République Algérienne Démocratique et Populaire
Ministère de L'enseignement Supérieur et de la Recherche Scientifique
Université de Djilali Bounaama Khemis Miliana

Mémoire de fin d'étude

* * * * *

Département des Mathématiques et Informatiques

spécialité : *Ingénierie de Logiciel et Systèmes Distribués*

MÉMOIRE

Pour obtenir

LE DIPLÔME DE MASTER EN INFORMATIQUE

Recommandation de publication dans un réseau social à base de classification

---

Réalisée par :

HAMOUMANE Nawal

MEBARKI Asma

Soutenu le 21 /09/ 2019, devant les membres du jury :

Président du jury : Mr HANICHE FAYCEL

Encadreur : Mme BOUDALI FATIHA

Examinateur 1 : Mr GOUDJIL Mouhammed

Examinateur 2 : Mr MEGHATRIA RIYAD

Année Universitaire : 2018/2019

Remerciement

*En introduction, nous voudrions remercier **allah** qui nous a aidés et nous a donné patience et courage au cours de cette année.*

*Tout d'abord, nous voudrions remercier **Mme Fatiha Boudali** , qui nous a permis de bénéficier de sa supervision. Les conseils qu'il nous a donnés ainsi que la patience et la confiance qu'il nous a témoignées ont été utiles pour mener à bien nos recherches.*

*Nous voudrions remercier tous ceux qui nous ont aidés, d'une manière ou d'une autre spécialement **Voisinage Pc** , tout au long de cette étude et de ce travail de recherche.*

*Nous remercions sincèrement **les membres du jury** pour l'intérêt qu'ils portent à nos recherches en acceptant d'étudier notre modeste travail et en l'enrichissant de leurs suggestions.*

*Nous remercions également tous nos **enseignants** pendant les années scolaires.*

Merci

Dédicace

Tous les mots ne sauraient exprimer la gratitude, l'amour, le respect, la reconnaissance, c'est tous simplement que : Je dédie cette mémoire de Master à :

A Ma très chère Mère ***kheira :*** *Tu représente pour moi la source de tendresse et l'exemple de dévouement qui n'a pas cessé de m'encourager. Tu as fait plus qu'une mère puisse faire pour que ses enfants suivent le bon chemin dans leur vie et leurs études.*

Pour les mots ne peuvent pas avoir raison qu'il est mort, a mon très cher Père ***Ibrahim*** *et Dieu ait son âme dans son paradis.*

A mon cher frère : ***Ismail.***

A mes chère sœurs : ***Hanane, Chahrazed.***

A mes familles et mes amis qui par leurs prières et leur encouragements.

.

NAWAL

Dédicace

Tous les mots ne sauraient exprimer la gratitude, l'amour, le respect, la reconnaissance, c'est tous simplement que : Je dédie cette mémoire de Master à :

A Ma très chère Mère ***karima :*** *Tu représente pour moi la source de tendresse et l'exemple de dévouement qui n'a pas cessé de m'encourager. Tu as fait plus qu'une mère puisse faire pour que ses enfants suivent le bon chemin dans leur vie et leurs études.*

A Mon très cher Père ***Abdelkader :*** *Aucune dédicace ne saurait exprimer le dévouement et le respect que j'ai toujours pour vous. Rien au monde ne vaut les efforts fournis jour et nuit pour mon éducation et mon bien être.*

A mes chère sœurs : ***fatima***, ***safaa***,***safia***, ***maria***, ***salsabil***, ***rifka***

A mes familles et mes amis qui par leurs prières et leur encouragements.

.

ASMA

Résumé

Les systèmes de recommandations sont des systèmes automatiques qui permettent, par des techniques de recommandation, de fournir à des utilisateurs des suggestions qui répondent à leurs exigences. Un grand nombre de systèmes de recommandation existent dans divers domaines. Leur objectif est de filtrer et d'adapter les informations pour chaque utilisateur. Les méthodes généralement utilisées pour le calcul de la recommandation sont soit basées sur le contenu soit sur la similarité de l'utilisateur avec les autres utilisateurs (approches collaboratives). cette dernière consiste à suggérer à un utilisateur des ressources pertinentes susceptibles de l'intéresser en se basant sur l'historique des utilisateurs qui partagent avec lui les mêmes centre d'intérêts.
Mots clés : Système de recommandation, Réseau social, Approches collaboratives, filtrage basé sur le contenu, suggestion, PHP.

Abstract

Recommendations systems are automatic systems that by «Recommendation techniques» provide to users suggestions that meet their requirements. Many recommendation systems exist in various areas. Their aim is to filter and adapt the information for each user. The method generally used for the calculation of the recommendation is based on the content or on the similarity of the user with other users (collaborative approaches). the latter is to suggest to a user relevant resources likely to interest him based on the history of users who share with him the same interests.
Keywords : recommendation system, Social network, Collaborative approaches, filtering based on content, suggestion, PHP.

Table des matières

Table des figures

Liste des tableaux

Introduction

Nous vivons, l'ère de l'information au cours des dernières années ; les réseaux sociaux ont gagnésen popularité, il nous permet à la fois l'accès et le partage de l'information. Les systèmes d'information spécifiquement les réseaux sociaux sont caractérisé pour leur volume croissant, leur hétérogénéité et qu'ils ne sont pas suffisamment adaptés aux besoins d'utilisateur ceci rend l'accès à l'information pertinente, un vrais challenge. Les besoins des utilisateurs sont difficile a traiter , car ils sont évolutifs et qu'ils ne sont pas formulés explicitement. Les systèmes de recommandation ont pour objectif remédier ce problème, en facilitant le filtrage et l'adaptation de l'information pour chaque utilisateur.
Vu le flux important des informations sur les réseaux sociaux et qui ne cesse d'augmentation, les utilisateurs se trouvent confrontés en permanence à un véritable de luge d'information, on palifie pour le développement de la technologie et l'utilisation croissante des appareils mobile tel que (smartphones, tablettes.....) Afin de réduirel'impact de cette surcharge et facilite l'accès a l'information pertinente, nous nous intéressons à fournir à l'utilisateur des recommandations qui lui faciliter l'accès, le filtrage, l'analyse et l'exploitation rationnel des informations sur les réseaux sociaux .
L'objectif de ce projet est d'étudier les challenges liés à la forte utilisation du web social pour mieux appréhender les problématiques des réseaux sociaux. Nous nous intéressons particulièrement un réseau social numérique. d'autre part Dans notre projet , nous avons en premier temps pris la préoccupation de construire un profil utilisateur dans le réseau social numérique et d'essayer en deuxième temps de l'enrichir avec un jeu de données du web plus important publiées en qualité dataset afin de construire les nouveaux centres d'intérêts de l'utilisateur. Et en dernier lieu on a proposé des recommandations de publication en se basant sur ses centres d'intérêts. Ce mémoire est organisé comme suit : Après l'introduction qui présente le contexte du projet, la problématique étudiée et notre contribution, nous présentons le premier chapitre, qui parle des notions et principes du système recommandation.
Le deuxième chapitre, s'intéresse au web social ainsi que ses différentes technologies.
Le troisième chapitre met en lumière les moyens de traiter et de catégoriser les données pertinentes, ainsi que leur utilité dans le contexte du réseau social.

Le quatrième chapitre « conception» représente la Vue fonctionnelle du système et Schéma général de système.

Le cinquième chapitre « réalisation » représente l'architecture de notre approche, les technologies et les outils utilisés pour l'implémentation et la mise en œuvre de notre application qui repose sur les points suivants :

• création réseau social.

• Enrichissement du profil obtenu en exploitant les données liées a chaque utilisateur.

• Recommandation de publication.

Et enfin, nous clôturons ce rapport par une conclusion générale tout en ouvrant une fenêtre sur les futurs travaux dans le contexte de ce projet de fin d'études.

Chapitre 1

Système de recommandation

1 introduction

Face à la masse très importante d'information qui ne cesse de s'accroitre avec le temps, les internautes ne peux guerre trouver ce qu'il cherche dans le temps voulus, l'idée de système recommandation est apparue, pour faciliter la tâche à ces derniers, ont leurs permettent d'accéder, rapidement et efficacement à ce qui les intéresses. Dans ce premier chapitre nous essayons d'étudier et de définir qu'est ce qu'un système recommandation, ainsi que son principe.

2 système de recommandation

Les systèmes de recommandation peuvent être définit de plusieurs façons, vue la diversité des classifications proposées pour ces systèmes, mais il existe une définition générale de Robin Burke [Bruke, 2002] qui les définit comme suit[kar,2014] : "Des systèmes capable de fournir des recommandations personnalisées permettant de guider l'utilisateur vers des ressources intéressantes et utiles au sein d'un espace de données important". Les deux entités de base qui apparaissent dans tous les systèmes de recommandations sont l'usager et l'item. L'«usager» est la personne qui utilise un système de recommandation, donne son opinion sur diverses items et reçoit les nouvelles recommandations du système. "L'item" est le terme général utilisé pour désigner ce que le système recommande aux usagers. Les données d'entrée pour un système de recommandation dépendent du type de l'algorithme de filtrage employé. Généralement, elles appartiennent à l'une des catégories suivantes :

- **Les estimations :** (également appelées les votes), expriment l'opinion des utilisateurs sur

les articles (exemple : 1 mauvais à 5 excellent).

• **Les données démographiques :** se réfèrent à des informations telles que l'age, le sexe, le pays et l'éducation des utilisateurs.

• **Les données de contenu :**qui sont fondées sur une analyse textuelle des documents liés aux éléments évalués par l'utilisateur. Les caractéristiques extraites de cette analyse sont utilisées comme entrées dans l'algorithme de filtrage afin d'en déduire un profil d'utilisateur.[Margaritis et Vozalis, 2003].

3 l'historique de Système de Recommandation

La capacité des ordinateurs pour faire des recommandations à des utilisateurs a été reconnue assez tôt dans l'histoire de l'informatique. Grundy [Rich, 1979], un système bibliothécaire, était une première étape vers des systèmes de recommandation automatiques. Ce système était assez primitif. Il classait les utilisateurs en "stéréotypes" en se basant sur une courte interview, et utilisait ces stéréo types pour produire des recommandations de livres. Ce travail constituait une première tentative intéressante dans le domaine des systèmes de recommandation. Cependant, son utilisation est restée très limitée. Au début des années 1990, le filtrage collaboratif apparait comme une solution pour faire face à la surcharge d'information. L'année 1992 voit l'apparition du système de recommandation de documents Tapestry [Goldberg et al.. 1992], ainsi que la création du laboratoire de recherche Goup Lens, qui travaille explicitement sur le problème de la recommandation automatique le cadre des forums de news de Usenet. Tapestry avait pour but de recommander à des groupes d'utilisateurs des documents issus des news groups susceptibles de les intéresser. L'approche utilisée était de type "plus proches voisins" à partir de l'historique de l'utilisateur. On parle alors de filtrage collaboratif manuel, comme une réponse au besoin d'outils pour le filtrage de l'information énoncé la même époque .La recommandation résulte d'une action collaborative des utilisateurs qui recommandent à d'autres utilisateurs des documents en leur attribuant des notes d'intérêt selon certains critères. Les systèmes de filtrage collaboratif automatiques apparaissent en suite. Group Lens [Resnicketal., 1994] utilise cette technique pour identifier les articles d'Usenet susceptibles d'être intéressants pour un utilisateur donné. Les utilisateurs doivent seulement attribuer des notes ou effectuer d'autres opérations observables (par exemple, lire un article) ; le système combine alors ces données avec les notes ou les actions d'autres utilisateurs pour fournir des résultats person-

nalisés. Avec ces systèmes, les utilisateurs n'ont aucune connaissance directe des opinions des autres utilisateurs, ni des articles présents dans le système. Au cours de ces dernières années, les systèmes de recommandation deviennent un sujet d'un intérêt croissant dans les domaines de l'interaction homme-machine, de l'apprentissage automatique ainsi que la recherche d'information. En 1995 apparaissent successivement Ringo [Shardanand et Maes, 1995a], un système de recommandation de musique, basé sur les appréciations des utilisateurs et Bellcore [Hill et al. 1995], un système de recommandation de vidéos. La même année, GroupLens crée la société Net Perceptions dont le premier client a été Amazone. De nos jours, les systèmes de recommandation sont devenus des composantes incontournables pour la plupart des sites du e-commerce.

4 Les applications dans le domaine des systèmes de recommandation

Les systèmes de recommandation se basent sur plusieurs domaines de recherche tels que la recherche d'information, le tableau suivant représenté les applications qu'y i permet de faire une recommandation a des domaines différents :

Domaine	application
livre	Amazon, fnac…
films	Netfix, MovieLens, allocine…
musique	Deezer…
jeux vidéo	Steam, microsoft xbox live…
amis	Facebook, twitter, google+…
tourisme	Expedia …
e-shopping	Cdiscount, eBoy, amazon, fnac…

Tableau 1 : les application qui utilise les systèmes de recommandation .

.

Chaque un de nous aime les films, jeux, sport, ...etc., c-à-dire s'amuser. Même les sites qui offrent ça utilisent de système de recommandation pour faciliter aux utilisateurs de trouver ce qu'ils veulent selon leurs goûts ces système basent sur le filtrage collaboratif. nous avons donner des exemples sur la recommandation des films :

• *Netflix :* Netflix recommande des films et des séries selon les goûts., il propose de service selon contenu basée sur l'historique de l'utilisateur, mais également il analyse les contenus préférés des personnes ayant les mêmes goûts.

• *MovieLens :* MovieLens est un site communautaire de recommandation de films. Les utilisateurs de ce site notent des films de 1 à 5. Ils peuvent également demander des suggestions de films étant donnés les notations qu'ils ont fournies.

5 Comment fonctionnent les systèmes de recommandation ?

Les systèmes de recommandation ont pour premier rôle d'identifier le sous groupe d'utilisateurs auquel appartient un utilisateur afin de lui proposer des résultats susceptibles de l'intéresser. L'identification de sous groupes d'utilisateurs auquel appartient un utilisateur se fait généralement en fonction de l'historique d'utilisation du service par cet utilisateur. Le système de recommandation peut toute fois s'appuyer des caractéristiques connues sur l'utilisateur (son âge, sa catégorie socioprofessionnelle, son sexe, son secteur professionnel..) ou sur une combinaison de ces caractéristiques et de son historique. Il ne reste alors au système de recommandation qu'à trouver les autres utilisateurs partageant le plus de points communs avec cet utilisateur, analyser les items les plus commandés, partagés ou plébiscités par ces utilisateurs afin de pouvoir proposer une sélection personnalisée d'items recommandés.

6 les approches de système de recommandation

Les Systèmes de Recommandation sont des outils visant à suggérer automatiquement aux utilisateurs des éléments utiles. La nature de cet élément peut être très différent (document textuel, image, vidéo, produit,..). Le terme de «item » est ainsi employé pour dénoter de manière générale tous ces éléments. Le figure de ci-dessous représente les techniques de système de recommandation :

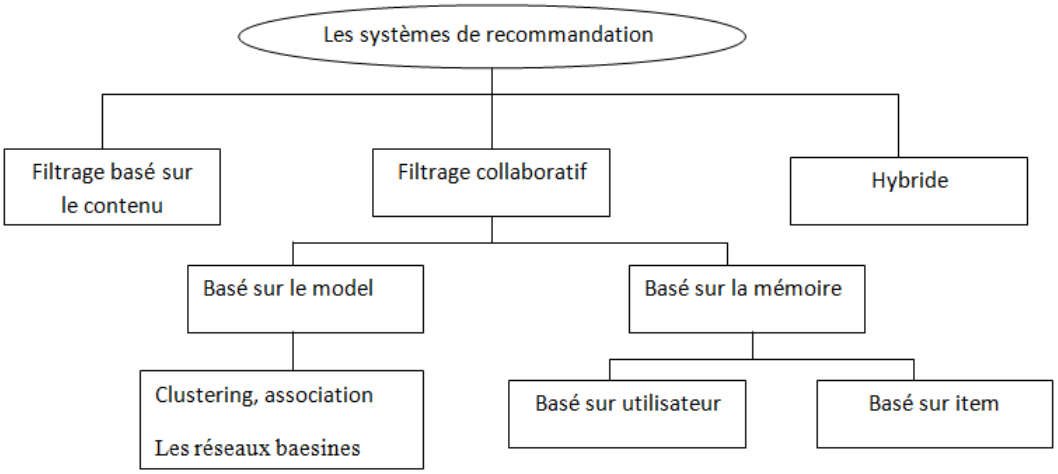

FIGURE 1.1 – les techniques de système de recommandation

Les stratégies employées pour identifier les informations à recommander sont nombreuses. Elles sont classées en trois catégories :

- *Le filtrage basé sur le contenu*
- *Le filtrage collaboratif*
- *Le filtrage Hybride*

6.1 le filtrage basé sur le contenu

Les systèmes de recommandation basés sur le contenu fonctionnent en analysant les caractéristiques des objets à recommander (produits, etc.) puis en les regroupant. Par la suite, le système va suggérer aux utilisateurs ayant acheté/consommé un produit quelconque par le passé, les objets/produits estimés similaires[Ricci et al. 2011].

FIGURE 1.2 – Systèmes de recommandation basés sur le contenu[Rahila, 2015].

.

L'architecture générale d'un système de recommandation basé sur le contenu s'articule autour de 3 modules principaux [Dahmani, 2014] :

• l'analyseur de contenu :

Selon la nature des données à recommander (texte, éléments multimédia, pages Web, produits commerciaux, etc.) une étape de pré-traitement est nécessaire an de décrire les objets à recommander et d'en extraire les caractéristiques .Le module d'analyse de contenu est responsable de produire une description structurée de ces objets. Cette description va servir d'élément d'entrée aux autres modules.

• Le module d'apprentissage de profils :

Ce module est responsable de l'analyse des interactions passées de l'utilisateur sur les objets du système. En utilisant des méthodes empruntées au monde de l'apprentissage, ce module construit une description des préférences des utilisateurs.

• Le module de filtrage :

A partir des prols utilisateurs et des descriptions des objets à recommander, ce module construit des listes de suggestions à présenter aux utilisateurs.

Les prols utilisateurs y sont déduits à travers les étiquettes (tags) issues de l'activité d'annotation exécutée par les utilisateurs sur le système. Ce système se base par la suite sur ces prols pour générer des recommandations de collaborateurs potentiels.

Représentation d'un item :

Dans la plupart des systèmes basés sur le contenu, la description de l'item est sous forme de texte, extrait de pages web, email ou fiche produit dans un site de e-commerce.[Idir,2017] Contrairement aux données structurées, il n'y a pas d'attributs avec des valeurs bien définies. Cela rend plus difficile l'apprentissage du profil utilisateur, en raison de l'ambigüité du langage naturel, et notamment de la polysémie (multiples significations pour un mot) et de la synonymie (plusieurs mots qui ont le même sens). Un pré-traitement peut alors être nécessaire afin d'extraire l'information pertinente et de la structurer sous forme d'un ensemble d'attributs.

