

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE



UNIVERSITE IBN KHALDOUN - TIARET
FACULTE DES SCIENCES DE LA TECHNOLOGIE ET DE LA MATIERE
DEPARTEMENT D'INFORMATIQUE

Mémoire de fin d'études

Pour l'obtention du diplôme de Master II
Spécialité : Réseaux et Télécommunication

Thème

**Classification Simultanée du Big Data
Dans les Réseaux Sociaux**

Réalisées par :

- HIRECHE Sara
- ELASSA Nour El Houda

Encadré par :

M. Laouni Djaafri

Année universitaire 2015 / 2016

Remerciements

*Louange à **DIEU**, nôtre grand seigneur, de nous avoir éclairé le chemin du savoir, et de nous avoir donné la bonne volonté et la patience pour mener à terme ce travail. Que le salut de **DIEU** soit sur son prophète **MOHAMED**, nôtre éducateur et guide dans le bon chemin.*

Nous adressons ensuite nos vifs remerciements et nôtres sincères gratitude à :

M. DJAAFRI, d'avoir accepté nôtre encadrement, et de nous avoir bien conseillées et orientées, ce qui nous a aplani les différents obstacles rencontrés tout au long de la préparation de ce mémoire.

Par ailleurs, nous remercions nos familles, et notre guide ABD EL KADER , et nos frères et amis, et particulièrement nos parents, pour leur soutien, leur patience et leur présence indéfectible.

Enfin, nous remercions toutes les personnes qui ont contribué de près ou de loin à ce modeste travail et que nous sommes contraints de ne pas mettre leur nom.

Dédicaces

Je Dédie ce travail à :

Ma mère

Mon père, celle qui a toujours veillée sur moi et qui a été ma source d'énergie.
Je reconnais qu'elle m'a appris à assumer mes décisions et défendre mes idées.

Mes frères, et sœurs : Mohamed, Khaled, Djalal, Aya, Amel, bilal

Mes chères amies : Rachida, Khaira, Safia, Bouchra

Mon binôme et chère amie houda

Toutes les personnes que je connais et que je n'ai pas pu citées

Et à tous ceux que j'aime et qui m'aiment.

Hireche

Dédicaces

Je Dédie ce travail à :

Mon père :

Celui qui m'a appris à assumer mes décisions et défendre mes idées.

Ma mère :

Celle qui a toujours veillée sur moi et qui a été ma source d'énergie

Ma chère marie et ma petit ALAA :

Celui qui m'a aidé, encouragé et qui restait toujours à coté de moi.

*Mon cher bébé que j'aime beaucoup **ADAM,***

*Mes frères, sœurs et cousins : **Asma, Hadjer, Khaira, et ma copine Sihem***

Toute la famille « Dieb » et « Elassa ».

*Mes chères amies : **Sara, Rachida, Khaira , Safia, Bouchra , Imene, Soumia.***

Toutes les personnes que je connais et que je n'ai pas pu citées.

Et à tous ceux que j'aime et qui m'aiment.

ELASSA

Sommaire

Chapitre I : réseaux sociaux

Introduction.....	2
I.1 Histoire et diversité des réseaux sociaux	2
I.2 Qu'est-ce un réseau social.....	3
I.3 La nécessité de suivre les réseaux sociaux	4
I.4 Objectifs des réseaux sociaux.....	5
I.5 typologie des réseaux sociaux	5
I.5.1 les réseaux généraliste	5
I.5.2 les réseaux thématique	6
I.5.3 les réseaux sociaux.....	6
I.5.4 les réseaux sociaux de microbloggin	6
I.6 les fonctionnalités caractéristiques des réseaux sociaux	7
I.7 les avantages et les inconvénients des réseaux sociaux	8
I.7.1 les avantages.....	8
I.7.2 Les inconvénients	10
I.8 Quelques conseils pour les réseaux sociaux	11
I.9 Analyse sur l'avenir des réseaux sociaux.....	12
I.10 Réseau social Second Life	12

chapitre II: les méthodes de classification & BIG DATA

Introduction.....	14
II.1 L'analyse des réseaux sociaux (SNA).....	14
II.2 Définition de l'analyse de donnée	14
II.3 les différentes étapes de l'analyse des données	15
II.3.1 Décrire les données.....	15
II.3.1.1 description d'une variable	15
II.3.1.2 analyse bivariee	15
II.3.2 simplifier les données :.....	16
II.4 Comment faire l'analyse des données	16
II.5 La classification [14]	17
II.5.1 Définition	17

II.5.2 Déroulement & Fonctionnement	17
II.6 présentation des méthodes de classification de données	17
II.6.1 Formalisation générale du problème de la classification	17
II.7 Les techniques de la classification automatique.....	18
II.8 Méthodes hiérarchiques (Méthodes divisives ou agglomératives)	20
II.9Méthode divisive et segmentation non supervisée.....	20
II.10 Classification Ascendante Hiérarchique (CAH)	21
II.11 Les big data	24
II.11.1 definition	
II.11.2 Un lien entre les réseaux sociaux et Big Data	24
II.11.3 Les caractéristiques du Big Data	24
II.11.4 les technique du big data	25
II.11.5 Analyser les données.....	25
CONCLUSION	26
chapitre III: conception et implémentation	
Introduction :.....	27
III.1 Langage de modulation UML :	27
III.2 Vues statiques du système :	28
III.3 Vues dynamiques du système :	29
III.4 Conception du projet :	29
III.5 Description du cas d'utilisation « créer un compte » :.....	30
III.5.1 Sommaire d'identification :	30
III.6 Présentation de l'environnement de (ECLIPSE, PHP, MYSQL) :	33
III.6.1 MYSQL :.....	33
III.6.1.1 Qu'appelle-t-on SQL?	33
III.6.1.2 SQL est un langage de définition de données :	33
III.6.2 PHP (web) :.....	34
III.6.3 Eclipse :	34
III.7 Présentation du prototype réalisé :	34

Table de figure

I.1 : tableau qui présente des exemples sur les réseaux sociaux	11
II.1 : Techniques de classification.....	24
III.1 : Axes de modélisation des diagrammes d'UML.....	34
III.2 : Diagramme de cas d'utilisation.....	36
III.3 : diagramme de séquence.....	38
III.4 : Interface d'authentification.....	41
III.5 : Interface d'entrée.....	42

Résumé

La collecte continue des données massives circulantes dans le réseau web , conduit à des problèmes qui sont causés par les propriétés de volume, la variété et la vitesse du Big Data, ainsi le contrôle des applications de réseautage social (permettent de relier des amis, des associés, ou d'autres individus moyennant principalement des outils logiciels (sites web) .

L'objectif du travail est de développer une approche de classification simultanée basée sur les méthodes de classification non supervisée notamment la méthode de classification hiérarchique ascendante qui permet de limiter les difficultés et les problèmes posés dans les réseaux sociaux et leurs applications.

Abstract

The continuous collection of circulating massive data in the web network, leading to problems that are caused by the volume properties, variety and velocity of Big Data and control of social networking applications (used to connect friends, associates, or other individuals primarily through software tools (websites).

The aim of the work is to develop a simultaneous classification approach based on clustering methods including hierarchical cluster method which limits the difficulties and problems in social networks and their applications.

Keywords: classification, social networking, big data.

Introduction générale :

Le monde d'aujourd'hui vit une véritable révolution dans le domaine de l'informatique, surtout quand il s'agit de communiquer via Internet, on parle sérieusement des réseaux sociaux comme un outil de communication et de travail pour les particuliers ainsi que pour les entreprises.

Le Web 2.0 a amené l'évolution des réseaux sociaux. Ils sont de plus en plus utilisés par les internautes qui naviguent sur Internet, et ils touchent un public extrêmement large. Les étudiants et les adolescents ont été les premiers utilisateurs de ce genre de sites, faisant d'eux les précurseurs des réseaux sociaux actuels. Il faut également savoir qu'actuellement, pour beaucoup d'internautes, utiliser ces sites est considéré comme une activité sociale à de communication capable de changer radicalement certaines approches de l'informatique en réseau d'internet.

L'utilité de notre approche réside dans la classification de données volumineuses, de manière automatique et simultanément, basée sur des algorithmes de classification, à pour rôle de prédire afin d'écouter les usagers suivant leurs besoins et de répondre sur tous les problèmes. Ce doit être une méthode efficace pour classer, filtrer et analyser les données avant l'utilisation, En conséquence, nous devons proposer la création d'un analyseur robuste spécialisé dans les big data.

Nous avons structuré notre mémoire selon le plan décrit ci-dessous :

Le premier chapitre présente les réseaux sociaux, leurs apparitions, caractéristiques, typologies des réseaux sociaux, principe de fonctionnement, comparaison entre quelques réseaux sociaux.

Le deuxième chapitre se concentre sur le traitement du big data, en commençant par l'analyse de données et leurs méthodes notamment la classification de données, ainsi que leurs approches, et en terminant par une conclusion.

Le troisième chapitre fournira une analyse complète de notre projet de fin d'étude par une conception modélisée avec UML. Il présente aussi son implémentation sous netbeans. Finalement on terminera notre mémoire de fin d'étude avec une conclusion générale et les perspectives ouvertes par le travail présenté dans ce projet.

Introduction

Nous connaissons actuellement surtout les réseaux sociaux par des ensembles d'individus ou d'organisations comme Facebook, Twitter, LinkedIn, pour but de créer des interactions sociales entre les utilisateurs.

Ce chapitre consiste une synthèse, qui d'une part portera sur des définitions et statistique relatifs aux réseaux sociaux et l'effet de ces sites soit coté Négative ou positive.

