667	1.600000
21	5	3	0	5	0	4.284444	3.764444	2.106667	4.244444	2.306667
25	4	2	0	5	0	3.885630	3.584296	1.407111	4.082963	1.553778

CHAPITRE 4. IMPLÉMENTATION ET RÉSULTATS

---

28	4	2	0	2	4	3.592375	3.905620	2.627141	4.138864	1.770252
29	3	3	1	3	5	4.039492	3.660375	2.041809	3.675924	2.384683
31	4	2	0	2	4	3.735966	3.644025	1.536121	3.978395	1.758979
32	3	3	2	2	5	3.915731	3.709602	1.369075	4.131893	1.517265
36	4	3	1	3	4	3.527715	3.713973	1.691272	3.942126	1.701151
39	5	3	3	3	5	3.235181	4.247598	2.446085	3.862808	2.913410
47	4	2	0	1	5	4.282345	3.816507	2.163072	4.257521	2.394227
58	4	5	1	5	1	3.418823	3.521100	1.410871	3.550501	2.492948
67	4	5	0	5	0	3.694588	3.501407	1.560725	3.636700	1.832863
70	5	5	2	5	0	3.712973	3.500094	1.570715	3.642447	1.788858
72	3	4	1	5	2	3.580865	3.700006	1.771381	3.709496	1.852591
77	3	4	1	4	1	3.772058	3.646667	1.584759	3.713966	1.923506
79	4	5	1	5	2	3.784804	3.976444	1.572317	4.314264	1.328234
81	4	4	2	5	1	3.185654	3.665096	1.504821	3.754284	2.021882
82	3	5	2	5	1	3.479044	3.711006	1.700321	3.916952	1.734792
87	4	5	0	4	1	4.098603	4.247400	1.513355	4.261130	1.982319
88	3	5	2	5	0	3.806574	3.749827	1.434224	3.950742	2.132155
99	1	4	4	3	3	4.320438	3.783322	2.095615	4.263383	2.342144

103	2	4	3	1	3	3.754696	3.652221	1.606374	4.084226	1.822810
114	2	3	2	1	4	3.583646	3.910148	2.640425	4.138948	1.788187
115	1	2	0	1	5	3.972243	3.794010	2.109362	3.675930	2.585879
117	1	3	2	2	4	3.931483	3.852934	2.807291	3.978395	1.905725
127	2	3	4	5	1	4.195432	3.856862	2.453819	3.531893	2.793715
140	5	5	4	5	2	4.213029	4.323791	1.830255	4.302126	2.319581
142	4	5	4	5	0	3.080869	4.288253	1.655350	3.420142	3.154639
143	5	5	5	5	2	3.138725	4.085884	1.577023	3.694676	2.676976
148	4	5	4	5	2	3.009248	4.272392	1.638468	3.379645	3.178465
151	3	5	3	5	2	3.867283	3.884826	1.909231	3.825310	2.611898
161	3	3	0	3	3	3.857819	3.658988	2.460615	3.121687	2.974127
163	3	3	0	1	4	4.057188	3.777266	2.297374	3.341446	2.864942
166	2	3	0	3	4	3.937146	3.451818	1.819825	3.422763	2.790996

TABLE 4.2: valeurs observées et prédites pour l'ensemble de test

#### 4.3.4.2 Factorisation matricielle probabiliste

##### Entraînement et test

Pour entraîner l'ensemble de données, Nous avons commencé à nous entraîner avec 150 itérations.

```
PMF training executed in 1.5050000001792796e-05
RMSE of PMF 1.8485817278868741
```

FIGURE 4.5 – les valeurs RMSE et le taux d’exécution de PMF

### Test de recommandation

La figure 4.6 montre les préférences de l’utilisateur 350, pour lequel nous n’avons que les informations présentées sur le tableau (il a noté que 3 annonces). Comme le montre la figure 4.7, les recommandations des autres annonces pour le même utilisateur.

userid	itemid	type	rating
350	1627579613	TechAds	5
350	1629453577	ClothesAds	5
350	1630647333	FoodAds	3
350	1628921964	JobAds	1
350	1630353274	NewsAds	0

FIGURE 4.6 – les préférences de l’utilisateur 350

UserID	itemid	type	Prediction
350	1629453577	ClothesAds	4.963481
350	1627579613	TechAds	4.951065
350	1630647333	FoodAds	2.536510
350	1630353274	NewsAds	0.776940
350	1628921964	JobAds	0.067486

FIGURE 4.7 – recommandation de publicités à l’utilisateur 350

#### 4.3.4.3 Résultats expérimentaux

Pour comparer les deux algorithmes nous avons calculé le temps d’exécution et le RMSE.

