**République Algérienne Démocratique et Populaire Ministère de l'Enseignement**

**Supérieur et de la Recherche Scientifique**

**UNIVERSITE SAAD DAHLEB DE BLIDA**

**Faculté des Sciences**

**Département d'Informatique**

**MEMOIRE DE FIN D'ETUDES**

**Pour l'obtention**

**D'un Diplôme de Master en**

**Informatique Option :ingénierie logicielle**

**et systèmeD'informatique et réseaux**

**THEME**

# *La prédiction des intérêts des utilisateurs Dans les réseaux sociaux*

**ORGANISME D'ACCUEIL : CERIST**

**Réalisé par:**

**BEN HANIA Asma**

**TIRAOUI Sarra Khawla**

**Président : Mme OUKID Saliha**

**Examinateur : Mme CHERGUENE**

**Promotrice : Mme BENBLIDIA Nadjia**

**Encadreur : Mme BOULKRINAT Nour El Houda**

***Promotion2020/2021***

# *Résumé*

Les réseaux sociaux fournissent un environnement d'échange et reposent principalement sur les utilisateurs dont le rôle est de créer, d'annoter le contenu des ressources et de construire des relations avec d'autres utilisateurs. Cependant l'augmentation rapide du nombre d'utilisateurs, partageant les mêmes domaines d'intérêt, dans ces réseaux nécessite la prédiction de profils en vue de répondre à la multiplicité et même à la divergence de leurs besoins en information.

L'objectif de notre travail consiste à proposer un système d'analyse prédictive des intérêts des utilisateurs non actifs on exploitant les techniques de prédiction les plus utilisés et les plus efficaces, tel que les arbres de décision, la régression logistique, k-plus proches voisins ou les réseaux de neurones. A cet effet, nous avons réalisé un système de prédiction basé sur le Deep learning.

**Mots-clés :** Réseau social, Profil social, utilisateur non actif, Analyse prédictive, Deep Learning, prédiction des centres d'intérêts.

# *Abstract*

Social networks provide an environment for exchange and rely primarily on users whose role is to create, annotate the content of resources and build relationships with other users. However, the rapid increase in the number of users, sharing the same areas of interest, in these networks requires the prediction of profiles in order to meet the multiplicity and even the divergence of their information needs.

The objective of our work is to propose a system of predictive analysis of the interests of non-active users by exploiting the most used and efficient prediction techniques, such as decision trees, logistic regression, k-closest neighbors or neural networks. To this end, we have produced a prediction system based on deep learning.



**1. Keywords:** Social Network, Social Profile, Inactive User, Predictive Analytics, Deep Learning, Interests Prediction.



# *الملخص*

توفر الشبكات الاجتماعية بيئة للتبادل وتعتمد بشكل أساسي على المستخدمين الذين يتمثل دورهم في إنشاء محتوى الموارد والتعليق عليه وبناء علاقات مع مستخدمين آخرين. ومع ذلك، فإن الزيادة السريعة في عدد المستخدمين، الذين يتشاركون نفس مجالات الاهتمام، في هذه الشبكات تتطلب التنبؤ بالملفات الشخصية لتلبية التعددية وحتى التباين في احتياجاتهم من المعلومات.

الهدف من عملنا هو اقتراح نظام للتحليل التنبئي لمصالح المستخدمين غير النشطين من خلال استغلال تقنيات التنبؤ الأكثر استخدامًا وكفاءة ، مثل أشجار القرار ، أو الانحدار اللوجستي أو أقرب الجيران أو الشبكات العصبية. تحقيقا لهذه الغاية قمنا بإنتاج نظام تنبؤ قائم على التعلم العميق.



**1. الكلمات المفتاحية**

لشبكة الاجتماعية ، الملف الشخصي الاجتماعي ، المستخدم غير النشط ، التحليلات التنبؤية ، التعلم العميق ، توقع الاهتمامات .

# *REMERCIEMENT*

*En préambule à ce mémoire, nous remercions ALLAH qui nous a aidé et donné la patience et le courage durant cette longue année d'étude.*

*Nous souhaitons adresser nos remerciements les plus sincères aux personnes qui nous ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire ainsi qu'à la réussite de cette formidable année universitaire.*

*Ces remerciements vont tout d'abord à notre encadreuse Mme Nour El HoudaBoulekrinat Pour sa disponibilité tout en long de la réalisation de ce Mémoire, Ainsi pour son inspiration, aide et son suivi.*

*Nous remercions très chaleureusement aussi, Mme BENBLIDIA Nadjia, Notre promotrice, pour sa confiance et ses encouragements.*

*Nos remerciements iront également vers tous ceux qui ont accepté avec bienveillance de participer au jury de ce mémoire.*

*On n'oublie pas nos parents pour leur contribution, leur soutien et leur patience.*

*Enfin, nous adressons nos plus sincères remerciements à tous nos proches et amis, qui nous ont toujours encouragées au cours de la réalisation de ce mémoire.*

*Merci à tous et à toutes*

# *Dédicaces*

*Je dédie ce modeste travail*

*À mes parents qui depuis mon plus jeune âge ont toujours fait leur maximum, en consacrant leurs temps et argent, pour m'éveiller et m'encourager dans mes passions.*

*C'est grâce à vous et pour vous que J'ai fait mon mémoire.*

*Aucun mot sur cette page ne saurait exprimer*

*Ce que je vous dois, ni combien je vous aime. Qu'Allah vous bénisse,*

*Vous assiste, vous vienne en aide*

*A mon binôme khawla*

*A mon cher frère, sœurs*

*A mon encadreur Mme Nour El HoudaBoulkrinatA*

*ma promotrice MmeBENBLIDIA Nadjia*

*A tous mes collègues, et toute la section Master 2 IL on témoignage de mon amitié sincère;*

*À ma grand-mère, la personne la plus importante de ma vie qui nous a quittées récemment et le véritable supporter de moi dans toute ma carrière, je vous dédie ce mémoire*

*En témoignage de mon amitié sincère;*

*A tous ceux qui m'ont soutenu, qu'ils trouvent ici l'expression de ma profonde*

***ASMA***

# *Dédicaces*

*Je dédie ce travail à :*

*A mes chers parents,*

*Sans qui je ne serais jamais arrivé jusqu'ici*

*Particulièrement à ma mère qui m'a encouragé et aidée pour achever mes études*
*Tout en espérant voir le fruit de ses sacrifices,*
*Surtout à mon père qui nous a*

*Quittés avant de me pouvoir voir le fuit de mon travail*
*Qu'Allah l'accueille en son vasteparadis*
*A mes frères, sœurs qui ma soutenue pendant toute la durée de ce*
*travail Qu'Allah les garde pour moi sains etsaufs*
*Pour leur contribution et leur encouragement*

*A mon encadreur MmeNour El HoudaBoulkrinatA*
*ma promotrice MmeBENBLIDIA Nadjia*

*A mes chères amies: Imad, Djoumana,Soumia*

*Ma source d'encouragement quotidienne, pour leur soutien moral, et avec qui je partage mon bonheur*

*A mon Binôme Asma que je lui souhaite toute la réussite dans sa vie professionnelle*

*A mes proches du cœur, qui ont été toujours là pour moi, avec beaucoup d'encouragement*

***KHAWLA***

# 1. Table des matières

# Liste des figures

# Liste des tables



# Glossaire

| Abréviation | Signification |
|---|---|
| KNN | K-plus proche voisin(K-nearest neighbors) |
| LSTM | Long Short-Term Memory(longue mémoire à court terme) |
| CSV | Comma-separated values |
| MVC | Modèle vue contrôleur (Model View Controller) |

# *Introduction Générale*

**Contexte**

Les réseaux sociaux fournissent un environnement d'échange et reposent principalement sur les utilisateurs dont le rôle est de créer, d'annoter le contenu des ressources et de construire des relations avec d'autres utilisateurs. Les intérêts d'un utilisateur correspondent à une des caractéristiques les plus fréquentes dans un profil utilisateur surtout dans les systèmes de recommandation. Les intérêts de l'utilisateur peuvent être classés en deux types : les intérêts à court terme et à long terme **[Ramiandrisoa et al, 2017]**. Le profil est à long terme si l'utilisateur est toujours intéressé par le sujet et que cet intérêt ne change que très rarement. Il est à court terme si l'utilisateur est intéressé par le sujet durant une période limitée, que cet intérêt soit éphémère ou change plus fréquemment Il est très important de trouver la méthode adéquate pour extraire les intérêts de l'utilisateur à partir des réseaux sociaux dans lesquels beaucoup de données peuvent être utilisées telles que le commentaire, la notation, le partage, etc. La base de l'extraction des intérêts est d'identifier, à partir des données des utilisateurs (cela peut être les messages, le nombre de clics, le nombre de visites, etc.). Prédire les intérêts futurs des utilisateurs dans leurs réseau social est très utile pour les systèmes de recommandation afin de leur recommander de façons fiable et efficace: des amis, groupes, pages, ...etc.

**Problématique**

Il existe une panoplie d'utilisateurs dans les réseaux sociaux tel que : lurkers, socialistes, actif, non actif. Ce dernier n'interagit pas (ou rarement) avec les réseaux sociaux et ne possède aucun ami (ou très peu). Les facteurs de son inactivité peuvent être des causes de santé, familiales, faux profils, ou des comptes malveillants (piratage). Cependant, Nous avons contacté à travers les travaux existants pour la prédiction des réseaux sociaux que les techniques de prédiction sont peu exploitées pour l'analyse sociale, aussi elles ne sont pas utilisées pour la prédiction des intérêts des utilisateurs non actifs.

Donc le problème qui se pose : quelle est la meilleure technique de prédiction a adapté pour prédire les intérêts des utilisateurs non actifs ? et comment prédire ses intérêts à partir seulement de son historique de recherche ?

**Objectifs**

Nous nous intéressons dans notre travail à la prédiction des intérêts des utilisateurs inactifs. Dans ce contexte, nous avons proposé une solution basée sur les requêtes de l'utilisateur (historique de recherches).

Nous avons dans un lieu appliqué quatre techniques de prédiction à savoir : l'arbre de décisions, k-plus proches voisins, régression logistique et le réseau de neurones, après une étude comparative notre choix s'est porté sur le Deep Learning. A cet effet, nous avons conçu et implémenté d'une application pour prédire des intérêts des utilisateurs inactifs en utilisant son historique de recherches et le Deep Learning.

**Organisation du mémoire**

Nous structurons ce présent mémoire en quatre chapitres afin d'aborder les principales étapes du projet. Il est organisé comme suit :

**Chapitre 1**, présente les réseaux sociaux, les éléments qu'il les constitue et s'appuie sur la notion du profil social.

**Chapitre 2**, présente les techniques de prédiction et les travaux existant sur la prédiction dans les réseaux sociaux.

**Chapitre 3**, concerne la conception de notre système de prédiction des intérêts utilisateur.

**Chapitre 4**, l'implémentation et représentation de notre application.

Enfin, ce mémoire se termine avec une conclusion générale.

## 1. Introduction

Le développement récent des ordinateurs et des réseaux informatique qui relie un tirée grand nombre de système, et avec l'arrivé du web 2.0, a amené l'évolution des médias sociaux qui sont devenus un véritable phénomène depuis quelques années sur Internet, ils se sont développé pour toucher à travers le monde des millions d'internautes. Cette révolution technologique, permet et facilite les interactions et les connexions entre les internautes (réseaux sociaux, blogs, forums).À l'heure actuelle les médias sociaux sont devenus un des usages les plus intéressants du web, ils sont utilisés pour définir tout site web ou application qui crée des discussions virtuelles dans un esprit collaboratif. Un média est social s'il permet aux internautes d'interagir les uns avec les autres et de partager des contenus, des informations et des idées.

Les réseaux sociaux aujourd'hui constituent un sous-ensemble des médias sociaux **[Priscille et al. 2017]**, ils correspondent aux applications dont l'objectif premier est la réelle mise en relation et ils sont de plus en plus utilisés et surtout par un public très large. Ces sites sont maintenant largement répandus, afin de toucher toutesles catégories de la population, ce qui les ronds un outil de communication et de liberté par excellence, car il nous permet d'accéder à une base mondiale d'information et de communiquer virtuellement de diverses manières : messagerie électronique, messagerie instantanée, salon ou groupe de discussion, ou à travers les blogs….

## 2. Définition des réseaux sociaux

Un réseau social est un espace virtuel où les gens de même affinités peuvent se rencontrer et interagir, il permet évidement d'échanger entre membre et de partager des informations. Plusieurs définitions du réseau social existent dans littérateur, nous citons les exemples ci-dessus :

**[LENDREVIE et al,2006]**définissent les réseaux sociaux comme étant: «des applications internet, généralement sous forme de site web, qui permettent de relier amis, associés ou visiteurs, d'échanger messages et documents, de participer à des communautés en ligne plus ou moins informelles ».

Les réseaux sociaux sont : « des services Web qui permettent aux individus: 1- de construire un profil public ou semi-public au sein d'un système, - 2 de gérer une liste des utilisateurs avec lesquels ils partagent un lien, -3 de voir et naviguer sur leur liste de liens et sur ceux établis par les autres au sein du système »**[Boyd et al, 2007]**.

« Le réseau social se définit comme une plateforme permettant de créer son profil pour construire des relations avec d'autres membres, y former des groupes d'intérêts communs et échanger. Il rend possible un dialogue ou une conversation, dans un cadre certes contrôlé et organisé, mais débarrassé des contraintes physiques de la proximité et de la synchroniser »**[BOURSIN et al, 2011]**.

## 3. Historique des réseaux sociaux

Dans le web 2.0, les applications sont plus accessibles et disposent des interfaces interactives permettant aux utilisateurs de produire, modifier et partager des informations qui sont par la suite enrichies par d'autres utilisateurs **[O'Reilly, 2005].**


**Figure 1:** Développent des réseaux sociaux **[Webself, 2021]**

Lancements de « friendster » en 2002 comme le premier réseau social aux sens actuel du mot. C'était un site destiné aux jeux qui utilisait pour la première fois la notion de cercles et réseaux d'amis. Ensuite, d'autres plateformes ont été mises en service ,on peut citer « MySpace » lancé en 2003 qui dépassa « friendster » en avril 2004 en termes de nombre de pages affichées. Les réseaux sociaux professionnels tels que « LinkedIn » et « XING » ont vu le jour en 2003. En 2004, « facebook » a été lancé pour servir à la communication entres les étudiants de l'université de Harvard mais son succès fait de lui le réseau social le plus populaire jusqu'à nos jours. Plusieurs réseaux de partage de contenu sont ensuite apparus : « Flickr » en 2003, « Youtube » en 2005 et « SlideShare »en 2006. Google a essayé de joindre

ce monde en offrant 30 millions dollars pour acheter « friendster » et on achetant « youtube » en 2006 et en lançant son propre réseaux « Google + » en juin 2011. L'invention des smartphones et le développement des applications mobile a permet aussi de créer des réseaux sociaux pour cette environnement. Nous pouvons citer: « WhatsApp » en 2009, « Instagram » en 2010.

## 4. Types de réseaux sociaux

Ces dernières années, les réseaux sociaux ont envahi le web, aujourd'hui il existe un grand nombre de réseaux sociaux dans le monde et d'énormément propositions sur ce sujet. Les réseaux sociaux peuvent être divisés en plusieurs catégories, selon les différents avis, plusieurs classifications sont proposées, on peut citer les suivantes qui sont basées sur le fait que les réseaux sociaux fournissent des outils qui facilitent le processus de mise en relation au tour d'un centre d'intérêt commun **[Torloting, 2006].**

### 4.1 Les réseaux sociaux plate-forme de partage

Ce type de réseau permette de diffuser du contenu qui est souvent multimédias (vidéos, sons,..) aux internautes, la mise en ligne et le partage de ces données multimédias, deviennent plus facile grâce à ce genre de réseaux qui sont aujourd'hui accessibles pour tous (ex: Youtube[1], Dailymontion[2])

### 4.2 Les réseaux personnels et généralistes

Souvent orientés autour d'un centre d'intérêt (musique, lecture, etc.), le but de ce type de réseaux n'est autre que de faire partager ses passions au reste de la communauté. Les mises en relation directes sont rares sur ce type de réseaux. Exemples: MySpace[3], Skyblog[4], Friendster[5], etc.

### 4.3 Les réseaux personnels et thématiques

Ils fonctionnent généralement de la même façon que les réseaux généralistes, sauf que ceux-ci sont orientés sur une thématique donnée comme les voitures, la technologie, la cuisine,…etc.

[1] www.youtube.com
[2] https://developer.dailymotion.com/
[3] https://myspace.com/
[4] https://www.skyrock.com/
[5] https://www.flickr.com/

### 4.4 Les réseaux professionnels

Les réseaux professionnels sont ceux les plus aboutis. Ils donnent la possibilité de mise en relation entre utilisateurs ainsi que le partage d'informations (informations sur l'entreprise, coordonnées, CV). Exemples: Via deo, LinkedIn…

## 5. Les caractéristiques des réseaux sociaux

Certains réseaux sociaux proposent également des composantes principales et des fonctionnalités pour faciliter l'interaction entre les utilisateurs, tels que les groupes, chat room et la messagerie instantanée. Nous définissons ci-dessous quelques concepts :

- Profil

Les profils peuvent être la base principale des réseaux sociaux, ils contiennent des informations démographiques de l'utilisateur tel que son nom, son sexe, sa ville natale et son emplacement actuel…etc. En plus de ça la plupart des réseaux sociaux encouragent l'utilisateur d'écrire une courte biographie sur eux-mêmes et de partager leurs goûts et leurs intérêts.