• **les avantages et les inconvénients de cette approche :**
- avantage : - Permit des recommandations de nouveaux items.
- Un nouvel utilisateur peut recevoir des recommandations des ses premières interactions avec le système.
- inconvénient : - La nécessité de disposer connaissance des objets dans le profil.

6.2 le filtrage collaboratif

La deuxième grande famille de systèmes de recommandation est basée sur l'hypothèse que les utilisateurs qui ont aimé des articles similaires par le passé ont un goût similaire et vont donc apprécier les mêmes articles dans le futur. Un des exemples les plus connus d'un tel système a été popularisé par le site de commerce en ligne Amazon.com et son algorithme de « Item-to-item Collaborative Filtering qui se traduit sur le site par la fonctionnalité "Les gens qui ont acheté le produit "x" ont aussi acheté le produit "y" [Lindenetal., 2003]. L'avantage principal de cette approche est qu'elle ne nécessite pas de description précise des objets à recommander. Les recommandations étant basées sur l'ensemble des interactions des utilisateurs avec les objets/produits, cette méthode permet de recommander des objets complexes sans avoir à les analyser. La plupart des services de recommandation de musique en ligne fonctionnent sur ce mode (ex. last. fm10) car les fichiers multimédia sont difficiles à analyser. Pour pouvoir fonctionner le système a besoin de collecter des données sur les utilisateurs et leurs préférences, cette collecte peut se faire de deux façons :

• **Collecte Explicite :** Dans ce cas, les utilisateurs sont sollicités pour émettre leur sa vis sur des produits/objets .Il peuvent le faire via un système de notation (ex. une grille de 5 étoiles,

un questionnaire de satisfaction), ou bien en publiant leurs avis sur un élément donné (Ex : La fonction "j'aime" sur le réseau social Facebook permet aux utilisateurs d'exprimer leur intérêt pour un élément donné).

• **Collecte Implicite :** La collecte implicite s'intéresse aux interactions des utilisateurs sur le système. Les exemples de cette collecte incluent la surveillance du nombre de visites sur une page, le nombre de vues sur une vidéo, le temps passé sur une section donné ou de l'historique des achats sur une plateforme de e-commerce.
[$Dasetal.$, 2007] décrit la plateforme de recommandation utilisé par Google News. L'approche basée sur le filtrage collaboratif a permis au système d'être indépendant vis-à-vis du contenu des publications suggérées et à la technique d'être adaptée à d'autres applications ou de gérer d'autres langues à moindre coût.

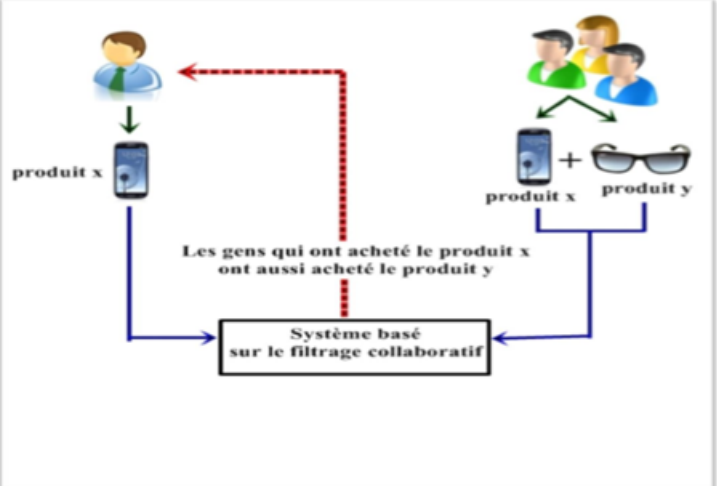

FIGURE 1.3 – Les systèmes de recommandation basé sur le filtrage collaboratif [Dahimene, 2014]

.

6.2.1 Filtrage collaboratif basé sur le modèle

Le deuxième type d'algorithmes, est comme le nom l'indique bases sur des modèles, supposes réduire la complexité. Ces modèles peuvent être probabilistes et utiliser l'espérance de l'évaluation pour calculer la prédiction. Comme ils peuvent être bases sur des classificateurs permettant de créer des classes pour réduire la complexité.

1• Modèle de clustring

Les méthodes de Clustering permettent de limiter le nombre d'individus considères dans le calcul de la prédiction. Le temps de traitement sera donc plus court et les résultats seront potentiellement plus pertinents puisque les observations porteront sur un groupe le plus proche de l'utilisateur actif. Autrement dit, au lieu de consulter l'ensemble de la population, nous estimons la préférence d'un groupe de personnes ayant les mêmes gouts que l'utilisateur.

2• K means

La méthode des plus proches voisins K-Means consiste dans un premier temps a choisir aléatoirement k centres dans l'espace de représentation utilisateurs/ressources. Ensuite, chaque utilisateur est mis dans le cluster du centre le plus proche. Quand les groupes de personnes sont formes, nous recalculons la position des centres pour chaque cluster et réitérons l'opération depuis le début jusqu'à obtenir un état stable ou les centres ne bougent plus. L'algorithme est certes simple a mettre en œuvre mais présente certains inconvénient, lie a la criticité du choix des clusters initiaux, pouvant influencer sur la qualité de la classification.

3• arbre de décision

Arbre de décision est un algorithme de filtrage collaboratif appelé l'arbre de recommandation (Recommandation Tree). L'algorithme de l'arbre de décision fractionne les données dans des cliques d'utilisateurs approximativement semblables. L'objectif est de maximiser les similarités entre les membres d'une même clique et a minimiser celles entre les membres de deux cliques différentes.

6.2.2 Filtrage collaboratif basé sur la mémoire

L'idée de base des systèmes basés sur la mémoire est la suivante :

• Le système maintient un profil utilisateur, c'est-à-dire un enregistrement des intérêts (aussi bien positifs que négatifs) de l'utilisateur pour certains objets.

• Puis il compare ce profil avec les profils des autres utilisateurs, et pèse chaque profil en fonction de son degré de similarité avec le profil de l'utilisateur considéré.

• Enfin, il considère un ensemble des profils les plus similaires (ou les plus opposés), et utilise l'information qu'ils contiennent pour recommander à l'utilisateur (ou mettre en garde contre) des objets qu'il n'a pas encore évalués.

Pour prédire la pertinence d'un item pour un utilisateur, on calcule donc la moyenne des notes

données aux items par les utilisateurs ayant les mêmes goûts (ou des goûts opposés), en utilisant des poids différents selon la mesure de similarité entre utilisateurs.[Laurent, 2001] Le tableau de données pour le filtrage collaboratif se présente souvent comme suit :

	item1	item2	item3	item4
U1		2	7	8
U2	4	1		7
U3	3	8		4

Tableau 2 : Une matrice des notes attribuées aux items par les utilisateurs.

.

• Filtrage collaboratif basé sur la mémoire (utilisateurs)

Cette technique de recommandation se base sur le principe de trouver des utilisateurs similaires à l'utilisateur courant puis d'utiliser leurs évaluations pour prédire ce que l'utilisateur courant peut aimer. Les utilisateurs similaires à l'utilisateur courant, appelés voisins de cet utilisateur, sont ceux qui ont un comportement d'évaluation similaire à celui de l'utilisateur courant.
$[Herlockeretal.1999]$ Présente les 3 étapes de cette technique de recommandation :
1. Calculer la similarité entre l'utilisateur courant et tous les utilisateurs du système.
2. Sélectionner un sous ensemble d'utilisateurs à utiliser comme un recommandeur. Il s'agit des utilisateurs voisins les plus proches.
3. Calculer les prédictions en utilisant une combinaison pondérée des évaluations appartenant aux voisins sélectionnés.

L'évaluation prédite r (u, i) de l'item i par l'utilisateur u dépend de :
• la similarité entre cet utilisateur et ses voisins les plus proches notée par sim(u, v).
• l'évaluation de l'utilisateur v sur l'item i notée par r(v, i), et d'un facteur de normalisation k donné par l'équation.
v appartient à l'ensemble N qui représente les voisins les plus proches de l'utilisateur u. r(u, i) est décrit par la formule suivante : .

$$r(u,i) = k \sum_{v \in N} r(v,i) \,.\, sim(u,v)$$

$$k = 1 \Big/ \sum_{v \in N} | \, sim(u,v) \, |$$

• Filtrage collaboratif basé sur la mémoire (items)

Le filtrage collaboratif basé sur les utilisateurs souffre de problèmes de montée en charge si la base d'utilisateurs est importante. La technique du filtrage collaboratif basé sur les items a été développée pour répondre à cette problématique. Cette technique est utilisée lorsqu'il s'agit de trouver des items similaires à l'item courant. Cette technique utilise les similarités entre les patterns des évaluations des items. Si deux items ont tendance à avoir les mêmes utilisateurs qui les aiment et les mêmes utilisateurs qui ne les aiment pas, alors ces items sont similaires[8] Les utilisateurs ont des préférences similaires pour les items similaires. Comme défini par [Gabrielsson et Gabrielsson.2006], cette technique se compose de 3 étapes :

1. Calculer la similarité entre l'item courant et tous les items du système.

2. Sélectionner les voisins les plus proches de l'item courant. Il s'agit des items les plus proches.

3. Calculer les prédictions en utilisant un algorithme basé sur l'évaluation par l'utilisateur courant des items appartenant au voisinage de l'item courant. L'évaluation prédite r (u, i) de l'item i par l'utilisateur u dépend de :

• La similarité entre cet item et les items évalués par l'utilisateur u notée par sim (i, j).

• L'évaluation de l'utilisateur u sur l'item j notée par r (u, j), et d'un facteur de normalisation k donné par l'équation.

j appartient à l'ensemble I qui contient les items, voisins les plus proche de l'item i, évalués par l'utilisateur u. r(u, i) est décrit par la formule suivante :

$$r(u,i) = \mathrm{k} \sum_{j \in I} r(u,j) \,.\, sim(i,j)$$

$$k = 1 \Big/ \sum_{j \in I} | \, sim(i,j) \, |$$

.

• **Mesure de similarité**

Plusieurs mesures de similarité entre utilisateurs et entre items ont été proposées dans la littérature. Selon Beliakov et al.,[2011] les deux types de mesures de similarité les plus populaires sont le coefficient de corrélation de Pearson et la similarité basée sur le cosinus. Il existe d'autres mesures de similarité telles que la différence moyenne quadratique [Shardanand et Maes, 1995] ou le coefficient de corrélation de Spearman [Herlocker et al. 2002] mais ils n'ont pas connu d'adoption significative par rapport aux deux mesures précédemment citées.[Charif.2014]

• **Coefficient de Corrélation de Pearson :** La corrélation de Pearson est une méthode issue des statistiques. Elle est aussi très utilisée dans le domaine des systèmes de recommandation pour mesurer la similarité entre deux utilisateurs. S'il s'agit de calculer la similarité entre deux utilisateurs, la corrélation entre eux est mesurée à l'aide des deux lignes, appartenant aux deux utilisateurs, de la matrice d'évaluations. Les colonnes des items non évaluées par les deux utilisateurs sont ignorées. Seuls les items co-évalues sont utilisées dans ce calcul.[Bresse et al.1998, Bank et Cole.2008, lu et al.2012] Ce coefficient se situe entre -1 et 1. Une similarité proche de -1 signifie une corrélation négative et inversement, une similarité proche de +1 signifie une corrélation positive. Il n'existe pas de corrélation entre les deux utilisateurs si la similarité est autour de 0. La similarité sim(u, v) entre les utilisateurs u et v est donnée par l'équation Eq sim(u,v).r(u, .) Est la moyenne des évaluations de l'utilisateur u. I est l'ensemble des items co-évalués par u et v. La formule suivante, nous donne cette valeur pour deux utilisateurs u et v : .

$$sim(u,v) = \frac{\sum_{i \in I}(r(u,i) - \bar{r}(u,.)).(r(v,i) - \bar{r}(v,.))}{\sqrt{\sum_{i \in I}(r(u,i) - \bar{r}(u,.))^2} \cdot \sqrt{\sum_{i \in I}(r(v,i) - \bar{r}(v,.))^2}}$$

La similarité sim(i, j) entre les items i et j est donnée par l'équation sim(i,j). r(., i) est la moyenne des évaluations de l'item i. U est l'ensemble des utilisateurs qui ont co-évalué les items i et j.

• **Similarité basée sur le cosinus :** Dans la matrice d'évaluation, les lignes associées aux utilisateurs sont considérées comme des vecteurs d'évaluation. Ce type de mesure de similarité est calculé en utilisant l'angle cosinus entre deux vecteurs d'évaluation. Cet angle est mesuré dans un espace à N dimensions où N est le nombre d'items co-évalués entre les deux utilisateurs. Cette similarité se situe entre 0 et 1 où le 0 signifient aucune similarité et 1 une forte similarité. Cette similarité entre les utilisateurs est décrite par la formule sim(u,v) et entre

$$sim(i,j) = \frac{\sum_{u \in U}(r(u,i) - \bar{r}(.,i)).(r(u,j) - \bar{r}(.,j))}{\sqrt{\sum_{u \in U}(r(u,i) - \bar{r}(.,i))^2} . \sqrt{\sum_{u \in U}(r(u,j) - \bar{r}(.,j))^2}}$$

les items par la formule sim(i,j)[Tadlaoui.2018] :

$$\mathrm{sim}(u,w) = \cos(\overrightarrow{x_u}, \overrightarrow{x_v}) = \frac{\sum_{i \in I_{uw}} v_{ui} \times v_{wi}}{\sqrt{\sum_{i \in I_{\mathrm{uw}}} v_{ui}^2} \sqrt{\sum_{i \in I_{\mathrm{wm}}} v_{wi}^2}} \tag{1.1}$$

$$\mathrm{sim}(i,j) = \cos(\vec{x}_i, \overrightarrow{x_j}) = \frac{\sum_{u \in U_{ij}} v_{ui} \times v_{uj}}{\sqrt{\sum_{u \in U_{ij}} v_{ui}^2} \sqrt{\sum_{u \in U_{ij}} v_{uj}^2}} \tag{1.2}$$

6.3 Filtrage Hybrides

Les systèmes hybrides permettent de résoudre les problèmes posés par l'utilisation de l'une des deux approches citées ci-dessus. Par exemple la première approche nécessite un riche historique d'interaction avec les objets du système et une description détaillé de ces derniers ce qui n'est pas évident dans certain cas (utilisateur fraichement inscrit). Par contre la deuxième, ont besoin de l'existence d'une large base d'interactions sur l'ensemble du catalogue d'objets du système afin de pouvoir calculer des rapprochements entre les utilisateurs.[Rahila.2015] La plupart des systèmes de recommandation existants privilégient le modèle hybride au vu des avantages qu'ils présentent. Parmi les systèmes hybrides les plus connus, le système de recommandation mis en place par le géant Américain de la vidéo à la demande sur Internet Netflix.

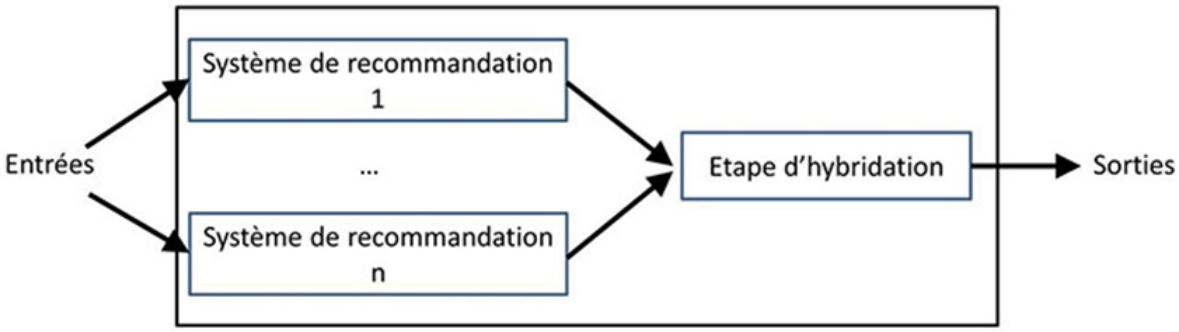

FIGURE 1.4 – Filtrage Hybrid [Dahimene , 2014]

.

Il existe plusieurs manières de faire de l'hybridation et aucun consensus n'a été défini par la communauté des chercheurs. Toutefois, Burke [Burke, 2002] a identifié quelque manière différente de faire l'hybridation[Idir.2017] :

• **Pondérée (Weighted) :** le score ou la prédiction obtenu par chacune des deux techniques est combiné en un seul résultat.

• **Par sélection (Switching) :** le système bascule entre les deux techniques de recommandation en fonction de la situation.

• **Mixte (Mixed) :** les listes des recommandations issues des deux techniques sont fusionnées en une seule liste.

• **Par combinaison des propriétés (Feature combination) :** les données issues des deux techniques sont combinées et transmises à un seul algorithme de recommandation.

• **Par augmentation de propriétés (Feature augmentation) :** le résultat d'une technique est utilisé comme entrée de l'autre technique.

• **En cascade :** Dans ce type d'hybridation, une technique de recommandation est utilisée pour produire un premier classement des items candidats et une deuxième technique affine ensuite la liste des recommandations.

7 Problèmes et limites des systèmes de recommandation

Les systèmes de recommandation se basant sur les techniques précédemment expliquées ont certaines limites. Plusieurs approches ont été proposées dans la littérature pour pallier ces limites. Ci-dessous les problèmes les plus importants sont décrits :

• **Démarrage à froid :** Les systèmes de filtrage collaboratif dépendent des évaluations des items par les utilisateurs. Ainsi, un nouvel item ne peut pas être recommandé tant qu'aucun utilisateur ne l'a évalué. Dans les systèmes de recommandation basés sur le filtrage collaboratif et les systèmes basés sur le contenu, il est impossible de prédire les préférences des utilisateurs sans connaitre leurs historiques d'évaluations d'items. Ainsi, les nouveaux utilisateurs ne recevront pas de recommandations précises avant d'avoir évalué un certain nombre d'items.[Bambi et al. 2011]

• **Sparsity :** Un système de recommandation souffre de la sparsity quand le nombre d'items évalués par les utilisateurs est très faible par rapport au nombre d'items total présent dans le système. Ce fait conduit à avoir une très faible densité dans la matrice d'évaluation utilisateurs/items. Cela a des conséquences sur la capacité du système de recommandation à recommander toutes les items disponibles et sur l'exactitude des recommandations générées.

• **serendipity :** Vu que les systèmes de recommandation basés sur le contenu ne recommandent que les items correspondants au profil de l'utilisateur, ce dernier ne recevra que des recommandations similaires à celles qu'il a déjà rencontrées. Il n'aura aucune chance de recevoir des recommandations inattendues. Cela peut amener l'utilisateur à se lasser des recommandations.

• **Problème du mouton gris :** Les utilisateurs d'un système de recommandation peuvent avoir des goûts particuliers et des préférences très inhabituelles par rapport aux autres. Ces utilisateurs sont à la frontière entre deux ou plusieurs clusters d'utilisateurs. Il leur est donc difficile de trouver des utilisateurs similaires et des recommandations pertinents.