**Le RMSE (Root Mean Square Error) [35]**

C'est la racine carrée de la variance des résidus. Il indique l'ajustement absolu du modèle aux données, c.à.d à quel point les points de données observés sont proches des valeurs prédites du modèle.

$$\text{RMSE} = \sqrt{\sum_{i=1}^n \frac{(p_i - o_i)^2}{n}} \quad (4.1)$$

$p$  est la valeur prédite.

$o$  est valeur observée.

$n$  est le nombre d'observations.

Des valeurs plus faibles de RMSE indiquent un meilleur ajustement. Le RMSE est une bonne mesure de la précision avec laquelle le modèle prédit des items, et c'est le critère le plus important pour l'ajustement si l'objectif principal du modèle est la prédiction [32].

	RMSE	Taux d'exécution (seconde)
Notre KDTree algorithme	1.51	0.78
PMF algorithme	1.84	1.5

TABLE 4.3 – comparaison entre KDTree algorithme et PMF algorithme

Comme on peut le constater dans le tableau 4.3, l'algorithme d'arbre KD est meilleur en matière de temps d'exécution et il a également un RMSE plus faible. Alors l'approche de recommandation basée sur les KD-Tree est plus fiable que celle basée sur les PMF.

## 4.4 Conclusion

Dans ce chapitre, nous avons présenté notre implémentation du site Hazmit dans le but de collecter des données après avoir effectué le traitement avec certains algorithmes. Notre objectif principal est d'améliorer l'efficacité de notre système de recommandations d'annonces utilisant un algorithme de classification ou une architecture arborescence pour gagner du temps. Les résultats présentés sont très encourageants par rapport aux expériences dans le domaine.

# Conclusion Générale

Dans ce mémoire, nous nous sommes intéressés au domaine de la recommandation du contenu adéquat aux préférences des différentes communautés d'intérêts sur les réseaux sociaux. En effet, dans ce projet un système de recommandation d'annonces intéressantes et spécifiques pour chaque utilisateur afin de garantir la satisfaction de l'utilisateur et l'annonceur à la fois.

Alors, 3 méthodes différentes ont été développés : une méthode de classification en utilisant k-means, une méthode basée sur une structure arborescente pour la création des communautés à savoir : les kd-trees en plus d'une méthode basée sur la factorisation matricielle probabiliste. L'évaluation a été faite sur deux dataset collectées à partir de notre réseau social hazmit développé dans le cadre de ce projet. En effet, le système développé génère des résultats très encourageants.

Finalement, ce travail nous a permis d'apprendre beaucoup de choses et d'explorer de nouvelles méthodes dans ce domaine de recherche et de lire plusieurs articles intéressants.

# Perceptives

Comme perspectives de recherches futures, nous envisageons de :

Utiliser des méthodes de classification floues et des systèmes d'apprentissage en profondeur.

Utiliser une autre structure arborescente comme VP-Tree et essayez de la combiner avec d'autres algorithmes de recommandation.

Atteindre à travers hazmit de nouveaux utilisateurs et les convaincre de passer plus de temps et de nouer plus de relations avec leurs amis. Ceci afin de collecter plus de données sur eux et sur les publicités qu'ils préfèrent pour collecter plus d'informations et améliorer ainsi la précision du système.

Faire des recherches psychologiques sur ce que chaque communauté algérienne aime et les diviser de façon plus précise selon de nouvelles critères.



# Bibliographie

- [1] Comment choisir le bon nombre de clusters dans l'algorithme k-means? <https://ichi.pro/fr/comment-choisir-le-bon-nombre-de-clusters-dans-l-algorithme-k-means-159844816308064>.
- [2] Google colab. <https://research.google.com/colaboratory/faq.html?hl=fr>.
- [3] Javascript. <http://www.toutjavascript.com/cours>.
- [4] Mysql. <http://dev.mysql.com>.
- [5] Numpy. <https://en.wikipedia.org/wiki/Numpy>.
- [6] Pandas. <https://en.wikipedia.org/wiki/Pandas>.
- [7] Php. <http://www.phpclasses.org>.
- [8] Recommender systems. <https://builtin.com/data-science/recommender-systems>.
- [9] Scikit-learn. <https://en.wikipedia.org/wiki/Scikit-learn>.
- [10] Understanding k-dimensional trees. <https://towardsdatascience.com/tagged/kd-tree?p=9746debcd940>.
- [11] Understanding k-means clustering in machine learning. <https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>.
- [12] w3school. <https://W3Schools.com>.
- [13] La publicité sur les réseaux sociaux, Aug 2015.
- [14] La liste des principaux réseaux sociaux en 2021, Jul 2021.
- [15] Gediminas ADOMAVICIUS et Alexander TUZHILIN : Toward the next generation of recommender systems : A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749, 2005.