I.1 Histoire et diversité des réseaux sociaux :

Depuis que les hommes existent, les réseaux sociaux ont été présent, des groupes sociaux se sont formés a propos de divers thèmes tels que la religion, les arts ou encore la vie de tous les jours. Ces réseaux informelles existent dès que deux personnes ou plus partage leurs avis. Tout au long de l'histoire, ces réseaux se sont multipliés et diversifié, des réunions aux thermes pendant l'époque romaines au syndicat de nos jours en passant par les salons littéraires du XVIIème siècle , les hommes ont toujours chercher a se réunir pour partager leurs opinions. A notre époque, on voit l'apparition de réseaux sociaux a caractères personnel tels que les agences de rencontrent mais aussi a caractères professionnels avec les agences de chasseur de têtes. [1]

Avec le développement d'Internet, des réseaux sociaux plus importants apparaissent, le premier d'entre eux fut classmates, qui a pour but de permettre de retrouver des anciens camarades de classes, collègue ou soldat rencontré a l'armée, il est créé en 1995 par Randy Conrads. C'est en 2001 que les réseaux en ligne sont véritablement apparue et en 2002, ils deviennent de plus en plus populaires notamment avec Friendster créée en mars de cette même année par Jonathan Abrams. Il est basé sur le concept de cercles d'amis, pour s'inscrire, il faut être invité par quelqu'un. Actuellement, le site Myspace a enregistré un plus haut taux de pages visitées que le moteur de recherche Google. Ces réseaux sociaux en ligne sont entré dans une nouvelle ère avec se qu'on appelle le Web2.0, c'est-à-dire, des sites ou l'utilisateur n'est pas passif, c'est un participants actifs du réseau. [1]

Le concept de réseau social a été réellement popularisé par l'arrivée du site Myspace en 2002, dans lequel l'utilisateur se crée un espace personnel à son image et partage ses passions avec les personnes qu'il a accepté comme « ami » pour ce site.[1]

La fin des années 80 se voit assortir d'une révolution dans le domaine de l'informatique et des développements Internet. Tim Berners-Lee, un membre du CERN de Genève, propose de développer un système hypertexte organisé en « Web » grâce à la rédaction d'un code HTML (un langage informatique permettant de rédiger des données pour représenter les pages sur internet), afin d'améliorer la diffusion des informations internes : c'est la création du « World Wide Web » (aussi appelé « WWW », « Web » ou « W3 ») et la naissance d'un tout nouveau mode de communication et d'échanges de données et d'informations.[2]

S'en est suivi la création d'une multitude de sites web (dont la plupart sont encore utilisés aujourd'hui), de l'élaboration de nouveaux langages informatiques, de la création des premiers logiciels et des premiers navigateurs tels que NCSA Mosaic, Lynx, Microsoft Internet Explorer, Netscape ou encore Mozilla. Le Web enregistre 26 sites en 1992, plus de 600 sites en 1993, plus de 10000 sites en 1994 et pas moins de 45 millions de sites 2004. Le Web est devenu une plate-forme virtuelle où les entreprises rachètent les logiciels développés par d'autres pour plusieurs milliards de dollars et où il règne une ambiance très concurrentielle.[2]

Vers le milieu des années 2000 apparaît le concept du « Web 2.0 ». Répandu par Tim O'Reilly en 2004, ce nouveau web est censé faciliter l'accès à l'information aux utilisateurs ayant peu de connaissances techniques de s'approprier les nouvelles fonctionnalités du web. En effet, les notions de communication deviennent populaires, notamment grâce à l'utilisation répandue des blogs, des sites dédiés à l'information libre appelés « pages wiki » (Encarta en 1993 puis Wikipédia en 2001) et bien entendu grâce à l'apparition des réseaux sociaux tels que Myspace en 2003, Facebook en 2004, etc. [2]

I.2 Qu'est-ce un réseau social :

Un « réseau social » est aujourd'hui défini comme étant constitué d'organisations ou d'individus reliés entre eux par des liens qui sont créés à l'occasion d'interactions sociales. Le terme réseau social (social network) a été mentionné pour la première fois par l'anthropologue anglais J. A. Barnes en 1954 (Balagué & Fayon, 2010). J.A Barnes a contribué à l'avancement des études sur les sciences sociales et la notion de « réseau », mais cet avancement n'aurait pu

être possible sans les nombreuses recherches réalisées par ses prédécesseurs au cours du vingtième siècle. [3]

De façon différente un Réseau social représente des liens sociaux, autrement dit des liens permettant des interactions sociales. Bien que médiatisés en raison de l'essor rapide de leur forme numérique sur le Web, la structure sociale que représente un réseau social existe sous une forme plus ou moins élaborée, plus ou moins dynamique et évolutive, depuis que l'homme interagit avec ses semblables. [4]

Avec précision. Christine Ballagué et David Fayon observent que : « Un réseau social permet d'être en contact avec ses amis et ses relations, de retrouver des personnes perdues de vue, de faire de nouvelles connaissances en fonction d'affinités, d'intérêts communs, de savoir ce que font vos contacts (via les newsfeed). Le fait de partager des données (informations, photos, vidéos, etc.) peut constituer aussi un moyen (réseaux de rencontres) ou une fin (réseaux dans lesquels l'internaute est actif) dans cette optique de partage.[4]

Alors les réseaux sociaux ont été un terreau fertile pour le développement de l'économie du partage, ou l'accès à prime sur la possession, la collaboration sur l'individualité.

I.3 La nécessité de suivre les réseaux sociaux :

Aujourd'hui, les réseaux sociaux sont devenus très présents, puisque la quasi-totalité des personnes de notre âge (jeune) possède un compte dans au moins un réseau social.

Un réseau social est un site internet permettant à l'internaute de s'inscrire et d'y créer une carte d'identité virtuelle. Le réseau est dit social parce qu'il permet d'échanger avec les autres membres inscrits sur le même réseau : des messages publics ou privés, des liens hypertextes, des vidéos, des photos, des jeux...

L'ingrédient fondamental du réseau social reste cependant la possibilité d'ajouter et gérer une liste de contacts. Cependant, il n'y pas que les adolescents ou les personnes de notre génération qui possèdent un compte dans les réseaux sociaux. Les adultes sont de plus en plus présents, principalement sur les réseaux sociaux spécialisés dans l'entreprise et le monde du travail.[5]

I.4 Objectifs des réseaux sociaux :

Les réseaux sociaux ont pour but de créer des interactions sociales entre les utilisateurs c'est ce qui explique leur succès :

- ✓ Augmenter votre notoriété
- ✓ Surveiller ce que les gens disent de vous
- ✓ Créer une communauté engagée et fidèle
- ✓ Augmenter votre exposition
- ✓ Ecouter les besoins et les problèmes de vos clients

I.5 typologie des réseaux sociaux :

Après avoir expliqué la notion de réseau social et les notions fondamentales de ces nouveaux médias, nous allons tenter de classer ces réseaux de manière à vous donner aidé à bien choisir le réseau qui correspondra le mieux à votre besoin; En effet, comme cela est le cas pour tous les secteurs, la croissance et la maturité s'accompagnent d'une spécialisation.

I.5.1 les réseaux généraliste : [6]

Avec plus de 3 milliards d'internautes à travers le monde, les réseaux sociaux se sont multipliés, diversifiés voire même spécialisés.

Les réseaux sociaux tels qu'on les connaît avec Facebook , Twitter et tagged, tentent à évoluer au fil des ans. Grâce à cette diversité, plusieurs autres réseaux sociaux se sont créés comme Hello, le petit frère de Twitter.

Nous avons choisi d'analyser les réseaux sociaux, principalement Facebook, assez vaste car c'est une sorte de monde virtuel extrêmement présent en ce moment, principalement pour notre génération, mais aussi chez les personnes plus âgées. Ce sujet nous intéresse beaucoup car nous aimerions en savoir plus concernant ces sites internet sur lesquels nous passons de nombreuses heures à retrouver nos amis à travers le monde.

I.5.2 les réseaux thématique : [7]

Construit autour d'une thématique particulière qui représente en général un intérêt commun pour ses membres, les réseaux sociaux spécialisés sont très nombreux et permettent d'orienter les interactions et contributions de chacun, les plus connus sont :

✓ **youtube :**

Très célèbre plateforme de publication de vidéos, Youtube est l'un des sites Internet les plus visités au monde. Il surfe sur l'évolution des habitudes de consommation des internautes de plus en plus tournés vers l'image et le son, donc la vidéo.

Youtube est aujourd'hui un média incontournable pour faire le succès sur Internet. Par contre le réseau a du mal à créer de l'interaction sociale et la qualité ou diversité du contenu qui y est publié est l'unique raison de son succès.

✓ **flicker :**

Beaucoup moins connu du grand public, Flickr est un réseau social de partage d'images, il est naturellement devenu un des sites favoris des photographes.

I.5.3 les réseaux sociaux :

Spécialement pensés pour la sphère professionnelle, ces réseaux sont de véritables outils de réseautage (ou Networking en Anglais). Ils permettent par exemple de :

- ✓ élargir ses relations d'affaires
- ✓ garder le contact avec les nouveaux et anciens collègues
- ✓ chasser les talents (recruter)
- ✓ prospecter de nouveaux clients / employeurs
- ✓ faire connaître son entreprise, marques et produits
- ✓ Les plus connus sont VIADEO en France et LINKEDIN dans le monde

I.5.4 les réseaux sociaux de microbloggin :

Après les blogs, il y a eu les micros blogs, ces messages courts qui permettent de partager rapidement et en quelques mots une idée, un avis, un évènement, twitter est

l'incarnation de cette catégorie de réseaux sociaux. Il permet via des "tweets" (messages ne dépassant pas les 140 caractères) de publier à l'ensemble de ses "suiveurs" une information en temps réel.

I.6 les fonctionnalités caractéristiques des réseaux sociaux : [8]

Les réseaux sociaux, pour favoriser l'interaction entre utilisateurs, proposent un ensemble de fonctionnalités qui permettent de les identifier :

- ✓ Un espace de présentation que l'utilisateur s'approprié. Il y précise son profil, y publie des éléments variant d'un site à l'autre : publications éditoriales, photographies, vidéos, articles, liens...
- ✓ Un outil de recherche pour identifier des membres partageant des zones d'interaction (passions, vie professionnelle, cursus scolaire, liens familiaux...) et élargir ses réseaux.
- ✓ Des solutions pour échanger et partager avec la communauté, comme par exemple une messagerie interne, ou encore la possibilité de commenter ou d'aimer une publication.
- ✓ Des outils collaboratifs élargissent souvent les possibilités d'interaction entre les membres d'un même réseau social. Des groupes permettent par exemple d'échanger sur un sujet commun, de construire un débat, tandis que les pages peuvent fédérer autour d'un projet.