• **Montée en charge :** Plus le nombre d'utilisateurs et d'items augmente dans le système de recommandation, plus les calculs nécessaires à la recommandation ne deviennent très couteux. Souvent des algorithmes de recommandation qui ont une énorme base d'utilisateurs et d'items préfèrent avoir des recommandations moins précises avec un temps de calcul rapide.

8 Conclusion

Dans ce chapitre, nous avons d'abord, présenté les systèmes de recommandation avec quelque exemple d'application, par la suite on a détaillé les trois approches les plus utilisée, à savoir : l'approche Filtrage Collaboratif, Filtrage basé sur le contenu et hybride. Ensuite, nous avons défini la notion de profil utilisateur avec ses deux facettes explicite et implicite. Nous avons également passé en revue les différentes mesures de similarité utilisée par les systèmes de recommandation. Enfin, nous avons terminé on citant quelques problèmes rencontrés par les systèmes de recommandation et leur principe. Dans le chapitre suivant, nous allons présenter les systèmes de recommandation pour les réseaux sociaux

Chapitre 2

Les Réseaux Sociaux

1 introduction

Depuis la création des réseaux sociaux, le nombre des utilisateurs ne cesse d'augmenter à ce jour des millions de gens utilise les réseaux sociaux (Facebook, twitter ...) pour ces interdépendances « les besoins entre les utilisateurs dans l'échange de données » (personnel, groupes sociaux, organisations). Afin de répondre à ces besoins qui ne cessent de s'accroitre, l'étude et le développement des réseaux sociaux deviennent une nécessité, pour cela nous étudierons dans ce chapitre, l'historique du web social et l'analyse des réseaux sociaux, toute on déterminant les avantage et les inconvenant des réseaux sociaux.

2 Les réseaux sociaux

2.1 Définition

Un réseau social est une structure comportant un ensemble d'acteurs qui sont impliqués sur certains types d'interactions. Un acteur est une entité sociale qui pourrait être une seule personne, un groupe ou une entreprise. Les acteurs sont reliés les uns aux autres par des liens qui peuvent désigner une ou plusieurs relations. Ces liens peuvent être de différents types, y compris des liens d'amitié, des liens de collaboration, des liens d'affaires, etc. Par conséquent, on peut distinguer :

• **Les réseaux hétérogènes :** des réseaux sociaux où plusieurs types d'acteurs ou plusieurs types de liens peuvent exister (par exemple, les acteurs des réseaux sociaux liés à différents types de liens, c'est à dire, les collègues et les amis).

• **Les réseaux homogènes :** ce sont les réseaux sociaux où il existe un seul type d'acteur avec un seul type de lien entre les acteurs (par exemple, les acteurs des réseaux sociaux connectés en utilisant uniquement des liens d'amitié).

Dans la pratique, les réseaux sociaux offrent aux internautes de nouveaux moyens et façons de se connecter, de communiquer et de partager des informations avec d'autres membres au sein de leurs plates-formes intéressantes. En théorie, ces réseaux sociaux sont composés de plusieurs éléments, peuvent contenir différents types de données, et avoir différentes représentations.

2.2 Média social

un média social est une plate-forme dont les activités intègrent trois éléments fondamentaux : la technologie, la création de contenus et les interactions sociales. De même, [Kaplan et Haenlein, 2010] définissent les médias sociaux comme un groupe d'applications en ligne qui se fondent sur la philosophie et la technologie d'Internet et permettent la création et l'échange de contenus générés par les utilisateurs.

2.3 Réseau social numérique (RSN)

les RSN sont des réseaux sociaux virtuels qui gagnent de plus en plus en popularité, non seulement dans la sphère économique ou publique mais aussi dans le monde académique. Un grand nombre de RSN est issu du développement des nouvelles technologies et de la popularité des médias sociaux qui permettent aux utilisateurs d'interagir, de se contacter, d'échanger des informations, de partager leurs intérêts en commun en ligne de manière générale sans limite de distance ni de temps.
Les RSN peuvent être une source concrète et très riche pour étudier les phénomènes liés aux comportements des réseaux sociaux eux-mêmes, ou bien aux comportements des utilisateurs. De manière générale, le terme réseau social est employé pour désigner indifféremment un réseau social traditionnel ou un réseau social numérique. Notons que si les définitions de média social et réseau social numérique semblent très proches, les réseaux sociaux numériques se construisent quasi uniquement sur les interactions entre utilisateurs ; la création de contenu étant une possibilité et non une exigence dans ce cas.[Sirinya,2017]

3 Développement historique du Web social

Le web des années 1990 ressemblait beaucoup à la combinaison d'un annuaire téléphonique et les pages jaunes et malgré la puissance de raccordement de liens hypertextes, il donne peu de sens à la communauté parmi ses utilisateurs.
Cette attitude passive à l'égard du Web a été brisée par une série de changements dans les habitudes d'utilisation et technologiques qui sont désormais désignés comme Web 2.0, un mot inventé par Tim O'Reilly [Tim O'Reilly]. Dans ce qui suit, nous résumons l'histoire et les caractéristiques qui définissent le Web 2.0. Les changements qui ont conduit à son niveau actuel d'engagement social en ligne n'ont pas été radicales ou individuellement significatifs. Néanmoins, ce jeu d'innovations dans l'architecture et les modèles d'utilisation du Web a conduit à un rôle tout à fait différent du monde en ligne comme plate-forme pour la communication et l'interaction sociale intense.
L'augmentation de notre capacité à obtenir de l'information et du soutien social en ligne peut être quantifiée. Une récente enquête d'envergure basée sur des entretiens avec 2200 adultes montre que l'Internet améliore de manière significative la capacité de maintenir leurs réseaux sociaux en dépit de craintes initiales concernant les effets de la diminution du contact avec la vie réelle. L'enquête confirme que non seulement les réseaux sont maintenus et étendus en ligne, mais ils sont également activé avec succès pour traiter les situations de la vie tels que l'obtention d'un soutien en cas de maladie grave, à la recherche d'emplois, et de s'informer sur les grands investissements, etc... La première vague de socialisation sur le Web était due à l'apparition des blogs, des wikis et d'autres formes de communication et de collaboration en ligne.
Les premiers réseaux sociaux en ligne (également appelés services de réseaux sociaux) entrés sur le terrain en même temps que les blogs et les wikis ont commencé à décoller. En 2003, le premier arrivant Friendster25 attiré plus de cinq millions d'utilisateurs enregistrés en l'espace de quelques mois, qui a été suivi par Google et Microsoft de départ ou d'annoncer des services similaires. Bien que ces sites disposent d'une grande partie de la même teneur qui apparaissent sur les pages Web personnelles, ils fournissent un point d'accès central et ajoute la structure dans le processus de partage des renseignements personnels et de la socialisation en ligne.

4 Le réseau traditionnel et le réseau social en ligne

Les individus se regroupent sur la base d'intérêts communs ou de valeurs partagées et le modèle de réseautage traditionnel s'est tout simplement transposé sur le Web. L'abolition de limites géographiques, temporelles et, jusqu'à un certain point, psychologiques semble être un facteur déterminant dans la propulsion du réseautage en ligne. Voici un aperçu des différences identifiées entre le réseau traditionnel et le réseau en ligne.

Réseau traditionnel	Réseau social en ligne
Selon une base géographique.	Sans frontières.
Basé sur des intérêts communs.	Basé sur des intérêts communs.
Limité par la classe sociale, la religion.	Sans limites (en principe).
Diffusion restreinte de l'information.	Diffusion en temps réel de l'information.
Pouvoir des leaders d'opinion limité à une présence dans les médias traditionnels ou à des actions en personne.	Présence des leaders d'opinions en ligne très importante. Influence en temps réel et exponentielle.
Diffusion et promotion de l'innovation et des nouveautés limitées par les lieux physiques ou par les médias traditionnels nécessaires à la communication.	Diffusion et promotion de l'innovation et des nouveautés en temps réel.
Information personnelle inexistante ou limitée au groupe d'appartenance.	Affichage en ligne d'information personnelle sur les membres.

FIGURE 2.1 – Comparaison entre le réseau traditionnel et le réseau social en ligne [Rahila.2015].

.

Ce tableau permet de constater deux grandes différences entre les réseaux traditionnels et les réseaux sur le Web. D'une part, le réseau en ligne est caractérisé par la notion d'instantanéité. D'autre part, il démontre la grande ouverture causée en partie par la réduction des limites physiques.

5 Analyse des réseaux sociaux

Pour analyser, étudier, extraire des informations à partir de réseaux sociaux, on a deux catégories principales :

5.1 SNA (social Network Analysis) :

a été développé par les sociologues à découvrir les propriétés des réseaux sociaux en mettant l'accent sur les relations sociales entre les acteurs d'un réseau. Plusieurs études ont été menées depuis les années 1930 l'exploitation de structures, d'identifier les acteurs de réseaux de liens mondiaux des rôles et des positions, et en examinant les modes d'interaction au sein des réseaux sociaux. Bien que les techniques d'analyse de réseaux sociaux révèlent des informations importantes sur les réseaux sociaux.

5.2 Link Mining :

a été introduit par des informaticiens pour extraire des motifs cachés à partir des données disponibles. Elle peut être considérée comme la tâche d'appliquer des techniques d'exploration de données sur les réseaux, tout en tenant compte explicitement sur les liens entre les acteurs des réseaux sociaux [L. Getoor and C.P.Diehl 2005]. L'objectif de l'exploration de données est de trouver des connaissances inconnues cachées et potentiellement utiles à partir d'une grande quantité de données. En fait, les techniques d'exploration de données sont devenues vitales pour découvrir des informations cachées de réseaux sociaux.

FIGURE 2.2 – Schéma d'analyse d'un réseau social [savadogo, 2018]

.

6 Types et caractéristiques des réseaux sociaux numériques

Le travail de (Kaplan et Haenlein, 2010) propose plusieurs catégories de médias sociaux. Nous nous sommes intéressés aux catégories qui possèdent, de manière explicite ou implicite, la caractéristique d'un réseau social dit numérique. La caractéristique « réseau social » sera dite explicite lorsque les liens entre utilisateurs sont construits explicitement par eux. La caractéristique « réseau social » sera dite implicite lorsque les liens entre utilisateurs ne sont pas explicites et peuvent être construits à partir des interactions ou actions des utilisateurs (annotations, réponses, etc.). En s'appuyant sur ce travail, nous listons ci-après les catégories principales de réseau social existantes.

Site de réseautage social (social networking site) il s'agit d'une application qui permet de créer un profil personnel, d'inviter d'autres utilisateurs qui auront accès à ce profil afin de communiquer, envoyer des messages publics ou privés. Cette application permet également de partager des contenus de ce profil sous la forme de textes, images, vidéos ou bien audio. On distingue différents types de réseaux sociaux en fonction du contexte et de leur utilisation. Les réseaux peuvent être qualifiés de :

- **généralistes :** ces sites permettent de créer et d'agrandir son cercle d'amis, les plus connus étant Facebook , Google+ , les plus spécifiques étant les sites de rencontre (ex. Meetic) ;

- **professionnels :** comme LinkedIn ou Viadeo qui sont devenus des outils indispensables dans la relation entre professionnels en permettant de construire des réseaux professionnels personnalisés (« réseautage » professionnel). Il existe aussi des réseaux sociaux professionnels spécialisés par métiers (avocat , marketing, finance…).

- **focalisés sur les intérêts :** comme la musique (MySpace , LastFM , Deezer , SoundClound), la littérature (Babelio , GoodReads), le cinéma (IMDb), … .

- **centrés sur les services et la vie quotidienne, sur sa vie de quartier (Peuplade) Blog :** un blog peut être considéré comme une sorte de page web personnelle sur laquelle une ou plusieurs personnes publient périodiquement des contenus. Contrairement au site web personnel, le blog bénéficie d'une structure éditoriale préexistante, sous la forme d'outils de publication plus ou moins formatés. Les utilisateurs peuvent ajouter des commentaires et entrer en conversation sur les billets (post) de leur blog. Les blogs ont un caractère polymorphe puisque toutes les formes d'expression sont utilisées (image, vidéo, texte, audio).

• **Micro-blog (microblogging service) :** il s'agit d'une nouvelle forme de média social, dont

la conception dérive de celle du blog, elle permet aux utilisateurs de publier de courts messages (tweet) destinés à leurs abonnées (followers). Le micro-blog a pour objectif de diffuser de l'information en temps réel. Il peut contenir non seulement du texte mais aussi des images, des vidéos embarquées ou bien des liens vers des sites web. Il est donc à mi-chemin entre le blog et la messagerie instantanée. Le micro-blog le plus populaire est Twitter mais il existe également d'autres plateformes comme SinaWeibo , Soup .

• **Communauté de partage d'informations :** l'objectif de ce type d'application est le partage de contenus multimédias entre utilisateurs. Dans le contexte du web 2.0, les utilisateurs peuvent créer, indexer, commenter et partager des contenus. Ce type d'application permet de partager des images (Flickr , Instagram , Pinterest ...), des vidéos (Youtube , Dailymotion , ...), des présentations (Slideshare), etc.

• **Forum de discussion :** un forum est un espace de discussion public qui permet aux utilisateurs d'échanger des points de vue sur les sujets qui les intéressent ou de poser des questions. Généralement les discussions dans le forum sont archivées et cela permet des communications asynchrones entre utilisateurs. Les sujets de discussion sont souvent affichés par ordre chronologique. Les discussions peuvent s'effectuer de manière privée ou publique. Il existe plusieurs forums de discussions en ligne orientés sur différents centres d'intérêt de l'utilisateur comme par exemple Reddit , 4chan , Usenet . Nous pouvons également citer les sites de questions/réponses (Q et A) comme Quora ou StackExchange qui rassemblent plusieurs forums de discussion spécialisés (par exemple, StackOverflow qui est orienté sur la programmation, MathOverflow qui traite de problèmes en mathématiques).

• **Wikis :** site permettent à un groupe de personnes de développer un site Internet de manière collaborative alors qu'ils n'ont aucune notion de HTML ou autre langage de programmation. N'importe qui peut modifier les pages. Le wiki le plus connu est l'encyclopédie en ligne :
Les réseaux sociaux dits numériques sont beaucoup étudiés dans le domaine du marketing. Plus généralement, les médias sociaux permettent aux organisations et aux entreprises d'avoir une interaction directe avec leurs clients. L'objectif poursuivi peut être professionnel, par exemple, les campagnes de marketing qui promeuvent des produits à travers les réseaux sociaux (ex. partager une photo de publicité d'un produit pour gagner un cadeau, partager des tags sur les produits, événements créés par l'organisation) ou encore en politique (ex. lors de l'élection du président des états Unis, Barack Obama a utilisé Twitter et Facebook pour présenter sa campagne). Ces types de campagne peuvent être appelés « Buzz ». Un buzz désigne un événement ou un phénomène qui attire l'attention des utilisateurs dans les media sociaux et implique un partage et une diffusion de l'information. Le buzz peut aussi induire la création de nouveaux

liens entre personnes (rencontres autour du buzz, ...). Après avoir donné une typologie rapide des réseaux sociaux numériques, la sous-section suivante établit le développement historique du web social et une comparaison, surtout une différenciation, entre réseaux sociaux numériques et traditionnels et développement historique du web social.

7 Techniques de recommandation basée sur un réseau social

Les techniques de recommandation qui utilisent les données issues d'un réseau social sont nommées SNBLs (Social Network-BasedRecommendation). La majorité de ces techniques se basent sur les approches traditionnelles de filtrage de contenus ou de filtrage collaboratif avec des améliorations et des extensions qui permettent d'intégrer facilement des données sociales. Dans ce qui suit, nous présentons les trois approches principales associées à ce type de recommandation.

7.1 Recommandation basée sur la confiance

Les réseaux sociaux permettent d'exploiter la confiance entre individus facilement ; on peut construire les systèmes qui donnent automatiquement des recommandations des amis dans un réseau social.En effet, l'apparition et la croissance des réseaux sociaux est la condition préalable pour le développement des techniques basées sur la confiance.
La confiance peut être définie dans un système de recommandation comme la similarité entre individus. C'est la similarité au niveau des préférences, par exemple, les nombre d'objets appréciés par deux individus est une métrique de la confiance entre eux.
La confiance est modélisée dans un réseau social par un graphe dont les nœuds représentent les individus et les liens représentent la relation de confiance. Chaque lien est caractérisé par un poids, un nombre réel qui représente le niveau de confiance entre les deux nœuds. La confiance entre deux individus n'est pas forcément symétrique.
Les approches pour construire un algorithme de recommandation basées sur la confiance sont très diverses. Néanmoins, toutes sont composées de deux parties principales :

- *Construire un modèle de confiance.*
- *un modèle de calcul et prévoir le niveau de l'intérêt d'un individu sur un objet.*

7.2 Exploitation des données textuelles dans le Web social

Parmi les données textuelles générées par les utilisateurs dans le Web social, les tags sont les plus utilisés pour la recommandation. Un tag (ou étiquette) est un mot-clé ou un terme associé à de l'information (par exemple une image, un article, ou un clip vidéo). Les tags sont habituellement choisis de façon personnelle par le créateur ou le consommateur de l'objet. De nombreux réseaux sociaux permettent aux utilisateurs d'ajouter des tags aux objets pour qu'ils puissent les retrouver facilement plus tard. Citons à titre d'exemple, les travaux qui ont montré que les tags sont de bonnes représentations des intérêts des utilisateurs et qu'ils peuvent être utilisés pour retrouver leurs préférences [Le Tran,2011].
Dans [Szomszor, Alani, Cantador, O'Hara et Shadbolt, 2008], la notion de folksonomie (folksonomy en anglais) est considérée comme un système de classification de tags représentant une vue de l'utilisateur sur l'ensemble des contenus d'un système. Dans ce contexte, les auteurs présentent une méthode pour la consolidation automatique des profils des utilisateurs, en fonction de leurs tags à travers plusieurs folksonomies. Cette méthode permet la construction de profil sémantique d'intérêt au travers de quatre étapes :

1. identification des comptes détenus par un individu particulier sur différents réseaux sociaux .

2. récolte de l'historique complet des tags relatif à cet individu au sein de chaque réseau.

3. filtrage des tags en éliminant les fautes d'orthographe, les synonymes,... .

4. la génération sémantique du profil d'intérêt des utilisateurs à partir des tags filtrés.

Carmagnole et al ont utilisé un modèle de la connaissance (knowledge-based model) où le profil d'intérêt de l'utilisateur est représenté par une ontologie. L'objectif de ce modèle est d'utiliser des annotations « sociales » (commentaires et tags) comme un moyen pour déduire des connaissances sur les utilisateurs [Carmagnola, Vernero et Grillo, 2009]. D'autre part, certains systèmes de recommandation utilisent les tags sans considération de l'aspect sémantique. Dans ce cas, seuls les tags sur les objets à recommander sont utilisés. La méthode proposée par Lui est basée sur le graphe tripartite utilisateur-objet-tag pour calculer le niveau d'intérêt d'un utilisateur pour un objet à partir de ce graphe .

