

Université de Blida 1 – SAAD DAHLAB



Faculté des sciences

Département d'Informatique

Mémoire présenté par :

M. TEMZI Abdelali

M. OUHDADOU Billal

Pour l'obtention du diplôme de Mastère

Domaine : Mathématiques et Informatique

Filière : Informatique

Spécialité : Ingénierie des Logiciels

Sujet :

**Essai d'application des Algorithmes
d'Apprentissage Automatique pour la
Classification des parties se prévalant de
l'Islam « al-tawâ'if al-islâmiyya »**

Soutenu le : 21/10/2021 devant les membres du jury de thèse :

M. HAMMOUDA	Université de Blida 1	Président
Mme. FERDI	Université de Blida 1	Examinatrice
M. CHERIF ZAHAR	Université de Blida 1	Promoteur
Mme. MEZZI	Université de Blida 1	Promotrice

Résumé

Dans ce projet, nous nous sommes intéressés particulièrement aux factions se prévalant de l'islam. Le cas d'étude se base sur un hadith authentique, le Messager d'Allah a dit : « La communauté juive s'est scindée en soixante et onze ou soixante-douze factions, toutes iront en enfer. La communauté chrétienne s'est également scindée en soixante et onze ou soixante-douze factions, toutes iront en enfer. Ma communauté-ci se divisera en soixante-treize factions, toutes iront en enfer sauf une : celle qui a la même religion que moi et mes compagnons avons aujourd'hui. ».

Ce hadith très connu a fait l'objet d'études de savants religieux pour essayer de déterminer les sectes vouées au feu de l'enfer d'entre celles se prévalant de l'islam. Toutefois, ces mêmes oulémas ont buté sur un problème : trouver les traits distinctifs qui permettent de retrouver soixante-treize parties.

Afin d'essayer de résoudre ce problème informatiquement, nous nous sommes basés sur le livre de la différence entre les sectes (الفرق بين الفرق للعالم عبد القاهر بن طاهر بن محمد) (البغدادي الإسفرائيني التميمي) et établis quelques critères qui nous permettrait de trancher sur les traits distinctifs entre les sectes. Nous avons alors créer une plateforme qui permet la saisie des sectes et que leurs caractéristiques ainsi que la recherche de ces dernières par des experts du domaine.

Pour ce faire, nous avons dans un premier temps, collecter un petit ensemble de données sur lequel nous avons appliqué des algorithmes de classification : K-means et arbre de décision (dcision tree) Et les résultats obtenus étaient que l'algorithme de K-means était plus fiable, rapide et donne les résultats souhaités.

Mots clés:

Apprentissage automatique, classification automatique, k-means, arbre de décision, sectes musulmanes.

Abstract

In this project, we were particularly interested in factions that use Islam. The case study is based on an authentic hadith, Allah's Messenger said, "The Jewish community has split into seventy-one or seventy-two factions, all will go to hell. The Christian community has also split into seventy-one or seventy-two factions, all of which will go to hell. My community will be divided into seventy-three factions, all of them will go to hell except one: the one that has the same religion that I and my companions have today. "

This well-known hadith has been the subject of studies by religious scholars to try to determine which sects doomed to hellfire among those which claim Islam. However, these same ulemas came up against a problem: to find the distinctive features which allow to find seventy-three parts.

In order to try to solve this problem by computer, we based ourselves on the book of the difference between the sects (الفرق بين الفرق للعالم عبد القاهر بن طاهر بن محمد) (البغدادي الإسفرائيني التميمي) and established some criteria on which we could decide on the traits. distinctive between sects. We then created a platform that allows the seizure of sects and their characteristics as well as the research of the latter by experts in the field.

To do this, we first have to collect a small data set on which we have applied classification algorithms: K-means and decision tree and the results obtained were that the algorithm of K- means was more reliable, faster and gives the desired results.

Key words :

Machine learning, machine classification, k-means, decision tree, Muslim sects.

مُلخَص

خدمة لديننا الإسلام كانت انطلاقتنا من حديث الرسول صلى الله عليه وسلم الذي يتحدث فيه عن حال أمته في آخر الزمان وما يحدث لها من شتات وافتراق قال الرسول صلى الله عليه وسلم قال: افتترقت اليهود على إحدى وسبعين فرقة، وافتترقت النصارى على اثنتين وسبعين فرقة، وستفترق هذه الأمة على ثلاث وسبعين فرقة كلها في النار إلا واحدة، قيل: من هي يا رسول الله؟ قال: من كان على مثل ما أنا عليه وأصحابي. وفي بعض الروايات: هي الجماعة. رواه أبو داود والترمذي وابن ماجه والحاكم، وقال: صحيح على شرط مسلم.

كان هذا الحديث المشهور موضوع دراسات لعلماء الدين لمحاولة تحديد الطوائف المحكوم عليها بالنار من بين أولئك الذين يدعون الإسلام. لكن هؤلاء العلماء أنفسهم واجهوا مشكلة: العثور على السمات المميزة التي تسمح بإيجاد ثلاثة وسبعين جزءاً.

لمحاولة حل هذه المشكلة عن طريق الكمبيوتر ، اعتمدنا أنفسنا على كتاب الاختلاف بين الطوائف (الفرق بين الفرق للعالم عبد القاهر بن طاهر بن محمد البغدادي الإسفرائيني التميمي) ووضعنا بعض المعايير التي يمكننا على أساسها اتخاذ قرار بشأن السمات . مميزة بين المذاهب. ثم أنشأنا منصة تتيح الاستيلاء على الطوائف وخصائصها وكذلك البحث عن هذه الأخيرة من قبل خبراء في المجال.

للقيام بذلك، تَعَيَّنَ علينا أولاً جمع مجموعة صغيرة من البيانات التي طبقنا عليها خوارزميات التصنيف: خوارزمية التجميع (K-means) وشجرة القرارات (Decision tree) وكانت النتائج التي تم الحصول عليها من طرف خوارزمية التجميع هي الأفضل في حالتنا حيث أعطتنا النتائج التي كنا نرغب بها وبسرعة أكبر.

كلمات مفتاحية:

التعلم الآلي، التصنيف الآلي، الوسائل ، خوارزمية التجميع، شجرة القرارات ،المذاهب الإسلامية.

Remerciements

Ce mémoire a été réalisé dans le cadre de fin d'études pour l'obtention du diplôme de Master en informatique. Nous tenons à remercier d'abord ALLAH de nous avoir donné force et santé pour mener ce travail à terme. Nous tenons à remercier nos parents pour leurs sacrifices, soutien et compréhension durant toutes nos années d'études. Nous tenons à exprimer nos profonds remerciements à **Mme Mezzi. M** et **M. CHERIF ZAHAR** Pour leurs conseils précieux, leurs remarques et leurs disponibilités lors de l'élaboration de cette étude.