Des Exemples sur les medias sociaux [9]

Réseaux sociaux	Date de fondation	Nombre d'utilisateur	objectifs
	2004	600 millions	Moyen de communiquer très efficace, très facile de trouver des membres de son entourage.
	2006	200 millions	Suivi l'actualité en direct, très rapide

	2005	1 milliards	Permet aux internautes de monde entier envoyer évaluer, regarder, commenté et partager des vidéos et aux détenteurs d'un compte youtube gratuit.
	2003	200 millions	Myspace a perdu son rôle de réseau social, et les utilisateurs discutent beaucoup moins qu'à ses débuts.
	2003	100 millions	La communauté est importante et cela permet d'avoir son CV en ligne sans avoir à créer un site Internet ou un blog.
	2004	100 millions	Peut servir de portfolio pour les photographie.une incroyable BD photos grâce aux photos libre de droit.
	2010	400 millions	Partager ses photos graphies et ses vidéos avec son réseaux d'amis. D'aimer et de laisser des commentaires sur les clichés déposés par les autre utilisateur.

Figure I.1 : tableau qui présente des exemples sur les réseaux sociaux

I.7 les avantages et les inconvénients des réseaux sociaux :

I.7.1 les avantages :

L'utilisation des réseaux sociaux englobe plusieurs avantages qui sont largement connus :

✓ Des outils de communication

Les réseaux sociaux sont rapidement devenus un outil de promotion pour les entreprises commerciales et les fournisseurs de service. Ils sont aussi des vecteurs de changement, de mobilisation. C'est en communiquant via ces

derniers que les acteurs du printemps arabe ont renversé trois gouvernements en 2011 (Tunisie, Égypte, Libye).

✓ **Des outils d'information**

C'est également via les réseaux sociaux que les entreprises de presse et les organes de communication informent leurs membres des derniers développements dans l'actualité.

✓ **Des outils de rapprochement**

Ces mêmes réseaux sociaux permettent de retrouver des relations perdues, d'anciens partenaires, des copains de classe jadis appréciés, voir des parents éloignés. Ils demeurent des outils efficaces pour obtenir des nouvelles de proches. Certains y ont même appris des naissances et des décès survenus dans leur propre famille.

✓ **Des outils de partage**

De nombreux réseaux sociaux, YouTube, Facebook et Dailymotion, et autres, permettent la diffusion des messages, des images, des vidéos...etc. qui peuvent changer le monde ou entraîner des interventions dans certains points chauds du globe.

✓ **Des outils valorisants**

Ces mêmes réseaux sociaux contribuent également, pour plusieurs, à renforcer leur ego. Ils y étalent leurs bons coups, leurs succès, y minimisent leurs échecs. Bref, les adeptes peuvent s'y montrer sous un jour avantageux, en plus de bonifier le sentiment d'appartenance à un groupe défini.

✓ **Des outils relationnels**

Chaque réseau social offre la possibilité de multiplier les relations, d'en développer de nouvelles, de fantasmer sur des individus qui acceptent nos demandes d'amitié, dénicher des amis lointains .

✓ **Un remède à la solitude**

Mal du siècle, la solitude mine le bien-être de nombreuses personnes. Or, les réseaux sociaux comblent une partie de ce vide relationnel et, parfois, freinent les pensées dépressives.

✓ **Partager avec le monde**

Sur ces sites, il est facile de partager avec les autres de vos ressources, connaissances et renseignements précieux. Vous pouvez donc aider et influencer les autres et faire une différence dans leur vie.

✓ **Aider d'autres gens**

Il peut être simple de faire une différence dans la vie des gens qui ne vont pas bien. Seulement en leur écrivant un petit message de sympathie ou d'empathie, vous pourriez contribuer à aider cette personne.

I.7.2 Les inconvénients :

Bien que plusieurs avantages découlent des réseaux sociaux, les inconvénients et les dangers qui en résultent sont aussi nombreux que variés.

Voici des exemples des principaux problèmes rencontrés :

✓ **L'intimidation**

L'intimidation trouve sa place dans les réseaux sociaux, des personnes mal intentionnées s'en servent pour harceler des collègues, des confrères ou consoeurs de classe. Jeunes et moins jeunes rédigent des messages insultants ou haineux et s'en donnent à cœur joie, sans censure.

✓ **La vengeance**

Pour certaines personnes, ces réseaux sont parfois utilisés comme outil de vengeance.

✓ **La vie professionnelle**

Pour de nombreux employeurs, ces réseaux sont une véritable mine de renseignements pour recruter ou congédier du personnel. Quelques photos

intimes suffisent parfois à écarter un candidat. Dans la même veine, une réflexion désobligeante sur votre travail peut vous valoir une réprimande, ou un congédiement.

✓ La vie privée

Informez vos « amis » de certaines facettes de votre vie privée peut entraîner des problèmes.

✓ La cybercriminalité

Les pirates informatiques sont passés maîtres dans le vol d'identité via les réseaux sociaux. Certains s'en servent pour avoir accès à vos renseignements personnels.

D'autres pénètrent dans votre ordinateur pour voler des mots de passe et des numéros importants : assurance sociale, carte bancaire, cartes de crédit, etc.

Et plus encore :

- ✓ Envoi de virus sur des comptes.
- ✓ hacking de comptes pour des sociétés ou gros organismes.
- ✓ vente des profils ou informations contre de l'argent.

I.8 Quelques conseils pour les réseaux sociaux :

Plusieurs précautions sont à prendre sur les réseaux sociaux pour éviter les risques que ceux-ci peuvent engendrer. Voici donc quelques conseils de sécurité :

- ✓ Réglez les paramètres de confidentialité afin de protéger votre identité.
- ✓ Choisissez soigneusement les personnes que vous acceptez comme amis ;
 - ↳ Si jamais vous acceptez une personne comme ami, cette personne a accès à toutes les informations que vous partagez avec vos amis. Par contre, il est possible à tout moment de supprimer une personne afin qu'elle n'est plus accès à vos informations.

- ✓ Montrez à certains amis une version limitée de votre profil;
 - ↳ Vous pouvez choisir de limiter l'accès qu'ont certains amis à votre profil (« profil limité »). Cette option est utile si vous avez des collègues auxquels vous ne désirez pas attribuer le statut complet d'ami ou avec lesquels vous ne souhaitez pas échanger certaines informations personnelles.[10]

- ✓ Donnez le minimum d'informations personnelles sur internet.
- ✓ Réfléchir avant de publier quelque chose.
- ✓ Se déconnecter avant de quitter l'ordinateur, car d'autres personnes pourraient avoir accès à votre compte, alors elles pourraient modifier votre profil ou publier des choses non désirables.
- ✓ Respectez les autres! Modérez vos propos dans vos publications.
- ✓ Ne jamais publier de photos gênantes de toi ou des autres.
- ✓ Vérifiez ce que les autres publient de toi sur le web.
- ✓ Utilisez un pseudo si possible.
- ✓ Ne pas choisir un mot de passe trop simple.
- ✓ Ne pas communiquer votre mot de passe.

I.9 Analyse sur l'avenir des réseaux sociaux :

Difficile aujourd'hui de dire quel avenir connaîtront les réseaux sociaux étant donné leurs jeunes âges et l'évolution rapide des nouvelles technologies. Actuellement les réseaux sociaux s'invitent déjà sur certains téléphones portables avec des services dédiés comme l'application Facebook sur l'Iphone de Apple. L'évolution logique voudrait que l'immersion soit plus totale avec une interface plus développée et pourquoi pas, un monde virtuel en trois dimensions dans lequel on se baladerait et discuterait avec ses amis.[11]

I.10 Réseau social Second Life : [11]

Ce concept existe déjà, il s'appelle "Second Life". Il s'agit d'un univers virtuel en 3D sorti en 2003 qui permet à l'utilisateur (le résident) de vivre une "seconde vie" virtuelle. Il est possible de se balader dans ce monde, de discuter avec les personnes que l'on rencontre, mais

aussi de se marier, de créer une entreprise ou de s'acheter des vêtements. Il existe un système monétaire, une monnaie virtuelle que l'on peut convertir en réels dollars américains.

Certains résidents y ont gagné des centaines de milliers de vrais dollars par exemple dans l'immobilier virtuel en vendant des parcelles de terrain ou des appartements. Second Life n'est donc pas un jeu, mais plutôt une expérience visant à être aussi variée que la vraie vie.

A son apogée, Second Life a séduit jusqu'aux partis politiques qui ont investi dans la construction d'espaces réservés aux membres avec la possibilité de se renseigner sur les prochaines votations, de débattre sur des sujets d'actualité etc.

Conclusion :

En conclusion puisque les réseaux sociaux jouent un rôle important dans les rapports sociaux des individus, malgré que nous avons montré les facettes négatives de ce système, à cause de les réseaux sociaux sont utilisés presque partout, à la maison, à l'école, au travail, avec les téléphones cellulaires, etc. Au sein du chapitre suivant nous allons parler du Big Data et les méthodes de classification.

Introduction

Dans ce chapitre on y insiste sur le traitement pratique de données, ce terme nous conduit directement sur la notion de classification qui est une étape relative à l'analyse de données et qui est consacré à la méthodologie statistique avec ses différentes techniques.

II.1 L'analyse des réseaux sociaux : (SNA)

Est une approche à la fois graphique et analytique consistant en une représentation cartographique d'entités reliées entre elles. Les entités peuvent être de toutes sortes, des individus, des entreprises, des groupes, des téléphones ou des ordinateurs etc..... Ces entités composent les nœuds d'un diagramme rattachés par des liens qui représentent les relations ou des flux entre chacun de ces nœuds. [12]

L'analyse des réseaux sociaux repose sur la théorie des graphes et d'algorithmes d'optimisation permettant d'évaluer la position relative de chaque entité au sein d'un réseau et la densité des liens entre ces dernières. On mesure la centralité des nœuds. Cet indicateur permet de connaître le rôle d'une entité et d'évaluer la proximité entre ces dernières pour déterminer des communautés.

II.2 Définition de l'analyse de donnée :

L'analyse des données est un ensemble de méthodes mathématiques qui permettent de traiter une masse d'information sous forme d'un tableau.

C'est un outil de visualisation. On cherche par ces méthodes à donner les liens pouvant exister entre les différentes données et à en tirer une information statistique qui permet de décrire de façon plus succincte les principales informations contenues dans ces données.

On peut également chercher à classer les données en différents sous-groupes plus homogènes.