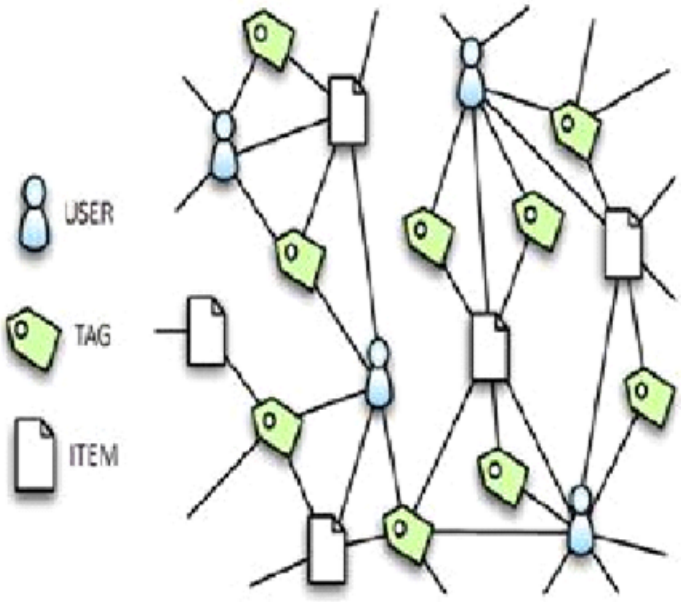

FIGURE 2.3 – Le graphe tripartite utilisateur-objet-tag [Liu, Zhang, Zhou, 2010]

.

Le système de recommandation de Bank et Franke utilise les évaluations des consommateurs pour découvrir le niveau de satisfaction d'un consommateur sur une caractéristique du produit. Cette analyse se fait en trois étapes :

1. le pré-traitement basé sur une analyse syntaxique et linguistique.

2. la détection de l'aspect du produit à évaluer.

3. l'analyse du sentiment permettant d'obtenir le niveau de satisfaction du consommateur.

Les études présentées ci-dessus montrent la possibilité d'utiliser les données textuelles pour enrichir le profil de préférences. Ces études ne rencontrent pas le problème de démarrage à froid mais elles sont limitées par leurs complexités (analyse textuelles, linguistiques, modèle de connaissances, ontologies,...).

7.3 Exploitation du profil déclaratif

La plupart des sites Internet permettent aux utilisateurs, lors de leurs inscriptions, de déclarer un profil, qui est constitué de données démographiques et de centres d'intérêts. Ces informations peuvent être utilisées pour remplir le profil de préférences des utilisateurs dans un système de recommandation. Cependant, il n'y a pas de systèmes de recommandation existants qui reposent uniquement sur des données démographiques. Un système peut difficilement donner des recommandations pertinentes parce qu'il n'y a pas une corrélation significative entre le goût d'un individu et ses informations démographiques. Néanmoins, on peut utiliser ces données comme une source supplémentaire pour améliorer la performance du système. Citons à titre d'exemple le projet présenté. Dans celui-ci, les auteurs proposent un filtrage collaboratif des données démographiques afin de recommander de la musique à partir des notes (ratings) attribuées à chaque chanson.

D'autre part, ce type de recommandation est peu précis à cause du profil déclaratif des utilisateurs, mais il présente une bonne solution pour le problème de démarrage à froid car les données démographiques sont toujours disponibles et peuvent être collectées facilement.

8 Conclusion

L'interaction des utilisateurs sur le web, la masse importante des données, la désorganisation de cette dernière est devenue un obstacle. Le web doit suivre l'évolution. La révolution technologique qui a donnée naissance de réseaux sociaux, ainsi que l'apparition de la technique des recommandations qui a permet aux utilisateurs des réseaux sociaux un travail rational et performant.
Pour cette raison nous proposons des recommandations qui ont pour objectifs l'amélioration de l'exploitation des données et facilite l'interaction des utilisateurs, Ces objectifs serrent détaillés dans les chapitres suivants.

Chapitre 3

classification

1 introduction

La Catégorisation de textes (C.T) est aujourd'hui un domaine de recherche bien établi et très actif. Les travaux portent depuis une quinzaine d'année sur les systèmes avec apprentissage des catégories à partir de corpus pré- étiquetés. Dans ce chapitre, nous présentons d'abord une définition de la catégorisation des textes ; ainsi que le processus de C.T qui est constitué de deux étapes : la pondération des termes, les méthodes de classification.

2 Apprentissage automatique(Machine Learning)

2.1 Définition

Machine Learning est une branche d'intelligence artificielle qui permet à une machine d'analyser un système, de comprendre pas à pas son fonctionnement et comme résultat d'effectuer ou simuler des différentes taches de ce système. L'algorithme d'apprentissage a pour objectif d'apprendre le fonctionnement du système étudié de manière active. Il connaît les entrées possibles du système et compose des séquences qu'il soumet au système (requêtes) pour observer ses réponses (séquences autorisée/refusée, valeurs renvoyées, etc.).[Orkhan ,lafarov et all .2012] D'autre part, nous pouvons définir le terme Machine Learning comme un ensemble d'algorithmes qui permettent d'apprendre le fonctionnement d'un système en observant régulièrement les tâches qu'il réalise, puis prédire son comportement et ses décisions.

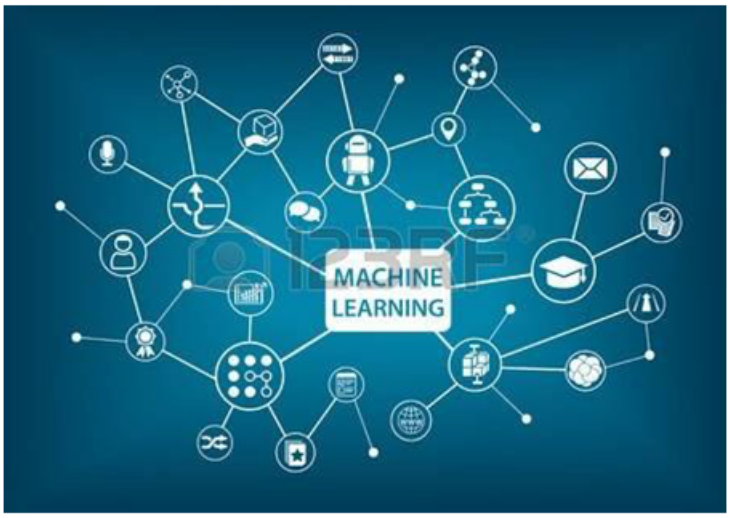

FIGURE 3.1 – Machine Learning

.

2.2 Modèles et types de Machine Learning

Dans le domaine de la machine Learning, il existe deux principaux types de tâches : supervisées et non supervisées.

1.2.1 Apprentissage non supervisé : consiste à ne disposer que des données d'entrée (X) et pas de variables de sortie correspondantes. Les problèmes d'apprentissage non supervisés peuvent être regroupés en problèmes de clustering et d'association.

1.2.2 Apprentissage supervisé : La majorité des experts machine learning utilisent un apprentissage supervisé. L'apprentissage supervisé consiste en des variables d'entrée (x) et une variable de sortie (Y). Vous utilisez un algorithme pour apprendre la fonction de mappage de l'entrée à la sortie. Y = f (X) .

Les problèmes d'apprentissage supervisé peuvent être regroupés en problèmes de régression et de classification. Pour la régression ce que nous souhaitons prédire est une valeur numérique continue (par exemple, le prix d'un appartement) alors que pour la classification, on cherchera à déterminer une valeur discrète et finie (par exemple, à quelle espèce appartient une fleur). Pour pouvoir faire ces prédictions, les algorithmes effectuent des calculs plus ou moins complexes sur des valeurs numériques [Christophe Thovex. ,2012]. Il est donc nécessaire de transformer les caractéristiques des éléments à analyser en un tableau numérique représentant ces caractéristiques. Afin d'obtenir de bonnes prédictions, il est nécessaire de rassembler toutes les caractéristiques importantes (pour le résultat qu'on souhaite établir) et de bien les modéliser.

Par exemple, si on cherche à déterminer le prix de vente d'un appartement, il faudra probablement prendre en compte sa superficie, son emplacement, son entretien, s'il est meublé...etc. Mais peut être aussi d'autres caractéristiques auxquelles on ne pense pas forcément au premier abord. Il est donc essentiel de très bien connaître le domaine métier pour pouvoir modéliser correctement les éléments que l'on souhaite traiter. Une fois toutes ces caractéristiques transformées en valeurs numériques, on peut appliquer un algorithme de machine Learning à nos données pour pouvoir construire un modèle prédictif. Un algorithme très simple est par exemple de faire une régression linéaire sur les caractéristiques des éléments. La valeur à prédire sera donc une combinaison linéaire des caractéristiques. On peut aussi essayer des régressions logarithmique ou polynomiale mais il suffit simplement de créer de nouvelles fonctionnalités pour pouvoir se ramener à une simple régression linéaire.

3 Apprentissage profond (deep Learning)

deep Learning est un ensemble de techniques d'apprentissage automatique qui a permis des avancées importantes en intelligence artificielle dans les dernières années. Dans l'apprentissage automatique, un programme analyse un ensemble de données afin de tirer des règles qui permettront de tirer des conclusions sur de nouvelles données. L'apprentissage profond est basé sur ce qui a été appelé, par analogie, des « réseaux de neurones artificiels », composés de milliers d'unités (les neurones) qui effectuent chacune de petites opérations simples. Les résultats d'une première couche de neurones servent d'entrée aux calculs d'une deuxième couche et ainsi de suite. Par exemple, pour la reconnaissance visuelle, des premières couches d'unités identifient des lignes, des courbes, des angles… des couches supérieures identifient des formes, des combinaisons de formes, des objets, des contextes… Les progrès de l'apprentissage profond ont été possibles notamment grâce à l'augmentation de la puissance des ordinateurs et au développement de grandes bases de données (big data).

4 Déférence entre machine Learning et deep Learning

Machine Learning	deep Learning
l'apprentissage automatique est une approche pour atteindre l'intelligence artificielle	l'apprentissage en profondeur est une technique de mise en œuvre de l'apprentissage automatique
L'apprentissage automatique utilise des algorithmes pour analyser des données, tirer des leçons de ces données et prendre des décisions éclairées en fonction de ce qu'il a appris.	structures d'apprentissage en profondeur algorithmes en couches pour créer un "réseau de neurones artificiels" capable d'apprendre et de prendre des décisions intelligentes par ses propres moyens.
apprentissage automatique : prend relativement moins de temps à former, allant de quelques secondes à quelques heures.	les algorithmes d'apprentissage en profondeur prennent en compte un si grand nombre de paramètres qu'ils nécessitent généralement beaucoup de temps pour se former.
il y a quelques milliers de points de données utilisés normalement pour l'analyse, fonctionne en bas de gamme.	Un million de points de données sont normalement utilisés pour l'analyse, le travail est effectué sur des machines haut de gamme.
L'apprentissage automatique utilise des algorithmes tels que : KNN,naïve bayes	l'apprentissage en profondeur interprète les données à l'aide de réseaux de neurones

Tableau 3 : Déférence entre machine Learning et deep Learning.

.

5 Classification (technique descriptive)

Est une méthode qui permet de regrouper des objets (personne, intérêts...etc.) en groupes, ou familles de sorte que les objets d'un même groupe se ressemblent le plus possible, et ceux de groupes distincts diffèrent le plus possible.
Le nombre des groupes est parfois fixés, ils ne sont pas prédéfinis mais déterminés au cours de l'opération. Dans le cas d'extraction de données à partir des réseaux sociaux, cette méthode nous permettra de regrouper les intérêts d'une personne par classe (personnage, produit, achat, consommation, activité...etc.). D'une autre façon nous pouvons dire que cette méthode descriptive permet de décrire de façon simple une réalité complexe en la résumant.

5.1 Catégorisation de textes (CT)

Plusieurs définitions de la CT ont vus le jour depuis son apparition, nous citons dans ce contexte les deux définitions suivantes :
Définition : La CT est une relation bijective qui consiste à "chercher une liaison fonctionnelle entre un ensemble de textes et un ensemble de catégories (étiquettes, classes)".[Radwan JALAM,2003]

5.2 La pondération des termes

La pondération des termes permet de mesurer l'importance d'un terme dans un document. Cette importance est souvent calculée à partir de considérations et interprétations statistiques. L'objectif est de trouver les termes qui représentent le mieux le contenu d'un document. Pour calculer la pondération on distingue les méthodes suivantes : [LAHLOU OUCHIHA,2016]

5.2.1 Le sac à mots : La façon la plus simple et la plus évidente pour la représentation d'un document texte par un document vecteur, est d'utiliser les mots comme descripteurs. Ainsi, nous construisons un sac à mots (bag of words) de tous les mots qui apparaissent au moins une fois dans le corpus. Cette méthode, qui conserve le sens naturel des descripteurs, est loin de répondre à toutes les attentes de la classification de texte, à cause, notamment, de certaines anomalies liées à la variation des fréquences par rapport à la longueur du document, au problème des mots composés, et principalement, au fait que l'ordre d'apparition des mots dans les phrases du document n'est pas pris en considération. Les mots sont regroupés en vrac et traités d'une manière indépendante, ce qui nuit considérablement à la sémantique du texte.

5.2.2 Les N-grams : Cette méthode de représentation de documents texte consiste à partager ce dernier en séquences de n caractères. En effet, si nous considérons seulement les lettres de l'alphabet comme caractères et dans le cas ou "n" égal à 1, c'est à dire la séquence contient juste une seule lettre, est ce que c'est une bonne manière de représenter un document dans le but de le classifier ? Certainement non, même si dans certains cas cela semble très efficace, notamment dans la reconnaissance de la langue. C'est pour cette raison d'ailleurs que le "n" est toujours supérieur à 1. Prenons l'exemple de la phrase "La classification supervisée " et essayons de la représenter en ngrams caractères. - si n=2 nous aurons "La","a "," c","cl","la","as","ss","si","if",etc. - si n=3 nous aurons "La ","a c"," cl","cla","las","ass","ssi","sif","ifi", etc. - si n=4 nous aurons "La c","a cl"," cla","clas","lass","assi","ssif","sifi","ific", etc Dans la littérature, le consensus s'est porté sur n=3, car n<3 la représentation est très élémentaire, tandis que n>3 on génère beaucoup de colonnes. Les

trigrammes de caractères semblent très efficaces et sont utilisés dans plusieurs applications . [LAHLOU OUCHIHA,2016]

5.2.3 Fréquence des termes (TF) : On désigne par TF la fréquence d'un mot (descripteur) dans un texte donné. C'est un calcul de fréquence très simple, mais qui s'avère efficace et pratique. Nous l'utilisons souvent en association avec d'autres fréquences. Nous dénombrons plusieurs manières de calcul de la TF : TF absolue : c'est le nombre de fois qu'un terme apparait dans un texte donné.

$$\mathrm{TF} = NT \tag{3.1}$$

ou NT est le nombre de fois où le terme est apparu dans le texte.
• **TF relative :** C'est le rapport entre le nombre de fois qu'un terme est apparu dans le texte sur le nombre de tous les termes du texte. Cette dernière est utilisée généralement pour limiter l'impacte de la longueur des textes. En effet, un terme qui apparait 6 fois dans un texte de 100 termes n'a pas le même degré de discrimination que celui qui apparait le même nombre de fois (6 fois) dans un texte de 20 termes.

$$\mathrm{TF} = NT * ST \tag{3.2}$$

où NT est le nombre de fois que le terme est apparu dans le document. et, ST est le nombre de tous les termes du document. TF booléenne : se contente juste de la présence ou de l'absence du terme dans le texte. TF = 1 ou 0 Le principal inconvénient de la fréquence des termes est le fait qu'il est possible, et d'ailleurs c'est un cas très probable en pratique, qu'un terme apparait avec une fréquence assez grande dans tous les documents d'un corpus. Dans ce cas, le terme en question perd toute sa notion de discrimination relative au degré de présence. Une autre notion vient rectifier ce cas exceptionnel, nommée IDF(Inverse documents frequancies).
5.2.4 Fréquence documents inverses (IDF) : Elle mesure en quelque sorte le degré de rareté d'un terme, non pas dans un document, mais dans tous les documents d'un corpus elle est définie par cette équation :

$$\mathrm{IDF}\,(t_i) = \log \frac{D}{\mathrm{DF}\,(t_i)} \tag{3.3}$$

Où N Doc est le nombre de documents dans le corpus, et DocT est le nombre de documents dans lesquels le terme est apparu. Si le terme est très présent dans tout le corpus alors le rapport sera égal à 1 et IDF = 0 donc le terme est neutralisé. Si par contre il apparait dans un seul document la valeur est maximale.

$$\mathrm{IDF} = \log NDoc \tag{3.4}$$

Cette pondération à elle seule ne définie nullement le degré de discrimination d'un terme dans un document puisque elle est relative au corpus. Mais l'association de IDF avec TF donne des résultats intéressants. **5.2.5 TFIDF :** Nous avons vu que la fréquence d'un terme dans un document joue un rôle important dans le calcul de son degré de discrimination. En revanche, la représentation d'un texte, dans le but de le classifier, ne dépend pas seulement de son contenu, mais elle est liée étroitement au corpus auquel le texte appartient, la rareté de ce terme au sein des autres documents du corpus s'avère aussi importante que sa fréquence (abondance) dans le document en question. Cette combinaison judicieuse de ces deux principes (abondance particulière et rareté générale) a engendré la pondération dite TFIDF. Elle est calculée avec la formule suivante :

$$\text{TFIDF}\,(t_i, d_j) = TF\,(t_i, d_j) * IDF\,(t_i) \tag{3.5}$$

où TF est relative ou absolue.

6 Méthodes de classification

La catégorisation de textes comporte un choix de technique d'apprentissage (ou classificateur) disponibles. Parmi les méthodes d'apprentissage les plus souvent utilisées figurent : l'analyse factorielle discriminante, la régression logistique, les réseaux de neurones, les plus proches voisins, les arbres de décision, les réseaux bayésiens, les machines à vecteurs supports et, plus récemment, les méthodes dites de boosting.

6.1 •Machine à vecteur support (SVM) :

Le but de SVM est de trouver un classificateur qui sépare au mieux les données et maximise la distance entre ces deux classes. Ce dernier est un classificateur linéaire appelé hyperplan. Comme montré dans la Figure(3.2) cet hyperplan sépare les deux ensembles de points.

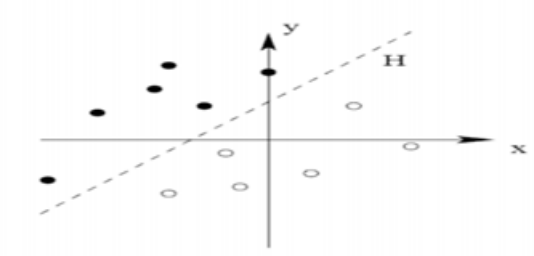

FIGURE 3.2 – la séparation du l'hyper plan par les SVM [Lahlou, 2016]

.

Les points les plus proches, qui seuls sont utilisés pour la détermination de hyperplan, sont appelés v ecteurs de support (voir Figure (3.3)).