Nous souhaitons exprimer nos remerciements à tous les membres du jury pour avoir bien voulu accepter de participer à l'expertise de notre modeste travail, prouvant ainsi l'intérêt qu'ils portent à ce travail. Nous exprimons notre gratitude à l'ensemble du corps enseignant, technique et administratif du département d'informatique de l'université de Blida, pour leur disponibilité et leur gentillesse et finalement, nos sincères remerciements à tous ceux qui ont contribué de près ou de loin à la réalisation de ce modeste travail.



Dé dicace...

Je tiens à remercier en premier lieu mes parents

*A mon symbole de sacrifice, écoles de mon enfance qui était
mon ombre durant toutes mes années d'études, toi mon cher
père qui a tant investi pour assurer mon Avenir.*

*À la source d'amour et tendresse à celle qui a tout donné à toi
ma chère **maman**.*

*A notre promotrice et chère sœur **Mme Mezzi.M***

*A mes chers frères **Abdeldjalil** et **Abderrezak** et mes chères
sœurs **Soumia**, **Akila**, **Ihssene** et **Hayaa***

*Et à tous les membres de ma famille .Je termine cette dédicace
à mes neveux et nièces : **Sabrina**, **Yahia**, **Mohamed Yasser** et
Warda.*

Abdelali...





Dédicace...

Je tiens à remercier en premier mes parents.

*A mon frère **Hamza** et ma sœur **Marwa**.*

*A notre promotrice **Melyara Mezzi** et ceux qui ont partagé avec moi tous les moments d'émotion lors de la réalisation de ce travail.*

*A ma fiancée **Loubna** qui m'a toujours encouragé.*

A ma famille, mes proches et à ceux qui me donnent de l'amour et de la vivacité.

Billal...



Table des matières

Introduction générale et mise en contexte.....	1
I.1 Préambule.....	1
I.2 Divergences dans les critères d'identification des « parties ».....	2
I.3 N'y a-t-il que soixante-treize parties dont une seulement ira au Paradis ? ...	2
I.4 Les parties promises à l'Enfer y séjourneront elle éternellement ?.....	3
I.5 Ecole dogmatique et école de droit	4
I.6 Eléments essentiels de divergence des firaq.....	4
I.7 Problématique.....	5
I.8 Objectifs	6
I.9 Organisation du mémoire	6
Chapitre I : Apprentissage Automatique et Algorithmes de Classification	8
I.1 Introduction	9
I.2 Définition.....	9
I.3 Classification automatique	9
I.3.1 Apprentissage supervisé (prédictif).....	10
I.3.1.1 Distance euclidienne	10
I.3.1.2 Word2Vec	10
I.3.1.3 Quelques algorithmes de l'apprentissage supervisé	11
I.3.2 Apprentissage non-supervisé (descriptif).....	13
I.3.2.1 Quelques algorithmes d'apprentissage non-supervisé	14
I.3.3 Apprentissage semi-supervisé.....	17
I.3.4 Apprentissage partiellement supervisé.....	17
I.3.5 Apprentissage par renforcement	17
I.3.6 Apprentissage par transfert	18
I.4 Conclusion.....	18
Chapitre II : Conception et modélisation de la solution proposée	19
II.1 Introduction	20
II.2 Architecture globale	20
II.3 Description des données	20
II.3.1 Données en entrée	20
II.3.2 Source de données.....	21

II.4	Expression des besoins	26
II.4.1	Identification des acteurs.....	26
II.4.2	Diagrammes des cas d'utilisation.....	27
II.4.2.1	Diagramme de cas d'utilisation global	27
II.4.2.2	Diagramme de cas d'utilisation « Gérer secte »	29
II.4.2.3	Diagramme de cas d'utilisation « Gérer ressources ».....	31
II.4.2.4	Diagramme de cas d'utilisation « Gestion des utilisateurs »	32
II.5	Analyse des besoins.....	34
II.5.1	Diagramme de séquence « Authentification »	34
II.5.2	Diagramme de séquence « Création compte pour un expert ».....	35
II.5.3	Diagramme de séquence gestion des ressources	36
II.5.3.1	Diagramme de séquence ajouter une ressource	36
II.5.3.2	Diagramme de séquence supprimer une ressource	37
II.5.3.3	Diagramme de séquence modifier une ressource	37
II.5.4	Diagramme de séquence gestion des sectes	38
II.5.4.1	Ajouter secte	38
II.5.4.2	Modifier secte	39
II.5.4.3	Supprimer secte.....	40
II.6	Conception de l'application.....	41
II.6.1	Diagramme de classes	41
II.7	Conclusion.....	43
Chapitre III: Tests et implémentation		44
III.1	Tests.....	45
III.1.1	Introduction	45
III.1.2	Jeux de données.....	45
III.1.2.1	Résultats de l'algorithme de classification K-means.....	45
III.1.2.2	Résultats de l'algorithme arbre de décision	50
III.1.3	La différence entre K-means et arbre de décision	55
III.2	Implémentation	56
III.2.1	Environnement de développement.....	56
III.2.2	Environnement Matériel.....	56
III.2.3	Environnement et langages de programmation utilisés	57

III.2.3.1	Langages de programmation.....	57
III.2.3.2	Modules externe utilisé.....	57
➤	Flask :.....	57
III.2.3.3	Environnement de développement.....	57
III.2.4	Réalisation du projet	58
III.2.5	Schéma d'accessibilité de la plateforme	58
III.2.6	Les interfaces	60
III.2.6.1	Session Administrateur	60
III.2.6.2	Session expert	63
III.2.6.3	Session visiteur	67
III.3	Conclusion	67
	Conclusion générale.....	68
III.1	Synthèse	69
III.2	Perspectives	70
	Bibliographie.....	72

Liste des tableaux

<i>Tableau 1: Exemple de données pour réaliser un arbre de décision</i>	12
<i>Tableau 2: Caractéristiques retenues pour les sectes musulmanes</i>	22
<i>Tableau 3: Description du diagramme des cas d'utilisation global</i>	28
<i>Tableau 4: Description du diagramme de cas d'utilisation "Gestion des sectes"</i>	29
<i>Tableau 5: Description du cas d'utilisation "Gestion des ressources"</i>	31
<i>Tableau 6: Description du diagramme de cas d'utilisation "Gestion des utilisateurs"</i>	33
<i>Tableau 7: Description du diagramme de classes</i>	42
<i>Tableau 8: Quelques lignes de dataset avant la classification</i>	47
<i>Tableau 9: La différence entre K-mens et Arbre décision</i>	56