L'élément fondamental de l'analyse de données est le tableau de données. Un tableau de donnée est en général à deux dimensions. Les plus fréquents des tableaux sont les tableaux

individu(Xvariables).

Chaque ligne d'un tel tableau contient les informations relatives à un individu. [13]

II.3 les différentes étapes de l'analyse des données :

II.3.1 Décrire les données :

La description des données est une étape importante de la démarche d'analyse. La plupart des enquêtes se limitent à cette étape, qui donne un premier niveau de lecture des résultats ou l'identification de certaines relations entre des variables de l'étude. [13]

II.3.1.1 description d'une variable

On appelle « variable » l'ensemble des valeurs observées sur les différents individus pour une caractéristique donnée ;

- ✓ décrire une variable qualitative.
- ✓ décrire une variable quantitative.

Une variable est qualitative dès lors qu'elle a pour valeur des modalités ; elle peut être nominale (lorsque l'ensemble des modalités ne possède pas de structure particulière) ou ordinale (lorsque l'ensemble des modalités est ordonné).

Une variable est considérée comme quantitative ou métrique lorsque ses modalités peuvent être mesurées (par exemple l'âge, la valeur d'une action, etc.).

II.3.1.2 analyse bivariée

L'examen de variables uniques permet une première lecture intéressante des résultats mais elle ne présente pas de véritable intérêt en termes d'analyse. Les descriptions faites sur les variables soulèvent toute une série de questions sur leurs relations, qui devront être mises en lumière en les rapprochant deux à deux dans des analyses bivariées.

Analyse bi variée pour objectif major :

- tester la signification, ou dissymétries entre les deux variables.

II.3.2 simplifier les données :

La simplification ou l'agrégation des données est fondamentale, elle sert à identifier les différentes dimensions d'un concept. par exemple L'analyse factorielle est utilisée pour décrire les données en un nombre agrégé de facteurs. Elle traduit une matrice de nombres difficile à lire par une série de tableaux plus simples, représentés sous forme de graphiques. [13]

II.4 Comment faire l'analyse des données :

La statistique classique nous a habitués à étudier les variables les unes après les autres, de construire autant d'histogrammes que des variables.

- ✓ découvrir la structure, éventuellement compliquée, d'un tableau de nombres à plusieurs dimensions et de traduire par une structure plus simple et qui la résume au mieux.
- ✓ Représenter la structure la plus simple graphiquement,
- ✓ décrire, réduire, classer et clarifier les données en tenant compte de nombreux points de vue et d'étudier,
- ✓ extraire les grands traits, les liaisons, les ressemblances ou les différences entre les variables ou groupes de Variables.

L'analyse des données recouvre principalement deux ensembles de techniques:

- les premières qui relèvent de la géométrie euclidienne et conduisent à l'extraction de valeurs et de vecteurs propres, sont appelées « analyses factorielles ».
- Les secondes, dites de « classification automatique » sont caractérisées par le choix d'un indice de proximité et d'un algorithme d'agrégation ou de désagrégation qui permettent d'obtenir une partition ou arbre de classification.

Parmi ces deux techniques, les premières occupent une place de choix, car elles sont utilisées soit seules, soit conjointement avec les secondes, alors que ces dernières sont rarement appliquées seules.

On s'intéressera surtout dans les paragraphes suivants à la classification automatique.

II.5 La classification : [14]

II.5.1 Définition :

La classification est une méthode d'analyse des données qui vise à regrouper en classes homogènes un ensemble d'observations. Le but de La classification est de diviser un ensemble des données en plusieurs classes homogènes.

- Le nombre des groupes est parfois fixé.
- objets d'un même groupe se Ressemblent le plus possible.
- Deux objets de groupe distincts différent le plus possible.

↳ Cette tâche n'est pas facile en cas des données complexes.

II.5.2 Déroulement & Fonctionnement : [15]

C'est une méthode d'analyse de données.

- ✓ Regrouper des objets similaires selon certains critères. Les diverses techniques de classification visent toutes à répartir n individus, caractérisés par p variables X_1, X_2, \dots, X_p en un certain nombre de sous-groupes aussi homogènes que possible, chaque groupe étant bien différencié des autres.
- ✓ Produit une série de groupe «clusters» emboîtés soit par agglomération des individus par la méthode de CHA (Classification Hiérarchique Ascendante) soit par division du tout : CHD (Classification Hiérarchique Descendante) soit par d'autres méthodes.

II.6 présentation des méthodes de classification de données :

II.6.1 Formalisation générale du problème de la classification :

- Soit I un ensemble des individus (objet) ; $I = \{W_1, W_2, \dots, W_n\}$.

↳ le problème consiste à réaliser une partition de $P = \{C_1, C_2, \dots, C_k\}$

Telle que : $C_1 \cup C_2 \dots \cup C_k = P$.

↳ L'intersection entre $C_i, C_j =$ un ensemble vide avec $i, j = 1, 2, \dots, k$.

↳ Suivant la nature de problème posé, on se fixe à l'avance ou non le

nombre k des classes C_i désirée.

II.7 Les techniques de la classification automatique :

Il existe des différentes techniques de classification :

➤ Algorithme hiérarchique :

- classification hiérarchique Ascendante (CHA) avec (CURE, ROCK, BIRCH).

- classification hiérarchique Descendante (CHD) avec (Williams and Lambert, TSVQ).

➤ Algorithme par partition :

- K-means (centres mobiles, Nuées Dynamiques)

- K-médoides (PAM, CLARA, CLARANS)

- TEXY

➤ Algorithme basés sur la densité :

- Approche basée sur la connexité de densité (OPTICS, DBSCAN, DBCLASD).

- Approche basée la fonction de densité (DENCLUE).

➤ Algorithme de grille :

(STING, CLIQUE, Wave Cluster)

➤ Autres méthodes

(COBWEB, ANTCLASS)

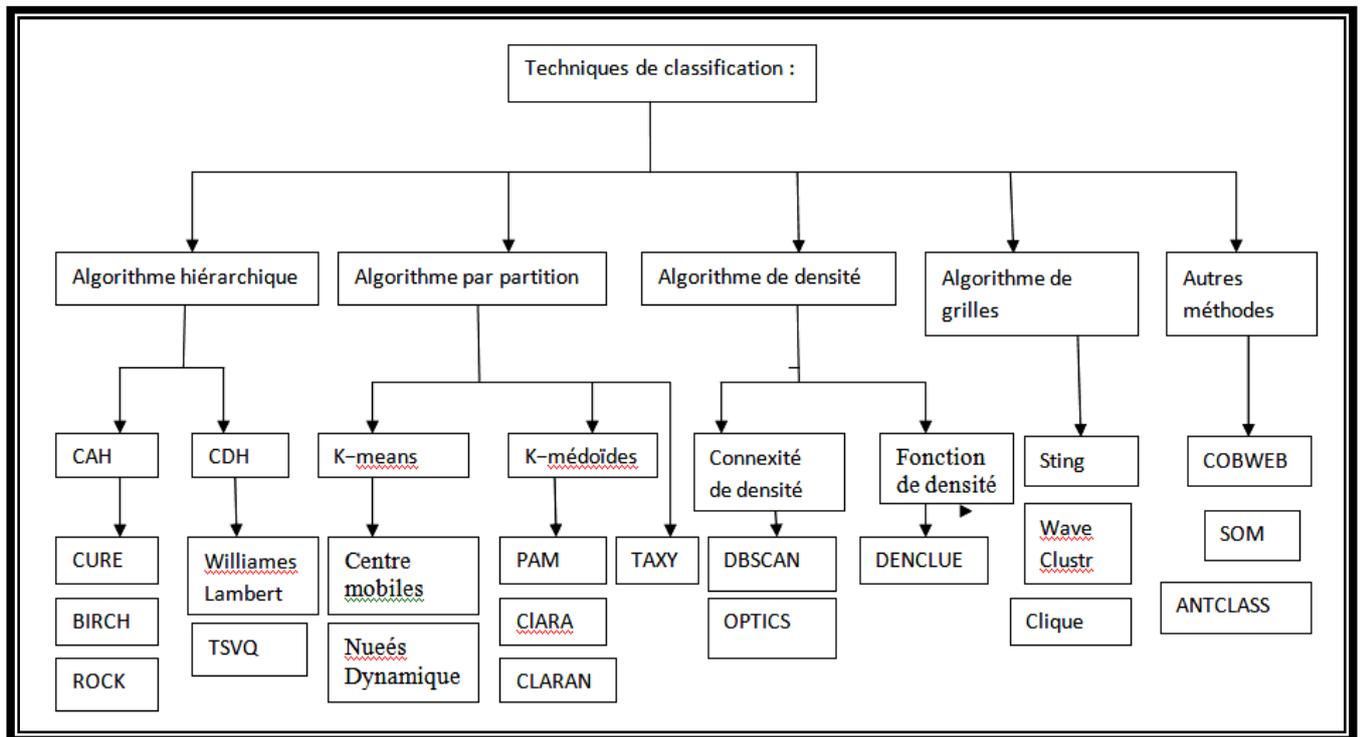


Figure II.1 : Techniques de classification.

Après avoir exposé quelques définitions sur la classification des données de façon générale, nous allons présenter et détailler dans le paragraphe suivant les méthodes hiérarchiques notamment « classification hiérarchique Ascendante(CHA) ».

II.8 Méthodes hiérarchiques (Méthodes divisives ou agglomératives) :

En classification hiérarchique, on peut procéder par une méthode soit descendante, soit ascendante.

Dans une méthode descendante ou (divisive), on part de l'ensemble de tous les éléments que l'on fractionne en un certain nombre de sous-ensembles. Ces derniers sont eux-mêmes fractionnés récursivement jusqu'à ce que l'on arrive aux éléments individuels :

(Règne → Embranchement → Classe → ... → Genre → Espèce).

Au contraire, dans une méthode ascendante ou agglomérative, on part des éléments individuels que l'on regroupe en sous-ensembles avant d'appliquer récursivement les regroupements sur les sous-ensembles jusqu'à obtenir un seul ensemble contenant tous les éléments :

(Espèce → Genre → ... → Embranchement → Règne) .

II.9 Méthode divisive et segmentation non supervisée :

DIV : une méthode divisive Cette méthode (Chavent 1997, 1998) divise à chaque étape une classe en fonction d'une question binaire et du critère d'inertie.