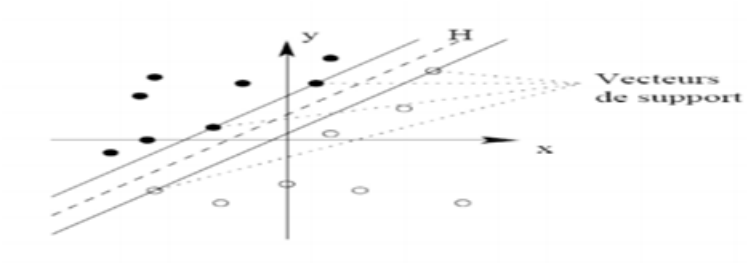

FIGURE 3.3 – Les vecteurs de support [Lahlou, 2016]

.

6.2 •k plus proches voisins :

C'est une méthode très connue dans le domaine de la catégorisation des textes. L'idée de K-plus proches voisins est de représenter chaque texte dans un espace vectoriel, dont chacun des axes représente un élément textuel (peut être un mot sous sa forme brute ou sous une forme lemmatisée). [Simon JAILLET.et al.2005]
L'algorithme de catégorisation de K-plus proches voisins est présenté comme suit :

Algorithme : **algorithme de classification par K-PPV**
Paramètre : le nombre K de voisin
Contexte : un échantillon de T textes classés en C=c_1, c_2,…c_n classes
Début
  Pour chaque texte T **faire**
    Transformer le texte T en vecteur T = (x_1, x_2, …, x_m),
    Déterminer les K plus proches textes du texte T selon une métrique de distance,
    Combiner les classes de ces K exemples en une classe C.
  Fin pour
Fin
Sortie : le texte T associé à la classe C.

FIGURE 3.4 – L'algorithme de K-PPV [Simon Jaillet.et al, 2005]

.

Le choix du paramètre K est primordial pour le bon fonctionnement de cette méthode. [Simon JAILLET.et al.2005]

6.3 •Naïve bayes :

Cette méthode se base sur le théorème de Bayes permettant de calculer les probabilités conditionnelles. Dans le cas de la CT, la méthode Naïve bayes est utilisée comme suit : on cherche la classification qui maximise la probabilité d'observer les mots du document. Lors de la phase d'entraînement, le classificateur calcule les probabilités qu'un nouveau document appartient à telle catégorie à partir de la proportion des documents d'entraînement appartenant à cette catégorie. Il calcule aussi la probabilité qu'un mot donné soit présent dans un texte, sachant que ce texte appartient à telle catégorie. Quand un nouveau document doit être classé, on calcule les probabilités qu'il appartienne à chacune des catégories à l'aide de la règle de Bayes[Simon RÉHEL.2005]
La formule :

$$P\left(c_k|\vec{d_j}\right) = \frac{P\left(c_k\right) P\left(\vec{d_j}|c_k\right)}{P\left(\overline{d}_j\right)} \tag{3.6}$$

.

• **Avantages :** l'hypothèse d'indépendance des descripteurs du classificateur Naïf Bayes le rend simple et efficace. Son entrainement ne nécessite pas beaucoup de documents, il a fait ses preuves dans la classification de documents courts, notamment les courriels (Ham/Spam) Inconvénients : Contrairement aux documents courts, les documents longs posent un grand problème pour le classificateur Naïf Bayes, un riche vocabulaire favorise les dépendances entre les descripteurs(termes).

6.4 • Arbre de décision :

Les arbres de décision sont composés d'une structure hiérarchique en forme d'arbre. Un arbre de décision est un graphe orienté sans cycles, dont les nœuds portent une question, les arcs des réponses et les feuilles des conclusions ou des classes terminales Un classificateur de texte basé sur la méthode d'arbre de décision est un arbre de nœuds internes qui sont marqués par des termes, les branches qui sortent des nœuds sont des tests sur les termes et les feuilles sont marquées par catégories. [Karima ABIDI,2011]

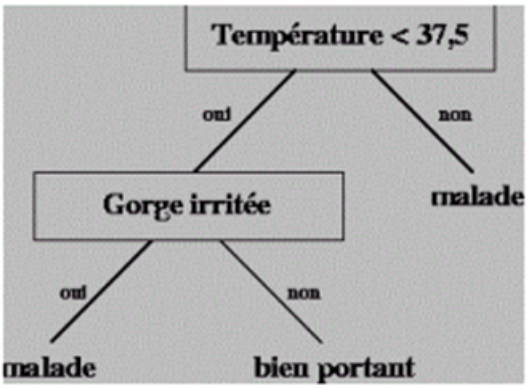

FIGURE 3.5 – Exemple d'arbre de décision [Lahlou, 2016]

.

Une méthode pour effectuer l'apprentissage d'un arbre de décision pour une catégorie Ci consiste à vérifier si tous les exemples d'apprentissage ont la même étiquette. Dans le cas contraire, nous sélectionnons un terme Tk, et nous partitionnons l'ensemble d'apprentissage en classes de documents qui ont la même valeur pour Tk, et à la fin on crée les sous arbres pour chacune de ces classes. Ce processus est répété récursivement sur les sous arbres jusqu'à ce que chaque feuille de l'arbre généré de cette façon contienne des exemples d'apprentissage attribués à la même catégorie Ci, qui est alors choisie comme l'étiquette de la feuille. L'étape la plus importante est le choix du terme de pour effectuer la partition

6.5 • réseaux de neurone :

Les réseaux de neurones artificiels sont habituellement utilisés pour des tâches de classification. Par analogie avec la biologie, ces unités sont appelées neurones formels. Un neurone formel est caractérisé par :

• Le type des entrées et des sorties.
• Une fonction d'entrée.
• Une fonction de sortie.
Le connexionnisme peut être défini comme le calcul distribué d'unités simples, regroupées en réseau. Un réseau de neurone est un ensemble d'éléments ou unités extrêmement simples (neurones) se comportant comme des fonctions de seuil, suivant une certaine architecture ; Chaque neurone prend en entrée une combinaison des signaux de sortie de plusieurs autres neurones, affectés de coefficients (les poids) ; L'apprentissage s'effectue sous le contrôle des associations prédéfinies entre documents (entrées du réseau) et classes (sorties du réseau) qui

fixent le comportement du réseau souhaité. La différence entre le comportement réel et désiré est une erreur qui sera à la base de l'apprentissage sous la forme d'une fonction de coût ou d'un signal d'erreur. Dans ce cas, l'apprentissage s'effectue en réajustant chaque fois les poids Wi. Donc les algorithmes d'apprentissage permettent de calculer automatiquement les poids qui correspondent en réalité à des paramètres permettant de définir les frontières des classes. Une structuration en couches effectue en cascade différents traitements sur un ensemble de données. Ces données sont présentées sur une couche terminale, appelée couche d'entrée ; elles sont ensuite traitées par un nombre variable de couches intermédiaires ou couches cachées. Le résultat est exposé sur l'autre couche terminale, la couche de sortie. [Maaath, 2011] Le principe général d'une approche neuronale est présenté ci-dessous. :

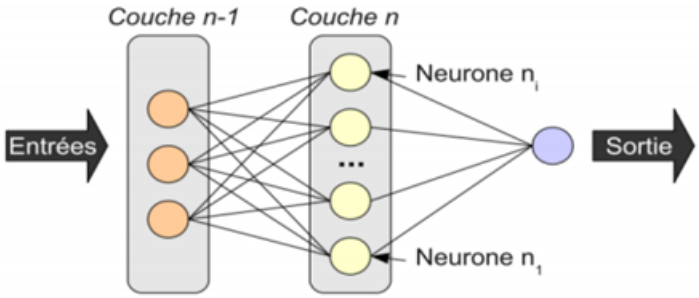

FIGURE 3.6 – Architecture générale d'un réseau de neurones artificiels [Lahlou, 2016].

.

Un réseau de neurones artificiels est composé d'une ou de plusieurs couches se succédant dont chaque entrée est la sortie de la couche qui la précède comme illustré sur la figure(3.6).

7 Conclusion

Dans ce chapitre nous avons présenté quelques techniques de la catégorisation automatique des textes et Apprentissage automatique.
La catégorisation de texte a essentiellement progressé ces dix dernières années grâce à l'introduction des techniques héritées de l'apprentissage automatique qui ont amélioré très significativement les taux de bonne classification.Le chapitre suivant présente la conception de notre système.

Chapitre 4

Conception

1 introduction

Les réseaux sociaux sont devenus nécessaires dans la vie, ils doivent donc être développés. La recommandation des publications aux utilisateurs est l'une tâche importantes des réseaux sociaux, elle consiste à proposer à l'utilisateur les publications susceptible de l'intéresser. Cette proposition peut être réalisée selon plusieurs facteurs et critères entre autre l'historique des publications de l'utilisateur, qui est l'objet de notre travail.
Dans ce chapitre, nous allons d'abord présenter Les diagrammes de cas d'utilisation, puis nous allons lister les fonctionnalités du système via des diagrammes de séquences ainsi que Le modèle entité-association et nous allons finir par présenter la stratégie de recommandation proposée.

2 Vue fonctionnelle du système

Dans cette partie nous allons reproduire les déférentes tâches du système sous la forme de diagrammes UML.
UML est un langage de modélisation graphique à base de pictogrammes. Il est apparu dans le monde du génie logiciel, dans le cadre de la conception orientée objet [Waad Gasmi,2011]. Au nal, le langage UML est une synthèse de tous les concepts et formalismes méthodologiques les plus utilisés, pouvant être utilisés, grâce à sa simplicité et à son universalité, comme langage de modélisation pour la plupart des systèmes devant être développés.

2.1 Les diagrammes de cas utilisation

ce diagramme qui résume les fonctionnalités de l'application.

FIGURE 4.1 – Figure : Diagramme de cas d'utilisation [Savadogo, 2018]

.

Nous nous intéressons principalement à la tâche de la publication qui est présente dans la Figure ci-dessous :

Diagramme Cas d'utilisation Ajouter un ami : Permet à chaque membre d'ajouter un nœud à son réseau.

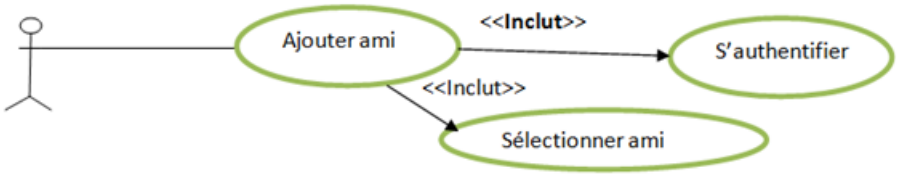

FIGURE 4.2 – Cas d'utilisation Ajouter un ami[Savadogo, 2018]

.

Diagramme Cas d'utilisation de la publication : Interface de publication : Cette interface permet au membre de : *• Publier des messages textuels • Publier du contenu hypermédia (articles ou fichiers ou vidéos ou photos) • Identifier éventuellement des amis dans une*

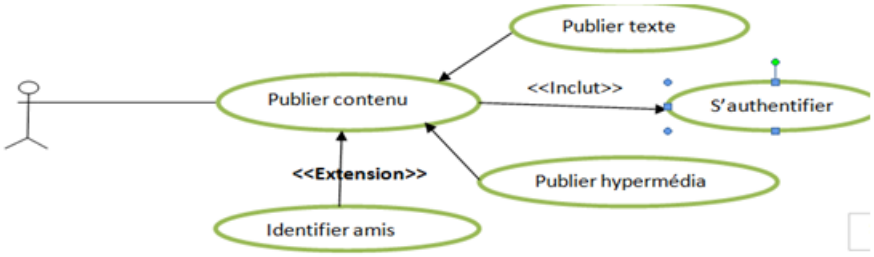

FIGURE 4.3 – Cas d'utilisation de la publication[Savadogo, 2018]

.

2.2 Les diagrammes de séquences

Les diagrammes de séquences sont la représentation graphique des interactions entre les acteurs et le système selon un ordre chronologique dans la formulation Unified Modeling Language. Ce formalisme de représentation nous permettra de mettre en évidence les interactions entre les membres et le système afin de bien traiter celles-ci. Dans le diagramme de séquence, les interactions se font à travers des messages dits synchrones et asynchrones. Un message synchrone nécessite de la part du système une réponse et est représente par une flèche pleine. Tandis qu'un message asynchrone est représenté par une flèche discontinue et ne nécessite aucune réponse système. Lors de la connexion d'un membre à notre application de réseaux social par exemple, il publie à ce dernier une publication synchrone. Le système traitera cette pubication et renverra à son tour une réponse.

• Diagramme de séquence du cas « inscription » Lorsqu'un utilisateur demande l'autorisation de s'inscrire sur le réseau social, il doit tout d'abord remplir un formulaire qui lui est fourni par le système. Après l'envoi des informations saisies par l'utilisateur, celles-ci sont traitées afin de déterminer si les informations fournies respectent un certain formatage. Le système doit par exemple vérifier que le format de l'email fourni par l'utilisateur respecte au moins l'un des formats de la plus part des emails.

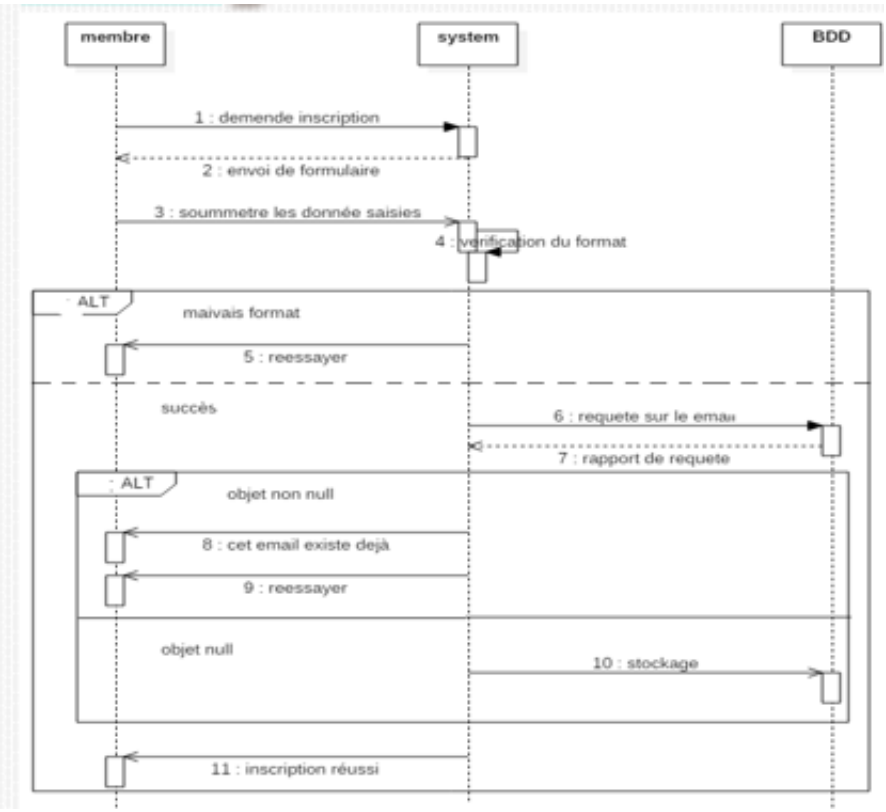

FIGURE 4.4 – Diagramme de séquence de l'inscription[Savadogo, 2018]

.

• **Diagramme de séquence du cas « authentification »** Lorsque l'utilisateur demande l'accès à l'application, il doit tout d'abord s'identifier par son e-mail et son mot de passe via le système qui prend en charge de vérifier et consulter la base de données. S'il est accepté, donc il aura l'accès au système et aux applications du menu correspondant. Sinon, le système lui affiche un message d'erreur afin de rectifier ses données.

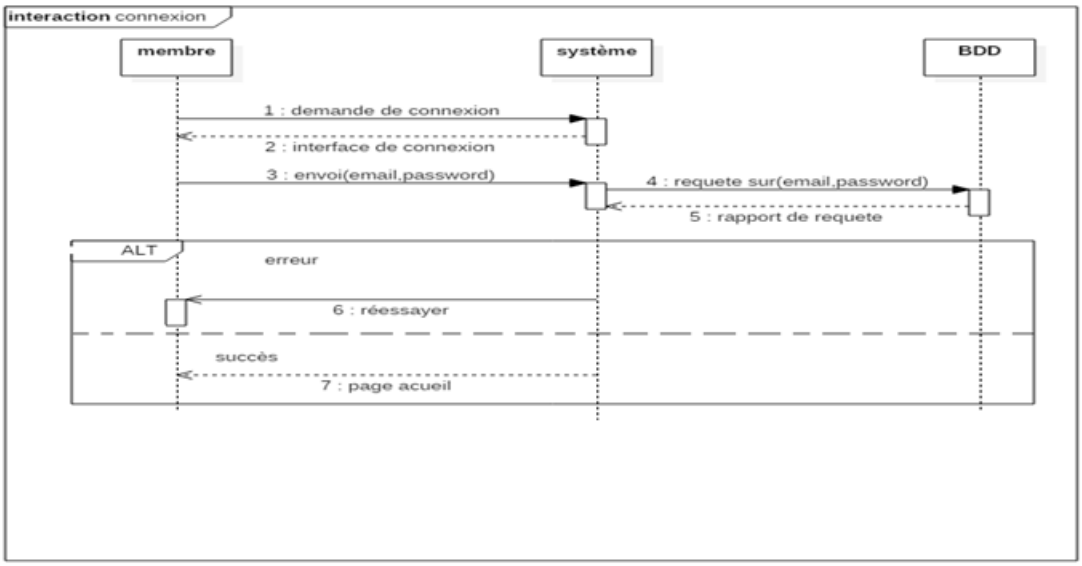

FIGURE 4.5 – Diagramme de séquence du cas « authentification »

.

• **Diagramme de séquence du cas « inviter ami »**

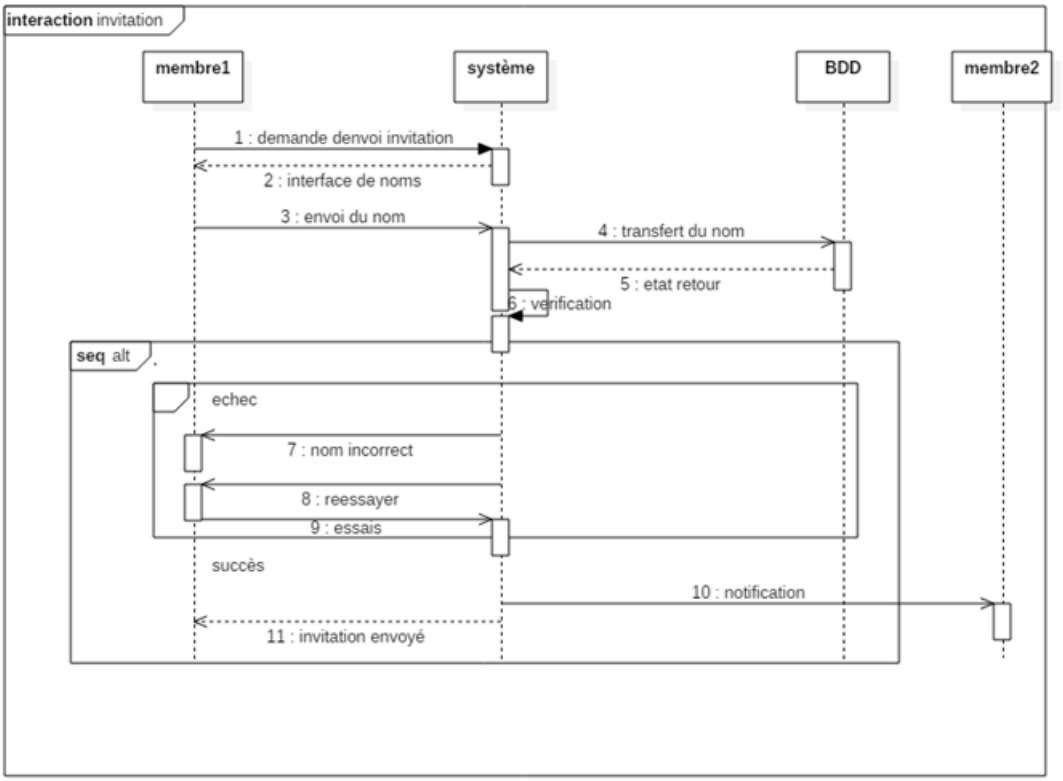

FIGURE 4.6 – Diagramme de séquence du cas « inviter ami »[Savadogo, 2018]

.