- Les amis (individus)

L'ami est un concept fondamental très important et le plus utilisé dans le monde virtuel. Un « Ami » peut être un ami, un membre de la famille, une connaissance, un ami d'un ami, ou même quelqu'un que l'utilisateur n'a jamais rencontré auparavant, sauf dans les réseaux sociaux. Les réseaux sociaux ont généralement une fonctionnalité de recherche qui peut aider l'utilisateur à trouver de nouveaux amis. Par exemple, les utilisateurs peuvent rechercher des amis partageant les mêmes centres d'intérêts, qui appartiennent à un certain groupe d'âge, ou qui vivent dans la même région.

- Les groupes (communauté)

Plusieurs utilisateurs regroupés autour d'une thématique ou d'un centre d'intérêt bien défini. Les membres ne sont pas forcément des amis mais ils peuvent publier librement dans l'espace réservé au groupe. Nous pouvons différencier une sous-catégorie de membres dans le groupe : les administrateurs ou les modérateurs, ils disposent de quelques privilèges par rapport aux autres membres. Il intervient pour modifier le contenu et le paramètre du groupe comme ils peuvent décider l'ajout ou la suppression des membres.

- Les évènements

C'est une fonctionnalité permettant aux « Amis » de savoir les événements à venir dans leur communauté ainsi que d'organiser des rassemblements sociaux.

- Les Tags (commentaire, annotation)

Selon Golder et Huberman**[Golder et al., 2005]**, le terme « tag » (en français « étiquette ») représente un mot-clé ou une expression associée ou assignée aux ressources. Il décrit ainsi l'objet et lui permet d'être retrouvé par navigation, par filtrage ou par recherche. Ces dernières années, ce processus a gagné beaucoup en popularité sur le Web.

- Flux d'actualité (News Feeds)

Les flux d'actualité sont des outils utiles pour rester en contact avec les « Amis ». Par exemple, les mises à jour de profil, les messages sur le blog, les photos et vidéos publiées.

## 6. Des exemples des réseaux sociaux

Il existe beaucoup de réseaux sociaux sur internet comme : Facebook, Flickr, LinkedIn, Viadeo, MySpace,…etc. Parmi les plus connait on trouve : Facebook, Twitter et Google Plus :

- Facebook[6] :

Facebook est un site où vous êtes susceptible de trouver des amis, des collègues et des parents. Bien que Facebook soit principalement axé sur le partage de photos, de liens et de la vie quotidienne.

- Twitter [7]:

Peut-être la plus simple de toutes les plateformes de médias sociaux, Il est devenu une des principales sources d'information en temps réel. Tous les événements sont visibles et commentés à la seconde, ce qui confère à cette plateforme autant de puissance que de risques. C'est une plateforme de micro-blogging, ça veut dire que vos posts sont limités en caractère – vous avez 160 symboles pour faire un message.

- Google+[8] :

Google a donné aux utilisateurs un réseau social qui a un petit quelque chose de tout le monde. Vous pouvez ajouter du nouveau contenu, mettre en surbrillance des sujets avec des has tags et même séparer des contacts en cercles.

## 7. Présentation d'un réseau social

La première personne à avoir représenté un réseau social est Jacob Levy Moreno au début des années 1930 **[MORENO, 2021].** Son objectif était de visualiser graphiquement un réseau

[6]www.facebook.com
[7]https://www.twitter.com/
[8] https://www.google.com/

social, en représentant les personnes par des points et une relation entre deux personnes par des flèches. Cette représentation est depuis désignée par le terme sociogramme. Les mathématiciens ont fait le rapprochement entre les représentations sociogrammes et la théorie des graphes au sens mathématique.

Le graphe est devenu par la suite la représentation adoptée par toutes les sciences manipulant l'analyse des réseaux sociaux, dont la sociologie, les mathématiques et l'informatique. Un réseau social est souvent représenté sous forme de graphe. Ce dernier est composé des nœuds (sommets) qui décrivent les personnes et les liens (arêtes) qui décrivent les connexions/relations sociales entre ces personnes. Un graphe sert à identifier les personnes selon différents critères à savoir: les personnes les plus reliées entre elles, les amis explicites d'un utilisateur ou même les personnes partageant des caractéristiques en communs. Dans ce contexte, il existe des travaux qui visent à analyser certains types de réseau)**[mezghani, 2015].**

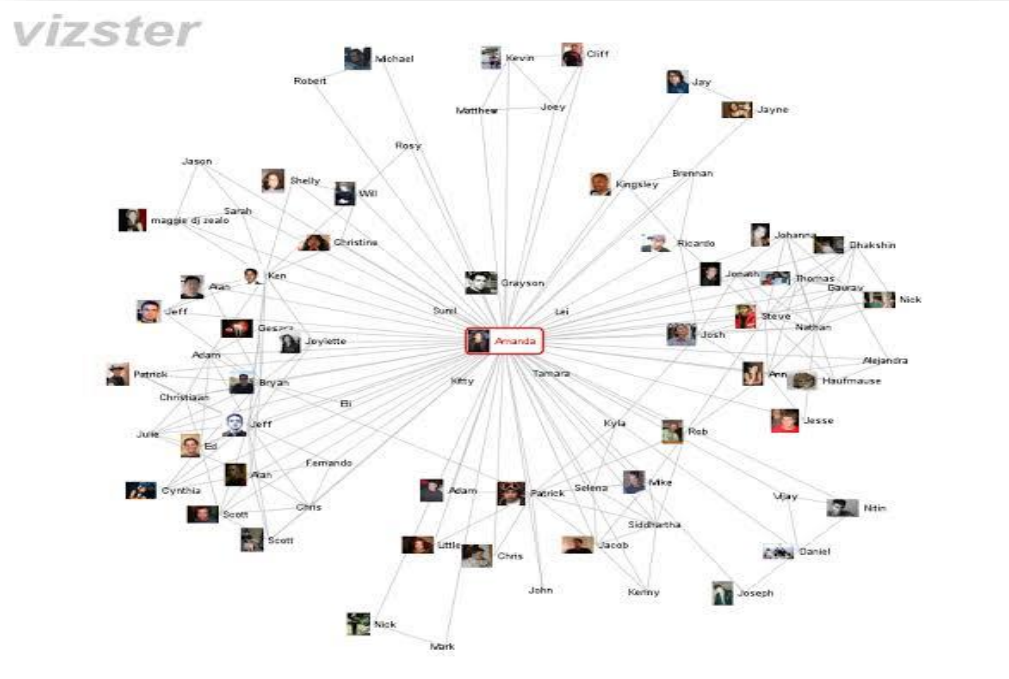


**Figure 2** : Représentation des réseaux sociaux **[Smith M, 2021]**

## 8. Les statistiques d'utilisation des réseaux sociaux

L'exploitation des réseaux sociaux est en augmentation depuis leurs apparitions jusqu'à maintenant, le nombre des utilisateurs s'accroît considérablement. Plusieurs pages dédiées aux statistiques des réseaux sociaux sont régulièrement mises à jour afin de montrer des chiffres réels des utilisateurs.

D’après les statistiques qui ont été faites, les résultats suivants ont été délivrés: avec plus de 2,7milliards d'utilisateurs actifs par mois, le réseau social de Mark Zuckerberg (Facebook) domine largement ce classement, Youtube totalise tout de même 2.2 millions d'utilisateurs actifs, arrivant ainsi à la deuxième place. C'est le WhatsApp qui s'assure la troisième place sur le podium avec 2.0 millions d'utilisateurs.

Ce graphique (**Figure 3**) montre le nombre d'utilisateurs actifs des réseaux sociaux en 2021 en millions **[Statista, 2021].**

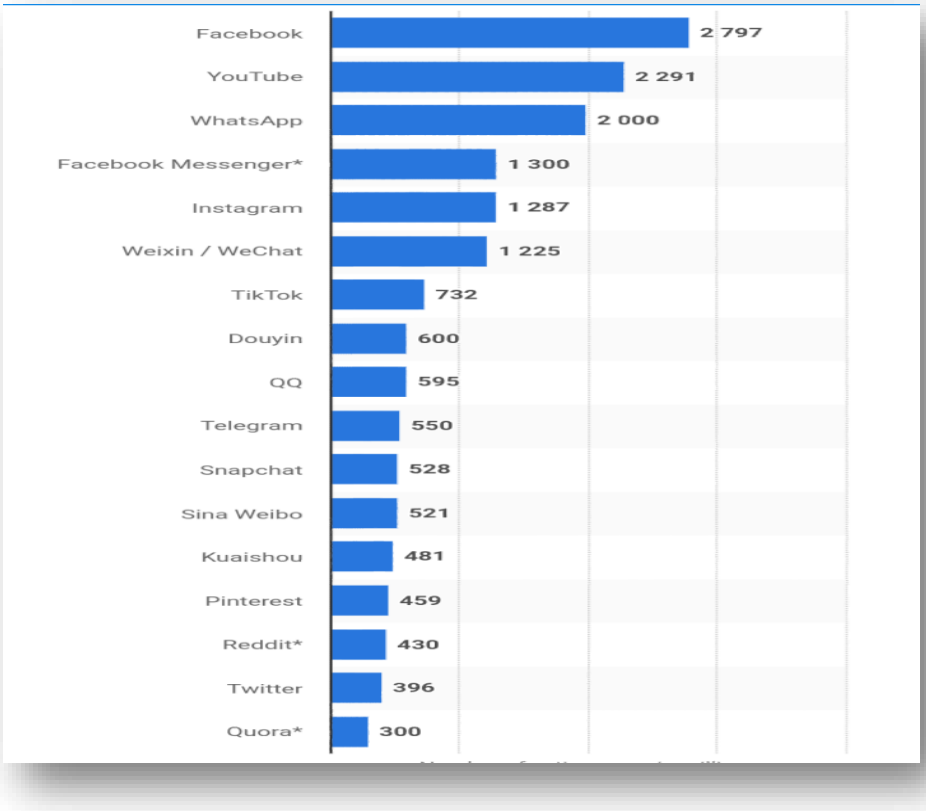


**Figure 3:** Représentation de la statistique des réseaux sociaux les plus utilisés a 2021

## 9. Les types des utilisateurs dans réseaux sociale

**[Brandtzaeg, 2011]** classe les utilisateurs comme suit :

- **Lurkers (27 %)**:les Lurkers constituent la plus grande catégorie d'utilisateurs. Ils sont nommés "lurkers" car ils sont assez faible participation et ils participent à des activités qui sont davantage liées aux loisirs. Ces les utilisateurs sont quelque peu impliqués dans plusieurs activités, mais seulement passivement ou dans une faible mesure.
- **Socialistes (25 %)**:ces utilisateurs sont les prochains plus grands utilisateurs et sont étiquetés socialises. Ils se classent haut sur écrire des lettres et des messages", "contacter les autres" et "chercher un nouvel ami".

- **Débatteurs (11 %):**les débiteurs sont aussi élevés que les socialistes en termes de niveau de participation, caractérisé par fortement impliqué dans les discussions, la lecture et la rédaction de contributions en général.
- **Actifs (18 %):**"Actives" sont ainsi étiquetées parce que ces utilisateurs sont engagés dans presque toutes sortes les activités de participation au sein de la communauté, ce qui inclut le fait d'être membre pour "publier et partager des images" la majorité de ce groupe sont des jeunes filles **[Brandtzaeg, 2011]**.

## 10.Définitions profil social

Il existe plusieurs définitions dans la littérature du profilage social (profil social), parmi ces définitions nous trouvons :

- Profil social désigne un profil dans lequel les informations sont extraites à partir des voisins sociaux de l'utilisateur **[Tchuente, 2013]**.
- Le profilage social désigne le fait d'extraire ou d'enrichir des informations sur les individus ou groupes d'individus en utilisant leur environnement social comme source d'information **[Sirinya, 2017]**.

## 11.La construction du profil social

La construction du profil reflète un processus qui permet d'instancier sa représentation à partir de diverses sources d'information **[Tamine, 2005**]. La construction du profil implique la collecte et l'utilisation de sources de données et d'informations pertinentes pour les représenter. La collecte de ces sources d'information comprend la spécification des types de données pertinentes à collecter, et les modes explicites ou implicites d'acquisition des données **[Daoud, 2009].**Ces deux méthodes sont décrites en détail dans les deux sous sections suivantes **[Kechid, 2009]** :

- Acquisition des données explicites

La construction explicite est basée sur une collecte d'information directement fournis par l'utilisateur via l'interface du système. Les informations exploitées pour la construction explicite sont généralement :

- Jugement explicite sur la pertinence des termes, documents.
- Définition domaine d'intérêts.
- Acquisition de données implicites

La construction implicite largement motivées par les travaux actuels dans le domaine, repose sur un procédé d'interface du contexte et préférence de l'utilisateur via son comportement lors

de l'utilisation du système ou d'autre application quotidiennes. Les informations exploitées pour la construction implicite sont généralement :

- Dernières pages visités.
- Durée de lecture des documents.

## 12. Représentation de profil social

Le modèle du profil constate a spécifier sous quelle forme les données du profil doivent être représentés, ils existent plusieurs méthodes pour représenter et structurer les profils **[Ramiandrisoa et al, 2017] :**

- Représentation sous forme de vecteurs

Il s'agit de l'une des représentations de profil la plus utilisée, en particulier pour sa simplicité. Habituellement, le vecteur profil correspond à un ensemble de caractéristiques avec pour chacune son poids (la valeur de la coordonnée pour une dimension). La façon de calculer le poids varie selon l'application.

- Représentation sous forme de concepts hiérarchiques

Par rapport à la représentation sous forme de vecteur, l'utilisation de concepts hiérarchiques permet au système de généraliser le profil. Par exemple, un utilisateur intéressé par les jeux olympiques, est probablement également intéressé par le sport. Le niveau de la hiérarchie des concepts peut être fixé **[Trajkova et al, 2004]** ou être dynamique **[Tchuente et al, 2013]**. Ici le concept peut remplacer les caractéristiques (intérêts, émotions, compétences, etc.).

- Représentation sous forme d'ontologies

L'ontologie permet d'avoir une représentation plus sémantique en associant des liens entre les termes ou les items du profil de l'utilisateur **[Trajkova et al, 2004], [Hernandez et al, 2007]**

- Représentation sous forme de graphes

Cette méthode est très utilisée pour avoir une analyse des relations entre les paires de caractéristiques (intérêts, émotions, etc.) où un nœud représente une instance de caractéristiques et un arc la relation entre deux nœuds . Cette représentation permet d'analyser des relations sémantiques **[Valafar et al, 2009]** et est proche de la représentation sous forme d'ontologies lorsque le graphe est un arbre. La plupart du temps, cette technique est utilisée pour analyser les relations entre les utilisateurs dans un réseau social où un nœud représente un utilisateur. Il y a d'autres types de représentation d'un réseau comme les réseaux Baesines, etc.

## 13.La prédiction des intérêts des utilisateurs dans les réseaux sociaux

La prédiction précise des intérêts futurs des utilisateurs sur les réseaux sociaux, en étudiant comment les utilisateurs réagiront si certains sujets émergeant à l'avenir.

**[Bao et al. 2013]** ont proposé un modèle probabiliste de factorisation matricielle pour prédire les intérêts futurs des utilisateurs dans les services de microblogging (délivrer des contenus courts). Ils supposent que l'ensemble de sujets de l'avenir est connu à priori et composé seulement de l'ensemble des sujets qui ont été observés dans le passé, ce qui semble être une hypothèse limite irréaliste, parce que les sujets d'intérêt des utilisateurs sur les réseaux sociaux, les sujets d'intérêt pour les utilisateurs sur les réseaux sociaux sont changées avec fonction des événements du monde réel **[Abel et al, 2011].** Par conséquent, une telle approche ne peut pas prédire les intérêts des utilisateurs en ce qui concerne les nouveaux sujets puisque ces sujets n'ont été jamais reçus.

A cet effet, dans les travaux de **[Zarrinkalam et al,2019]** ils représentent des méthodes pour prévoir les intérêts des utilisateurs en ce qui concerne les futurs sujets non observés, lorsque les intérêts et les sujets eux-mêmes sont autorisés à varier dans le temps. Cela permet d'effectuer le futur planifié en étudiant comment les utilisateurs réagiront si certains sujets émergent à l'avenir.

## 14.Conclusion

A travers ce premier chapitre, nous avons mis en évidence la notion de réseau social en donnant la définition, les types et la représentation de réseaux sociaux. Ensuite, nous sommes passés à la notion de profil, sa représentation, et la prédiction des intérêts du profil social. Dans le prochain chapitre nous étudions les techniques de prédiction.

# 1. Introduction

Quelque temps auparavant, la prédiction de l'avenir était quelque chose d'absurde et impossible à réaliser, cela était due au faite du maque de l'information nécessaire à la réalisation de cette dernière. Aujourd'hui, l'exploitation des immenses données collectés dans des domaines différents ont permis non seulement la résolution de problèmes complexes, ou encore découvrir de nouveaux savoirs, mais aussi la prédiction des intérêts des futurs utilisateurs. Et cela est arrivé grâce aux méthodes de prédiction utilisé dans le Data Mining.