Table des figures

<i>Figure 1: Exemple sur un arbre de décision sur construit à partir de tableau 1</i>	12
<i>Figure 2: Exemple sur l'algorithme de classification K-NN</i>	13
<i>Figure 3 : Exemple d'identification les directions principales</i>	16
<i>Figure 4 : Architecture global du système</i>	20
<i>Figure 5: Les principaux courants de l'Islam</i>	21
<i>Figure 6: Capture d'écran du premier dataset</i>	22
<i>Figure 7: liste des caractéristiques dans notre base de données</i>	25
<i>Figure 8: Quelques lignes de notre fichier Csv</i>	26
<i>Figure 9 : Diagramme de cas d'utilisation global</i>	28
<i>Figure 10 : Diagramme de cas d'utilisation « Gestion des sectes »</i>	29
<i>Figure 11 : Diagramme de cas d'utilisation « Gestion ressources »</i>	31
<i>Figure 12 : Diagramme de cas d'utilisation « Gestion caractéristique »</i>	32
<i>Figure 13 : Diagramme de cas d'utilisation « Gérer utilisateurs »</i>	33
<i>Figure 14 : Diagramme de séquences « Authentification »</i>	34
<i>Figure 15 : Diagramme de séquences « Création d'un compte »</i>	35
<i>Figure 16 : Diagramme de séquence ajouter ressource</i>	36
<i>Figure 17: Diagramme de séquence supprimer une ressource</i>	37
<i>Figure 18 : Diagramme de séquence « Modifier une ressource »</i>	38
<i>Figure 19 : Diagramme de séquence « Ajouter une secte »</i>	39
<i>Figure 20: Diagramme de séquence « Modifier une secte »</i>	40
<i>Figure 21 : Diagramme de séquence « supprimer une secte »</i>	41
<i>Figure 22: Diagramme de classes de l'application</i>	42
<i>Figure 23: résultat 1 de l'algorithme K-means</i>	46
<i>Figure 24: Graphe nombre de sectes par rapport aux clusters</i>	47
<i>Figure 25: Quelques lignes de Dataset après la classification</i>	48
<i>Figure 26: Code pour spécifier le profondeur de l'arbre dans l'algorithme DT</i>	51
<i>Figure 27: résultat de l'arbre et valeur d'accuracy par rapport à D3</i>	51

<i>Figure 28: résultat de précision de l'arbre de décision par rapport à une autre caractéristique.....</i>	<i>52</i>
<i>Figure 29: Arbre de notre Dataset par rapport à une autre caractéristique.</i>	<i>53</i>
<i>Figure 30: résultat de classification par rapport au DT algorithm.</i>	<i>53</i>
<i>Figure 31 : Schéma d'accessibilité</i>	<i>59</i>
<i>Figure 32: Page d'authentification.</i>	<i>60</i>
<i>Figure 33 : Administrateur page d'accueil 1</i>	<i>60</i>
<i>Figure 34 : Administrateur page d'accueil 2.....</i>	<i>61</i>
<i>Figure 35 : Administrateur page d'accueil 3 -statistiques-.....</i>	<i>61</i>
<i>Figure 36: Administrateur page 4 ajouter un administrateur.</i>	<i>62</i>
<i>Figure 37 : Page Messagerie</i>	<i>62</i>
<i>Figure 38 : page d'accueil expert 1.</i>	<i>63</i>
<i>Figure 39: taches que l'expert peut les faire.....</i>	<i>63</i>
<i>Figure 40: Quelques statistiques que l'expert peut consultées.</i>	<i>64</i>
<i>Figure 41 : interface pour envoyer un message.....</i>	<i>64</i>
<i>Figure 42 : ajouter des caractéristiques et valeur pour des caractéristiques</i>	<i>65</i>
<i>Figure 43: Ajouter une nouvelle secte.</i>	<i>65</i>
<i>Figure 44: Afficher tous les sectes ajoutées par l'expert.....</i>	<i>66</i>
<i>Figure 45: La liste de toutes les sectes de la base de données</i>	<i>66</i>
<i>Figure 46: Interface 1 de la recherche de visiteur.</i>	<i>67</i>
<i>Figure 47: Interface de l'affichage les résultats de la recherche.....</i>	<i>67</i>

Introduction générale et mise en contexte

Introduction générale et mise en contexte

I.1 Préambule

Lorsque le prophète de l'islam décède le 12 du mois de Rabî' al Awwal de l'an 11 de l'Hégire, il laisse à sa communauté un héritage lourd : Le Livre d'Allah et la Parole de Son Messenger et il leur enjoint de ne point se diviser à leurs sujets. C'est ainsi qu'il le leur rappelle dans le sermon dit de l'adieu, survenu deux mois avant sa mort au cours duquel il leur dira : « *Je vous ai laissé ce dont si vous vous y cramponnez, vous ne vous égarerez nullement : le Livre d'Allah et la Sunna de son Messenger* ». Dans un verset du Coran, on lit : « Et voilà Mon chemin dans toute sa rectitude, suivez-le donc; et ne suivez pas les sentiers qui vous écartent de Sa voie. » Voilà ce qu'Il vous enjoint. Ainsi atteindrez-vous la piété » (Coran, al-an'âm (les troupeaux), v. 153). Le prophète, en expliquant ce verset, avait dessiné sur le sable une ligne droite et longue et sur les côtés des lignes tordues et courtes en disant : La ligne droite est le sentier qui mène à Allah et les petites lignes à sa droite et à sa gauche sont les [autres] sentiers » en allusion au fait que ce sont des sentiers qui ne mènent nulle part et qui écartent du sentier qui mène à Allah.

Le prophète n'a jamais cessé de prévenir ces Compagnons et -après eux- toute la communauté à venir des dangers qui les guettent, le plus dangereux étant celui de l'éloignement de l'enseignement d'Allah et de la Sunna de son Messenger. Ainsi est-il allé jusqu'à les instruire que sa communauté allait se diviser. C'est ainsi qu'il avait dit : « *Les juifs se sont divisés en soixante et onze parties, les chrétiens en soixante-douze parties. Ma communauté se divisera en soixante-treize parties : toutes sont promises à l'Enfer sauf une : Celle qui adopte la même religion que moi et mes compagnons avons aujourd'hui.* »

De ce hadith et de ses semblables, les docteurs musulmans dès le début de l'islam se sont penchés sur le recensement des parties qui se sont distancées de l'islam originel et le catalogue de leurs déviations d'autant que les premières parties commencent à apparaître déjà du vivant même des Compagnons du prophète, tel étant le cas des *kharidjites*, des *chiïtes* et des *qadariyya*.