• **Diagramme de séquence du cas « publication »** Lorsqu'un membre demande de publier du contenu, le système récupère ce contenu et le stocke dans la base de données de l'application.

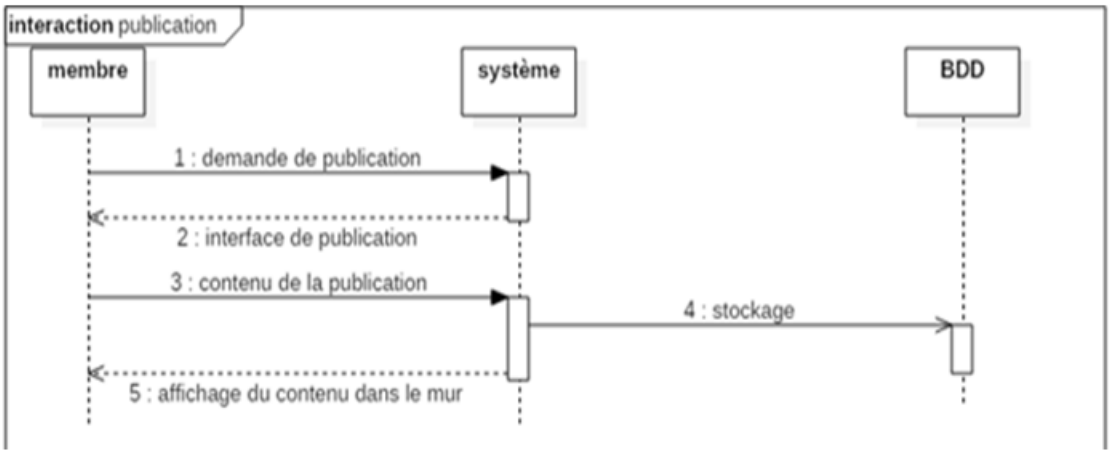

FIGURE 4.7 – Diagramme de séquence du cas « publication »[Savadogo, 2018]

.

• **Diagramme de séquence du cas « recommandation d'une publication** Lorsqu'un membre publie des publications, la publication est enregistrée dans la base de données. Ensuite le système récupère les données (les catégories des publications) sur cette base il les recommande des publications aux utilisateur Selon leurs intérêts.

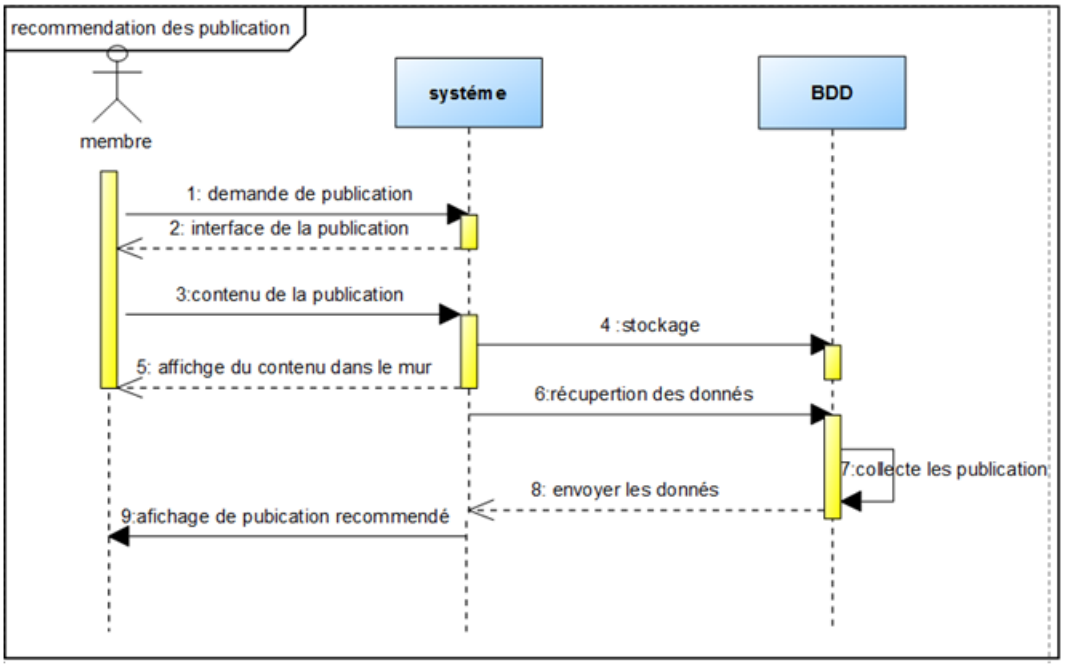

FIGURE 4.8 – Diagramme de séquence du cas « recommandation d'une publication »

.

2.3 Le modèle entité-association

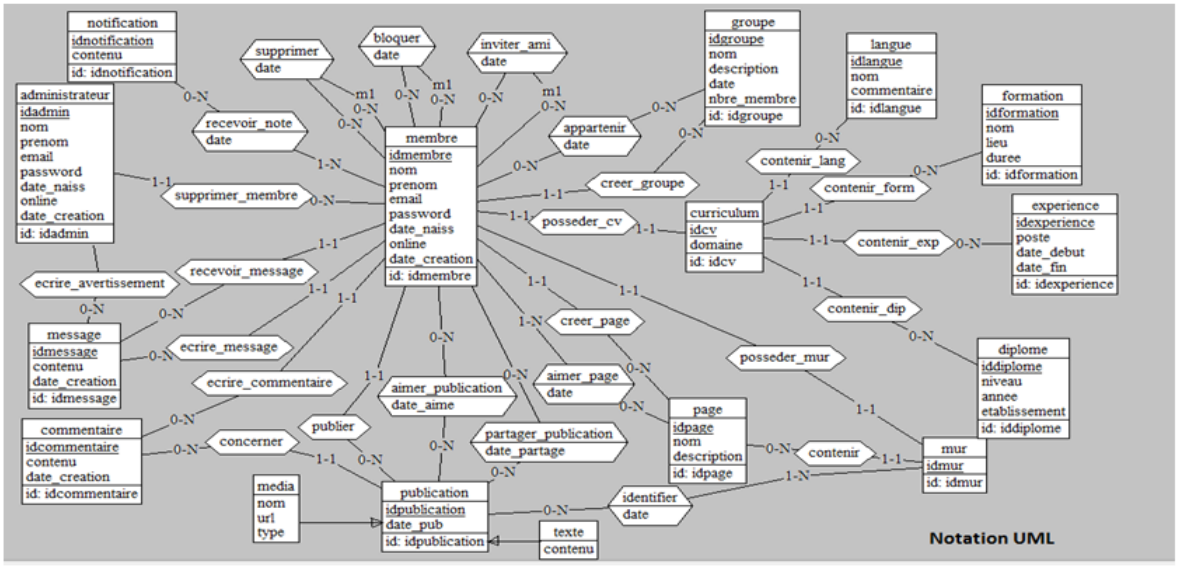

FIGURE 4.9 – Schéma du modèle entité-association [savadogo, 2018].

. Le modèle suivant est un extrait du modèle globale de réseaux sociaux professionnel. Dans ce

modèle on a ajouté deux entité la première est de nommée « catégorie-pub » qui sert à classifier les publications. La deuxième entité est appelé « suggestion » qui a pour objectif de faire le filtrage le classement des publications.

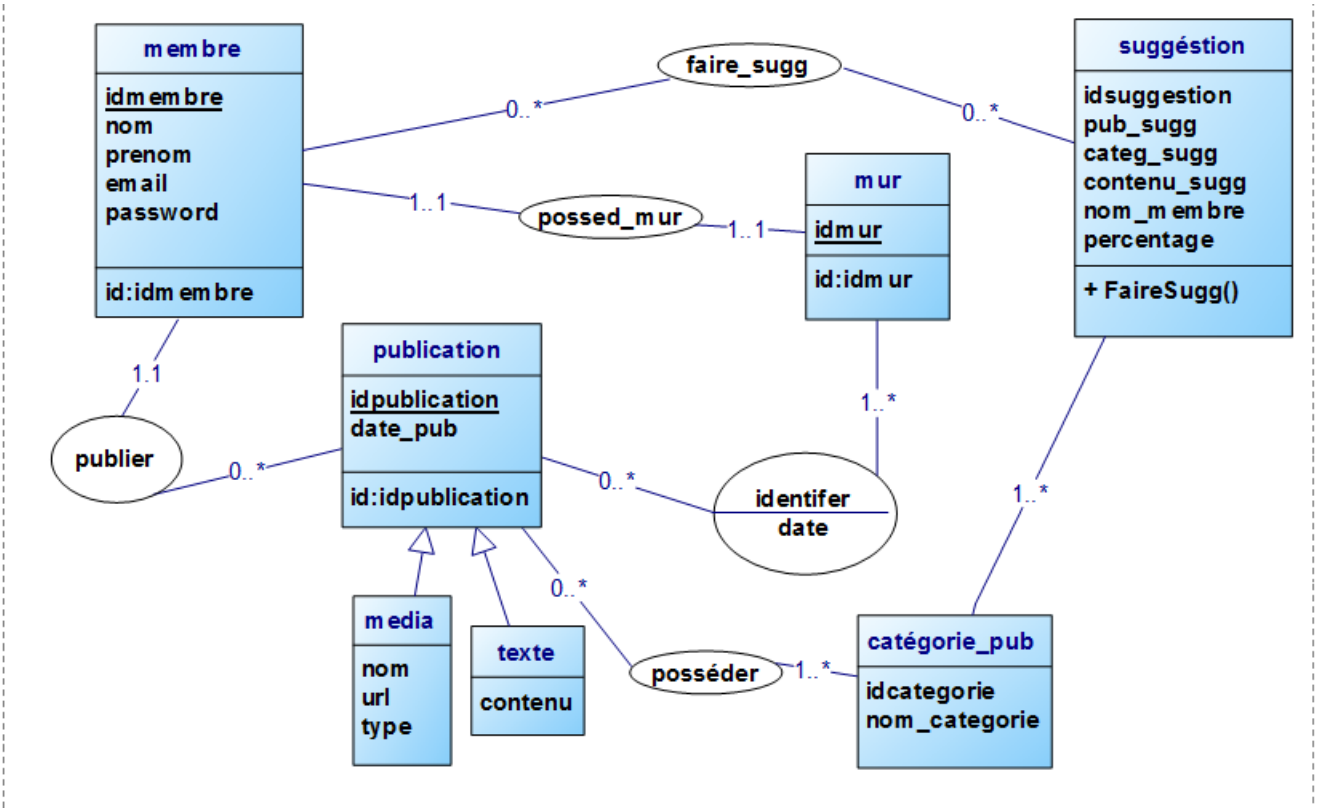

FIGURE 4.10 – Schéma extrait du modèle entité-association

.

2.4 Le modèle relationnel

Le modèle relationnel est une manière de modéliser les relations existantes entre plusieurs informations, et de les ordonner entre elles. Les informations sont présentées sous forme de tables contenant des attributs. Une table contient au moins une clé primaire et éventuellement des clés secondaires. Dans notre modèle relationnel, les clés primaires sont mises en gras tandis que les clés secondaires sont précédées du signe dièse [].

• membre(**idmembre**, nom, prenom, email, password, date-naiss, date-creation, idmur)

Attributs	Description
idmembre	Identifiant du membre
nom	Nom du membre
prenom	Prenom du membre
email	Email du membre
password	Mot de passe du membre
date-naiss	Date de naissance du membre
date-creation	Date de création du comte du membre
idmur	Identifiant du mur du membre

Tableau 4 : Description de la table «membre».

.

• publication(**idpublication**, contenu, date-pub, idmembre)

Attributs	Description
idpublication	Identifiant de la publication
contenu	Contenu de la publication
date-pub	Date de publication
idmembre	Identifiant du membre qui a publié

Tableau 5 : Description de la table «publication».

.

• suggestion(**idsuggéstion**,Pub-sugg ,pourcentage ,idmembre)

Attributs	Description
idcategorie	Identifiant de la catégorie
idmembre	Identifiant du membre qui a publié
Pub-sugg	Publication suggestion (recommandé)
pourcentage	Pourcentage de classification

Tableau 6 : Description de la table «suggestion».

.

• categorie(**idcategorie**, nom-cat)

Attributs	Description
idcategorie	Identifiant de la catégorie
nom	nom de catégorie

Tableau 7 : Description de la table «categorie».

.

• Inviter-ami(**idmembre**, date)

Attributs	Description
idmembre	Identifiant du membre qui a publié
date	Date d'inviter

Tableau 9 : Description de la table "amis".

.

3 Schéma général de système

Notre travail s'inscrit dans le cadre de l'apprentissage automatique plus précisément les systèmes de recommandations et leurs fonctionnements pour aider les utilisateurs à leur proposer des publications en relation avec leur domaine et leur centre d'intérêt. Pour atteindre cet objectif, nous proposons l'architecture générale représentée dans le schéma suivant :

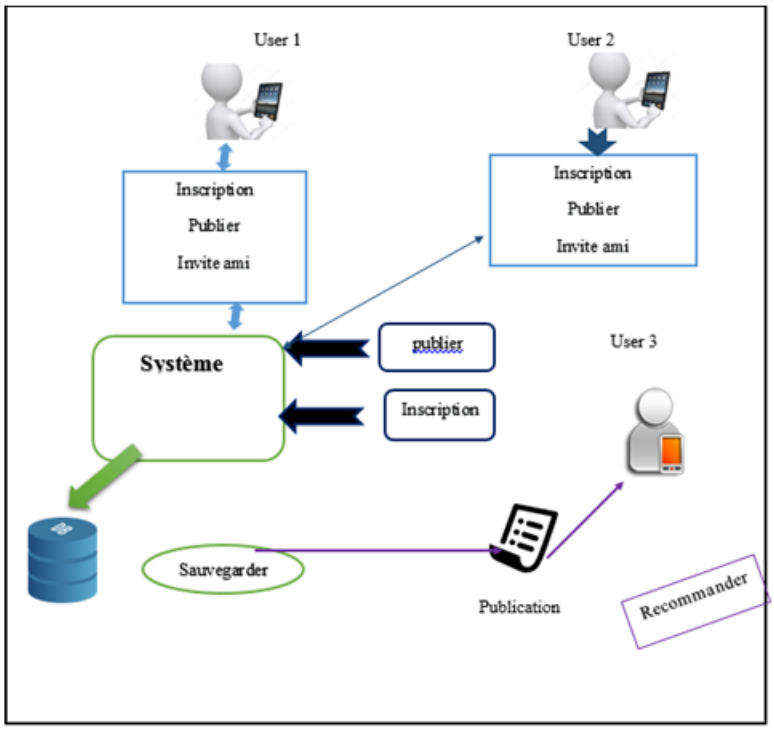

FIGURE 4.11 – Schéma général de notre système

.

Notre recommandation a pour objectif de classifier et filtrer les publications Selon le domaine et le centre d'intérêt des utilisateurs.

4 Stratégie de recommandation :

Qui se résume dans trois modules :
* module d'inscription.
* module de publication.
* module de recommandation.

4.1 module d'inscription :

Dans ce module, l'utilisateur introduit les informations de son profil personnel comme : Nom ; Prénom ; Email ; Date de naissance ; Mot de passe ... ec.
Le mot de passe a fin de sécuriser son profil une fois validé.
Il devient membre dans notre réseau social qui lui donne le droit de publier.

4.2 module de publication :

Après l'inscription de nouvel utilisateur a notre réseau social, et après qu'il soit un membre de ce réseau social il pourra donc publier ses réflexions forme de : (texte, image... ec).

4.3 module de recommandation :

A pour rôle la classification des différentes publications Selon le domaine et le centre d'intérêt des utilisateurs, Une fois la classification des publications terminées. Commence la phase de filtrage selon l'historique des publications d'utilisateur, on utilise les résultats de classification, on tenant compte du temps et nombre des publications de l'utilisateur.

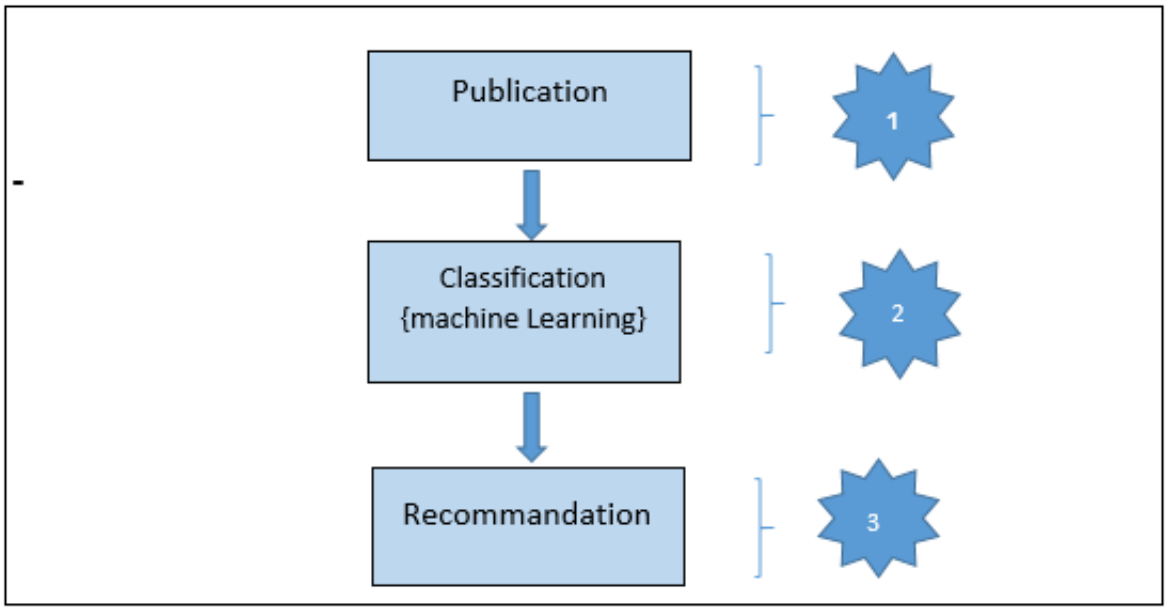

FIGURE 4.12 – module de recommandation

.