# 2. Datamining

Le data mining est un processus de mise à jour de nouvelles corrélations, tendances et de modèles significatifs par un passage au crible des bases de données volumineuses, et par l'utilisation de modèles d'identification technique aussi bien statistiques que mathématiques **[René et al,2001].** Il existe plusieurs méthodes de Data Mining classifiées selon deux catégories : supervisée et non supervisée (**Figure 4**).

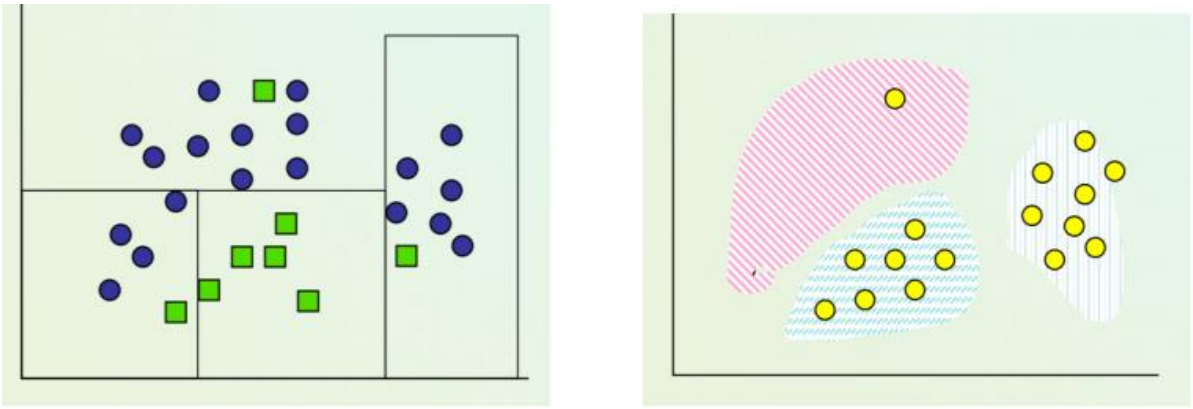

**Figure 4:** classification supervisée / classification non supervisée

## 2.1 Les Méthodes supervisées

Tel que les données sont constituées d'un ensemble de caractéristiques décrivant chaque individu et chaque individu possède une variable particulière **[Azé, 2003]**. Parmi ces méthodes nous trouvons la classification (permet de classer chaque élément à une classe prédéfinie on utilise un modèle obtenu par apprentissage), et l'estimation (elle porte sur des variables continues, on utilise l'estimation pour obtenir la valeur d'une variable).

## 2.2 Les Méthodes non supervisées

Dans ce type d'apprentissage la connaissance de la sortie désirée n'est pas nécessaire, c'est-à-dire que le réseau s'auto-organise et organise les entrées qui sont présentées comme vecteur d'entrée **[Tsiptsiset al, 2010].** Parmi ces méthodes nous trouvons le regroupement par similitude

(l'objectif est de déterminer quels objets vont naturellement ensemble).

### 3. Les techniques de prédiction

Il existe plusieurs techniques de data mining pour la prédiction. Parmi ces techniques, nous trouvons :

### 3.1Les arbres de décision

Les arbres de décisions sont des outils d'aide à la décision qui permettent selon des variables discriminantes de répartir une population d'individus en groupes homogènes en fonction d'un objectif connu. Les arbres de décision sont des outils puissants et populaires pour la classification et la prédiction. Un arbre de décision permet à partir des données connues sur le problème de donner des prédictions par réduction, niveau par niveau, du domaine des solutions.

Chaque nœud interne d'un arbre de décision permet de répartir les éléments à classifier de façon homogène entre ses différents fils en portant sur une variable discriminante de ces éléments (**Figure5**). Les branches qui représentent les liaisons entre un nœud et ses fils sont les valeurs discriminantes de la variable du nœud. Et enfin, les feuilles d'un arbre de décision représentent les résultats de la prédiction des données à classifier **[Calas, 2009]**.

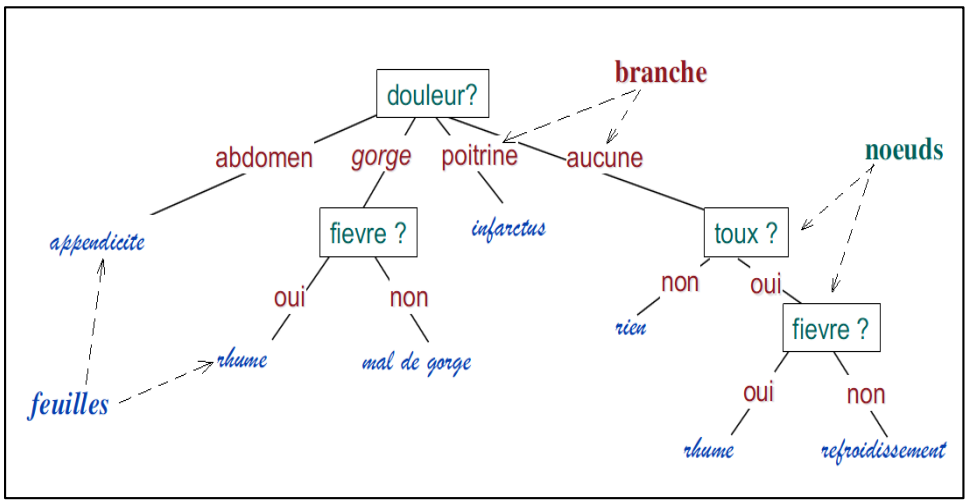


**Figure 5:** Exemple d'un arbre de décision

### 3.2Les K plus proches voisins

L'algorithme des k-plus proches voisins (k-ppv) ou k-nearestneighbors(K-NN) est une méthode d'apprentissage supervisé dédiée à la classification ,il est considéré comme l'algorithme d'apprentissage automatique le plus simple. Afin de prédire la catégorie d'un exemple donné, l'algorithme recherche les K voisins les plus proches de ce nouveau cas et prédit la réponse la

plus fréquente des K voisins les plus proches. Le principe de prise de décision est global: de cette façon, la distance entre l'échantillon inconnu et tous les échantillons fournis est calculée. Cet exemple est ensuite affecté à la classe majoritaire représentée dans les K échantillons. Cette méthode utilise deux paramètres: le nombre K et la fonction de similarité utilisée pour comparer le nouvel exemple avec la classification d'exemple existante **[Haykin, 1998].**

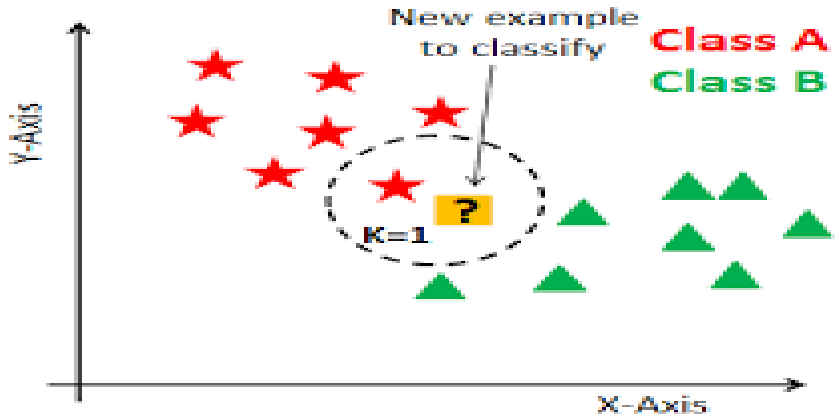


**Figure 6 :** Exemple Les K plus proches voisins

### 3.3 La Régression logistique

La régression logistique est une méthode d'analyse statistique qui consiste à prédire une valeur de données d'après les observations réelles d'un jeu de données. La régression logistique est devenue un outil important dans la discipline de l'apprentissage automatique. Cette approche permet d'utiliser un algorithme dans l'application d'apprentissage automatique pour classer les donnes entrantes en fonction des données historiques. Plus il y a de données pertinentes en entrée, plus l'algorithme est en mesure de prédire des classifications au sein des jeux de données **[Koudri, 2011]**.

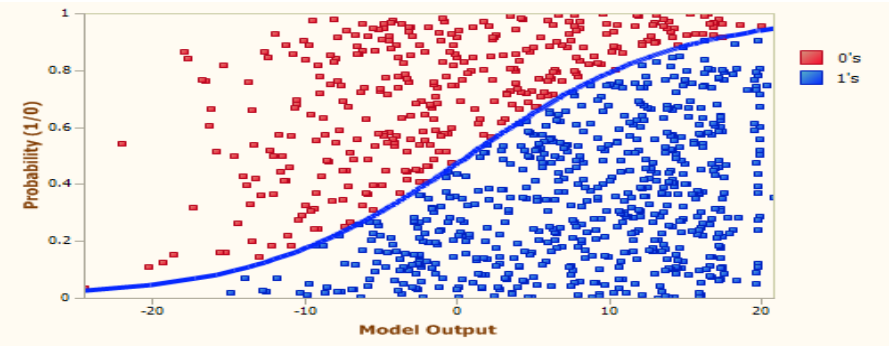


**Figure 7:**Exemple La régression logistique

### 3.4 Apprentissage profond (Deep learning)

L'apprentissage profond (en anglais: Deep learning) est apparu comme une nouvelle zone de recherche de l'apprentissage automatique. Au cours des dernières années, les techniques développées dans l'apprentissage profond ont déjà eu un impact sur les travaux de traitement des signaux et de l'information, compris les aspects de l'apprentissage automatique et l'intelligence artificielle **[Goodfellow, 2016]**.

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 réseaux de neurones**

Le réseau neuronal est un ensemble d'unités d'entrée/sortie connectées et chaque connexion à un poids présent avec elle au cours de la phase d'apprentissage, le réseau apprend en ajustant les poids de manière à pouvoir prédire la classe correcte étiquettes des tubes d'entrée. Les réseaux neurone sont la capacité remarquable de dériver du sens de complexes ou des données imprécises et peuvent être utilisés pour extraire des modèles et détecter des tendances trop complexes pour être remarquées par des humains ou d'autres techniques informatiques. Ils sont bien adaptés pour des apports et produits, par exemple la réorganisation des caractères manuscrits, pour former un ordinateur à la prononciation de l'anglais texte et de nombreux problèmes commerciaux réels et ont déjà été appliqués avec succès dans de nombreuses industries. Les réseaux neuronaux sont les meilleurs pour identifier les tendances ou les modèles de données et bien adaptés à la prédiction ou prévisions des besoins**[Bharati et al,2010]**.

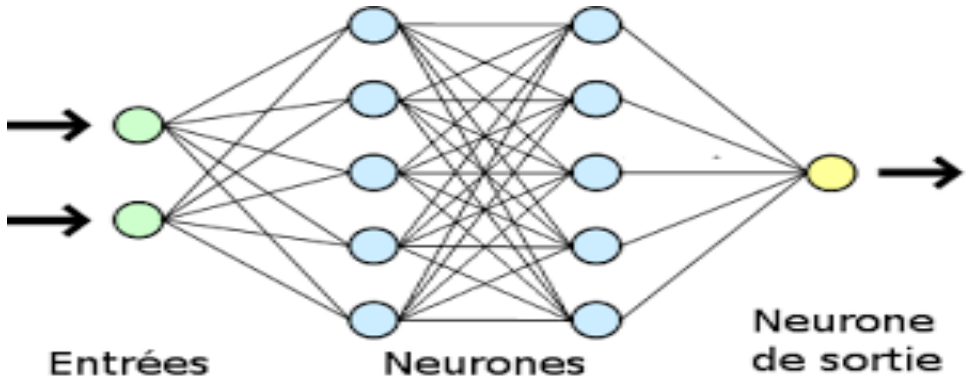


**Figure 8:** Exemple d'un réseau de neurones

## 4. Comparaison entre les techniques de prédiction

Pour comprendre le fonctionnement et la difficulté liée à l'emploi des quatre techniques précédemment présentées, nous allons présenter à travers le tableau suivant (**tableau 1**) les

avantages et les inconvénients de leurs utilisations.

| Techniques | Avantages | Inconvénients |
|---|---|---|
| **Les réseaux de neurones** | + Exprimer les résultats des données inconnues, capacité d'apprentissage automatique, cela permet de résoudre problème pas besoin d'écrire des règles compliqué.<br>+ Lisibilité Le résultat d'un réseau de cellules organisé selon la structure, classement du bruit<br>+ Efficace, peut être combinée avec autres méthodes. | - Ce sont de vraies boites noires qui ne permettent pas d'interpréter les modèles construits. En cas d'erreurs du système, il est quasiment impossible d'en déterminer la cause.<br>- Faut passer un grand nombre de fois tous les exemples de l'échantillon d'apprentissage avant de converger et donc le temps d'apprentissage peut être long. |
| **Les arbres de décision** | + Règles claires fournies Classement, faible indépendance avec des échantillons de données, facile comprendre et expliquer par l'utilisateur. Intuitif et disponible Représentation graphique, langage parlé et Facile à lire, accord Classement individuel. | - Impossible de détecter combinaisons de variables, et Besoin de fournir un Grand apprentissage.<br>- La performance est souvent Quand le nombre baisse Les cours deviennent trop importants. |
| **Les k plus proches voisins** | + Cette méthode résout une tâche non supervisée, aucune information n'est donc nécessaire Sur les données.<br>+ Technologie facile à mettre en œuvre.<br>+ applicable à tout type de données (même du texte), en sélectionnant une Bonne notion de distance | - La difficulté de trouver une bonne fonction de distance<br>- le choix du logarithme k est nécessaire, un mauvais choix de k produira de mauvais résultats résultat.<br>-Expliquer les difficultés de certains clusters. |

**Tableau 1:** Synthèse des quatre techniques de prédiction

## 5. Prédiction dans les réseaux sociaux

Plusieurs approches ont été proposées pour incorporer les techniques de prédiction l'analyse des réseaux sociaux. Dans ce qui suit, nous présenterons ces travaux de recherche et les résultats qui ont été obtenus.

### 5.1 Prédiction des clics dans le réseau social

**[Gharibshah et al, 2020]** Le but de cette recherche est la prédiction des intérêts des utilisateurs et la réponse aux clics des utilisateurs à l'aide de réseaux de neurones profonds basés sur LSTM. Les réseaux de mémoire à long court terme (LSTM) ont été introduits pour utiliser des cellules structurées à plusieurs portes spéciales pour remplacer les nœuds de la couche cachée, un moyen efficace de surmonter ces problèmes est composé d'une cellule chaque cellule LSTM comprend trois entrées de porte : porte d'entrée, porte d'oubli et portes de sortie, dans ce cas ils sont cités l'ensemble des catégories des pages

visitées comme des protes d'entrées, ils ont effectuait des comparaisons avec plusieurs méthodes. Les résultats de ce travail ont montré que la méthode LSTM est capable de coder des informations temporelles latentes utiles dans des séquences de requêtes pour prédire les réponses et l'intérêt des utilisateurs.

### 5.2 Prédiction des intérêts des utilisateurs dans le réseau social

**[Ben Sassi I et al., 2007]** Le but de cette étude est de développer une nouvelle approche pour la prédiction des intérêts des utilisateurs afin d'enrichir leurs requêtes et d'élargir leurs cercles sociaux. Cette approche s'appuie sur la technique de la classification dans le but de prédire les intérêts des utilisateurs, à partir de DBPEDIA (extrait des informations factuelles des pages Wikipédia). Le principe de cette approche permet de définir une représentation dynamique de la situation d'un utilisateur dans son environnement mobile. Ainsi, la situation physique(localisation exacte, date, heure),elle est transformée en une situation sémantique(la saison et la partie de la journée) basé sur les concepts de DBPEDIA. Ils ont comparé les résultats qu'ils ont obtenus avec ceux retournés par le moteur de recherche Google, ils ont calculé la précision des deux approches (approche d'enrichissement de requêtes vs celle de Google). Ils sont remarquent que le nombre total des ressources retournées par leur approche est largement inférieur à celles retournées par Google. Cette comparaison montre une amélioration au niveau de la précision des ressources obtenues grâce à processus d'enrichissement.

**[Lewenberget al, 2015]** Le but de cette étude est de développer un modèle d'utilisation des émotions pour prédire les centres d'intérêt des utilisateurs dans réseaux sociaux en ligne, ils sesont basés sur le modèle de régression logistique il est proposé d'utiliser six émotions qui représentent: joie, tristesse, la surprise, la colère, la peur, et le dégout ou bien saisir un mot-clé et à son tour associé à son émotion selon une liste de synonyme. Ils ont proposé une probabilité sur les émotions pour donner des résultats de la prédiction d'intérêt d'utilisateur. Les résultats de ce travail montrent qu'en utilisant les proportions d'émotions exprimes dans les médias sociaux il est possible de déterminer s'ils sont intéressés ou désintéressés par divers sujets et plusieurs corrélations intéressantes entre les centres d'intérêt d'un utilisateur.

## 3. Conclusion

A travers de ce chapitre, nous avons mis en évidence la notion de Data Mining qui est une phase très importante du processus de la prédiction. Nous avons par la suite défini chacune de ses tâches appartenant à deux catégories de méthodes : supervisées et non-supervisées, puis

présenté en détail les méthodes de prédiction: les arbres de décision, Knn, régression logistique ainsi que les réseaux de neurones les algorithmes génétiques. Nous avons présenté quelques travaux existants pour la prédiction dans les réseaux sociaux. Nous avons contacté dans un premier lieu que les techniques de prédiction sont peu exploitées pour l'analyse sociale, aussi elles ne sont pas utilisées pour la prédiction des intérêts des utilisateurs non actifs.