Parmi les auteurs qui se sont intéressés à ce sujet, il y a lieu de citer :

1. *Shahrastânî*, « al milal wa-n-nihal » (Les religions et leurs ramifications)

Introduction générale et mise en contexte

2. Ibn Hazm, « *al-fisal fi-l-milali wa-l-ahwâ'i wa-n-nihal* » (L'exposé clair sur les religions et leurs ramifications)
3. Abu-l-Hasan al-Ash'arî, « *maqâlât al-islâmiyyîn* » (les doctrines des parties se prévalant de l'islam)
4. Baghdâdî, « *al-farq bayn al firaq* » (Les différences entre les parties se prévalant de l'islam)

Ibn Taymiyya explique que « Ces auteurs de livres sur les parties se prévalant de l'islam sont souvent issus de parties elles même en faux avec les Compagnons du prophète et leurs disciples. ». Parmi les ouvrages modernes, il y a lieu de citer :

Muhammad 'alawiyy assaqâf, « *mawsû'at al-firaq* » (l'Encyclopédie des parties se prévalant de l'islam).

I.2 Divergences dans les critères d'identification des « parties »

Quel sont les critères qui discernent une partie d'une autre ? Qu'est-ce qui identifie une partie ?

Chacun des auteurs -dans des ouvrages non nécessairement spécialisés dans les sectes- a eu sa façon de voir, a fixé ses propres critères d'identification. Il en résulte naturellement des divergences dans la classification. L'exemple des *ibadites* en est un : pour la majorité, ces derniers sont une des ramifications des *kharidjites*. Pour d'autres, du fait qu'ils ont beaucoup d'affinités avec les *mu'tazila*, ils en sont considérés comme une ramification. Il en est de même des *zaydiyya*, sont-ils des chiites empreints de sunnisme ou des sunnites empreints de chiisme ? Les *ashaarites* et les *maturidites* sont-ils des sunnites ou autre que cela ? Cela va dépendre forcément des critères de différenciation mais aussi à la primauté de tel ou tel autre argument sur les autres.

I.3 N'y a-t-il que soixante-treize parties dont une seulement ira au Paradis ?

Tous les auteurs qui ont traité du sujet, qu'ils aient ou non dédié un ouvrage à cela, s'accordent à dire que le nombre de sectes, si l'on recense toutes les ramifications, se compte en centaines voire plus. Il est clair qu'au sein d'une même population, on peut trouver des divergences minimales. Prises en compte, elles donneraient lieu à des parties beaucoup plus nombreuses. Mais si on s'astreint à ne considérer que les éléments fondamentaux, on trouve qu'elles ne forment qu'une et une seule partie.

Introduction générale et mise en contexte

I.4 Les parties promises à l'Enfer y séjourneront elle éternellement ?

De l'avis d'Ibn Taymiyya, le Prophète n'a pas dit qu'elles étaient destinées à séjourner éternellement en enfer. Toutefois, il y a celles dont la croyance est une « incroyance manifeste ». Les auteurs de pareilles doctrines dites *bida' mukaffira*, qui décèdent s'en être repentis se trouvent de facto dans la catégorie des non musulmans. Mais ceux dont la *bid'a* n'est pas en opposition flagrante avec quelque croyance de l'islam ne sont pas voués à séjourner éternellement en Enfer.

Bien que nous sortions du cadre du mémoire, au vu de l'incompréhension qui peut en découler, il est nécessaire d'apporter un éclaircissement par l'exemple :

Dans des versets on ne peut plus clair et très nombreux, Allah met en avant la place incontestable des Compagnons du Prophète. Parmi ceux-là, le verset ou Allah dit : « *Ne sont point égaux ceux d'entre vous qui ont investi leurs biens [pour plaire à Dieu] avant la conquête de la Mekke et qui ont combattu. Ceux-là sont plus méritoires que ceux qui ont investi leurs biens après et ont combattus. Et Allah leur a tous promis al-Husnâ (le Paradis)* ». Dès lors, beaucoup de Docteurs musulmans ont penché pour le fait que les chiites qui portent atteinte aux Compagnons du Prophète ne sont pas croyants. Parmi ces docteurs, il y a l'illustrissime Mâlik ibn Anas, le fondateur éponyme de l'école malékite (al-madhhab el mâlikî).

A contrario, il y a des parties dont la divergence essentielle avec l'islam originel relève de l'incompréhension de certains versets. Tel est le cas des *nufât* qui ont été incapables de regrouper le verset : « Il n'y a rien qui Lui ressemble » avec des versets comme : « Ô Iblîs ! Qu'est-ce qui t'a empêché de te prosterner de ce que j'ai créé de Mes deux mains ? ». Pour eux, le fait que rien ne lui ressemble nous oblige à considérer que la main de Dieu, c'est sa force. Or ceci est faux. D'abord, le verset énonce « Mes deux mains » et il n'y aurait aucun sens à dire Mes deux forces ! Et d'autre part, les premiers musulmans n'ont jamais entendu pareille chose de la sorte. Lorsqu'un homme est venu à l'imâm Mâlik fils d'Anas lui dire : Comment Dieu s'est-il placé par-dessus son trône ? L'imâm lui répond : « Le fait que Dieu se soit placé par-dessus son trône est un fait certain, la foi en cela est une obligation et poser

Introduction générale et mise en contexte

des questions de la sorte est une bid'a (une chose nouvelles que les anciens ne connaissaient pas). »

I.5 Ecole dogmatique et école de droit

Les parties musulmanes (firaq) sont celles qui diffèrent sur des points de dogme, des points liés aux croyances. A l'opposé, les écoles de droit (madhahib fiqhiyya) s'intéressent à connaître le statut au regard de la loi musulmane de tel ou telle pratique. Ces dernières ne sont pas concernées par l'objet de l'étude du fait que le consensus unanime les en exclu. Certains compagnons du prophète ont divergé lors d'un voyage s'il fallait faire telle ou telle chose. Le prophète ne les teint pas pour autant pour coupable de quelque erreur que ce soit parce qu'ils avaient fait l'effort de connaître la vérité dans un cas de figure qu'ils ignoraient.

Les firaq, elles, se sont distancées volontairement du sentier d'Allah. Leur incompréhension résulte d'un entêtement résultant lui-même d'un rejet intérieur de la Vérité, non d'une incapacité à comprendre. C'est ainsi que certaines d'entre ces firaq disaient clairement : « Les premiers musulmans (salaf) sont plus savants alors que les derniers (khalaf) sont plus sages », sentence attestant clairement qu'ils connaissaient la vérité mais qu'ils s'en sont distancés volontairement.