Cette méthode de classification construit sa hiérarchie dans le sens inverse, en commençant par une grande classe contenant tous les objets. A chaque étape, elle divise une classe en deux classes

plus petites jusqu'à ce que toutes les classes ne contiennent qu'un seul individu.

Ceci veut dire que pour n individus, la hiérarchie est construite en $n - 1$ étapes.

- ✓ Dans la première étape, les données sont divisées en deux classes au moyen des dissimilarités.
- ✓ Dans chacune des étapes suivantes, la classe avec le diamètre le plus grand

se divise de la même façon.

- ✓ Après $n - 1$ divisions, tous les individus sont bien séparés.

Remarque :

Dans le cas d'une variable qualitative non ordonnée, on se heurte vite à un problème de complexité, le nombre de dichotomies du domaine d'observation étant alors égal à $2^m - 1$.

II.10 Classification Ascendante Hiérarchique (CAH) :

Le but du module de CAH est d'obtenir une classification automatique de l'ensemble d'individus.

Elle commence par :

- ✓ déterminer parmi les n individus, quels sont les 2 individus qui se ressemblent le plus par rapport à l'ensemble des p variables spécifiées.

Elle va alors :

- ✓ regrouper ces 2 individus pour former une classe. Il existe donc à ce niveau ($n - 1$) classes, les autres ne contenant qu'un unique individu.
- ✓ Le processus se poursuit en déterminant quelles sont les 2 classes qui se ressemblent le plus, et en les regroupant. Cette opération est répétée jusqu'à l'obtention d'une unique classe regroupant l'ensemble des individus.

Cette procédure est basée sur 2 choix :

- ✓ La détermination d'un critère de ressemblance entre les individus. La méthode laisse à l'utilisateur le choix de la dissimilarité.
- ✓ La détermination d'une dissimilarité entre classes : appelé un critère d'agrégation.

Les critères d'agrégation :

De nombreux critères d'agrégation les plus connus sont :

- ✓ Le critère du saut minimal : La distance entre 2 classes $C1$ et $C2$ est définie par la plus courte distance séparant un individu de $C1$ et un individu de $C2$.

$$D(C1, C2) = \min(\{d(x, y)\}, x \in C1, y \in C2)$$

- ✓ Le critère du saut maximal : La distance entre 2 classes C1 et C2 est définie par la plus grande distance séparant un individu de C1 et un individu de C2.

$$D(C1, C2) = \max(\{d(x, y)\}, x \in C1, y \in C2)$$

- ✓ Le critère de Ward : Ce critère ne s'applique que si on est muni d'un espace euclidien. La dissimilarité ρ entre 2 individus doit être égale à la moitié du carré de la distance euclidienne d . Le critère de Ward consiste à choisir à chaque étape le regroupement de classes tel que l'augmentation de l'inertie intra-classe soit minimal.

$$D(C1, C2) = n_{C1}n_{C2}$$

Avec :

g_{C1} : Le centre de gravité de C1

g_{C2} : Le centre de gravité de C2

- ✓ Le critère des centres de gravité : La distance entre 2 classes C1 et C2 est définie par la distance entre leurs centres de gravité.

$$D(C1, C2) = d(g_{C1}, g_{C2})$$

La difficulté du choix du critère d'agrégation réside dans le fait que ces critères peuvent déboucher sur des résultats différents. Le critère le plus couramment utilisé est celui du Ward.

Nous présentons dans la suite quelques algorithmes basés sur le principe de CAH :

- ✓ Clustering Using REpresentatives(CURE)
- ✓ Balanced Iterative Reducing and Clustering using Hierarchies(BIRCH)
- ✓ Robust Clustering using links (ROCK)

Exemple : Algorithme CURE

Son nom vient de Clustering Using REpresentatives, il a été proposé par « Guha et al » , cet algorithme utilise un échantillon représentatif de l'échantillon total pour réduire la complexité temporelle des calculs. Cet échantillon sera divisé en sous-ensembles qui sont regroupés en sous-classes. Les sous-classes seront agrégées hiérarchiquement en utilisant la

distance entre deux sous-classes C1 et C2, la plus petite distance entre un représentant de C1 et un représentant de C2 jusqu'à obtenir k de classes demandées.

↳ Le détail de cette méthode est décrit dans l'algorithme suivant :

Données : k le nombre maximum de classes désiré.

Début

(1) Extraire un échantillon « Is » de taille « s » de l'ensemble « I » d'individus

(2) Diviser l'échantillon « Is » en « p » sous ensemble de taille « s/p »

(3) Déterminer une partition partielle « q » de chaque sous-ensemble en « s/pq » sous-classes avec $q > 1$

(4) Eliminer les sous-classes d'effectif faible.

(5) Déterminer dans chaque sous-classes un ensemble de « c » points

(6) Agréger de façon hiérarchique les sous-classes

$d(C1, C2) = \min[d(x, y); \forall x \text{ représentant de } C1 \text{ et } \forall y \text{ représentant de } C2]$

(7) Arrêter la procédure d'agrégation quand on obtient k classes.

(8) Classer l'ensemble « I » total en utilisant les « c » points représentant les k classes obtenues après l'agrégation. Chaque objet est affecté a la classe possédant le représentant qui lui est le plus proche.

FIN

II.11 Les Big Data :

II.11.1 définition :

Les Big Data, littéralement les « grosses données », ou « Méga données », parfois appelées données massives, désignent des ensembles de données qui deviennent tellement volumineux qu'ils en deviennent difficiles à travailler avec des outils classiques de gestion de base de données ou de gestion de l'information.[16]

II.11.2 Un lien entre les réseaux sociaux et Big Data :

Un réseau social est à la base défini comme un ensemble d'identités sociales, telles que des individus ou organisations, reliées entre elles par des liens créés lors d'interactions sociales.

L'émergence des réseaux sociaux est donc liée aux nouvelles révolutions technologiques et techniques.

C'est donc la problématique que se posent les entreprises sur les enjeux du traitement des données des différents réseaux. Il s'agit donc de savoir si le « Big Data » et les réseaux sociaux sont une équation gagnante pour l'entreprise.

Dans un premier temps nous verrons l'aspect de ces nouvelles sources de données, avec les nouveaux enjeux technologiques et les ouvertures économiques.

Dans un second temps, nous verrons que les réseaux sociaux sont une approche complexe, avec des données difficile d'accès.

II.11.3 Les caractéristiques du Big Data : [17]

Big Data se caractérise par la problématique des 3V

✓ **Volume :**

La quantité de données générée est en pleine expansion et suit une loi quasi exponentielle. Le commerce électronique et les réseaux sociaux sont les grands contributeurs de cette profusion de données.

✓ **Velocity (rapidité) :**

La rapidité de renouvellement des données dans un monde connecté n'est plus à démontrer. Toutes les nouveautés et mises à jour sont stockées en respect de la devise du big data : "On garde tout!".

- ✓ Variety (variété) :

Les bases de données spécialisées sont en mesure de gérer la multiplicité des formats des données : numérique, texte, image.

II.11.4 les technique du Big Data : [17]

Le coût des technologies a drastiquement chuté. La mise en œuvre des solutions de stockage et de traitements massivement "parallélisé" pour traiter les "grandes" données, ainsi que le temps de traitement sur des bases de données géantes, plusieurs solutions peuvent entrer en jeu :

- ✓ Des bases de données NoSQL (comme MongoDB, Cassandra ou Redis) qui implémentent des systèmes de stockage considérés comme plus performants que le traditionnel SQL pour l'analyse de données en masse (orienté clé/valeur, document, colonne ou graphe).
- ✓ Des infrastructures de serveurs pour distribuer les traitements sur des dizaines, centaines, voire milliers de nœuds. C'est ce qu'on appelle le traitement massivement parallèle. Le framework Hadoop est sans doute le plus connu d'entre eux. Il combine le système de fichiers distribué HDFS, la base NoSQL HBase et l'algorithme MapReduce.
- ✓ Le stockage des données en mémoire (Memtables) permet d'accélérer les temps de traitement des requêtes.

II.11.5 Analyser les données :

Bien évidemment, stocker une telle quantité de données n'a d'autre finalité que de tirer des enseignements pertinents afin d'accroître notre connaissance sur notre environnement professionnel en l'occurrence. Les finalités des analyses big data sont multiples. Que ce soit pour des analyses prospectives afin de mieux comprendre les attentes des clients, la modélisation pour mieux fixer le prix d'une nouvelle gamme.

CONCLUSION

L'état des lieux que nous avons mené dans ce chapitre nous a permis de cerner l'apport de l'utilisation des méthodes de classification pour les Big Data.

Cependant, vu la diversité des méthodes, on regardera comment se présentent les résultats pour l'une d'entre elles : la classification hiérarchique ascendante, qui est la plus élaborée des méthodes de classification.

Dans le chapitre suivant en va faire une conception et une implémentation sur notre application.

Introduction :

Pour une meilleure compréhension du travail effectué, nous présentons dans ce chapitre l'étape d'analyse, conception et de réalisation. Nous commencerons par une brève présentation du langage de modélisation utilisé (UML). Ce dernier présente les diagrammes de conception adoptés lors de la phase d'élaboration.

En fin, nous décrirons les fonctionnalités qu'offre notre application, son architecture, les outils de développement que nous avons utilisés, ainsi qu'une vue générale sur l'application.

En ce qui concerne notre maquette logicielle, nous avons choisi MYSQL pour abriter la base de données, PHP, et ECLIPSE comme langage de programmation.