Dans le prochain chapitre nous montrons l'utilisation des techniques présentées ici, et la conception détailler de notre approche.

# 1. introduction :

Les intérêts des utilisateurs ont montré leur importance croissante dans la conduite du développement d'applications personnelles centrées sur l'utilisateur. Les études existantes pour prédire les intérêts des utilisateurs se concentrent sur les comportements de navigation et contenus de navigation (tels que les pages Web visualisées). Un des problèmes les plus difficiles à résoudre consiste à trouver des techniques de prédiction des intérêts des utilisateurs dans les réseaux sociaux.

A travers ce chapitre, nous allons présenter en détail notre solution pour la prédiction des intérêts des utilisateurs inactifs qui sont souvent connectés mais n'interagissent pas sur les réseaux sociaux à partir de son historique de recherche. Nous avons d'abord effectué une étude comparative entre les différent techniques proposées au niveau du chapitre deux. Ensuite, les trois techniques sont testés sur un data set pour choisir la meilleures techniques pour notre contribution.

## 2. Préparation des données :

Pour la bonne analyse et conception de l'outil prédictif, la phase de récupération de données et d'identification de leurs structures est très importante, voir critique. Effectivement, l'acquisition des données et de l'historique des recherches représente un atout majeur dans la conception de notre système, car toute méthode d'analyse prédictive se base sur un historique pour pouvoir prédire le comportement futur. Il est donc très important de se doter de données significatives et récentes.la phase préparation des données extraites à partir d'un réseau social, elle est divisé en deux étapes: description de data set et prétraitement. Nous allons exploiter, dans notre travail, le data set Delicious utilisé par **[Cantador et al, 2011]**.

### 2.1Description du Dataset :

Pour la bonne analyse et conception de l'outil prédictif, la phase de récupération de données et d'identification de leurs structures est très importante, voir critique. Effectivement, l'acquisition des données et de l'historique des recherches représente un atout majeur dans la conception de notre système, car toute méthode d'analyse prédictive se base sur un historique pour pouvoir prédire le comportement futur. Il est donc très important de se doter de données significatives et récentes.

Nous allons effectuer notre analyse prédictive sur les données du Dataset Delicious, qui est un ensemble de données publié dans le cadre du the ***2nd International Workshop on***

***Information Heterogeneity and Fusion in RecommenderSystems (HetRec 2011)*** à the 5th ***ACM Conference on RecommenderSystems (RecSys 2011)***http://recsys.acm.org/2011.

Le dataset Delicious offre des informations relatives aux réseaux sociaux, les signets (bookmarking) et les balises (tagging) à partir d'un ensemble d'utilisateurs (20K) du système de bookmarkingdu site social Delicious[9]. Le data set est organisé en un ensemble de fichiers : *user_contacts.dat, bookmarks.dat,user_taggedbookmarks.dat, user_contacts-timestamps.dat, bookmark_tags.dat,user_taggedbookmarks-timestamps.dat,*u*ser_taggedbookmarks.dat,* et*tags.dat.*

Pour le besoin de notre travail, nous allons utiliser le fichier ***user_taggedbookmarks.dat***, il contient1867 utilisateurs et chaque utilisateur possède un historique de tag et de bookmark, le nombre total de tag est 53388 et 104799 de bookmark. Il contient437592 lignes, les colonnes sont :

- userID : qui signifie l'identificateur de l'utilisateur ;
- bookmarkID : identificateur du bookmark;
- tagID : identificateur du tag ;
- day,month, year, hour, minute, second : le jour le moi et l'année l'heur minute second à laquelle le tag a été posté par l'utilisateur.
- Hypothèse : Dans le cadre de notre projet, pour la réalisation de la prédiction des centres d'intérêts des utilisateurs non actifs, il faut se doter de données sur l'historique de recherches de l'utilisateur. Cependant, vu que nous n'avons pas trouvé un dataset qui correspond à notre problématique, nous allons travailler avec le dataset Delicious ou nous supposons que les tags représentent dans notre projet les requêtes exprimés par l'utilisateur lors de ces recherches.

[9]http://www.delicious.com.

| | userID | bookmarkID | tagID | day | month | year | hour | minute | second |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 8 | 1 | 1 | 8 | 11 | 2010 | 23 | 29 | 22 |
| 1 | 8 | 2 | 1 | 8 | 11 | 2010 | 23 | 25 | 59 |
| 2 | 8 | 7 | 1 | 8 | 11 | 2010 | 18 | 55 | 1 |
| 3 | 8 | 7 | 6 | 8 | 11 | 2010 | 18 | 55 | 1 |
| 4 | 8 | 7 | 7 | 8 | 11 | 2010 | 18 | 55 | 1 |
| 5 | 8 | 8 | 1 | 8 | 11 | 2010 | 18 | 49 | 5 |
| 6 | 8 | 8 | 8 | 8 | 11 | 2010 | 18 | 49 | 5 |
| 7 | 8 | 8 | 9 | 8 | 11 | 2010 | 18 | 49 | 5 |
| 8 | 8 | 9 | 1 | 8 | 11 | 2010 | 18 | 37 | 13 |
| 9 | 8 | 9 | 10 | 8 | 11 | 2010 | 18 | 37 | 13 |

**Figure 9:** Extrait de dataset Delicious

- La figure 10est un exemple de données que nous pouvons avoir du dataset Delicious. Les données visibles sur cette figure concernent les informations sociales l'utilisateur dont l'id est égale à 8.

**2.2 Nettoyage de données :**

- Le nettoyage de données est une étape importante avant d'analyser ou de modéliser les données. Dans notre dataset il n'existe pas des valeurs manquant par contre nous avons observé (Figure 10)après le calcul des nombre de requêtes (tags) pour chaque utilisateur qu'il existe des utilisateurs qui possèdent un nombre de requêtes inférieur ou égale 5 et d'autres utilisateurs supérieur ou égale 1000, donc les données sont aléatoires et déséquilibrées et dans ce cas, il est impossible de vérifier l'exactitude des prédictions. A cet effet, nous avons éliminé ces utilisateurs par l'algorithme suivant (Algorithme 1) :

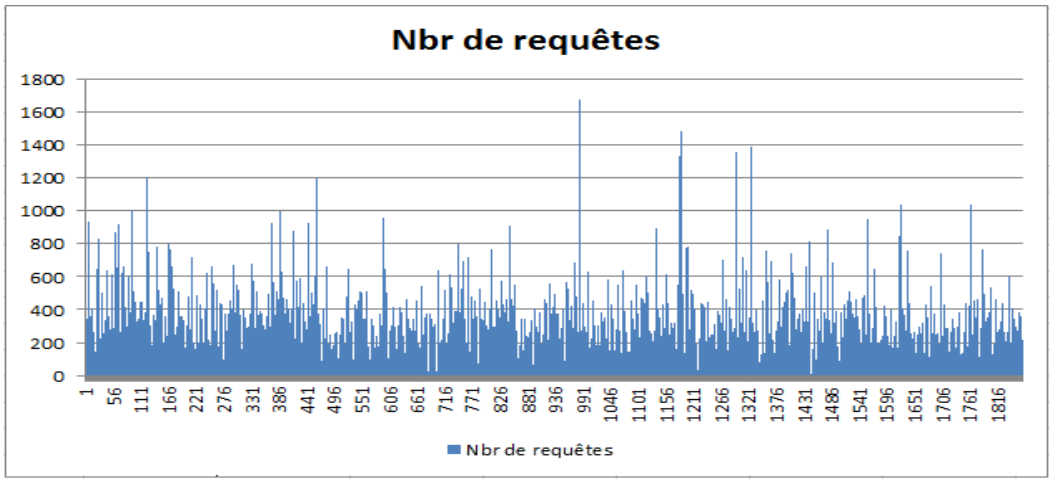


**Figure 10:** le nombre de requête

```
Algorithme : élimination
        list_utilisateur_éliminé  : list
        data_nbr_req, data :Data
        i=entier
Début:
1 :list_utilisateur_éliminé  ← list_vide
2 : Pour i←0 a nbr_ligne_data_nbr_reqfaire /*parcoure le dataset de nbr de requêtes  */
3 :Sidata_nbr_req [‘nbr_req’][i]<=5 ou data_nbr_req [‘nbr_req’][i]>=1000 Alors
4 :list_utilisateur_éliminé.ajouter(data_nbr_req  [‘userID’][i]) ;
                /*sauvegarde les utilisateurs qui on nbr de req<=5 ou >=1000*/
5 : finsi ;
6 :fait ;
7:Pour i ←0 a list_utilisateur_éliminé  .langueur () faire
8 : data. Suprimer  (data [data.userID =list_utilisateur_éliminé  [i]].index, inplace=True)
9 :fait;
```

**Algorithme 1 :** Elimination utilisateur

Après avoir appliqué cet algorithme, nous avons éliminé 77 utilisateurs et 15294 lignes.

### 2.3 Ajout une nouvelle colonne

Quoique le dataset contient des données consistantes mais il reste inadapté pour le besoin de la prédiction vu le manque d'information sur les données qu'il faut considérer comme des données futures (intérêts futures). Il fallait dans ce cas ajouter une nouvelle colonne qui représente l'intérêt prédit ou la sortie de l'analyse prédictivecàdlabel. A cet effet, nous considérons les 15 dernières requêtes comme intérêts futures pour chaque utilisateur et donc le label est égale 1. Par la suite, nous parcourant les lignes de chaque utilisateur et nous comparons ses requêtes avec ceux considères comme prédits (les 15 derniers) si elles sont égaux le label reçoit 1, sinon 0 (**Algorithme 2**).

```
Algorithme : ajouet_labels
        nb, j : entier ; tab : matrice  d'entier ; dataset, nouveau_dataset : Dataset
 Début:
1: nouveau_dataset=dataset. ajouter_colonne (‘intérêt’)
2 :Tab=new tab(1867,17)
3 :J←0 ;
4: Nbr←0;
5: Pour i←0 a nombre_line_datasetfaire     /*boucle pour parcoure tous les lignes de dataset */
6: Si (dataset [‘userID’][i] ≠ dataset [‘userID’][i+1]) alors
7: tab[i][0]=I;
9: tab[i][1]←dataset ['userID'][j];          tab[i][6]← dataset ['reqID'][j];
10: tab[i][2]←=dataset ['reqID'][j-1];    tab[i][7] ←dataset ['reqID'][j-2];
11: tab[i][4] ←dataset ['reqID'][j-3];     tab[i][8] ←dataset ['reqID'][j-4];
12: tab[i][3] ←dataset ['reqID'][j-5];     tab[i][9] ←dataset ['reqID'][j-6];
13: tab[i][5] ←dataset ['reqID'][j-7];     tab[i][10] ←dataset ['reqID'][j-9];
14: tab[i][4] ←dataset ['reqID'][j-10];   tab[i][11] ←dataset ['reqID'][j-11];
15: tab[i][3] ←dataset ['reqID'][j-12];   tab[i][12] ←dataset ['reqID'][j-13];
16: tab[i][5] ←dataset ['reqID'][j-14];   tab[i][13] ←dataset ['reqID'][j-15];
17: Finsi;
```

/* remplis le tableau par l'indice de chaque utilisateur et les 15 derniers tags on concéder ses tag comme des intérêts future */

**18** : **Fait pour** ;

**19** : **Pour** i←0 anombre_line_ **faire**/*Boucle pour parcoure tous les lignes denouveau dataset et remplir la colonne intérêt avec 0 ou 1 */

**20**: Tanque(j<1867 and tab[j][1]≠ nouveau_dataset ['userID'][i]) faire

**21**: j=j+1;

**22**: Fait Tanque ;

**23**: Si (tab[j][2] = nouveau _dataset ['reqID'][i ] ou tab[j][3] = nouveau_dataset ['reqID'][i ] ou tab[j][4] =nouveau_dataset ['reqID'][i ] ou tab[j][5] =nouveau_dataset ['reqID'][i ] ou tab[j][7]=nouveau_dataset ['reqID'][i ] ou tab[j][8] =nouveau_dataset ['reqID'][i ]ou tab[j][9] =nouveau_dataset ['reqID'][i ]ou tab[j][10] =nouveau_dataset ['reqID'][i ] ou tab[j][11] =nouveau_dataset ['reqID'][i ] ou tab[j][12] =nouveau_dataset ['reqID'][i ] ou tab[j][13] =nouveau_dataset ['reqID'][i ] ou tab[j][14] =nouveau_dataset ['reqID'][i ] ou tab[j][15] =nouveau_dataset ['reqID'][i ] ) alors

/*Si tag-id existe dans les 5 derniers tags alors la cellule prendre la valeur 1*/

**24** : nouveau_dataset ['nteret'][i ]←1;

**25** : Si non nouveau_dataset ['nteret'][i ]←0 ;

/*Si tag-idn'existe pas dans les 5 derniers tags alors la cellule prendre la valeur 1*/

**26** : Finsi;

**27** : Fait pour;

Fin.

**Algorithme 2 :** ajouet_labels

## 4. Choix de la méthode prédictive :

Lors de la conception d'un système prédictif, on se trouve généralement face à une multitude de techniques et méthodes prédictives, qui semblent à première vue toutes rapides et efficaces. Cependant, si l'on regarde de plus près, chaque méthode présente des avantages qui la rendent imbattable sur une catégorie de prédictions, et des inconvénients la rendant inutilisable dans d'autres catégories(***cf chap.2 §3.3.***). Cela veut dire que chaque problématique a sa propre technique de prédiction optimale. Pour cela, nous n'allons tester, dans un premier temps, les quatre méthodes de prédiction présentées dans le chapitres 2 (arbre de décisions, Knn, régression logistique, et réseau de neurones), nous n'allons retenir qu'une d'entre elles, nous avons exclus les autres pour leur faible de pertinence.

Avant de détailler chaque technique, nous déterminons les de variables d'entrée et de sortie :

- Les entrées : sont les deux colonnes userid, reqid (features) ;
- Les sorties : la colonne intérêt prédit (label).

**Remarque :** nous n'avons pas utilisé, dans notre travail, ***bookmarkid***car cette information social n'influence pas sur la prédiction des centres intérêts des utilisateurs non actifs, aussi vu que le temps dans le dataset est divisé en ***'année, mois, jour, heure, minute, seconde'*** le nombre d'entrées augmente et influence sur la pertinence des résultats (sur apprentissage[10]) pour cela nous avons supprimé les données sur le temps.

### 4.1 L'arbre de décisions

Nous avons implémenté l'algorithme de l'arbre de décisions à l'aide du modèle DecisionTreeClassifier présent dans la bibliothèque sklearn.treeet nous avons choisi le «score et accuracy[11]» comme mesure de performance.

### 4.2 KNN

Nous avons implémenté l'algorithme KNN à l'aide du modèle KNeighborsClassifier présent dans la bibliothèque sklearn, accompagné d'un ensemble de paramètres à spécifier afin d'obtenir le modèle le plus optimal possible. Ces paramètres sont les suivants :

▪ n_neighbors : Le nombre de voisins pris par le modèle c'est le paramètre (k) de l'algorithme, par défaut n_neighbors=5.

▪ metric: la métrique de distance à utiliser par défaut est Minkowski, la liste des métriques disponibles est consultable dans la documentation de "sklearn.neighbors.DistanceMetric", nous avons utilisé la métrique accuracy .

### 4.3 Régression logistique

Nous avons implémenté l'algorithme de la régression logistique à l'aide du modèle LogisticRegression présent dans la bibliothèque sklearn.linear_model et nous avons utilisé l'accuracy et le score comme mesures de performance.

### 4.4 Deeplearning (réseau de neurones)

Afin de créer notre modèle ou boîte noire, il abritera les neurones et les organisera sous la forme de couches de neurones connectées entre elles. Grâce à des relations logiques, chaque couche traite les données entrantes de la couche qui la précède, nous utilisons le module "keras" de TensorFlow, nous allons créer un modèle séquentiel, ce qui signifie que les canaux couche par couche du réseau de neurones seront en séquence les uns après les autres le

[10]L'overfitting intervient lorsque l'algorithme sur-apprend (*overfit*), autrement dit, lorsqu'il apprend à partir des données mais aussi à partir de patterns (schémas, structures) qui ne sont pas liés au problème, comme du bruit.

[11]L'accuracyest une mesure qui décrit généralement les performances du modèle dans toutes les classes. Il est utile lorsque toutes les classes sont d'égale importance. il s'agit du rapport entre le nombre de prédictions correctes et le nombre total de prédictions

modèle créé se compose d'une couche d'entrée couche cachée (Dense) de une couche de sortie.

nous allons dans ce qui suit le détailler.

Pour que le réseau soit opérationnel il faut passer par les étapes suivantes :

- Création du modèle et choix des hyperparamètres (nombre de couches, type du modèle...).
- Choix de la fonction de coût.
- Entrainement du modèle.