• **publication :** publier le contenu de publication par un utilisateur. En prenant l'exemple de texte.

• **classification :** pour faire la classification on a suivi les étapes suivants :

- découper le contenu en un ensemble de mots.

- après le découpage du texte, on a utilisé la méthode **TFIDF** pour calculer la fréquence des mots afin de déduire leur l'importance dans ce contenu et d'afficher le résultat en pourcentage.

- pour faire la classification on a utilisé l'algorithme de **Naïve Bayes**, cet algorithme repose sur le résultat précédent pour classer le contenu dans une catégorie à laquelle il appartient.

• **recommandation :**

- Après classification et connaissance de chaque publication et classification, nous sauvegardons ces publications dans la base de données, Chaque utilisateur doit publier plus de deux publications dans le même domaine pour connaître ses intérêts et lui proposer les publications de ses amis.

- Une suggestion se produit lorsqu'un ami ou les amis d'un utilisateur publient des publications en rapport avec leurs intérêts. Ils apparaissent sur la page d'accueil de l'utilisateur en tant que "publications recommandé".

•**TFIDF :**

est facile à implémenter et peuvent être utilisé pour détecter la similarité statistique entre deux documents, mais elle ne prennent pas en compte la position des mots dans le texte. L'autre inconvénient comme toutes les méthodes statistiques, il n'y a pas de sémantique associée. Pour tous nos vecteurs les attributs sont l'ensemble des mots uniques et les individus sont des vec-

teurs qui reflètent la présence des mots unique dans chaque publication. A la fin de cette étape on génère les fichiers associés à chaque type de transformation pour pouvoir les passer vers les algorithmes de classification.

• **Algorithme naïve bayes :** On à utiliser cette algorithme pour classifier les publications qui publie par l'utilisateur lui même.

Algorithm 1.1 Naive Bayes

Train($\mathbf{X}, \mathbf{Y}$) {reads documents $\mathbf{X}$ and labels $\mathbf{Y}$}
  Compute dictionary D of $\mathbf{X}$ with n words.
  Compute m, m_{ham} and m_{spam}.
  Initialize $b := \log c + \log m_{\text{ham}} - \log m_{\text{spam}}$ to offset the rejection threshold
  Initialize $p \in \mathbb{R}^{2\times n}$ with $p_{ij} = 1$, $w_{\text{spam}} = n$, $w_{\text{ham}} = n$.
  {Count occurrence of each word}
  {Here x_i^j denotes the number of times word j occurs in document x_i}
  for $i = 1$ to m **do**
    if $y_i =$ spam **then**
      for $j = 1$ to n **do**
        $p_{0,j} \leftarrow p_{0,j} + x_i^j$
        $w_{\text{spam}} \leftarrow w_{\text{spam}} + x_i^j$
      end for
    else
      for $j = 1$ to n **do**
        $p_{1,j} \leftarrow p_{1,j} + x_i^j$
        $w_{\text{ham}} \leftarrow w_{\text{ham}} + x_i^j$
      end for
    end if
  end for
  {Normalize counts to yield word probabilities}
  for $j = 1$ to n **do**
    $p_{0,j} \leftarrow p_{0,j}/w_{\text{spam}}$
    $p_{1,j} \leftarrow p_{1,j}/w_{\text{ham}}$
  end for
Classify(x) {classifies document x}
  Initialize score threshold $t = -b$
  for $j = 1$ to n **do**
    $t \leftarrow t + x^j(\log p_{0,j} - \log p_{1,j})$
  end for
  if $t > 0$ **return** spam **else return** ham

FIGURE 4.13 – algorithme de naïve bayes

.

• **Algorithme basé sur le modèle :** est comme le nom l'indique basés sur des modèles, supposés réduire la complexité. Ces modèles basés sur des classificateurs. On a utilisé la classification pour créer des classes qui nous permettent d'ordonner les publications selon leur catégories (le plus grand pourcentage). C'est-à-dire le système recommande les publications qui ont le plus grand pourcentage.

	P1	P2	P3
U1	OUI	NON	NON
U2	NON	OUI	NON
U3	OUI	OUI	NON
U4	OUI	NON	OUI
U5	OUI	OUI	NON
U6	NON	NON	OUI
percentage	0.4	0.3	0.2

Tableau 10 : un tableau montre la classification des publications selon pourcentage.

.

Il s'agit de recommander des publications sur la base du comportement passé de l'utilisateur, son historique c'est-à-dire ces actions : (publier), Qui sont enregistrées dans le système comme une catégorie. Historique catégorie de l'utilisateur est la base de notre système de recommandation pour recommander à l'utilisateur des publications selon un facteur :

```
Get publication  () ;
Classifie _par _spécialité () ;
Assigner _a _chaque_ utilisateur_ son_ publication () ;
```

.

5 Conclusion

Dans cette partie nous avons parlé de notre travail, ainsi que la conception que nous avons proposée pour réaliser notre travail, on a utilisé la classification l'une des techniques les plus utilisées dans les systèmes de recommandation. Pour clarifier comment fonctionne le système, on a proposé un schéma général et pour la modélisation on a utilisé UML.
dans le chapitre suivant nous allons en présenter la réalisation de notre système.

Chapitre 5

Réalisation

1 introduction

Vu l'importance des réseaux sociaux dans le quotidien des gens et le nombre d'utilisateurs qui ne cesse d'augmenter l'implication du système de recommandation est devenu une obligation, notamment la création de système de recommandation de publication qui fait l'objet de notre projet. Dans ce chapitre nous allons parler des grands axes de la réalisation de notre application dans le but de décrire le principe implémentant l'environnement de notre système de recommandations.

2 Technologies et outils de développement

2.1 • JSON (JavaScript Object Notation) :

est un format de données textuelles, générique, dérivé de la notation des objets du langage ECMAScript. Il permet de représenter de l'information structurée. Créé par Douglas Crockford, il est décrit par la RFC 4627 de l'IETF ..[json]

2.2 • JavaScript :

est un langage de script orienté objet principalement utilisé dans les pages HTML. A l'opposé des langages serveurs (qui s'exécutent sur le site), JavaScript est exécuté sur l'ordinateur de l'internaute par le navigateur lui-même. Ainsi, ce langage permet une interaction avec l'utilisateur en fonction de ses actions (lors du passage de la souris au dessus d'un élément,du

redimensionnement de la page etc). La version standardisée de JavaScript est l'ECMAScript.

• jQuery :

est un framework Javascript sous licence libre qui permet de faciliter des fonctionnalités communes de Javascript.

2.3 • Php (HyperText Preprocessor) :

PHP est un langage de programmation interprété 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 disposant depuis la version 5 de fonctionnalités de modèle objet complète.

2.4 • Css :

Vous permet de définir l'apparence des textes (comme la police, la couleur, la taille, etc...), ainsi que l'agencement de la page (comme les marges, l'arrière-plan, etc...). CSS définit donc la présentation du document. CSS est l'abréviation de Cascading Style Sheets. Un style définit la façon dont un élément HTML (par exemple <h1>) sera affiché. Ces styles peuvent être définis dans une feuille de style externe (un fichier .css). Une feuille de style peut être utilisée pour définir la présentation de plusieurs documents HTML, ce qui permet de gagner beaucoup de temps. HTML a été conçu pour définir la structure d'un document pas sa présentation. Par conséquent tout ce qui est lié à la présentation d'un document devrait être défini à l'aide de CSS. Typiquement, il faut préférer CSS à l'utilisation de balises HTML permettant de définir la présentation d'un document (comme par exemple<font color ='…'>).

2.5 • Html :

Définition HTML est un langage pour décrire des pages web. HTML est un acronyme pour Hyper TextMarkupLanguage (langage de balisage d'hypertexte). Voici un exemple de code HTML :

```
<html>
  <head>
    <title>Le titre de la page</title>
  </head>
  <body>
    <h1>Mon premier titre</h1>
    <p>Mon premier <b>paragraphe</b></p>
  </body>
</html>
```

FIGURE 5.1 – Un exemple de code HTML.

.

3 Les logicielles de développement

3.1 • WampServer :

Wamp est une plate-forme de développement Web sous Windows pour des applications Web dynamiques à l'aide du serveur Apache2, du langage de scripts PHP et d'une base de données MySQL. Il possède également PHPMyAdmin pour gérer plus facilement vos bases de données [Anthony NIZAC.2018].

3.1.1 MySQL : est un système de gestion de base de données relationnel, un langage de requêtes vers les bases de données exploitant le modèle relationnel et utilise le langage SQL comme langage de requête. SQL est un langage de manipulation de bases de données mis au point dans la années 70 par IBM, il permet d'effectuer trois types de manipulations :

. *La manipulation des tables :* Création, suppression, modification de la structure.

. *Les manipulations des données de la base :* Sélection, modification, suppression d'enregistrement.

. *La gestion des droits d'accès aux tables :* Contrôle des données, droit d'accès, validation des modifications.

3.1.2 PhpMyAdmin : Est une interface d'administration pour le SGBD MySQL. Il est écrit en langage PHP et s'appuie sur le serveur HTTP Apache. Il permet d'administrer les éléments suivants :

• Les bases de données.
• Les tables et leurs champs (ajout, suppression, définition du type).
• Les index, les clés primaires et étrangères.
• Les utilisateurs de la base et leurs permissions.
• Exporter les données dans divers formats (CSV, XML, PDF, Open Document, Word, Excel et Latex).

4 Méthode de classification applique

4.1 • Php-ml :

Est une Bibliothèque d'apprentissage automatique pour PHP Nouvelle approche de l'apprentissage automatique en PHP. Algorithmes, validation croisée, réseau de neurones, prétraitement, extraction de fonctionnalités et bien plus dans une bibliothèque. Actuellement, cette bibliothèque est en cours de développement, mais vous pouvez l'installer avec Composer [PHP-ml, 2017].

Naïve Bayes Classifier : Classificateur basé sur l'application de théorème de bayes de fortes hypothèses d'indépendance (naïve) entre les caractéristiques.
• ***Train :*** Pour former un classificateur il suffit de fournir des échantillons de train et des étiquettes (comme ARRAY). Exemple vous pouvez former le classificateur à l'aide de plusieurs ensembles de données. Les prédictions seront basées sur toutes les données d'apprentissage.

```
$samples = [[5, 1, 1], [1, 5, 1], [1, 1, 5]];
$labels = ['a', 'b', 'c'];

$classifier = newNaiveBayes ();
$classifier->train ($samples, $labels);
```

.
vous pouvez former le classificateur a l'aide de plusieurs ensembles de données. les prédictions seront basées sur toutes les données d'apprentissage.

• ***Prédire :*** pour Prédire la Predict méthode d'utilisation des étiquettes .vous fournir un échantillon ou un tableau d'échantillons :

```
$classifier->predict ([3, 1, 1]);

/ Return 'a'

$classifier->predict ([[3, 1, 1], [1, 4, 1]);

/Return ['a', 'b']
```

.

5 Architecture technique de notre système :

• Au lancement de l'application le script PHP s'exécutera.

• Le script PHP va récupérer les données depuis la base de données MySQL. Ensuite ces données seront envoyées au réseau social.

• Ensuite, l'application site web va obtenir ces données .Elle les analysera et les affichera sur navigateur (site web : réseau sociale).

On a utilisé **PHP**, du fait qu'il permet de créer des applications web, et q'il offre des modules de classification . on a utilisé **JSON** en raison de la facilité d'implémentation. En effet, il représente des objets sous forme d'une chaîne de caractères, en utilisant une notation compatible avec php.

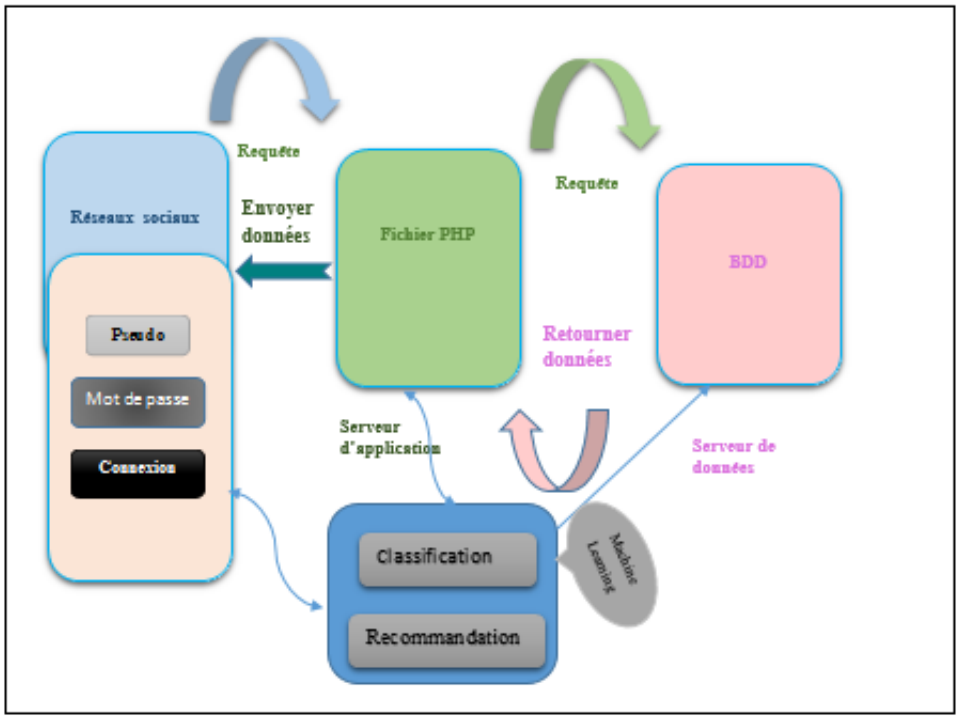

FIGURE 5.2 – Architecture technique de notre système.

.

6 Les fonctionnalités du système

• Onglet de "se connecter" :

Lorsque l'application se lance une interface comme décrit à la figure ci-dessous apparait offrant la possibilité à l'utilisateur de se connecter(login) ou de s'inscrire (sing up)s'il ne possède pas un compte utilisateur.

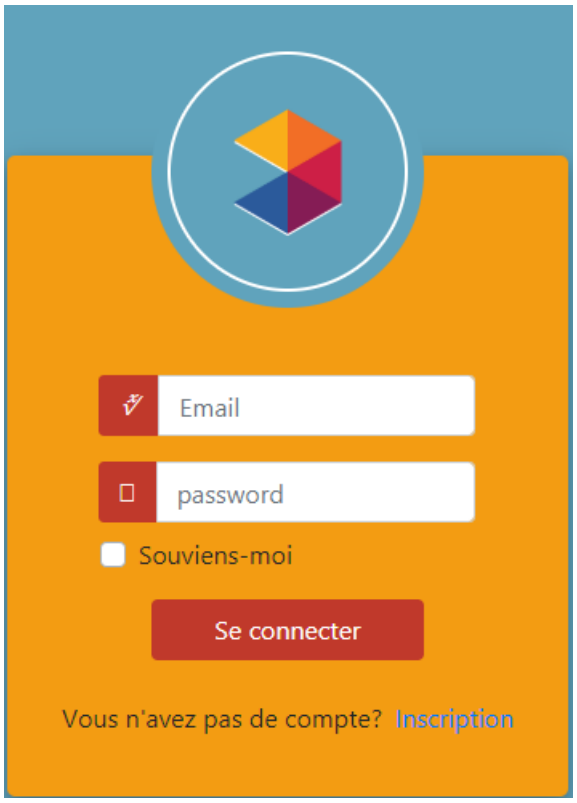

FIGURE 5.3 – Onglette de se connecter.

.

• **Onglet "d'inscription" :**

Lorsque l'utilisateur clique sur sing up de l'interface ci-dessus il est dirigé vers un formulaire lui permettant de créer un nouveau compte. Le compte est créé à partir du nom de l'utilisateur, son prénom, date de naissance, l'email et le mot de passe.

FIGURE 5.4 – Onglet d'inscription

.

Si les identifiants du membre sont corrects, il accède directement à l'interface d'accueil. C'est sur cette interfaces qu'on trouve les publications de l'utilisateur, de ses amis ainsi que des actualités du domaine d'activité du membre. L'ordre d'affichage de ces publications est basé sur un algorithme de sélection :

• **Etape1 :** Sélectionner les publications du membre.

• **Etape2 :** sélectionner les publications des amis du membre.

• **Etape3 :** afficher les publications sélectionnées par ordre de leur date de création.

• **Etape4 :** classification des publications Sélectionner selon le contenu.

• **Etape5 :** nos recommandations se font sur les résultats de la classification et sur l'historique de l'utilisateur lui-même afin de répondre aux préférences de l'utilisateur et ses domaines d'intérêt.

• **Onglet de "Mure" :**

Onglet d'informations sur le profil : montre les informations personnel du profil utilisateur .

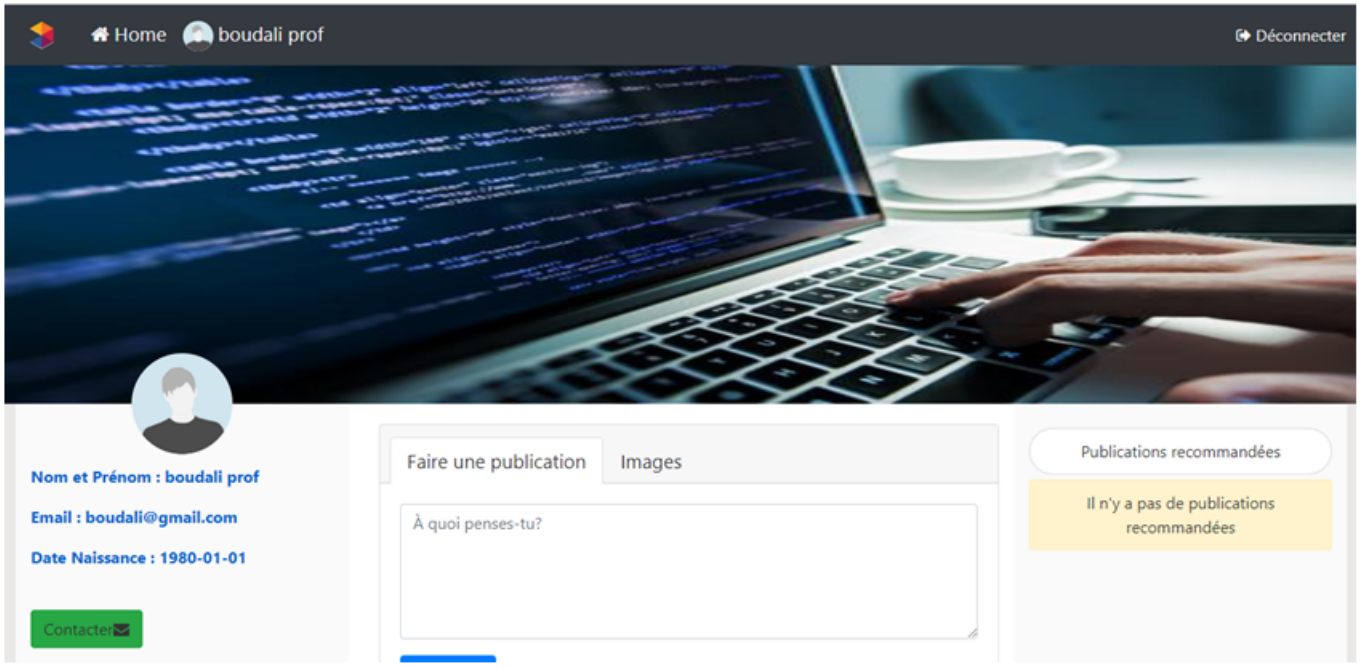

FIGURE 5.5 – Onglet d'inscription

.