I.6 Eléments essentiels de divergence des firaq

Safar al-Hawâlî, dans un petit opuscule intitulé : « usûl al-firaq wa-l-adyân wa-l-madhâhib al-fikriyya » (Les éléments essentiels des sectes, des religions et des écoles de pensée) énumère les critères suivants qui sont les principaux sur lesquels les firaq se sont divisées :

1. *La foi* : pour les premiers musulmans, c'est un acte et une parole et elle peut augmenter et diminuer. Pour les kharédjites et les mu'tazila, c'est la mise en pratique de toutes les ordonnances religieuses et l'abstention de toutes les interdictions. Elle ne peut ni augmenter ni diminuer. Dès lors, il en découle pour eux que celui qui n'accomplit pas un devoir religieux ou ne s'abstient pas de commettre un péché est un incroyant (kâfir). Pour les murdji'a, c'est juste la croyance du cœur. Pour eux, pécher ou faire de bonnes œuvres n'a aucun impact sur la foi.

Introduction générale et mise en contexte

2. *La croyance ne la prédestination* : les premiers musulmans estiment que la foi en la prédestination est un article de cette dernière. Les qadariyya ainsi que les mu'tazila et les chiaa la rejettent. Les djabriyya ne croient pas au libre arbitre de l'homme.
3. Les noms et attributs d'Allah : les premiers musulmans croyaient à tous les noms et attributs d'Allah qui sont dans le Coran et la Sunna sans qu'il n'est en cela un semblable. Ils ne déformaient pas le sens des mots. A l'opposé, les mushabiha prétendent que les attributs d'Allah sont pareils à ceux de ces créatures et les nufât en changent le sens de tout ou partie des attributs de peur de tomber dans l'erreur des mushabiha.
4. l'imama (ou le califat) : Pour les premiers musulmans, le califat est parvenu après le Prophète à Abû Bakr puis Omar puis Othman puis Ali. Il doit rester dans Qoreish. Les chiites se sont divisés : pour les zaydiyya, le califat des trois premiers est valable mais Alî le méritait davantage. Les autres chiites les maudissent et les excommunient pour avoir pris la place -selon eux- de Alî. Tous les chiites considèrent que la direction de la communauté doit échoir à des descendants d'Ali avec des divergences dans leur identification. Les khawaridj reconnaissent le caractère valide des califats des deux premiers et maudissent les deux derniers. Ils estiment que le califat peut échoir à tout musulman.

I.7 Problématique

Il résulte de ce qui précède que la classification des sectes musulmanes pose un certain problème quant à arriver à identifier les soixante-douze classes autres que celle qui a la même religion que le prophète et ses compagnons avaient.

Cela ne veut pas dire non plus que les musulmans ont été incapables de les identifier. Les catalogues des uns et des autres parmi les auteurs anciens et contemporains en témoignent.

La problématique -telle que nous la voyons- réside dans le fait de tester jusqu'à trouver le ou les bons critères de différenciation.

Le travail de différenciation a dû prendre un temps considérable aux auteurs postérieurs et même contemporains du fait qu'il s'agit d'un travail manuel.

Introduction générale et mise en contexte

L'utilisation de l'outil informatique et le substrat de l'intelligence artificielle permettrait assurément de faire des progrès en la matière voire de trouver les bons discriminants des différentes sectes parmi les soixante-douze qui sont dans l'erreur.

I.8 Objectifs

A travers ce travail, nous voulons mettre à profit les connaissances modernes en termes de classificateurs pour essayer de résoudre le problème cité ci-haut.

Il s'agira :

1. De fournir une application permettant de saisir les champs de la connaissance sur lesquels il y a des divergences entre firaq.
2. De fournir un moyen de saisir les paramètres (valeurs de chaque champ cité en 1) pour chaque secte
3. De faire un catalogue des différentes techniques utilisées en matière de classification supervisée, non supervisée ou semi-supervisée.
4. De choisir d'entre le précédent catalogue les techniques qui pourraient répondre le mieux au problème
5. D'appliquer les méthodes retenues sur les données recensées et à jour sur le poids de chaque discriminant jusqu'à retrouver le nombre « 73. »
6. De conclure sur les critères discriminants qui donnent le bon résultat.
7. De comparer avec les travaux de tel ou tel autre.

I.9 Organisation du mémoire

Afin d'atteindre les objectifs susmentionnés, nous avons choisi d'organiser notre mémoire en 3 chapitres :

- Le chapitre 1 est consacré quant à lui à la notion de l'apprentissage automatique. Nous y parlerons de Classification Automatique, ainsi que ses différents types : supervisé, semi-supervisé, partiellement supervisé, et non-supervisé. Nous y aborderons également les différentes manières de représentation des textes.
- Le chapitre 2 concerne la présentation de la conceptualisation et la modélisation de la solution proposée.
- Le chapitre 3 quant à lui abordera le côté technique de l'implémentation de la solution et comprendra une présentation de l'application web réalisée.

Introduction générale et mise en contexte

Chapitre I : Apprentissage Automatique et Algorithmes de Classification

Chapitre I : Apprentissage Automatique et Algorithmes de Classification

I.1 Introduction

La classification automatique des documents textuels arabes devient nécessaire à cause du volume de documents arabes échangés sur internet et stockés sur les supports électroniques.

Dans ce chapitre nous allons parler sur la notion d'Apprentissage Automatique (ou Apprentissage Statistiques ; en Anglais Machine Learning) et les algorithmes de classification (supervisée, semi-supervisée, et non supervisée).

I.2 Définition

Dans l'intelligence artificielle (IA) et dans le traitement automatique des langues naturel (TALN) on trouve plusieurs méthodes et différentes manières à utiliser afin de classifier ou traiter des textes, des images, des vidéos, etc...

TALN Les utilisations statistiques du traitement automatique du langage naturel reposent sur des méthodes stochastiques, probabilistes ou simplement statistiques pour résoudre certaines difficultés discutées ci-dessus, particulièrement celles qui surviennent du fait que les phrases très longues sont fortement ambiguës une fois traitées avec des grammaires réalistes, autorisant des milliers ou des millions d'analyses possibles. Les méthodes de désambiguïsation comportent souvent l'utilisation de corpus et d'outils de formalisation comme les modèles de Markov. Le TALN statistique comporte toutes les approches quantitatives du traitement linguistique automatisé, y compris la modélisation, la théorie de l'information, et l'algèbre linéaire. La technologie pour le TALN statistique vient principalement de l'apprentissage automatique et de l'exploration de données, lesquels impliquent l'apprentissage à partir des données venant de l'intelligence artificielle [1].

Dans ce chapitre nous allons représenter quelques méthodes et algorithmes utilisés dans le domaine de classification des textes et différents types d'apprentissage.