### 4.4.1 Création du modèle

#### 4.4.1.1 Architecture de réseau de neurones

La plupart des réseaux de neurones actuels sont organisés en couches de nœuds (c'est-à-dire des neurones) qui sont connus pour être composés d'une couche d'entrée, d'une couche cachée et d'une couche de sortie comme le montre la figure 11, et ils sont "feed-forward", ce qui signifie que les données se déplacent à travers eux dans une seule direction Un réseau de neurones se compose de milliers, voire de millions de nœuds de traitement simples qui sont densément interconnectés par des connecteurs appelés connexions pondérées. Les nœuds effectuer quelques opérations sur les entrées pour atteindre la sortie finale. L'ensemble des opérations effectué par un nœud est illustré à la Figure 3.3

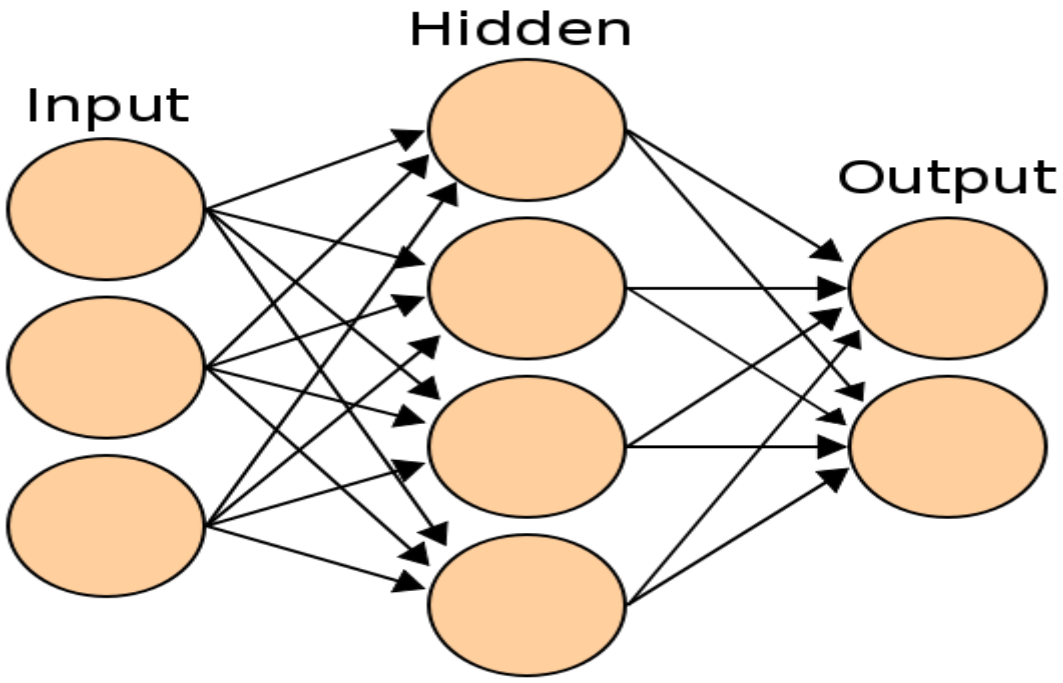


Figure 11 : Représentation des couches de réseau de neurones

Un nœud individuel peut être connecté à plusieurs nœuds de la couche avant lui, à partir desquels il reçoit des données, et plusieurs nœuds dans la couche après lui, auxquels il envoie des données (c'est-à-dire que la couche d'entrée est la couche inférieure). Lorsqu'un réseau de neurones est en cours de entraîner, chaque nœud est initialisé avec un ensemble de poids et de seuil à des valeurs aléatoires, qui stockeront et évalueront l'importance de l'une des entrées pour le résultat. Pendant l'entraînement, le réseau de neurones commence par fournir les données d'entraînement à la couche d'entrée. Ensuite, il passe à travers les couches successives, se multiplie et s'ajoute ensemble de manière complexe après avoir stocké les informations concernant l'importance de l'entrée, l'information passe par une fonction d'activation qui décide de passer ou non l'information au neurone suivant ; Puis, il arrive enfin radicalement transformé à la couche de sortie. Après avoir trouvé des modèles en corrélation avec la couche de sortie (c'est-à-dire une étiquette particulière), les poids et les seuils sont ajustés à plusieurs reprises jusqu'à ce que les données d'apprentissage portent les mêmes étiquettes produire systématiquement des résultats similaires grâce à l'utilisation d'une méthode d'optimisation. Ce dernier estime le gradient d'erreur pour l'état actuel du réseau de neurones à l'aide d'exemples à partir de l'ensemble de données d'apprentissage, puis il met à jour les poids en utilisant la rétro propagation des pertes.

#### 4.4.1.2 Création l'architecture de notre modèle

Pour créer notre modèle, qui il va loger des neurones et les organiser en couches de neurones interconnectés. Grâce à la relation logique, chaque couche traite les données entrantes des couches précédentes. Pour cela nous utilisons le module "keras" de TensorFlow (Keras est une API de réseaux de neurones de haut niveau, écrite en Python et interfaçable avec TensorFlow, CNTK et Theano. Elle a été développée avec pour objectif de permettre des expérimentations rapides).

Notre modèle se compose de 3 couches :

- **Première couche (couche d'entrée)** : spécification de la forme des données en entrée par paquet du réseau avec 300 nœuds et la fonction d'activation ' relu'. La fonction 'relu' (signifie unité linéaire rectifiée) est une fonction d'activation non linéaire qui a gagné en popularité dans le domaine de l'apprentissage en profondeur. Le principal avantage de l'utilisation de la fonction 'relu' par rapport aux autres fonctions d'activation est qu'elle n'active pas tous les neurones en même temps.

- **Deuxième couche (couche cachée)** : cette couche contient 128nœuds et la fonction d'activation 'relu'.

- **Troisième couche (couche de sortie)** : une couche de réseau de neurones densément connecté avec un seul nœud qui applique en sortie une fonction d'activation 'sigmoid'. Nous avons choisi comme fonction de transfert une fonction sigmoïde $f(x) = \frac{1}{1+e^{-x}}$ qui présente l'avantage d'être dérivable dont la dérivée est $f'(x) = f(x).(1 - f(x))$ ainsi que de donner des réels compris entre 0 et 1.

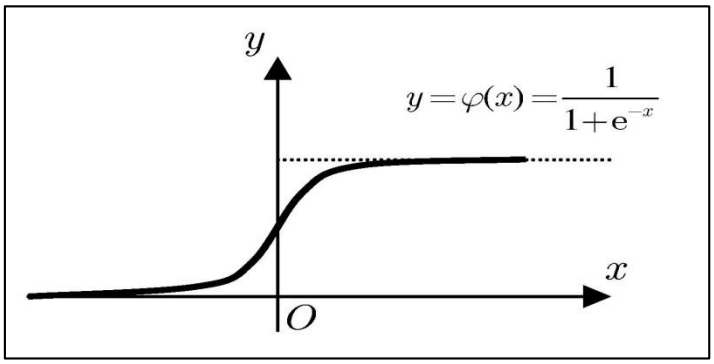


**Figure 11:** Fonction sigmoïde

### 4.4.2 La Compilation du modèle

Consiste à configurer le processus d'apprentissage qui comporte le choix de l'optimiseur (optimisation « RMSprop »), la fonction qui calcule la perte d'information (Binarycrossentrpy) et les métriques du modèle nous utilisons « accuracy ».

- L'**accuracy** est une métrique de performance qui évalue la capacité d'un modèle de classification à bien prédire à la fois les individus positifs et les individus négatifs.

- Optimiser(RMSprop): la fonction d'optimisation est un algorithme mathématique qui utilise des dérivés pour comprendre les changements survenus dans le réseau de neurones et suivre l'évolution. Les changements se feront par la fonction de perte qui à chaque itération modifiera les poids des connexions entre neurones, réduisant ou augmentant le poids entre neurones pour obtenir un meilleur résultat, notre choix est d'amener RMSprop, C'est un optimiseur très robuste.

### 4.4.3 Entrainement du modèle

Après la configuration du model pour la classification on l'entraine sur les données en utilisant la méthode « fit ».Le déroulement de la phase d'entrainement du réseau de neurones se fait en deux étapes :

**1) La première étape** est l'étape de propagation vers l'avant. Cette étape est effectuée tout en apprenant à partir des données, en traversant ou en traversant l'ensemble du réseau de neurones (toutes les couches du réseau de neurones), de sorte que chaque neurone de la couche traite les données d'entrée et transforme et passer les informations au niveau suivant jusqu'au dernier niveau du classement prédit.

**2) deuxième étape :** fonction de perte ou la fonction d'estimation d'erreur mesure la précision de la classification. Par conséquent, au fur et à mesure que le modèle est établi, le poids de connexion du neurone s'ajustera progressivement jusqu'à ce qu'une bonne prédiction soit obtenue (l'erreur est presque nulle), et l'étape d'ajustement est appelée rétropropagation.

**4.4.4 Tester et utiliser :**

Ce qui suit est un aperçu du travail d'amélioration des réseaux de neurones et des informations qu'ils génèrent.

-Réduire le taux de perte et les changements de qualité de l'information.

-Dans le processus de fonction de réseau de neurones, les résultats obtenus après chaque itération sont améliorés, le taux de perte est réduit et la classification est plus crédible.

**La figure 13** représente les changements de loss et accuracy dans le réseau de neurones.

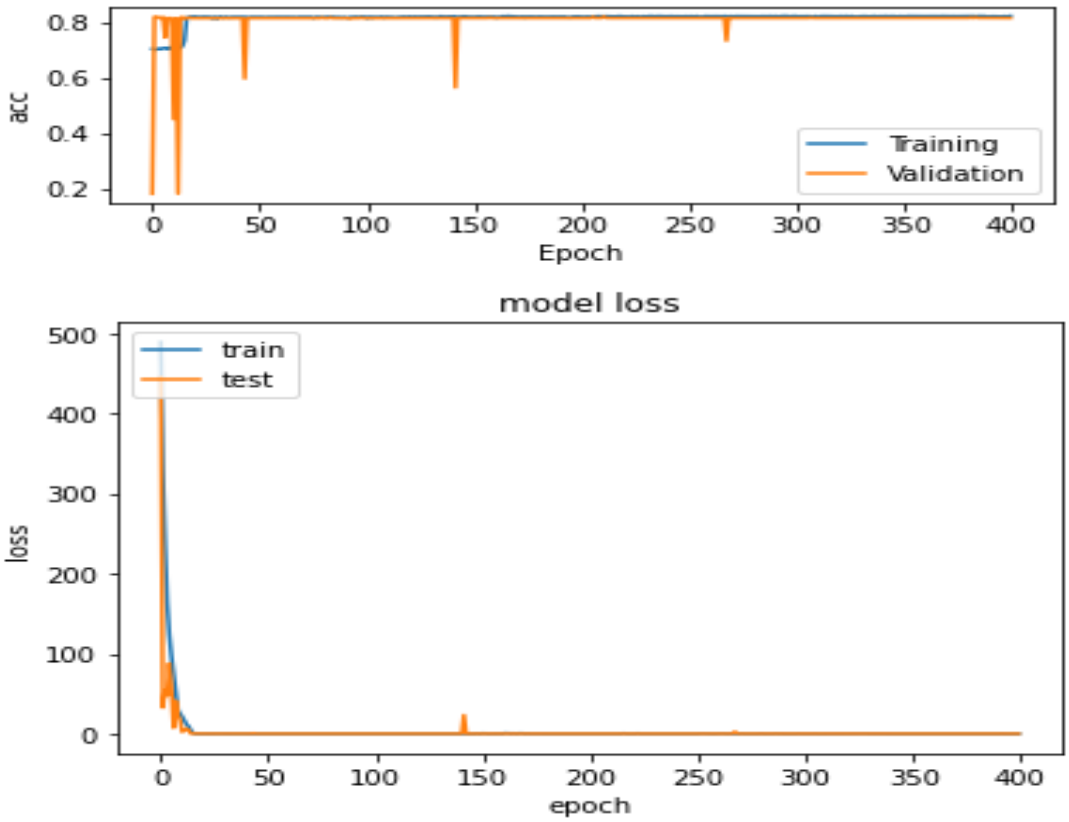


**Figure 12:**les changements de loss et accuracy

Le tableau suivant résume les paramètres utilisés dans notre réseau de neurones :

| Paramètres | Valeur |
|---|---|
| Activation Function | Relu |
| Optimizer | RMSprop |
| Loss | Binary crossentropy |
| Metric | accuracy |
| Epochs | 400 |
| Batch size | 1000 |
| Validation split | 0.5 |
| Number of layers | 3 |
| Size of the input | 2 |

**Tableau 2:** Les paramètres du réseau de neurones

### 4.5 Etude Comparative entre les techniques de prédiction :

Dans cette partie nous comparons les quatre techniques de prédiction précédemment décrites et implémentées. La comparaison est effectuée par rapport à la mesure Accuracy, les résultats sont présentés par le **tableau 2**et la **figure 12** qui illustrent la performance en terme d'évaluation pour le nombre d'utilisateurs @Nbr (Nbr = 100, 500, 1000, 1790) obtenue de l'arbre de décisions, le knn, la régression logistique, et le réseau de neurones.

| Accuracy / Technique | @100 | @500 | @1000 | @ 1790 |
|---|---|---|---|---|
| **Arbre de décisions** | 0.7072195186987534 | 0.7049747841695871 | 0.6899582332166874 | 0.6848891204592406 |
| **KNN** | 0.7230184654356376 | 0.723195714733453 | 0.7105029752785796 | 0.7029196146181832 |
| **Régression Logistique** | 0.8304779681354576 | 0.8257258455138616 | 0.8162342165099741 | 0.8109974588200881 |
| **Réseau de neurones** | 0.7856052930653096 | 0.8134300218522549 | 0.820450338870287 | 0.8124966748058796 |

**Tableau 3:**Comparaison entre les techniques de prédiction en terme d'Accuracy

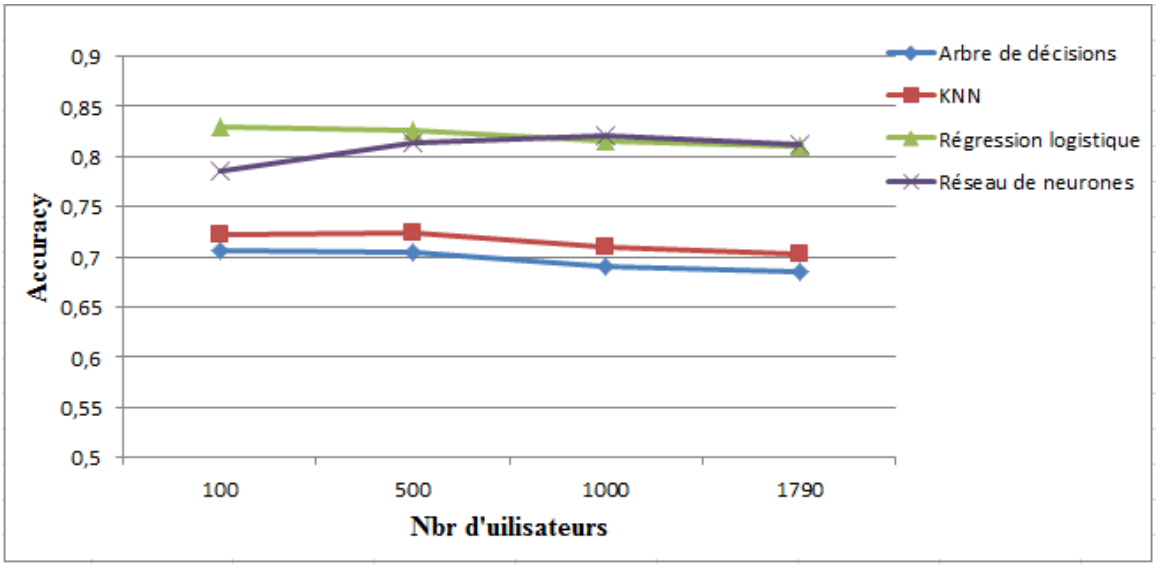


**Figure 13:** Comparaison entre les techniques de prédiction en termes d'Accuracy

Comme le montre le **tableau 3** et la **figure 14**, nous remarquons dans un premier lieu à travers les résultats obtenus que l'arbre de décisions et Knn sont moins performants que la régression logistique et le réseau de neurones, ils sont donc exclus pour la prédiction des centres d'intérêts. Nous pouvons exclure en second lieu la méthode de régression logistique, qui est une méthode d'apprentissage supervisé et de variables continues, certes, mais qui devient lente et imprécise quand il s'agit d'un gros volume de données (diminution de l'accuracy lors de l'augmentation du nombre d'utilisateurs).

Nous pouvons conclure que les réseaux de neurones (deeplearning) s'adaptent parfaitement à la problématique de prédiction d'intérêts vu que quand le volume de données augmente il y a des fois une faible basse de l'accuracy.

Apres étude comparative entre les différentes méthodes prédictives, et vu que le réseau de neurones soit le mieux adapté à notre problématique nous allons enregistrer le modèle dans un fichier .h5 pour l'utiliser en suit dans notre application

## 5. Conclusion

Nous avons présenté au cours de ce chapitre le cadre conceptuel de notre système. Nous avons effectués des tests sur le dataset delicious avec quatre techniques de prédication qui sont : arbre de décisions, régression logistique, K plus proches voisins et le deeplarning (les réseaux de neurones), pour répondre à l'objectif principal de notre projet: pour la prédication des intérêts d'utilisateur.