- [16] Yassine AFOUDI, Mohamed LAZAAR et Mohamed AL ACHHAB : Collaborative filtering recommender system. *In International Conference on Advanced Intelligent Systems for Sustainable Development*, pages 332–345. Springer, 2018.
- [17] Rishabh AHUJA, Arun SOLANKI et Anand NAYYAR : Movie recommender system using k-means clustering and k-nearest neighbor. *In 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pages 263–268. IEEE, 2019.
- [18] Jean-Marc ALLIOT et Nicolas DURAND : Algorithmes génétiques. *Centre d'Etudes de la Navigation Aérienne*, 2005.
- [19] Christophe ASSELIN : Facebook les chiffres essentiels en 2021 en france et dans le monde.
- [20] Alejandro BELLOGÍN, Iván CANTADOR, Fernando DÍEZ, Pablo CASTELLS et Enrique CHAVARRIAGA : An empirical comparison of social, collaborative filtering, and hybrid recommenders. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 4(1):1–29, 2013.
- [21] David BEN-SHIMON, Alexander TSIKINOVSKY, Lior ROKACH, Amnon MEISLES, Guy SHANI et Lihi NAAMANI : Recommender system from personal social networks. *In Advances in intelligent web mastering*, pages 47–55. Springer, 2007.
- [22] Jon Louis BENTLEY : Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.
- [23] Daniel BILLSUS et Michael J PAZZANI : User modeling for adaptive news access. *User modeling and user-adapted interaction*, 10(2):147–180, 2000.
- [24] Jesús BOBADILLA, Rodolfo BOJORQUE, Antonio Hernando ESTEBAN et Remigio HURTADO : Recommender systems clustering using bayesian non negative matrix factorization. *IEEE Access*, 6:3549–3564, 2017.
- [25] Robin BURKE : Hybrid recommender systems : Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370, 2002.
- [26] Robin BURKE : Hybrid web recommender systems. *The adaptive web*, pages 377–408, 2007.
- [27] Chohra CHEMSEDDINE et Khaledv ABDELAZIZ : Outil d'analyse des réseaux sociaux à base de web sémantique pour les environnements d'apprentissages. 2013.

- [28] Per-Erik DANIELSSON : Euclidean distance mapping. *Computer Graphics and image processing*, 14(3):227–248, 1980.
- [29] Joydeep DAS, Subhashis MAJUMDER et Prosenjit GUPTA : Spatially aware recommendations using kd trees. 2013.
- [30] Joydeep DAS, Subhashis MAJUMDER, Prosenjit GUPTA et Kalyani MALI : Collaborative recommendations using hierarchical clustering based on kd trees and quadtrees. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 27(04):637–668, 2019.
- [31] Mingsheng FU, Hong QU, Zhang YI, Li LU et Yongsheng LIU : A novel deep learning-based collaborative filtering model for recommendation system. *IEEE transactions on cybernetics*, 49(3):1084–1096, 2018.
- [32] Karen GRACE-MARTIN, MM ALI, KEWAL, SHAWN, ULRICH, Nishant MATTHUR, KRIS, HASHIM, DINA, YASH et et AL. : Assessing the fit of regression models, Nov 2020.
- [33] Folasade Olubusola ISINKAYE, YO FOLAJIMI et Bolande Adefowoke OJOKOH : Recommendation systems : Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3):261–273, 2015.
- [34] Dietmar JANNACH, Markus ZANKER, Alexander FELFERNIG et Gerhard FRIEDRICH : *Recommender systems : an introduction*. Cambridge University Press, 2010.
- [35] C KILGUS et W GORE : Root-mean-square error in encoded digital telemetry. *IEEE Transactions on Communications*, 20(3):315–320, 1972.
- [36] M Sandeep KUMAR et J PRABHU : A hybrid model collaborative movie recommendation system using k-means clustering with ant colony optimisation. *International Journal of Internet Technology and Secured Transactions*, 10(3):337–354, 2020.
- [37] Jie LU, Dianshuang WU, Mingsong MAO, Wei WANG et Guangquan ZHANG : Recommender system application developments : a survey. *Decision Support Systems*, 74:12–32, 2015.
- [38] Julian MCAULEY, Rahul PANDEY et Jure LESKOVEC : Inferring networks of substitutable and complementary products. *In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 785–794, 2015.