I.3 Classification automatique

La classification automatique consiste à réorganiser ou regrouper différents objets dans des classes (sous-ensembles). Dans le domaine du Machine Learning (apprentissage automatique), on trouve deux types principaux d'apprentissages : supervisé où les classes ont l'association sémantique par contre dans l'apprentissage non-supervisé il est plus difficile à déterminer l'association sémantique entre les différentes classes.

I.3.1 Apprentissage supervisé (prédictif)

Dans la classification supervisée les données entrent dans le processus de traitement sont déjà organisées et regroupées dans ce cas les algorithmes utilisés se servent à prédire des résultats.

Il existe plusieurs algorithmes et techniques : Naïve Bayes, K-NN, Arbres de classification et de régression... [2]

L'exemple suivant nous permettra d'éclairer la méthode de l'apprentissage supervisée. Nous donnons par exemple une base de données pour une application de gestion des locations automobiles. D'où chaque ligne représentant une observation (voiture, camion, motos, etc...) décrite par des variables dites prédictives (les colonnes: type de type véhicule, nombre de place, puissance, boîte auto ou manuelle pour le changement des vitesses, disponibilité de GPS, etc...), codées par des chaînes de caractères, des chiffres, des booléens, etc... Le but est de prédire le prix du véhicule souhaité. [3]

I.3.1.1 Distance euclidienne

En mathématiques, une distance est une application qui formalise l'idée intuitive de distance, c'est-à-dire la longueur qui sépare deux points. C'est par l'analyse des principales propriétés de la distance usuelle que Fréchet introduit la notion d'espace métrique, développée ensuite par Hausdorff. Elle introduit un langage géométrique dans de nombreuses questions d'analyse et de théorie des nombres.

La formule de la mesure de distance est comme suivante :

$$D(a, b) = \sqrt{\sum |a - b|^2} \quad (1)$$

I.3.1.2 Word2Vec

En intelligence artificielle et en apprentissage machine, Word2vec est un groupe de modèles utilisé pour le plongement lexical (word embedding). Ces modèles ont été développés par une équipe de recherche chez Google sous la direction de Tomas Mikolov. Ce sont des réseaux de neurones artificiels à deux couches entraînés pour reconstruire le contexte linguistique des mots. La méthode est implémentée dans la bibliothèque Python Gensim.

Deux architectures ont été initialement proposées pour apprendre les Word2vec, le modèle de sacs de mots continus (CBOW: continuous bag of words) et le modèle skip-gram2. Le CBOW vise à prédire un mot étant donné son contexte, c'est-à-dire étant donné les mots qui en sont proches dans le texte. Un tel contexte est par exemple les 5 mots à droite et les 5 mots à gauche du mot à prédire. Le skip-gram a

Chapitre I : Apprentissage Automatique et Algorithmes de Classification

une architecture symétrique visant à prédire les mots du contexte étant donné un mot en entrée.

En pratique, le modèle CBOW est plus rapide à apprendre, mais le modèle skip-gram donne généralement de meilleurs résultats.

1.3.1.3 Quelques algorithmes de l'apprentissage supervisé

Arbre de décision

La popularité de la méthode repose en grande partie sur sa simplicité. Il s'agit de trouver un partitionnement des individus que l'on représente sous la forme d'un arbre de décision. L'objectif est de produire des groupes d'individus les plus homogènes possibles du point de vue de la variable à prédire. Il est d'usage de représenter la distribution empirique de l'attribut à prédire sur chaque sommet (nœud) de l'arbre [4].

Concrètement, chaque nœud d'un arbre de décision contient un test (un IF...THEN) et les feuilles ont les valeurs Oui ou Non. Chaque test regarde la valeur d'un attribut de chaque exemple. En effet, on suppose qu'un exemple est un ensemble d'attributs/valeurs. Pour des textes, chaque attribut peut être un mot et la valeur sera par exemple 0 ou 1 selon que ce mot appartient ou non au document.

L'algorithme de l'arbre de décision suit les étapes ci-dessous :

Étape 1 : Sélectionnez la caractéristique (variable prédictive) qui classe le mieux l'ensemble de données dans les classes souhaitées et attribuez cette caractéristique au nœud racine.

Étape 2 : Descendez à partir du nœud racine, tout en prenant des décisions pertinentes à chaque nœud interne de sorte que chaque nœud interne classe au mieux les données.

Étape 3 : Revenez à l'étape 1 et répétez jusqu'à ce que vous affectiez une classe aux données d'entrée.

Exemple:

Fièvre	Douleur	Toux	Maladie
Oui	Abdomen	Non	Appendicite
Non	Abdomen	Oui	Appendicite
Oui	Gorge	Non	Rhume
Oui	Gorge	Oui	Rhume
Non	Gorge	Oui	mal de gorge
Oui	Non	Non	Aucune
Oui	Non	Oui	Rhume

Chapitre I : Apprentissage Automatique et Algorithmes de Classification

Non	Non	Oui	Refroidissement
Non	Non	Non	Aucune

Tableau 1: Exemple de données pour réaliser un arbre de décision.

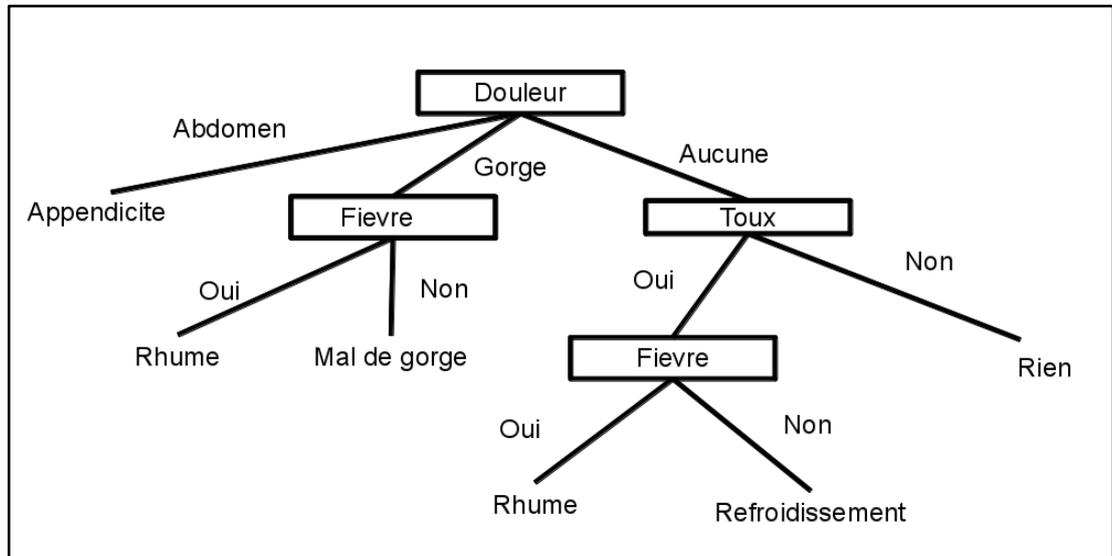


Figure 1: Exemple sur un arbre de décision sur construit à partir de tableau 1.