## 1. Introduction

Au cours du chapitre précédent nous avons abordé la conception de notre système, nous arrivons dans ce chapitre à entamer la mise en œuvre pour la création de l'application web pour la prédiction des intérêts des utilisateurs inactifs, qui va nous permettre d'arriver aux objectifs fixés précédemment, dans une nous avons conçu une architecture globale illustrant ses composantes principales et nous allons présenter le choix logiciels et matériels utilisé pour l'implémentation et dans la deuxième nous utilisons des outils de langage UML pour modéliser notre système et nous allons présenter les principales interfaces qui la composent à travers des fenêtres de capture.

## 2. Architecture du système

L'architecteur globale de notre système est représentée dans **(la figure15).** Le processus de l'application de système s'étale sur plusieurs étapes.

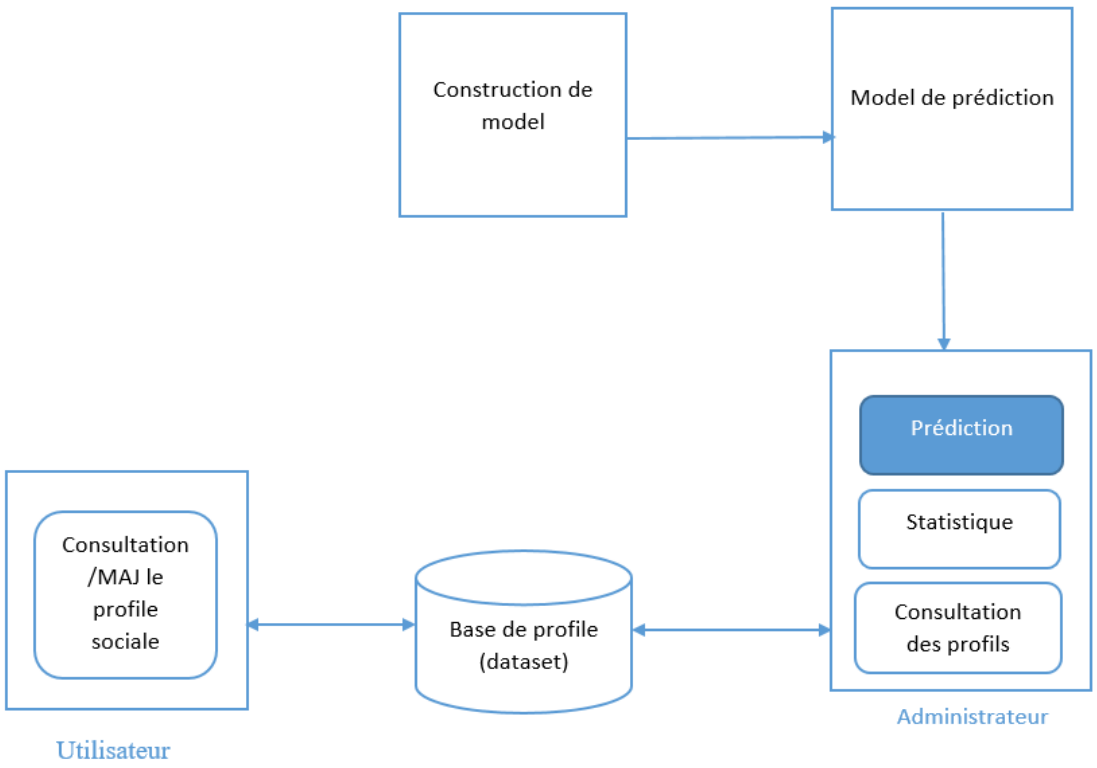


**Figure 14:** Schéma global du système

Notre système comprend les composants suivants :

### 2.1 Utilisateur

Dans notre travail nous nous intéressons à l'utilisateur non actif qui n'interagit pas dans les réseaux sociaux pour différents raisons : médicale, familiale, malicieuse, …etc. Ce type de profil manque beaucoup de données sociales, nous exploitons son historique de recherches dans le cadre de notre projet les rôles de cet utilisateur sont:

- ✓ Consulter les ressources.
- ✓ Modifier ses informations personnelles.

### 2.2 Administrateur

C'est le responsable de la gestion du système et les rôles de l'administrateur sont:

- ✓ Consulter les différentsprofils.
- ✓ Appliquer les techniques de prédiction.
- ✓ Visualiser les différentes statistiques.

## 3. Configuration matérielle du système

L'implémentation de notre application a été réalisée sur une machine virtuelle possédant les caractéristiques suivantes :

- ✓ Marque : HpEliteBook 830 G6.
- ✓ Mémoire(RAM) :8 Go.
- ✓ Processeur:Intel® Core™ i5-8265U 1.60GHZ 1.80GHZ.
- ✓ Système d'exploitation:Windows 10 professionnel (64bits).

## 4. Environnement du travail

Dans cette partie nous présentons toutes les technologies et les outils de développement que nous avons utilisé pour la réalisation de notre système depuis la récolte de données.

### 4.1 Python 3.7.4[12]

Python est un langage de programmation interprété à usage général, interactif, orienté objet et de haut niveau. Python combine une puissance remarquable avec une syntaxe claire. Il comporte des modules, des classes, des exceptions, des types de données dynamiques de très haut niveau et un typage dynamique. Il existe des interfaces pour de nombreux appels système et bibliothèques, ainsi que pour divers système de fenêtrage.

[12]https://www.python.org/downloads/release/python-394/

### 4.2 Flask

Flask est un micro framework open-source de développement web en python. Il est classé comme microframework car il est très léger. Flask a pour objectif de garder un noyau simple mais extensible. Il n'intègre pas de système d'authentification, pas de couche d'abstraction de base de données, ni d'outil de validation de formulaires. Cependant, de nombreuses extensions permettent d'ajouter facilement des fonctionnalités.

### 4.3Modèle MVC

Le Framework flask utilise l'architecture MVC est composé de trois types de modules ayant trois responsabilités différentes: les modèles, les vues et les contrôleurs.

- Un modèle (Model) contient les données à afficher.
- Une vue (View) contient la présentation de l'interface graphique.
- Un contrôleur (Controller) contient la logique concernant les actions effectuées par l'utilisateur.

### 4.4 Jupyter [13]

Jupyter est une application web utilisée pour programmer dans plus de 40 langages de programmation, dont Python, Julia, Ruby,R, ou encore Scala2. C'est un projet communautaire dont l'objectif est de développer des logiciels libres, des formats ouverts et des services pour l'informatique interactive. Jupyter est une évolution du projet IPython. Jupyter permet de réaliser des calepins ou notebooks, c'est-à-dire des programmes contenant à la fois du texte en mark down et du code. Ces calepins sont utilisés en science des données pour explorer et analyser des données.

### 4.5SGBD SQLite [14]

SQLite est un SGBD que contrairement aux autres SGBD, tel que MySQL il ne se base pas sur le modèle client-serveur mais il est directement intégrée aux programmes. L'intégralité de la base de données (déclarations, tables, index et données) est stockée dans un fichier qui est indépendant de la plateforme. Il est intégré dans les bibliothèques standards de beaucoup de langages comme PHP ou Python, il est connu aussi par son extrême légèreté (moins de 300 Ko).

[13]https://jupyter.org/install
[14]https://www.sqlite.org/download.html

### 4.6 Colab[15]

Nous avons aussi programme et teste les techniques de prédiction et le système développe dans le cloud Colab. Google Colab ou Colaboratory est un service cloud, offert par Google (gratuit), basé sur Jupyter Notebook et destiné à la formation et à la recherche dans l'apprentissage automatique. Cette plateforme permet d'entraîner des modèles de Machine Learning directement dans le cloud.

### 4.7 CSV

Comma-separated values, connu sous le sigle CSV, est un format texte ouvert représentant des données tabulaires sous forme de valeurs séparées par des virgules. Un fichier CSV est un fichier texte, par opposition aux formats dits « binaires ». Chaque ligne du texte correspond à une ligne du tableau et les virgules correspondent aux séparations entre les colonnes. Les portions de texte séparées par une virgule correspondent ainsi aux contenus des cellules du tableau.

## 5. Diagrammes UML

Diagrammes UML sont constitués de diagrammes qui servent à visualiser et décrire la structure et le comportement des objets qui se trouvent dans un système. Il permet de présenter des systèmes logiciels complexes de manière plus simple et compréhensible.

### 5.1 Présentation de cas d'utilisation

Les diagrammes de cas d'utilisation modélisent le comportement d'un système et permettent de capturer les exigences du système. Chaque usage que les acteurs font du système est représenté par un cas d'utilisation. Nous présentons le diagramme de cas d'utilisation suivant (**figure 13**):

-Accès au système.

-Gérer les utilisateurs.

-Faire des statistiques.

-Appliquer les techniques de prédiction.

-Consulter.

**Diagramme de cas d'utilisation**

[15]https://colab.research.google.com/

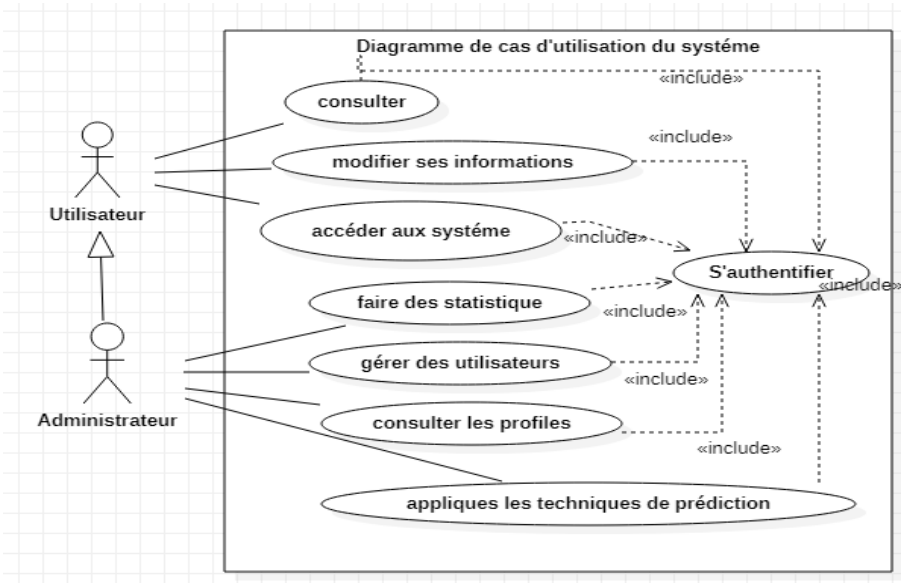


**Figure 15:**Diagramme de cas d'utilisation

## 5.2 Diagrammes de séquences

Les diagrammes de séquences sont la représentation graphique des interactions entre les acteurs et le système selon un ordre chronologique. Le diagramme de séquence permet de montrer les interactions d'objets dans le cadre d'un scénario d'un diagramme des cas d'utilisation.

### 5.2.1 Scénario d'authentification

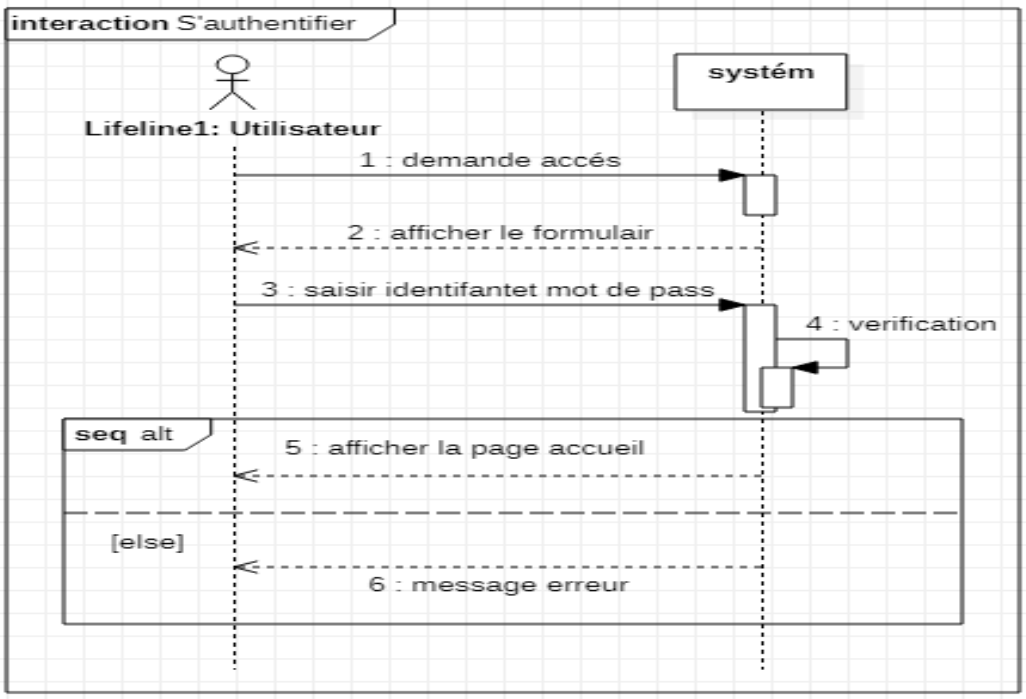


**Figure 16:**Diagramme de séquence pour s'authentifier

### 5.2.2 Scénario de modification le profil

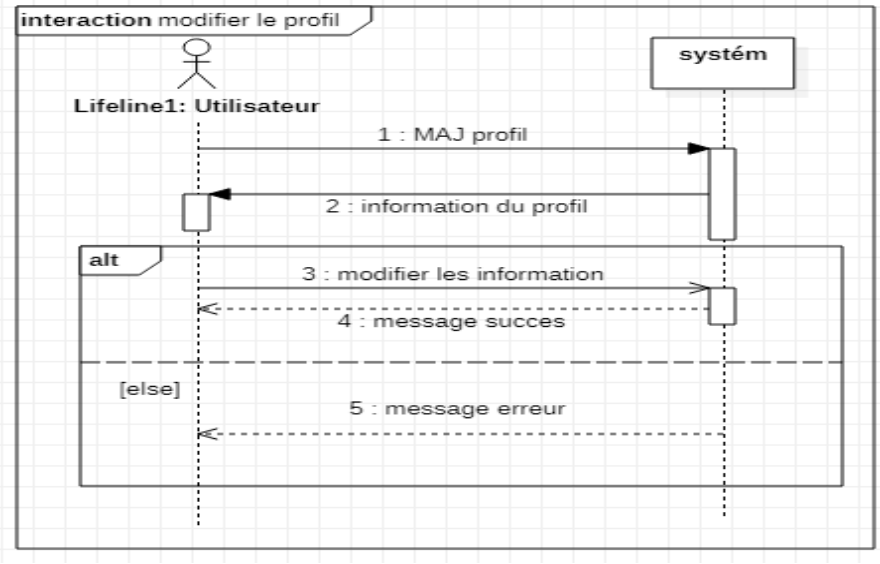


**Figure 17:**Diagramme de séquence pour modification de profil

### 5.2.3 Scénario de consultation/rechercher

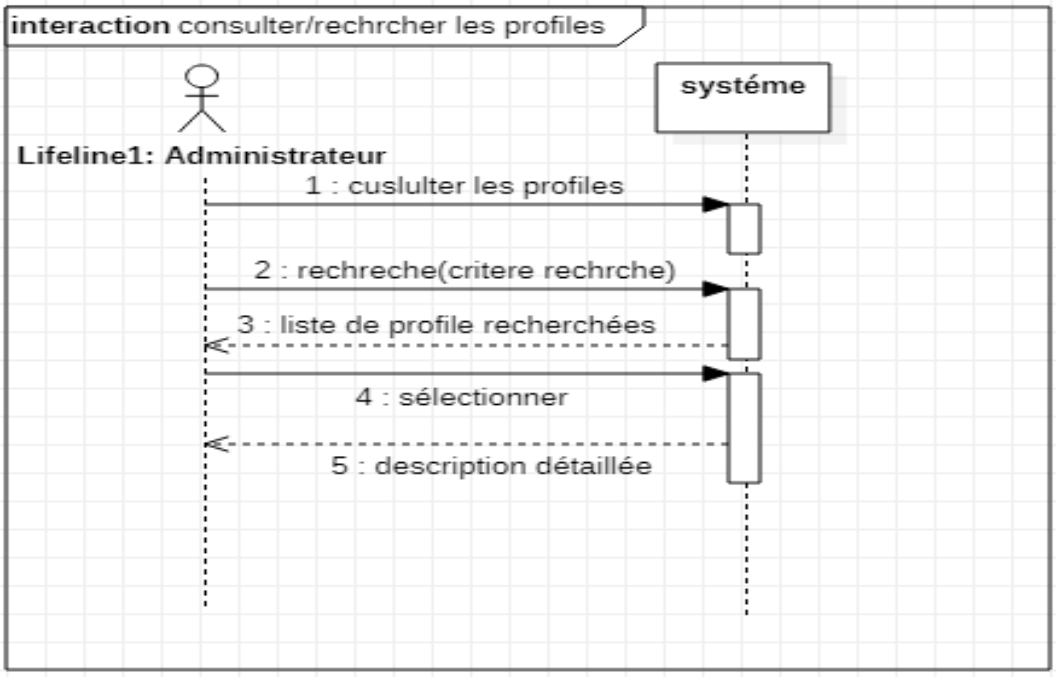


**Figure 18:**Diagramme de séquence pour consulter/rechercher

### 5.3 Diagramme de classes

Le diagramme de classe est considéré comme le plus important dans la modélisation orientée objet, il permet de fournir une représentation abstraite des objets du système qui vont interagir

pour réaliser les cas d'utilisation. Après l'identification des besoins concernant notre projet, nous avons proposé le diagramme de classes suivant :