- [39] Bamshad MOBASHER, Xin JIN et Yanzan ZHOU : Semantically enhanced collaborative filtering on the web. *In European Web Mining Forum*, pages 57–76. Springer, 2003.
- [40] Maryam Khanian NAJAFABADI, Mohd Naz’ri MAHRIN, Suriayati CHUPRAT et Haslina Md SARKAN : Improving the accuracy of collaborative filtering recommendations using clustering and association rules mining on implicit data. *Computers in Human Behavior*, 67:113–128, 2017.
- [41] Michael J PAZZANI : A framework for collaborative, content-based and demographic filtering. *Artificial intelligence review*, 13(5):393–408, 1999.
- [42] Christian ROBERT : *Le choix bayésien : Principes et pratique*. Springer Science & Business Media, 2005.
- [43] Millend ROY, Abhinav GAUTAM et Aayush SUGANDHI : A deep learning framework for enhancing maritime coastal security. *In 2021 2nd International Conference for Emerging Technology (INCET)*, pages 1–6. IEEE, 2021.
- [44] Tomasz RUTKOWSKI, Jakub ROMANOWSKI, Piotr WOLDAN, Paweł STASZEWSKI, Radosław NIELEK et Leszek RUTKOWSKI : A content-based recommendation system using neuro-fuzzy approach. *In 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8. IEEE, 2018.
- [45] Jiangbo SHU, Xiaoxuan SHEN, Hai LIU, Baolin YI et Zhaoli ZHANG : A content-based recommendation algorithm for learning resources. *Multimedia Systems*, 24(2):163–173, 2018.
- [46] Barry SMYTH et Paul COTTER : A personalised tv listings service for the digital tv age. *Knowledge-Based Systems*, 13(2-3):53–59, 2000.
- [47] Alessandro SUGLIA, Claudio GRECO, Cataldo MUSTO, Marco DE GEMMIS, Pasquale LOPS et Giovanni SEMERARO : A deep architecture for content-based recommendations exploiting recurrent neural networks. *In Proceedings of the 25th conference on user modeling, adaptation and personalization*, pages 202–211, 2017.
- [48] DERRY O’ SULLIVAN, Barry SMYTH et DAVID WILSON : Preserving recommender accuracy and diversity in sparse datasets. *International Journal on Artificial Intelligence Tools*, 13(01):219–235, 2004.

- [49] Ugo VOLLI : Pertinence sémiotique de la notion d'ensemble flou. *In A Semiotic Landscape. Panorama sémiotique*, pages 565–568. De Gruyter Mouton, 2019.
- [50] Sun-Chong WANG : Artificial neural network. *In Interdisciplinary computing in java programming*, pages 81–100. Springer, 2003.
- [51] David C WILSON, Barry SMYTH et Derry O' SULLIVAN : Sparsity reduction in collaborative recommendation : A case-based approach. *International journal of pattern recognition and artificial intelligence*, 17(05):863–884, 2003.
- [52] Yandi XIA, Giuseppe DI FABBRIZIO, Shikhar VAIBHAV et Ankur DATTA : A content-based recommender system for e-commerce offers and coupons. *In eCOM@ SIGIR*, 2017.
- [53] Rui XU et Don WUNSCH : *Clustering*, volume 10. John Wiley & Sons, 2008.
- [54] Xian-Da ZHANG : Machine learning. *In A Matrix Algebra Approach to Artificial Intelligence*, pages 223–440. Springer, 2020.
- [55] Zhijun ZHANG, Gongwen XU, Pengfei ZHANG et Yongkang WANG : Personalized recommendation algorithm for social networks based on comprehensive trust. *Applied Intelligence*, 47(3):659–669, 2017.
- [56] Cai-Nicolas ZIEGLER et Georg LAUSEN : Analyzing correlation between trust and user similarity in online communities. *In International Conference on Trust Management*, pages 251–265. Springer, 2004.