🌈 K Plus Proche Voisin (K-NN)

L'algorithme K-NN qui signifie k-voisins les plus proches utilise l'intégralité du data set en tant qu'entraînement, au lieu de diviser se dernier en un ensemble d'entraînement et un ensemble de test (training et testing sets en Anglais).

Quand un résultat est requis pour une nouvelle instance de données, l'algorithme KNN parcourt l'intégralité du data set pour rechercher les k-instances les plus proches de la nouvelle instance ou le nombre k d'instances les plus similaires au nouvel enregistrement, puis renvoie la moyenne des résultats ou le classe à laquelle appartient cette instance si c'est un problème de classification. L'utilisateur spécifie lui-même la valeur de k.

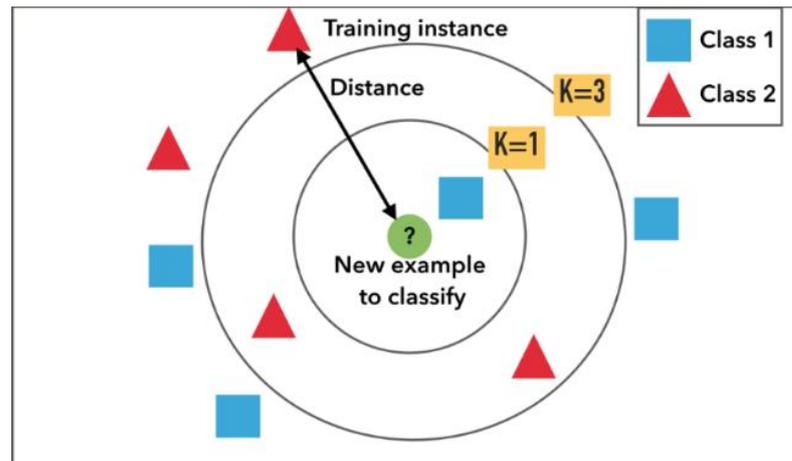


Figure 2: Exemple sur l'algorithme de classification K-NN.

✚ Classifieur Naïve Bayes

Il serait difficile et pratiquement impossible de classer manuellement une page Web, un document, un courrier électronique ou toute autre note textuelle volumineuse. C'est ici que l'algorithme d'apprentissage automatique du classificateur Naïve Bayes entre en scène. Un classificateur est une fonction qui attribue la valeur d'un élément de population à l'une des catégories disponibles. Par exemple, le filtrage du courrier indésirable est une application populaire de l'algorithme Naïve Bayes. Filtre anti-spam ici, est un classificateur qui attribue une étiquette «Spam» ou «Pas de spam» à tous les emails.

Naïve Bayes Classifier est l'une des méthodes d'apprentissage supervisé les plus populaires parmi celle utilisant les similarités, qui s'appuie sur le populaire théorème de probabilité de Bayes. En particulier pour la prédiction des maladies et la classification des documents. Il s'agit d'une classification simple de mots basée sur le théorème de probabilité de Bayes pour l'analyse subjective du contenu.

I.3.2 Apprentissage non-supervisé (descriptif)

En utilisant la classification non-supervisée on peut faire le classement des divers objets non classés dont on ne les connaît pas une classification à l'avance. Chaque objet doit appartenir à l'une des classes créées par l'algorithme de classification. On peut trouver deux types de classification non-supervisé sont : hiérarchique et non-hiérarchique [5].

Autrement dit, dans la CNS on recherche les motifs fréquents (Frequent Item Sets) puis on recherche sur les règles d'association entre eux.

✚ Classification hiérarchique (Hierarchical clustering)

Dans ce type de classification les sous-classes créés sont structurées de façon hiérarchique les uns dans les autres. Nous distinguons les Classifieurs Hiérarchiques (CH) descendantes (fractionnées) c'est de faire diviser l'ensemble de tous les individus en un certain nombre de sous-ensembles puis chaque sous-ensembles doit être découpé en un certain nombre de sous-ensembles, et ainsi de suite.

Et les CH ascendantes, dans ce cas on va commencer par des individus seuls et on les regroupe en sous-ensembles, puis on regroupe les sous-ensembles et ainsi de suite. On utilise le critère d'agrégation pour déterminer quelles classes on va fusionner.

✚ Classification non-hiérarchique

Dans ce cas les individus ne sont pas structurés de façon hiérarchique. On parle de partition si chaque individu fait partir que d'un sous-ensemble. Si chaque individu peut appartenir à plusieurs groupes, avec la probabilité p_i d'appartenir au groupe i , alors on parle de recouvrement.

1.3.2.1 Quelques algorithmes d'apprentissage non-supervisé

➤ **K-means clustering (K-moyenne)**

Algorithme des K-means ou des k-moyennes a été introduit pour la première fois par James MacQueen en 1967 [6]. C'est l'un des algorithmes de clustering les plus répandus. Il permet d'analyser un jeu de données caractérisées par un ensemble de descripteurs, afin de regrouper les données "similaires" en groupes (ou clusters). On peut résumer les traitements de cet algorithme en 4 étapes principales sont [7]:

- Étiquetez le nombre de clusters.
- Établir la coordonnée centroïde.
- Déterminez la distance de chaque objet par rapport au centroïde.
- Regroupez les objets en fonction de la distance minimale.

Habituellement, les critères de l'algorithme des k-moyennes utilisent une fonction de critère d'erreur au carré, définie comme suit [19]:

$$SSE = \sum_{i=1}^n (X_i - \bar{X})^2 \quad (2)$$

Où n est le nombre d'observations X_i est la valeur de l' $i^{\text{ème}}$ observation.

➤ **Réduction de la dimensionnalité (Dimensionality Reduction)**

Cet algorithme nous permet de convertir ou projeter des données d'un espace de grande dimension dans un espace de plus petite dimension. Cette opération est cruciale en apprentissage automatique pour lutter contre ce qu'on appelle le fléau des grandes dimensions (le fait que les grandes dimensions altèrent l'efficacité des

Chapitre I : Apprentissage Automatique et Algorithmes de Classification

méthodes). Pour que l'opération soit utile il faut que les données en sortie représentent bien les données d'entrée. Cette opération est vraiment très utile parce que les données de petite dimension peuvent être traitées plus rapidement.