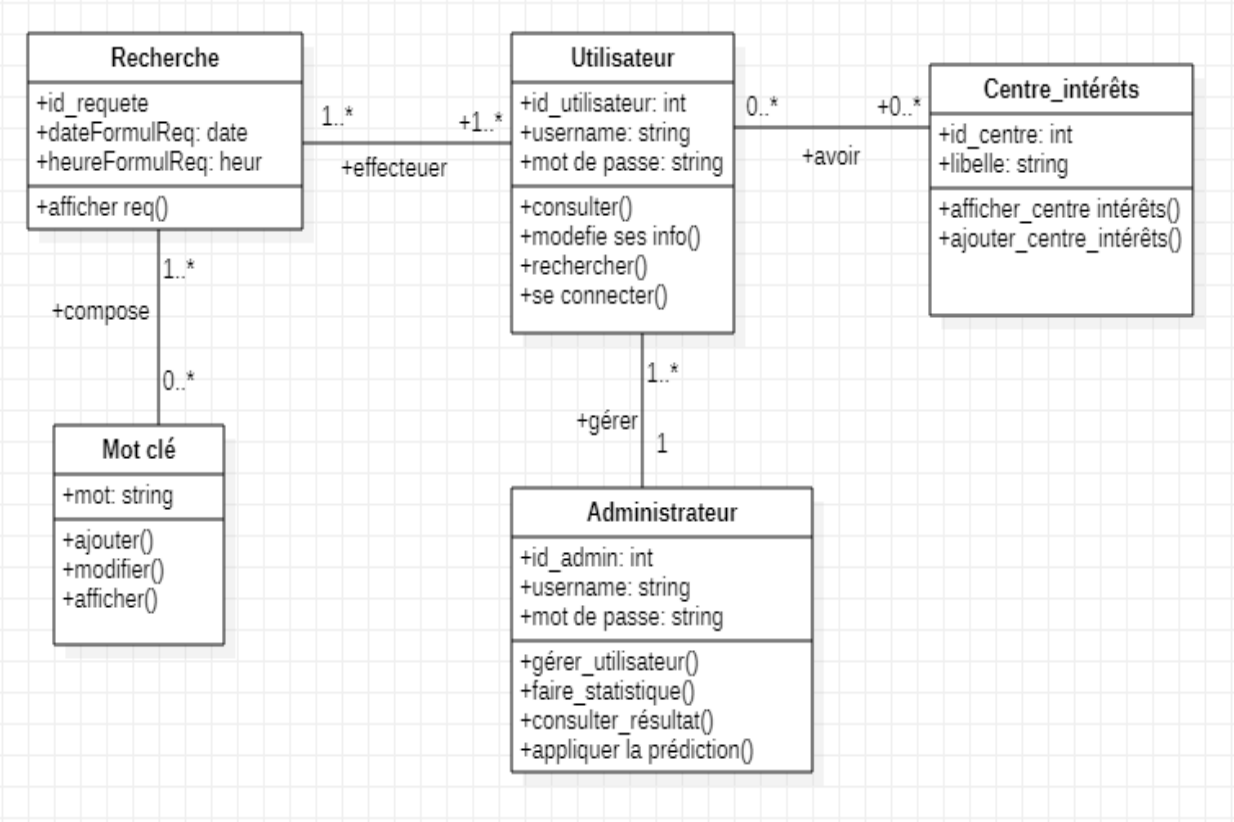


**Figure 19** : Diagramme de classes

## 6. Interfaces de l'application

Les interfaces de l'application établissent un dialogue entre la machine et l'homme, et facilite l'utilisation du système, nous présentons dans ce qui suit notre application web Butterfly travers ses principales interfaces et fonctionnalités.

### 6.1 Register

Dans le but d'observer la prédiction des intérêts des utilisateur défini principalement par ses centres d'intérêt et se basant sur ses requêtes, notre application offre au départ à l'utilisateur la possibilité de s'inscrire au système pour lui dédier un profil qui va lui permettre d'interagir avec le système et de bénéficier des différentes fonctionnalités.

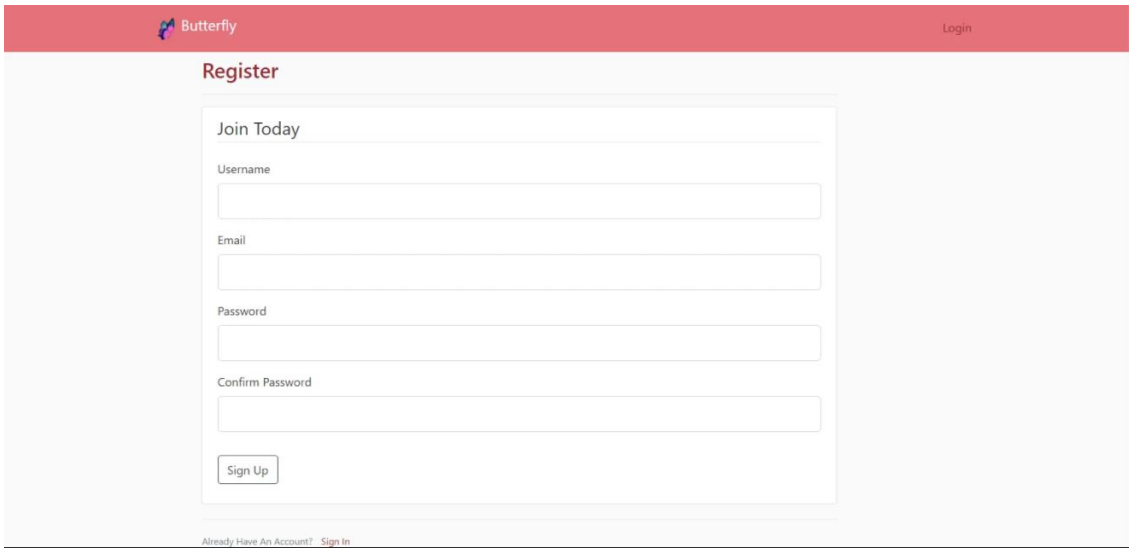


**Figure 20:** Page d'inscription

### 6.2 Login

Une fois inscrit, l'utilisateur pourra ensuite accéder directement à son espace via l'authentification qui nécessite l'introduction d'un email et password.

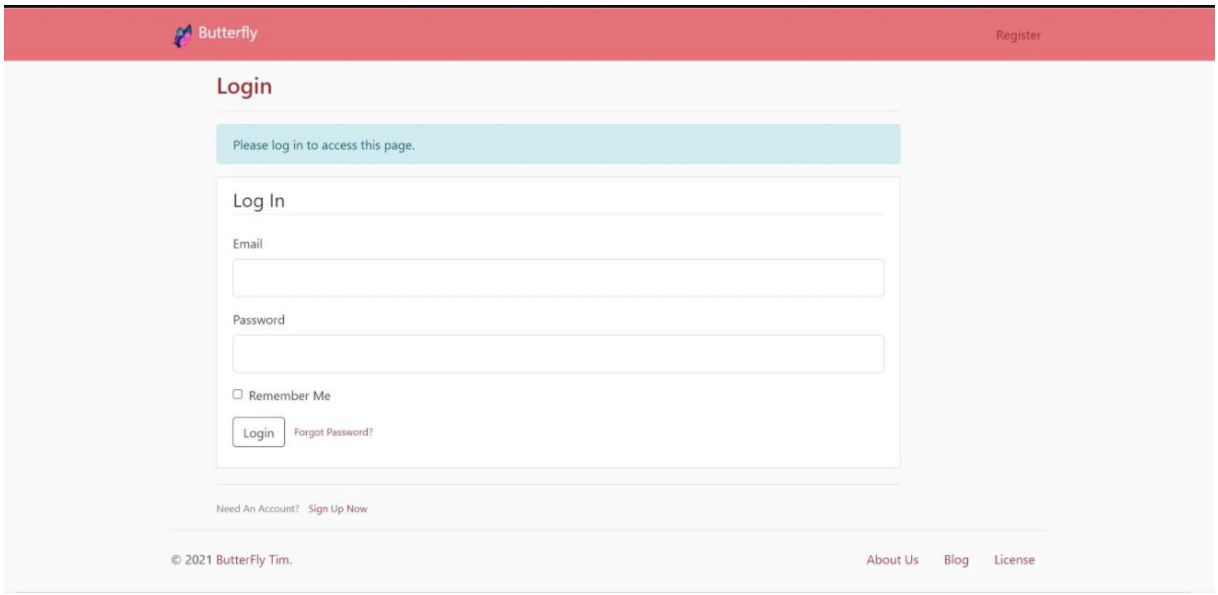


**Figure 21:**Authentification

### 6.3 Espace utilisateur

L'espace utilisateur représente comme son nom l'indique l'espace dans lequel un utilisateur se retrouve accédé à son profil et pourra ensuite interagir avec le système, notre application lui offre les fonctionnalités suivantes :

- Consultation le profil

Dans cette partie, l'utilisateur peut consulter ses informations personnelles (figure 20) et ses publications (figure 21).

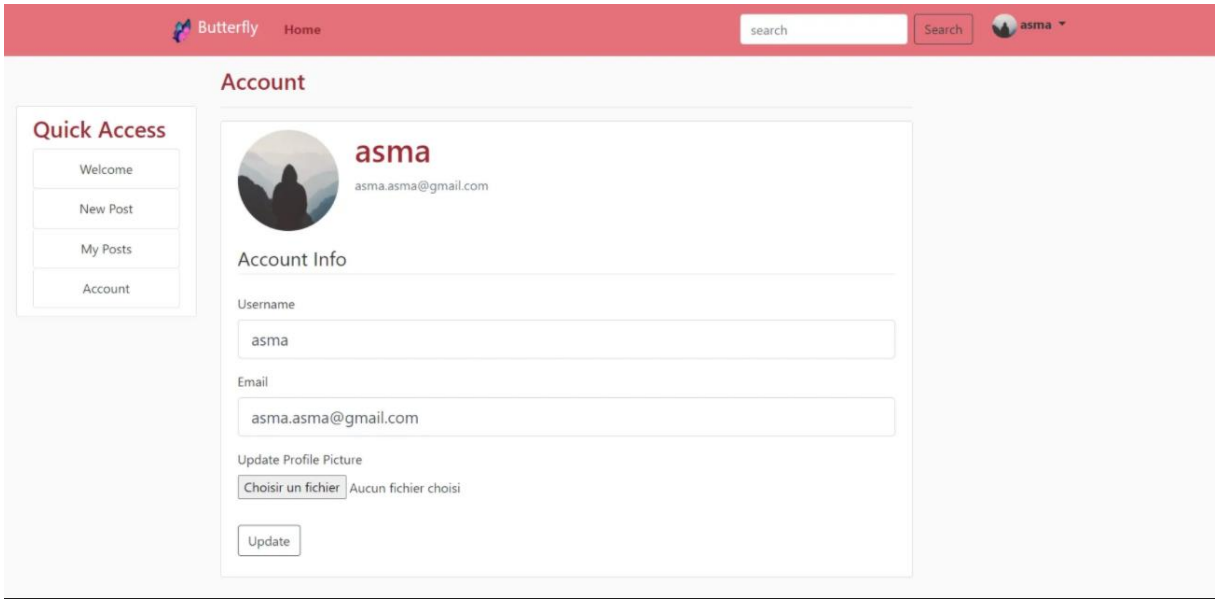


**Figure 22:** Consultation de profil

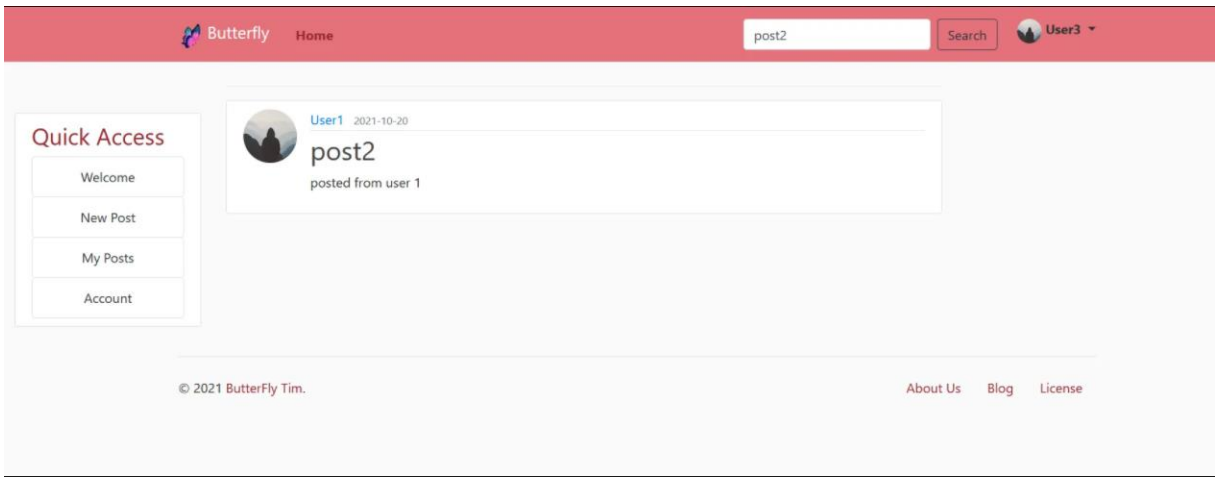


**Figure 23:** Consultation de la publication

- New post

Dans cette partie, l'utilisateur peut publier plusieurs publications.

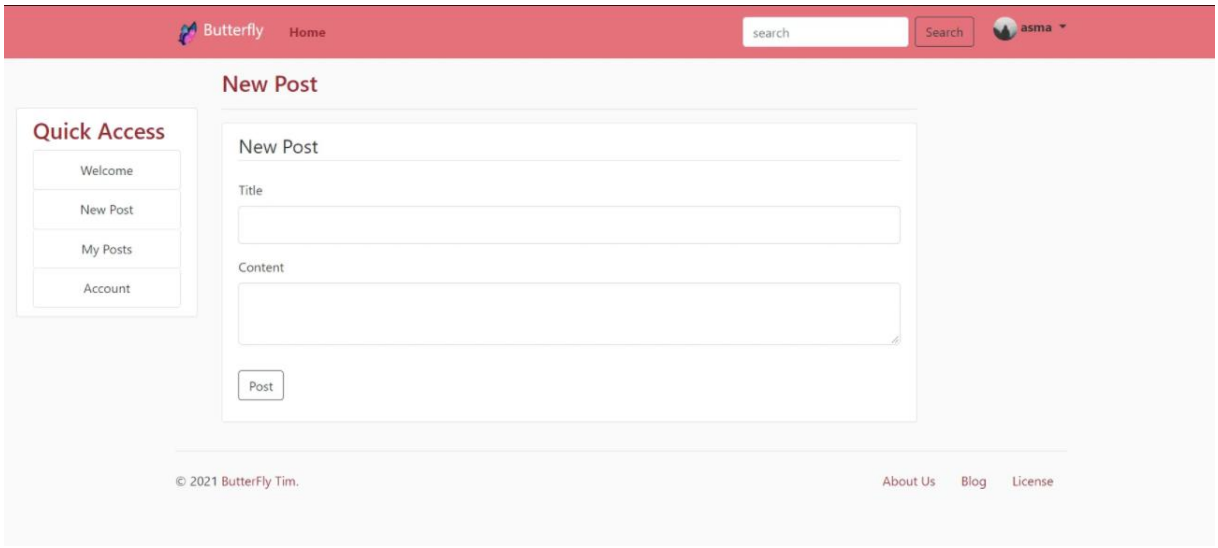


**Figure 24:** Ajout d'une nouvelle publication

- Lancement des requêtes

L'utilisateur peut exprimer ces besoins en information en émettant des requêtes.

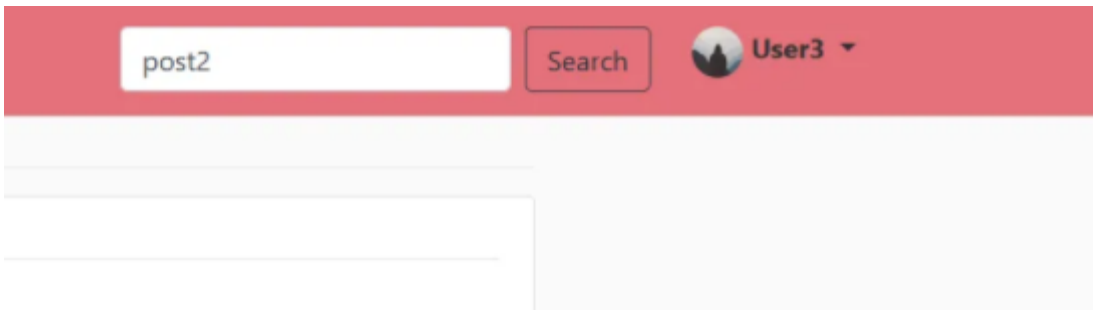


**Figure 25:** Lancement de requêtes

### 6.4 Espace administrateur

Un espace est consacré à l'administrateur, à travers lequel il pourra, d'une part, faire la prédiction, ainsi voir les centres d'intérêts des utilisateurs, d'autre part, avoir une vue globale sur tous les profils inscrits sur le site et accompagné des informations générales.