III.1 Langage de modulation UML : [18]

UML est un langage de modélisation objet, permettant de représenter, communiquer les divers aspects d'un système d'information. Il définit plusieurs diagrammes qui tournent autour de trois axes de modélisation :

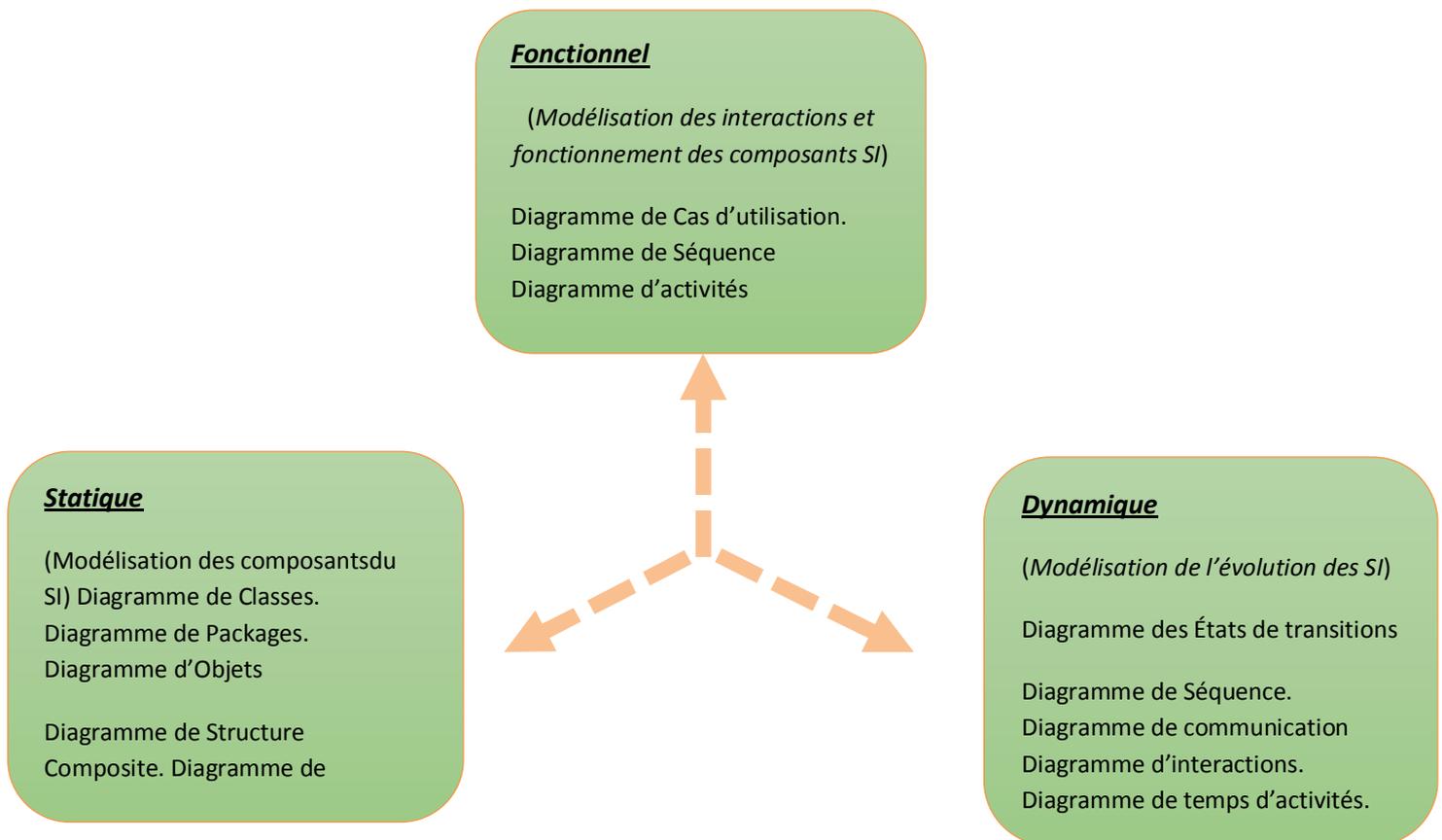


Figure III.1 : Axes de modélisation des diagrammes d'UML

III.2 Vues statiques du système : [1]

- Diagramme de cas de classes
- Diagramme de packages
- Diagramme d'objets.
- Diagramme de cas de classes.
- Diagramme de composants
- Diagramme de déploiement.

III.3 Vues dynamiques du système : [18]

- Diagramme des cas d'utilisation (use cases)
- Diagrammes de collaboration
- Diagramme de séquence
- Diagramme des états de transition
- Diagramme d'activité
- Diagramme de communication
- Diagramme d'interaction
- Diagramme de temps

III.4 Conception du projet :

Bien que le formalisme UML fournit jusqu'à présent 13 diagrammes, Dans ce qui suit nous présentons le diagramme de cas d'utilisation et diagramme de séquences.

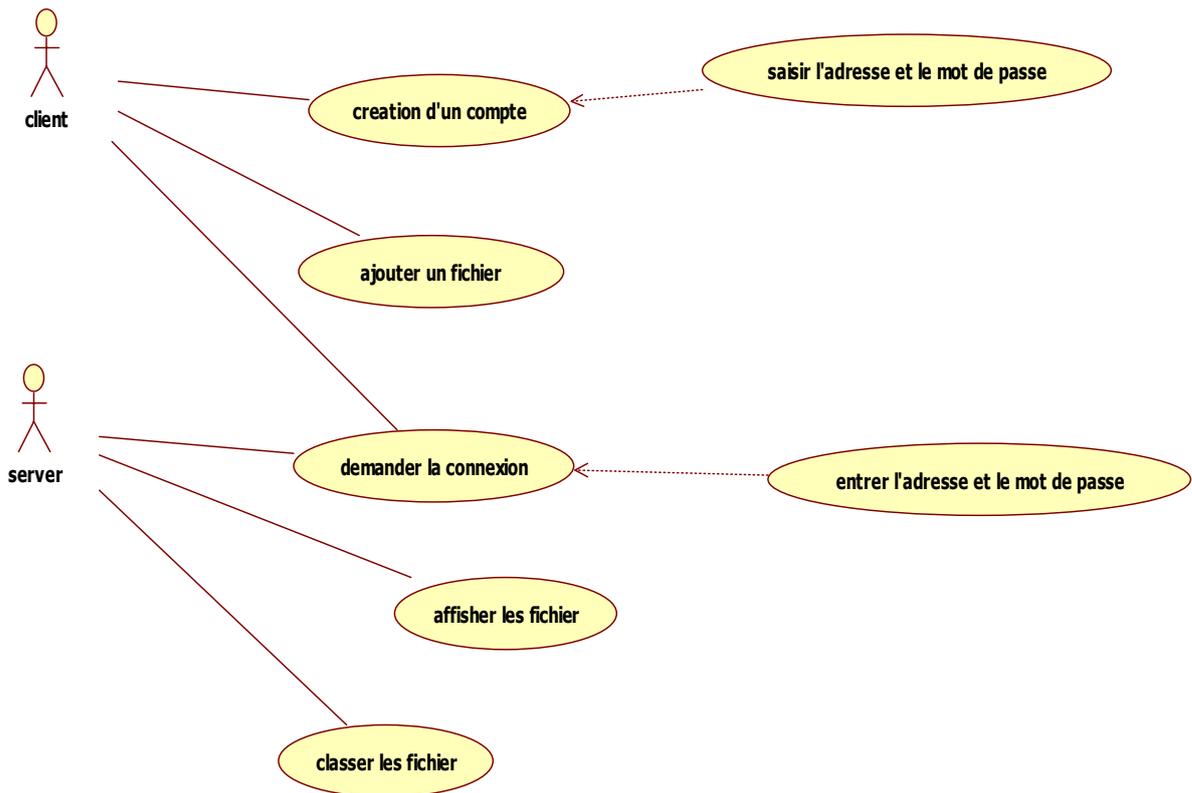


Figure III.2 : Diagramme de cas d'utilisation

III.5 Description du cas d'utilisation « créer un compte » :

Pré conditions : l'utilisateur souhaite connecter à l'interface client.

III.5.1 Sommaire d'identification :

✓ Scénario principal

1. l'utilisateur saisir l'adresse et le mot de passe
2. l'interface client vérifie l'adresse et le mot de passe.
3. L'utilisateur est authentifié et accède au système.

✓ Scénario alternatif

1. Erreur d'authentification
2. le système affiche un message d'erreur (adresse /mot de passe incorrect)

3. l'utilisateur quitte l'interface client

- ✓ Pré-conditions : La connexion est établit, l'utilisateur est identifié

Description des enchainements :

Titre : Saisir l'adresse et le mot de passe.

Résumé : Ce cas permet au l'utilisateur de créer un compte.

Acteurs1 :l'utilisateur (client).

1. choix de fichier (image, vidéo).
2. ajouter le fichier.