• **Onglet "des publications recommandées" :**

1- la liste des publications recommandées pour l'utilisateur est vide.

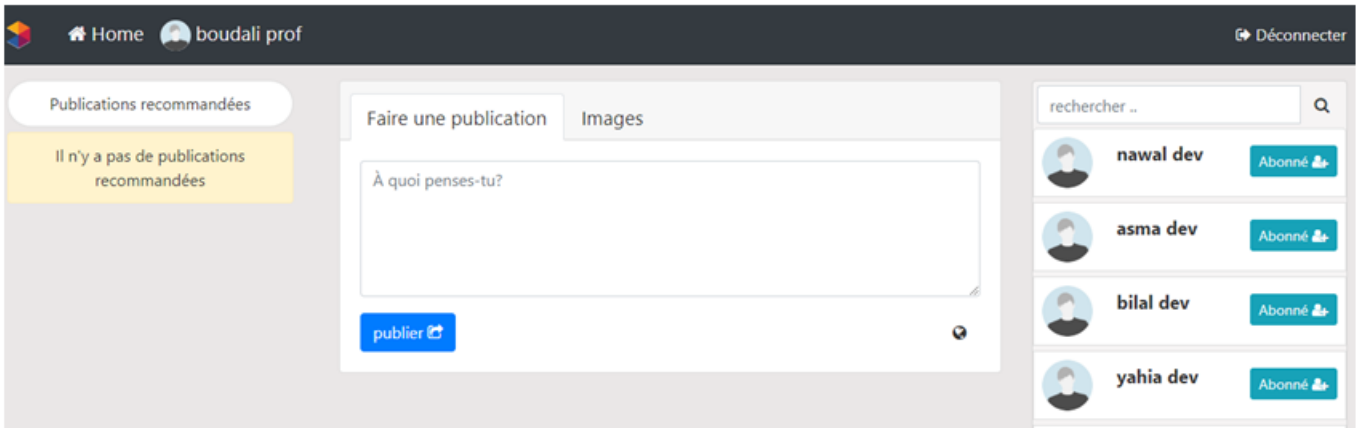

FIGURE 5.6 – Onglet "d'accueil (publication non recommander)".

.

2- elle permet d'afficher la liste des publications recommandées pour l'utilisateur.

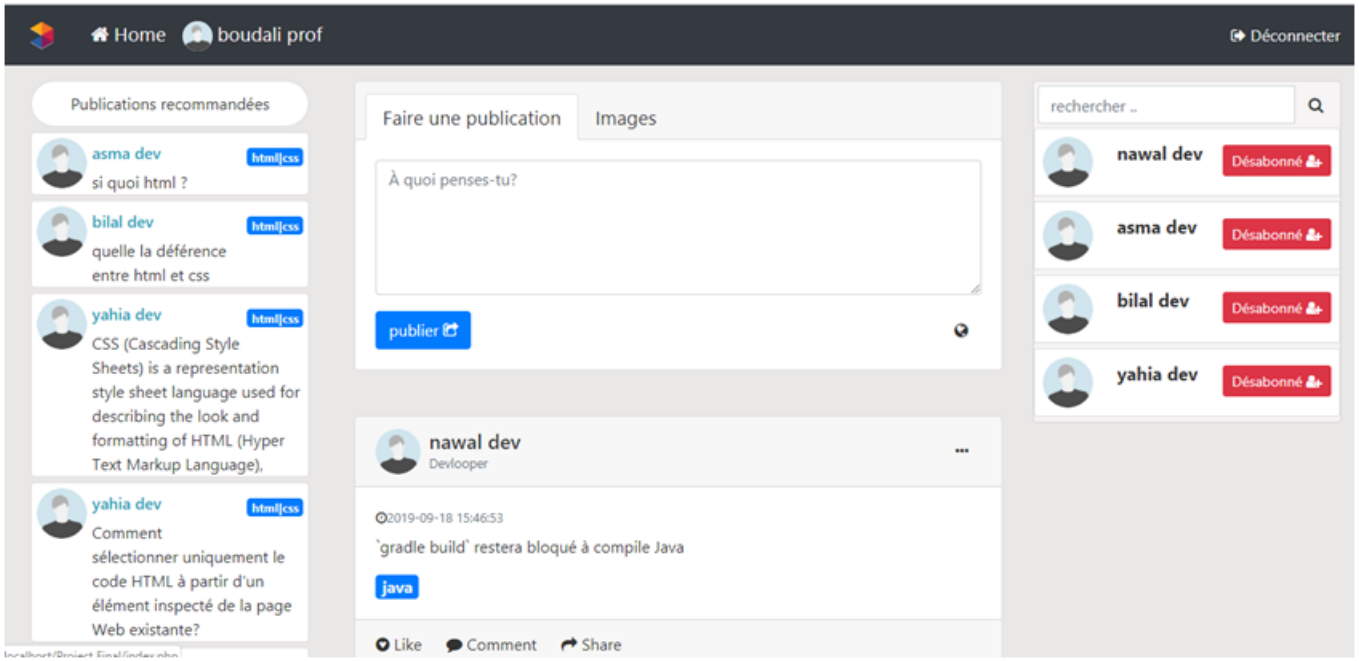

FIGURE 5.7 – Onglet "d'accueil (publication recommander)".

.

L'application offre aussi la possibilité à un membre de donner son avis sur une publication en commentant la publication ou en aimant la publication en cliquant sur le bouton « like ». Pour commenter une publication, le membre presse sur le bouton « comment » et une interface contenant les commentaires des autres membres de la même publication et une zone de texte

pour saisir un commentaire s'affiche.

7 Conclusion :

Dans ce chapitre, nous avons présenté les différentes phases par lesquelles nous sommes passés pour réaliser notre application. En effet, nous avons explicité toutes les étapes pour la construction du profil utilisateur sur un réseau social. En commençant par l'extraction des informations sur les activités de l'utilisateur ainsi que ses amis. Ces derniers sont utilisés comme point d'amorce dans la classification pour interroger les données publiées afin d'enrichir le profil utilisateur et faire une recommandation des publications pour chaque utilisateur.

Conclusion Général :

Le travail présenté dans ce mémoire a pour but d'améliorer l'expérience des utilisateurs des plateformes sociales via l'introduction de deux mécanismes complémentaires qui sont l'enrichissement du profil utilisateur ainsi que la recommandation de publication correspondants à ses centres d'intérêts. Nous nous sommes intéressés particulièrement de réseau social numérique, Dans le contexte de notre projet, nous avons développé un prototype qui permet de construire un profil utilisateur constitué d'une dimension utilisateur et d'une dimension sociale. Le résultat obtenu composé d'un ensemble de centres d'intérêts a été enrichi par l'analyse d'un data set Le dataset forme JSON utilisé a été interrogé pour enrichir la dimension obtenue dans le but de formuler certaines recommandations à l'utilisateur. Le développement réalisé repose sur l'agencement de plusieurs technologies et langages tels que php les techniques (TF/IDF), naïve bayes (php-ml) pour interroger le data set.

• **Perspectives :**

• *Nous travaillons sur l'amélioration de notre projet on ajoutant d'autre type de recommandation(recommandation des pages, des amis, des groupe...ect).*

• *Construire un moteur d'indexation automatique de publications.*

• *Amélioration du pourcentage de prédiction.*

Bibliographique :

[Adomavicius, G., Tuzhilin, A, 2005]. 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), (pp. 734-749).

[Amatriain, 2013]. Big personal : data and models behind Netflix recommendations. In Proceedings of the 2nd International Workshop on Big Data, Streams and Heterogeneous Source Mining : Algorithms, Systems, Programming Models and Applications, Big Mine 2013, Chicago, IL, USA, August 11, 2013, pages 1–6.

[Amit Sheth Ramesh Jain, 2007] : Amit Sheth Ramesh Jain, Social Networks and the Semantic Web. Amsterdam : Springer Science + Business Media, 2007.

[ATTIAS C., BRAYER C., BRUNO S., JACQUOT C., STRUL R., THOBELLEM A., VILLALBA A., 2010]. « Les médias sociaux », Paris, IAB France.

[Bambini et al, 2011]. Riccardo Bambini, Paolo Cremonesi, and Roberto Turrin. A recommender system for an IPTV service provider : a real large-scale production environment. In Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, editors, Recommender Systems Handbook, pages 299–331. Springer US, January 2011.

[Bank et Cole, 2008]. Jacob Bank and Benjamin Cole, Calculating the jaccard similarity coecient with ma produce for entity pairs in wikipedia, Wikipedia Similarity Team, 2008.

[Bank, M., Franke, J, 2010]. Social networks as data source for recommendation systems. Will Aalst, John My lopoulos, Norman M. Sadeh, Michael J. Shaw, Clemens Szyperski, Francesco Buccafurri et Giovanni Semeraro, editeurs : E-Commerce and Web Technologies, volume 61 de Lecture Notes in Business Information Processing, Springer , 49-60.

[Beliakov, G., Calvo, T., James, S, 2011]. Aggregation of preferences in recommender systems. In Recommender systems handbook (pp. 705, 734). Springer, Boston, MA.

[Berners-Lee 2000] : Berners-Lee, “Weaving the Web”. San Francisco, CA : Harper San Francisco.

[Breese et al, 1998]. John S. Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithm for collaborative ltering. In Proceedings of the 14 th Conference on Uncertainty in Articial Intelligence, page 43–52, 1998.

[Burke, 2002]. Hybrid recommender systems : Survey and experiments. User Modeling and User-Adapted Interaction, 12(4) :331–370.

[Carmagnola, F., Vernero, F., Grillo, P, 2009]. Sonars : A social networks based algorithm for social recommender systems. Geert-Jan Houben, Gord Mc Calla, Fabio Pianesi et Massimo Zancanaro, editeurs : User Modeling, Adaptation, and Persona-lization, volume 5535 de Lecture Notes in Computer Science, Springer Berlin/ Heidelberg, 10.1007/978-3-642-02247-0-22, 223-234.

[castragons.s , 2008]. « Modélisation de comportements et apprentissage stochastique non supervise de stratèges d'interaction sociales ou sein de système temps réel de recherche et d'accès l'information ». Thèse de doctorant de l'université Nancy .novembre 2008.

[Charif alchikh haydar, 2014]. Les systèmes de recommandation ‘a base de confiance, l'Université de Lorraine, 2014, France.

[Christophe Thovex. ,2012] :Christophe Thovex. Réseaux de Compétences : de l'analyse des Réseaux Sociaux à l'analyse Prédictive de Connaissances. Artificiel Intelligence.Université de Nantes, FRANCE, 2012.

[Das et al , 2007]. Das, A. S., Datar, M., Garg, A., and Rajaram, S. (2007). Google news personalization : Scalable online collaborative ltering. In Proc. Intl. World Wide Web Conference (WWW), pages 271–280.

[Diederich et Iofciu,2006]. Diederich, J. and Iofciu, T. (2006). Finding communities of practice from user profiles based on folk sonomies. In Proceedings of the EC-TEL06 Workshops, Crete, Greece , October 1-2, 2006.

[Ding Y., 2011]. « Scientific collaboration and endorsement : Network analysis of coauthor ship and citation networks », Journal of informatics, 5, 1, p. 187 203.

[FABRIZIO SEBASTIANI,2002] : FABRIZIO SEBASTIANI, «Machine Learning in AutomatedText Catégorisation», Conseil recherché National, Italie, Mars 2002

[Gabrielsson, S., Gabrielsson, S, 2006]. The use of Self-Organizing Maps in Recommender Systems, A survey of the Recommender Systems field and a presentation of a State of the Art Highly Interactive Visual Movie Recommender System. Mémoire de master, Uppsala Université.

[Goldberg et al., 1992]. Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. (1992). Using collaborative filtering to weave an information tapestry. Communications of the ACM, 35(12) :61–70.

[Herlocker, J. L., Konstan, J. A., Borchers, A., Riedl, J. 1999, August]. An algorithmic framework for performing collaborative filtering. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval (pp. 230-237).

[Herlocker, J. L., Konstan, J. A., Riedl, J, 2002]. An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. Information retrieval, 5(4), (pp. 287-310).

[Hill et al., 1995] : Hill, W., Stead, L., Rosenstein, M., and Furnas, G. (1995). Recommending and evaluating choices in a virtualcommunity of use. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 194–201. ACM Press/Addison-Wesley Publishing Co.

[Idir Benouaret, 2017] : Un système de recommandation contextuel et composite pour la visite personnalisée de sites culturels. Autre [cs.OH]. Université de Technologie de Compiègne, 2017. Français.

[Kaplan A.M., HAENLEIN M., 2010] : « Users of the world, unite ! The challenges and opportunities of Social Media », Business Horizons, 53, 1, p. 59 68.

[Karaouzere Meryem ,2015] : Système de recommandation des services web sémantiques,Université Abou BakrBelkaid ,2015.Tlemcen

[Karima ABIDI,2011] : Karima ABIDI, « La catégorisation de texte Multilingue », Mémoire de magistère, Ecole supérieur d'Informatique, Algérie, 2010-2011.

[L. Getoor C. P. Diehl, 2005]. "Link mining : a survey," ACM SIGKDD Explorations Newslette, pp. 3-12, 2005.

[LAHLOU OUCHIHA,2016] : CLASSIFICATION SUPERVISÉE DE DOCUMENTS ÉTUDE COMPARATIVE, UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS, Canada ,JANVIER 2016

[Laurent Candillier, 2001] : Apprentissage automatique de profils de lecteurs, Equipe GRAPPA, Lille 3 Laboratoire d'informatique Fondamentale de Lille,Juin 2001, France.

[Le Tran, D. K, 2011]. Conception et développement de fonctionnalités innovantes liées à Facebook pour un système de recommandation. Rapport bibliographique Dept. Logique des Usages, Sciences Sociales et de l'information Telecom Bretagne.

[Linden et al., 2003] : Linden, G., Smith, B., and York, J. (2003). Industry report : Amazon.com recommendations : Item-to-item collaborative ltering. IEEE Distributed Systems Online, 4(1).

[Liu, Z.-K. Z., Zhang, Y.-C., Zhou, 2010]. Solving the cold-start problem in recommender systems with social tags.

[Lü et al, 2012] : LinyuanLü, MatúšMedo, Chi Ho Yeung, Yi-Cheng Zhang, Zi-Ke Zhang, and Tao Zhou. Recommender systems. Physics Reports, 519(1) :1–49, 2012.

[Mataalah Hocine,2011] :mataalah Hocine, « classification automatique de textes Orienté Agent » faculté des sciences Algérie, 2010-2011

[Margaritis et Vozalis, 2003] : Margaritis, K. and Vozalis, E. (2003). Analysis of recommender systems' algorithms. In 6th Hellenic European Conference on Computer Mathematicsits Applications (HERCMA), Athens, Greece.

[Orkhan ,lafarovet all .2012] : Orkhan ,lafarov et SubhanGasimov, Machine Learning. rapport de projet dans le cadre dun master 2 informatique, Université de Franche-Comté,FRANCE. 2012

[O'Donovan, J.,Smyth, B, 2005]. Trust in recommender systems. Proceedings of the 10th

international conference on Intelligent user interfaces, IUI '05, New York,NY, USA, ACM, 167-174.

[Pazzani, M. J, 1999]. A framework for collaborative, content-based and demographic ltering. Artif. Intell. Rev., 13, 393-408.

[Rahila Abdelkader, 2015] :La recommandation dans les Réseaux Sociaux avec l'utilisation des Données Liées, Université Abou Bakr Belkaid, 2015.Tlemcen

[Radwan JALAM,2003] : Radwan JALAM, « Apprentissage automatique et catégorisation de textes multilingues », Thèse de doctorat, Université Lumière Lyon 2, France, Juin 2003.

[Resnick et al., 1994] : Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). Grouplens : an open architecture for collaborative filtering of netnews. In Proceedings of the 1994 ACM conference on Computer supported cooperative work, pages 175–186.

[Rich, 1979] : Rich, E. (1979) "User modeling via stereotypes". Cognitive science, 3(4) :329–354.

[Ricci et al., 2011] : Ricci, F., Rokach, L., Shapira, B., and Kantor, P. B., editors (2011). Recommender Systems Handbook. Springer.

[Sarwar, B., Karypis, G., Konstan, J., Riedl, J, 2001] : Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conferenceon World Wide Web, April, April (pp. 285-295).

[Shardanand et Maes, 1995] : Shardanand, U. and Maes, P. (1995a). Social information filtering : algorithms for automating ?word of mouth ? In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 210– 217. ACM Press/Addison-Wesley Publishing Co.

[Savadogo, 2018] : savadogo mahama, KABORE AbdulRazzaq et KABRE Bassirou," Conception et réalisation d'un système de réseautage social mobile sous Android", mémoire licence,université de djilali bounaama de khmis milaina, Algérie, 2018.

[Simon JAILLET.et al.2005] : Simon JAILLET, Maguelonne TEISSEIRE, Jacques CHAUCHE, Violaine PRINCE, « Classification automatique de documents, Le coefficient des deux écarts », Université Montpellier2, France, 2005.

[Simon RÉHEL.2005] : Simon RÉHEL, « Catégorisation automatique de textes et Concurrence de mots provenant de documents non étiquetés », Mémoire doctorat , Université Laval Québec, Canada, Janvier 2005.

[Sirinya, 2017]. Sirinya ON-AT "Temporalité et réseaux sociaux : prise en compte de l'évolution dans la construction du profil utilisateur", thèse doctorant,l'université de Toulouse,29/05/2017, France.

[S. Wasserman K. Faust, 1999]. "Social Network Analysis : Methods and Applications" Cambridge university presse, vol. 01,1999.

[Tim O'Reilly 2005] : O'Reilly Network ; What Is Web 2.0 : Design Patterns and Business Models for the Next Generation of Software. [Online]. http ://www.oreillynet.com/lpt/a/6228.

[Vozalis, M., Margaritis, K. G, 2006]. On the enhancement of collaborative ltering by demographic data. Web Intelli. and Agent Sys., 4, 117-138.
[WASSERMAN S., FAUST K., 1994] : Social network analysis : methods and applications, Cambridge ; New York, Cambridge University Press.
[PHP-ml, 2017].php-ai/php-ml quality Award Winner yegir256.com 2017.