La figure (24) représente l'espace d'accueil de l'administrateur après son authentification.

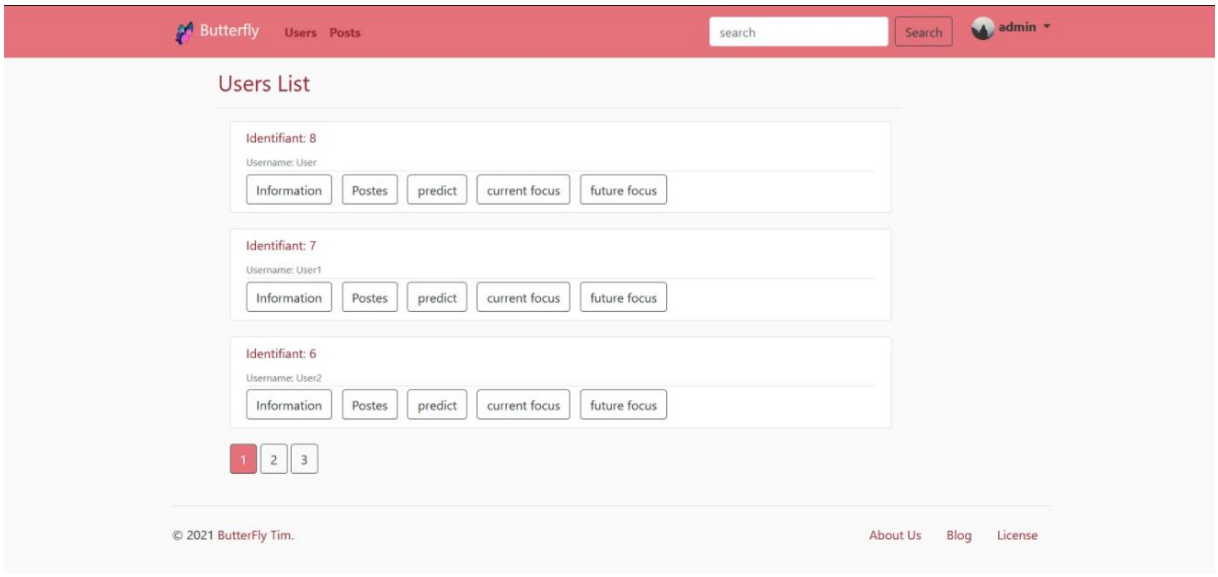


**Figure 26:** Espace administrateur

- Consultation des profils

Consultation des profils (cliquant sur le bouton Information) sur cet espace l'administrateur peut consulter les informations supplémentaires de chaque profil tel que ses informations personnelles.

- Lancement de prédiction

Lancement de prédiction (cliquant sur le bouton predict) c'est le cœur de partie administrateur de site pour lancer manuellement la prédiction des centres d'intérêt.

Une fois la prédiction effectuée, l'administrateur pourra visualiser:

La figure (25) représenteles actuels centres d'intérêts cliquant sur le bouton Current Focus.

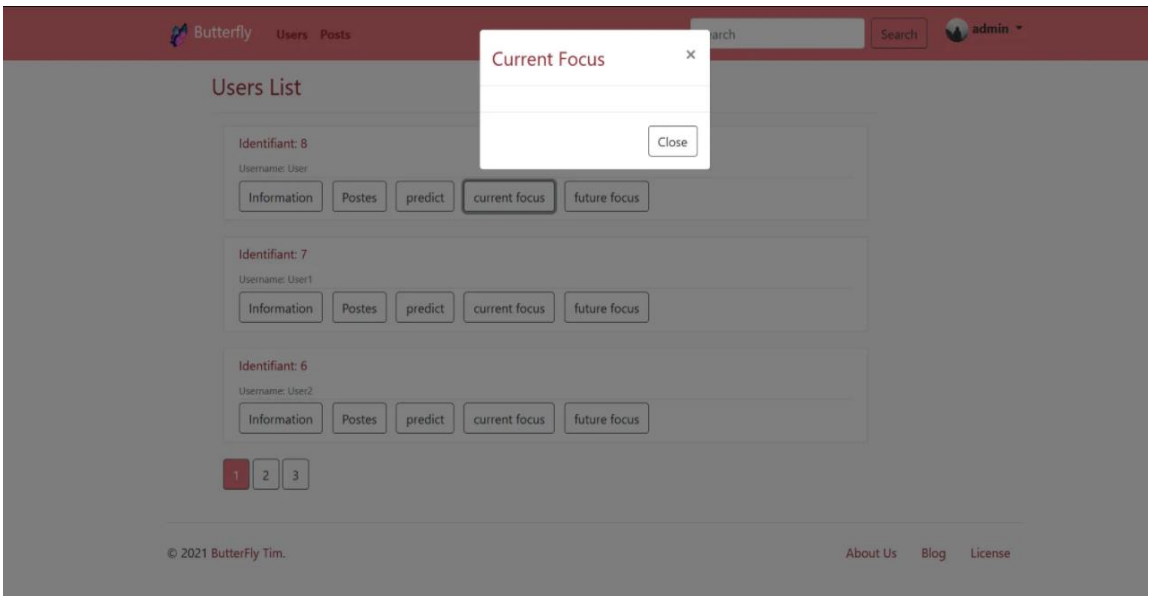


**Figure 27:** Centres d'intérêts actuels

- La figure 26les nouveaux centres d'intérêts cliquant sur le bouton Future Focus.

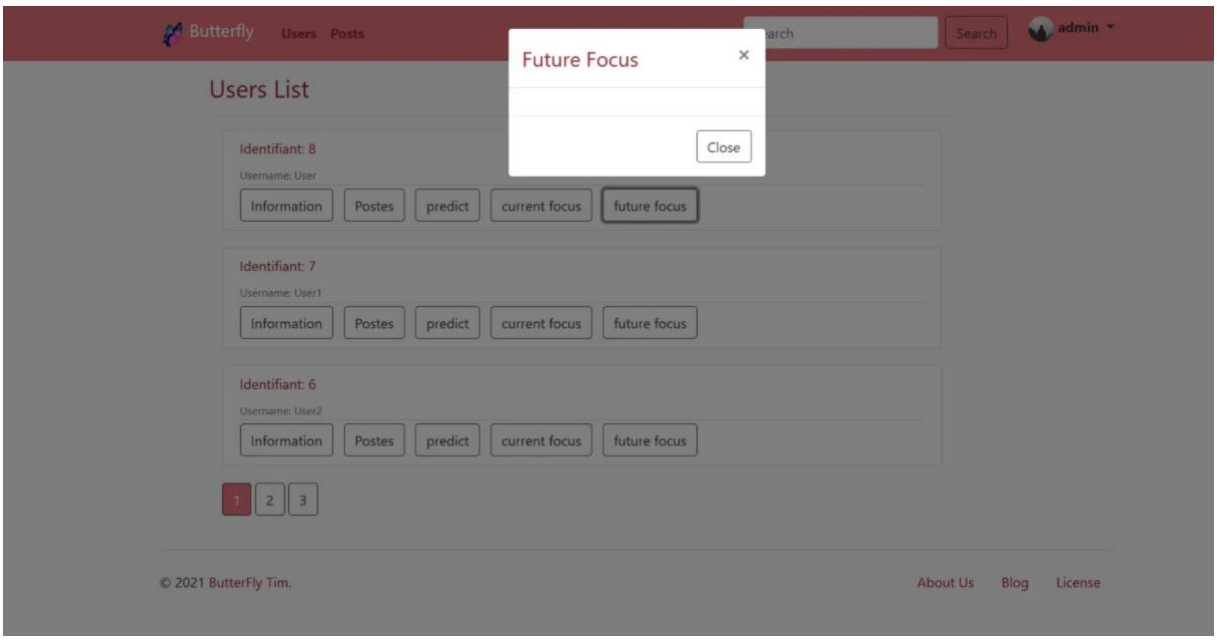


**Figure 28:** Centres d'intérêt futurs

## 7. Conclusion

Nous avons présenté à travers ce chapitre l'implémentation de notre système. Pour cela nous avons tout d'abord présenté une architecture globale et le choix logiciels et matériels ainsi que les différents outils utilisés pour atteindre notre but. En fin nous avons décrit les principales fonctionnalités de notre application, illustrant sa convivialité à travers les différentes interfaces qui la représentent.

# *Conclusion Générale*

Les intérêts d'un utilisateur changent avec le temps, en particulier dans le cas des intérêts extraits de puis les réseaux sociaux, pour un utilisateur, les intérêts qui sont extraites a une période donnée peuvent ne plus être significatifs ultérieurement. En effet dans les réseaux sociaux beaucoup de données peuvent être utilisées telles que le commentaire, la recherche, le partage, etc.

Dans le cadre de ce projet, nous nous sommes intéressés à la prédiction des intérêts des utilisateurs inactifs dans les réseaux sociaux à partir de son historique de recherche. En effet, la recherche d'utilisateur constitue une information importante qui reflète ses intérêts. A cet effet, le travail réalisé dans le cadre de ce mémoire se focalise sur des utilisateurs non actifs à travers ses centres d'intérêts.

Pour ce faire, nous avons opté pour une des techniques de prédiction connue pour sa puissance dans la classification qui est le réseaux de neurones, en élaborant une solution qui nous a permis d'effectuer la prédiction des intérêts des utilisateurs non actifs dans les réseaux sociaux.

Nous envisageons quelques perspectives qui permettent l'amélioration et la conformité de notre travail notamment:

- Améliorer notre système de prédiction par la prise en compte des centres d'intérêts actuels d'utilisateur ;
- Prédire les intérêts des utilisateurs à court terme ou langue terme ;
- Etendre de notre système par l'ajout des mots clés des ressources consultées par les utilisateurs inactifs
- Comparer les résultats retournés par notre système avec le Deep Learning LSTM.

# *Références Bibliographiques*


**[Abel et al, 2011]** Abel F, Araújo S, Gao Q, Houben G. "Analyzing Cross-System User Modeling on the Social Web". International Conference on Web Engineering, Paphos, Cyprus, 2011.

**[Azé, 2003]** Azé J,"Extractions des connaissances à partie des données numériques et textuelles ", Thèse de doctorat, Université Paris-Sud, 2003.

**[Bao et al. 2013]** Bao H, Liao S, Song S, Gao H. "Predicting the Future With Social Media", International Conference on Web Intelligence and Intelligent Agent Technology, USA, 2013.

**[Ben Sassi et al, 2007]** Ben Sassi I, Trabelsi C, Bouzeghoub, Ben Yahia S. "Recherche d'information contextuelle basée sur la prédiction des intérêts des utilisateurs et leurs relations sociales", no 1 DOI:10.3166/ISI.18.1.59-84 , Tunisien, 2013.

**[Bharati et al, 2010]** Bharati M, Ramageri B." DATA MINING TECHNIQUES AND APPLICATIONS", Indian Journal ofComputer Science and Engineering, 2010.

**[BOURSIN et al, 2011]** Boursin L, Ludovic P. "le media humain dangers et opportunités", Editions d'Organisation, 2011.

**[Boyd et al, 2007]** Boyd D, Ellison N. "Social Network Sites: Definition, History, And Scholarship", Journal of Computer-Mediated Communication, No.13, 2007.

**[Brandtzaeg, 2011]** Brandtzaeg P, "A typology of social networking sites users", International Journal of Web Based Communities, 2011.

**[Calas, 2009]** CALAS G. "Études des principaux algorithmes de data mining", EPITA, France, 2009, http://guillaume.calas.free.fr/data/Publications/DM-Algos.pdf [consulté le 28.08.2021]

**[Cantador et al, 2011]** Cantador I, Brusilovsky P, Kuflik T. "nd Workshop on Information Heterogeneity and Fusion in Recommender Systems", Proceedings of the 5th ACM conference on Recommender systems, USA, 2011.

**[Daoud, 2009]** Daoud, M. "Accès personnalisé à l'information: approche basée sur l'utilisation d'un profil utilisateur sémantique dérivé d'une ontologie de domaines à travers l'historique des sessions de recherche", thèse de doctorat, Université Paul Sabatier-Toulouse III, 2009.

**[Gharibshah et al, 2020]** Gharibshah Z ,Zhu X ,Hainline A. "Deep learning for user interest and response prediction in online display advertising", Datasci.ENG.5,http://doi.org/10.1007/s41019-019-00115-y,2020.

**[Golder et al, 2005]** Golder S, Huberman B, "The structure of collaborative tagging systems", Journal of Information Science, V 32, 2005.

**[Goodfellow, 2016]** Goodfellow I, Bengio Y, Courville A. "deep learning", Edition mit press, London, 2016

**[Han et al, 2012]** Han J, Kamber P. "Data mining: concepts and techniques", Elsevier, Etats-Unis, 2012.

**[Haykin, 1998]** Haykin S. "Neural Networks: A Comprehensive Foundation", prentice Hall, Etats-Unis, 1998.

**[Hernandez et al, 2007]** Hernandez N., Mothe J., Chrisment C., Egret D., " Modeling context through domain ontologies ", Information Retrieval, vol. 10, no 2, 2007.

**[Kechid, 2009]** Kechid S. "Intégration du modèle utilisateur dans un système de recherche d'information distribuée", Thèse de Doctorat, Algérie, 2009.

**[Koudri, 2011]** Koudri M. Modele de melange Gaussien.Application sur image cytologique , Memoire de Master : Université Abou Bakr Belkaid–Tlemcen, Algérie, 2011.

**[Lendrevie et al, 2006]** Lendrevie J, Levy J, Lindon D. "Mercator", 8ème édition, DUNOD, Paris, 2006.

**[Mezghani, 2015]** Mezghani M. "Analyse des réseaux sociaux : vers une adaptation de la navigation sociale", thèse de doctorat, Université Toulouse III - Paul Sabatier, 2015.

**[MORENO ,2021]** MORENO J." EncyclopædiaUniversalis [enligne]", URL : https://www.universalis.fr/encyclopedie/jacob-levy-moreno/. 2021.

**[Priscille et al, 2017]** Priscille R, Arnaud B. "La communication", Editions ECONOMICA, paris, 2017.

**[Ramiandrisoa et al, 2017]** Ramiandrisoa F, Mothe J. "Profil utilisateur dans les réseaux sociaux : État de l'art", Conference in: Rencontres Jeunes Chercheurs en Recherche d'Information (RJCRI 2017) in CORIA 2017, France, 2017.

**[René et al, 2001]** René L, Gilles V. "Le Data Mining", Editions Eyrolles, Mars 2001.

**[Lewenberg et al, 2015]** Lewenberg Y, Bachrach Y, Volkova S, "Using Emotions to Predict User Interest Areas in Online Social Networks", 2015 IEEE International Conference on Data science and Advanced analyticsDSAA),doi:10.1109/DSAA.2015.7344887,2015.

**[Sirinya, 2017]** Sirinya O. "Temporalité et réseaux sociaux: prise en compte de l'´évolution dans la construction du profil utilisateur". Thèse de doctorat, Université Paul Sabatier Toulouse. France, 2017.

**[Smith, 2021]** Smith M. "Atelier de 15 December: crée proper carte des réseaux socieux twitter". https://www.connectedaction.net/workshop-december-15th-create-your-own-twitter-social-network-map/,[Consulté le 12 juillet 20021].

**[Statista, 2021]** Statista, "Most popular social networks worldwide as of July 2021, Ranked by number of active users", https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/, [Consulté le 12 juillet 20021].

**[Tamine et al, 2005]** Tamine L, Boughanem M. "Accès personnalisés à l'information : Approches et techniques", Thèse de Doctorat, l'Institut de recherche en informatique de Toulouse, 2005.

**[Tchuente, 2013]** Tchuente D, Canut M.-., Jessel N, Peninou A, Sèdes F. " A community-based algorithm for deriving users' profiles from egocentrics networks : experiment on Facebook and DBLP ", Social Network Analysis and Mining, vol. 3, no 3, France, 2013.

**[Torloting, 2006]** Torloting P. "Enjeux et perspectives des réseaux sociaux", thèse de doctorat, France, 2006.

**[Trajkova et al, 2004]** Trajkova J, Gauch S, "Improving ontology-based user profiles", Coupling approaches, coupling media and coupling languages for information retrieval, Paris, 2004.

**[Tsiptsis et al, 2011]** Tsiptsis K, Chorianopoulos A. "Data mining techniques in CRM: inside customer segmentation", John Wiley & Sons, Etats-Unis, 2011.

**[Tsiptsiset al, 2010]** Tsiptsis K, Chorianopoulos A. "Data mining techniques in CRM : inside customer segmentation", Edition Wiley, Etats-Unis, 2010.

**[Valafar et al, 2009]** Valafar M., Rejaie R., Willinger W. "Beyond friendship graphs: a study of user interactions in Flickr ", Proc. of the 2nd ACM workshop on Online social networks, 2009.

**[Webself, 2021]** Webself. "Les réseaux sociaux au sein de la société", https://les-reseaux-sociaux-67.webself.net/developpement, 12 juillet 20021.

**[Zarrinkalam et al, 2019]** Zarrinkalam F, Kahani M, Bagheri E. "User interest prediction over future unobserved topics on social networks", Information Retrieval Journal, 2019.