Acteur2 : server

1. afficher les fichiers de client
2. classifier les fichiers

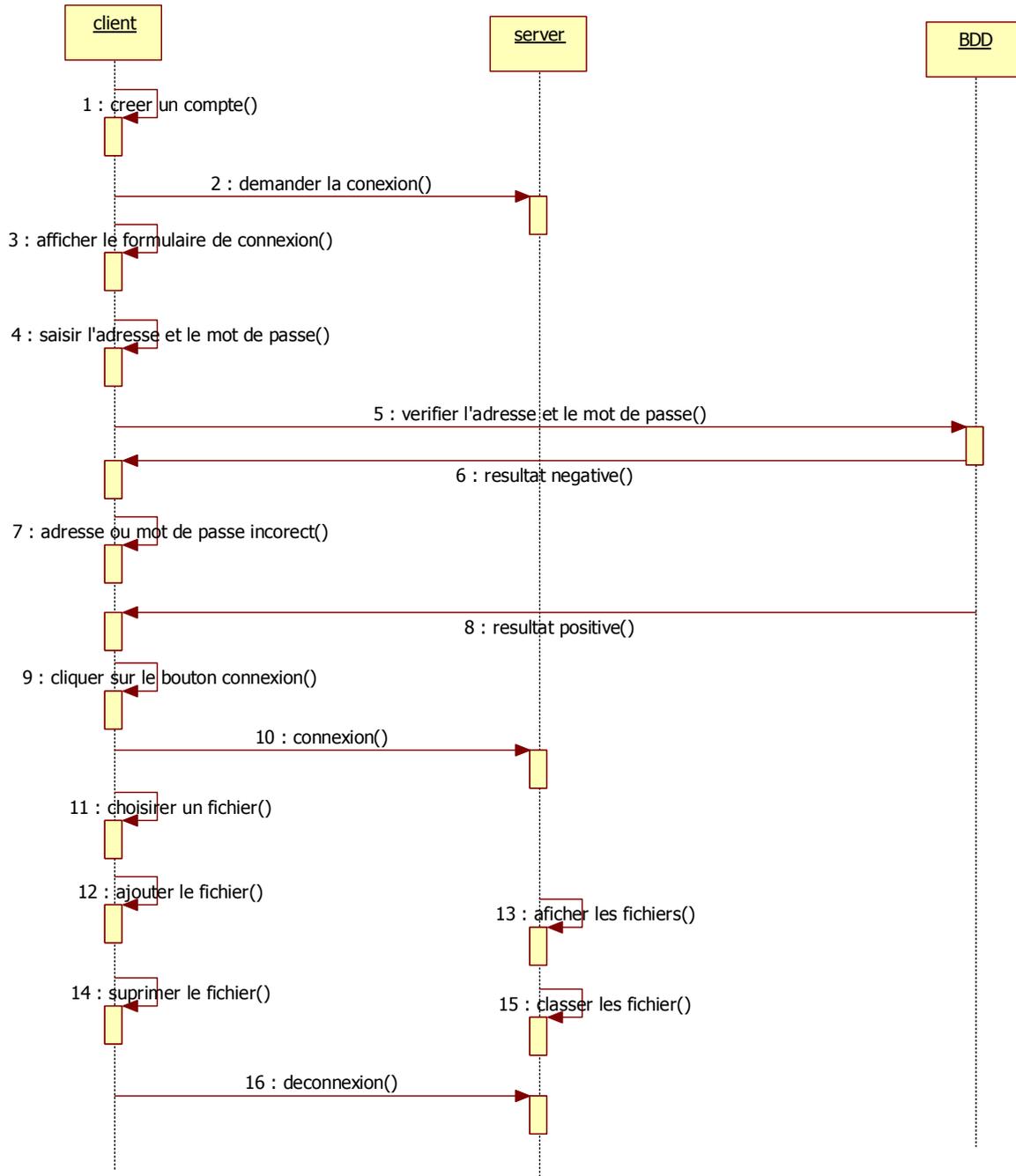


Figure III.3 :diagramme de sequence

III.6 Présentation de l'environnement de (ECLIPSE, PHP, MYSQL) :

Nous présenterons dans cette partie les outils que nous avons utilisés pour le développement de la solution en question.

Nous avons opté pour le système de gestion de base de données :

III.6.1 MYSQL : [19]

La base de données MySQL est la base de données open source la plus populaire au monde. Elle permet entre autres de développer à moindre coût des applications de bases de données Web et intégrées fiables, hautement performantes et évolutives.

III.6.1.1 Qu'appelle-t-on SQL : [20]

SQL (« Structured Query Language », traduisez Langage de requêtes structuré) est un langage de définition de données (LDD, ou en anglais DDL « Data Definition Language »), un langage de manipulation de données (LMD, ou en anglais DML, Data Manipulation Language), et un langage de contrôle de données (LCD, ou en anglais DCL, Data Control Language), pour les bases de données relationnelles.

III.6.1.2 SQL est un langage de définition de données :

✓ **SQL est un langage de définition de données (LDD) :**

C'est-à-dire qu'il permet de créer des tables dans une base de données relationnelle, ainsi que d'en modifier ou en supprimer.

✓ **SQL est un langage de manipulation de données :**

SQL est un langage de manipulation de données (LMD), cela signifie qu'il permet de sélectionner, insérer, modifier ou supprimer des données dans une table d'une base de données relationnelle.

✓ **SQL est un langage de protections d'accès :**

Il est possible avec SQL de définir des permissions au niveau des utilisateurs d'une base de données. On parle de **DCL** (Data Control Language).

III.6.2 PHP (web) :

Le **PHP** est un langage informatique utilisé sur l'internet. Le terme **PHP** est un acronyme récursif de « **PHP**: Hypertext Preprocessor ». Ce langage est principalement utilisé pour produire un site web dynamique. Il est courant que ce langage soit associé à une base de données, tel que MySQL. [21]

PHP Hypertext Preprocessor, plus connu sous son sigle **PHP** (acronyme récursif), est un langage de programmation libre, principalement utilisé pour produire des pages Web dynamiques via un serveur HTTP, mais pouvant également fonctionner comme n'importe quel langage interprété de façon locale. PHP est un langage impératif orienté objet.

PHP a permis de créer un grand nombre de sites web célèbres, comme Facebook Wikipédia, etc. Il est considéré comme la base de la création des sites Internet dits dynamiques.[22]

III.6.3 Eclipse : [23]

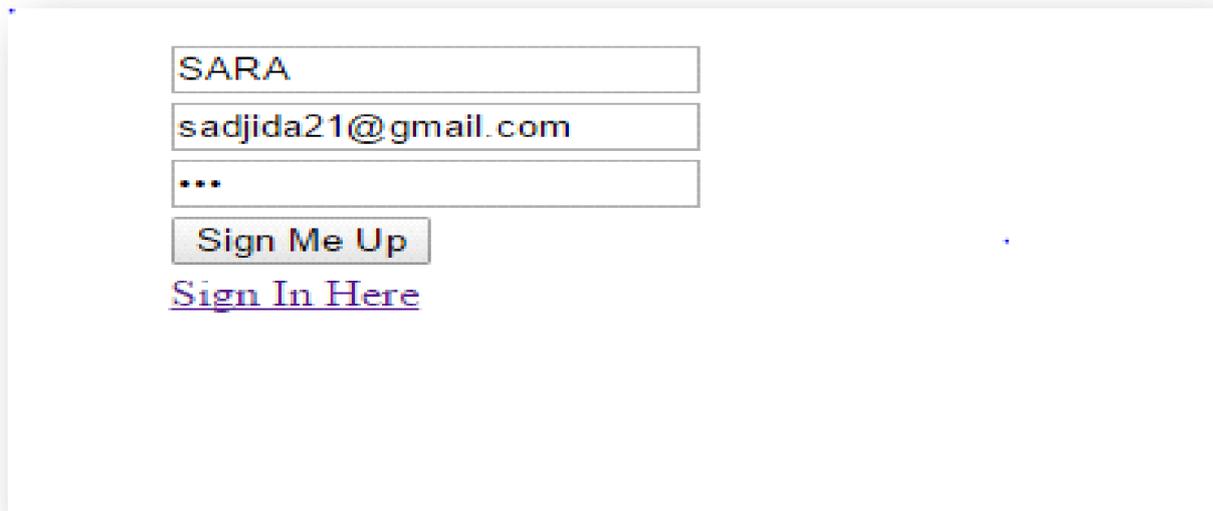
Eclipse est un environnement de développement intégré (IDE) utilisé dans la programmation informatique. Il contient un espace de travail de base et un système de plug-in extensible pour personnaliser l'environnement. Eclipse est écrite en Java et de son utilisation principale est pour le développement d'applications Java, mais il peut également être utilisé pour développer des applications dans d'autres langages de programmation par l'utilisation de plugins, y compris: Ada, ABAP, C, C ++, COBOL, Fortran, Haskell , JavaScript, Julia, Lasso, Lua, NATURAL, Perl, PHP, Prolog, Python, R, Ruby (y compris Ruby on Rails framework), Rust, Scala, Clojure, Groovy, Scheme, et Erlang. Il peut également être utilisé pour développer des packages pour le logiciel Mathematica. Les environnements de développement comprennent les outils de développement Eclipse Java (JDT) pour Java et Scala, Eclipse CDT pour C / C ++ et Eclipse PDT pour PHP, entre autres.

III.7 Présentation du prototype réalisé :

Pour la présentation de notre application, on va visualiser les différentes étapes de manipulation de l'application.

✓ **Coté Client :**

- **Interface d'authentification :** Avant d'accéder à notre page web, l'utilisateur doit d'abord s'authentifier (Saisir nom d'utilisateur et le mail et le mot de passe).



A screenshot of a web interface for user registration. It features four input fields stacked vertically. The first field contains the text 'SARA'. The second field contains the email address 'sadjida21@gmail.com'. The third field contains three dots '...', indicating a password field. Below the input fields is a button labeled 'Sign Me Up'. Underneath the button is a blue, underlined link that says 'Sign In Here'.

Figure III.4 : Interface d'authentification.

En cas d'accepte l'utilisateur reçoit le message suivant :

«Successfully registred »

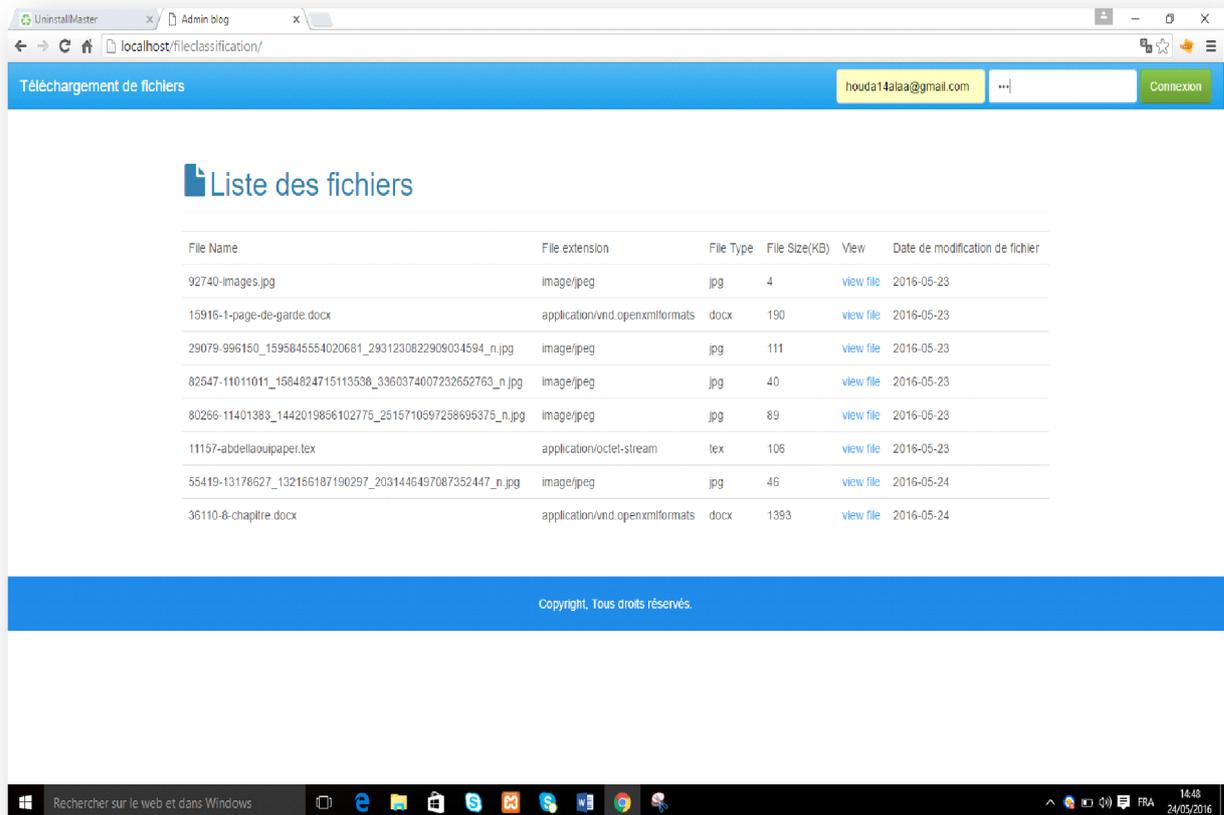
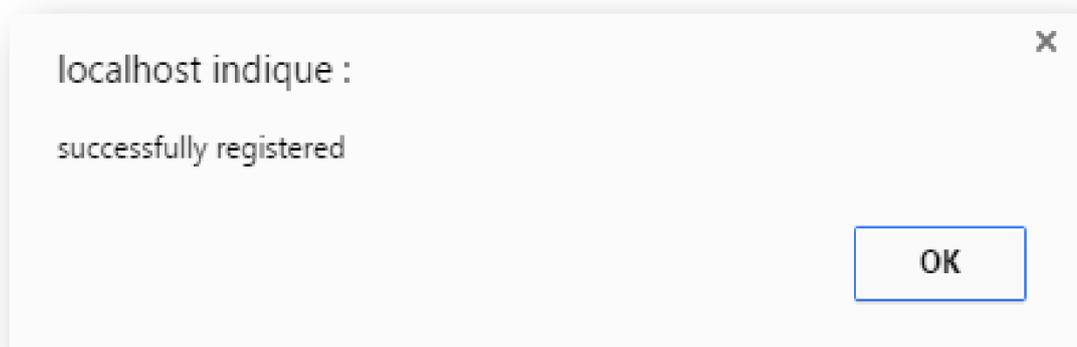


Figure III.5 : Interface d'entrée.

- **Interface D'entrée** : Ici l'utilisateur doit saisir son propre information pour accéder à la page web principale sur un clic u bottent connexion.



✓ *Interface Principale :*

Cette fenêtre contient les fichiers déjà télécharger, si on clique sur « view » on peut ouvrir ce fichier et on peut le supprimer, pour ajouter un fichier d'abord il faut faire le choix d'un fichier par le bottent « choisir un fichier » après en clique sur « télécharger ».

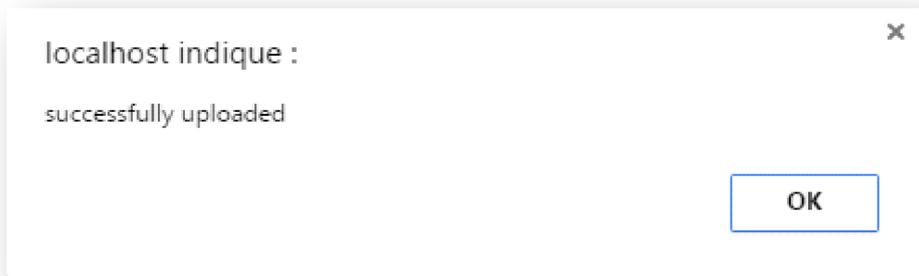
The screenshot shows a web browser window displaying a file management interface. The browser's address bar shows the URL `localhost/fileclassification/index.php`. The page has a blue header with the text "Téléchargement de fichiers" and a "Déconnexion" button. The main content area is divided into two sections:

- Ajouter un fichier:** This section contains a button labeled "Choisissez un fichier" and a message "Aucun fichier choisi". Below this is a text input field with the placeholder "Choisissez votre fichier" and an "Ajouter" button.
- Liste des fichiers:** This section features a table with the following data:

File Name	File extension	File Type	File Size(KB)	View	Supprimer le fichier	Date de modification de fichier
86665-3907_1538949826405772_7538487788777942749_n.jpg	image/jpeg	jpg	57	view file [X]		2016-06-01

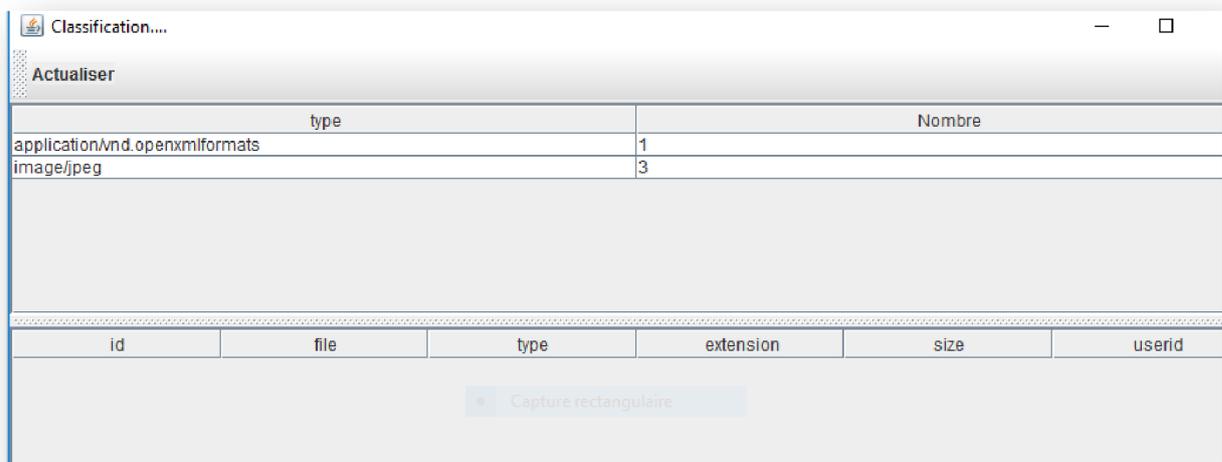
At the bottom of the page, there is a blue footer with the text "Copyright, Tous droits réservés." The Windows taskbar is visible at the very bottom, showing the time as 13:41 on 15/06/2016.

On cas de réussite l'utilisateur reçoit un message d'accord :



✓ **Coté serveur :**

Au niveau de serveur quand le client ajoute un fichier on peut voir cette modification au niveau de serveur



Actualiser	
type	Nombre
application/vnd.openxmlformats	1
image/jpeg	3

id	file	type	extension	size	userid
Capture rectangulaire					

Quand on clique sur une ligne dans le premier tableau on reçoit :

type		Nombre	
application/vnd.openxmlformats		1	
image/jpeg		3	

id	file	type	extension	size	userid
43	43687-1-page-de-garde...	application/vnd.openxmlf...	docx	190	2

Conclusion

Ce chapitre décrit et regroupe les résultats de la phase de conception et la réalisation de notre application, notamment les diagrammes des cas d'utilisation, de séquences, où nous avons abordés les technologies utilisées ainsi que les outils de développement, Nous avons achevé le travail par des captures d'écrans de l'application détaillant les différentes fonctionnalités.

Conclusion générale

Durant notre mémoire nous sommes intéressés à la classification du big data dans les réseaux sociaux , nous proposons une solution à cette problématique en générant une architecture client-serveur, doit être défini pour la génération du plan d'exécution à l'aide d'une approche de classification simultanée basée sur les méthodes de classification non supervisée notamment la méthode de classification hiérarchique ascendante.

Pour but d'informer les personnes de tout âge sur les impacts que peuvent avoir les réseaux sociaux. Ainsi que surveiller et analyser les données massives échangées, afin de prédire a pour rôle d'écouter les usagers suivant leurs besoins et de répondre sur tous les problèmes et d'améliorer QoS dans le web 2.0.

Bibliographie

[3] : Diawo K Memoire Online Les marques sur les réseaux sociaux et l'implication des consommateurs au sein des plateformes de marques page 1

[6] : Balagué Ch & Fayon D Francis, À quoi sert un réseau social ?, Social Networks History of Innovation and invotion .Springer, Italy(2014)

[7]: Ouaabi H, Introduction au concept d'innovation sociale,2012

[13] : Analyse de données avec SPSS. Manu Carricano. Copyright2009 Pearson Education France.

[14] : AMBAPOUR S Introduction à l'analyse des données. 53BAMSI REPRINT 04/2003 .

[15] : Contribution aux méthodes de Classification non supervisée via des approches pré - topologique et d'agrégation d'opinion. Annee2007.mounzer BOUBOU.Universitede

Webographie :

[1] : <http://tpereseauxsociaux.e-monsite.com/>, consulté le 2016.03.01

[2] : <http://www.andlil.com/>, Hugo Lauras, consulté le 2016.03.01

[4] : <http://www.e-marketing.fr/> , consulté le 2016.03.05

[5] : <https://www.ladissertation.com/> , consulté le 2016.03.05

[7] : <http://www.cc-lacqorthes.fr/>, consulté le 2016.03.20

[8] : <http://blog.kinoa.com/>, benjamin.thiers consulté le 2016.04.01

[9] : <http://www.paris-sorbonne.fr/toutes-les-soutenances>

[10] : <http://jaimemavieprivee.jimdo.com/>, consulté le 2016.04.05

[11]: <http://oseox.fr/ereputation/>, consulté le 2016.04.05

[12] : <http://www.statistique-2013.fr/>, consulté le 2016.04.02

[16] : <https://fr.wikipedia.org/wiki/>, consulté le 2016.04.03

[17] : <http://www.piloter.org/>, consulté le 2016.04.11

[18] : www.UML.com

[19] : <https://www.oracle.com/> consulté le 04.07.2016

[20] : <http://www.commentcamarche.net/> consulté le 04.07.2016

[21] : <http://glossaire.infowebmaster.fr/> consulté le 04.07.2016

[22] : <https://fr.wikipedia.org/> consulté le 04.07.2016

[23] : <https://en.wikipedia.org/> consulté le 04.07.2016