Pour réduire la dimension, on peut agir sur deux leviers :

- a. Supprimer des dimensions (ou descripteurs).
- b. Combiner les variables afin d'obtenir un plus petit nombre de nouvelles variables plus expressives et/ou moins redondantes.

On distingue plusieurs méthodes, Parmi les méthodes les plus connues, on cite [8] :

- ✓ l'ACP (Analyse en composantes principales) et ses variantes, qui consistent à identifier les principales directions (combinaisons de descripteurs), i.e. celles qui concentrent le plus de variance.
- ✓ l'ACI (Analyse en composantes indépendantes) qui cherche en plus à identifier des directions orthogonales, c'est à dire décarrelées les unes des autres.
- ✓ les régularisations de type L1 ou Lasso (cas par exemple des machines à vecteurs de support ou SVM - (Support Vector Machine) qui forcent le modèle à atteindre un équilibre entre performance et nombre de dimensions retenues.
- ✓ les algorithmes sparses, comme par exemple l'apprentissage par dictionnaire sparse, dont le but est de représenter un signal complexe comme une combinaison d'un nombre réduit de signaux élémentaires simples.

➤ Analyse des composants principaux (Principal Component Analysis)

Cet algorithme nous permet d'analyser et visualiser des données contenant des individus décrits par plusieurs variables quantitatives.

Comprendre les détails de l'ACP nécessite une connaissance de l'algèbre linéaire. Ici, nous n'expliquerons que les bases avec une représentation graphique simple des données.

Dans le Plot 1A ci-dessous, les données sont représentées dans le système de coordonnées X-Y. La réduction de la dimension est obtenue en identifiant les directions principales, appelées composantes principales, dans lesquelles les données varient.

L'ACP suppose que les directions avec les plus grandes variances sont les plus "importantes" (i.e., principales).

Dans la figure ci-dessous, l'axe PC1 est le premier axe principal le long duquel les échantillons présentent la plus grande variation. L'axe PC2 est la seconde direction la plus importante et orthogonal à l'axe PC1.

Chapitre I : Apprentissage Automatique et Algorithmes de Classification

Les dimensions de notre jeu de données peuvent être réduites à une seule dimension en projetant chaque échantillon sur le premier axe principal (Plot 1B) [9].

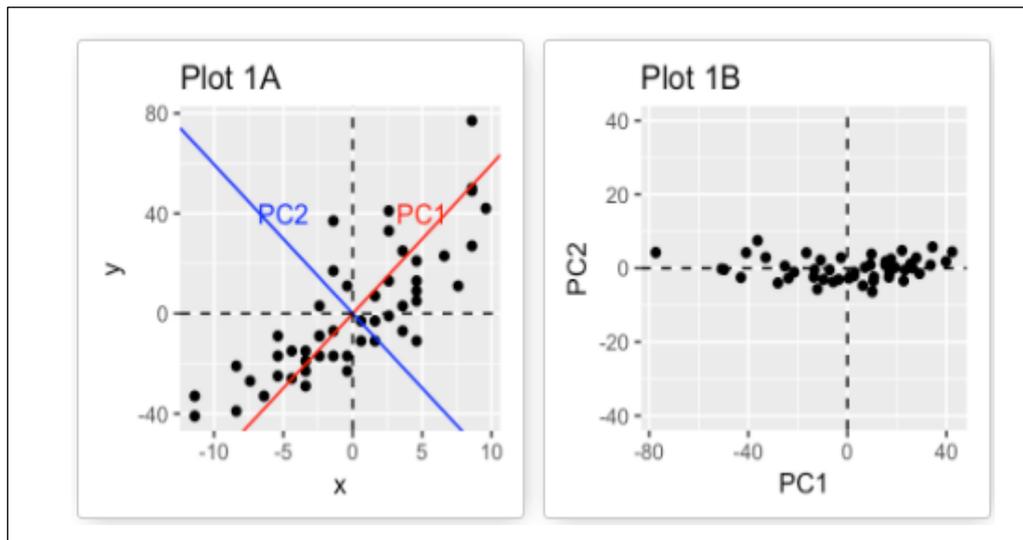


Figure 3 : Exemple d'identification des directions principales

➤ Décomposition en valeur singulière (Singular Value Decomposition)

SVD d'une matrice est un outil important de factorisation des matrices rectangulaires réelles ou complexes. Ses applications s'étendent du traitement du signal aux statistiques, en passant par la météorologie.

Le théorème spectral énonce qu'une matrice normale peut être diagonalisée par une base orthonormée de vecteurs propres. On peut voir la décomposition en valeurs singulières comme une généralisation du théorème spectral à des matrices arbitraires, qui ne sont pas nécessairement carrées.

On appelle valeur singulière de M toute racine carrée d'une valeur propre de M^*M , autrement dit tout réel positif σ tel qu'il existe un vecteur unitaire u dans K^m et un vecteur unitaire v dans K^n vérifiant :

$$M * u = \sigma v \text{ et } Mv = \sigma u \quad (3)$$

Les vecteurs u et v sont appelés vecteur singulier à gauche et vecteur singulier à droite pour σ , respectivement.

Dans toute décomposition en valeurs singulières,

$$M = U\Sigma V^* \quad (4)$$

Chapitre I : Apprentissage Automatique et Algorithmes de Classification

Les coefficients diagonaux de Σ sont égaux aux valeurs singulières de M . Les colonnes de U et de V sont, respectivement, vecteur singulier à gauche et à droite pour les valeurs singulières correspondantes.

I.3.3 Apprentissage semi-supervisé

L'apprentissage semi-supervisé utilise un ensemble de données étiquetées et non-étiquetées. Il se situe ainsi entre l'apprentissage supervisé qui n'utilise que des données étiquetées et l'apprentissage non-supervisé qui n'utilise que des données non-étiquetées. L'utilisation de données non-étiquetées, en combinaison avec des données étiquetées, permet d'améliorer de façon significative la qualité de l'apprentissage. Un autre avantage vient du fait que l'étiquette de données nécessite l'intervention d'un utilisateur humain. Lorsque les jeux de données deviennent très grands, cette opération peut s'avérer fastidieuse. Dans ce cas, l'apprentissage semi-supervisé, qui ne nécessite que quelques étiquettes, revêt un intérêt pratique évident et indiscutable [10].

I.3.4 Apprentissage partiellement supervisé

Probabiliste ou non, il est utilisé lorsque les données intégrées au départ dans le système sont partiellement étiquetées. Par exemple, des données sont définies appartenant à A , B ou C et d'autres n'appartenant pas à A mais peut-être à B ou C .

Entraînement d'un modèle avec des données où seulement certains des exemples d'apprentissage sont étiquetés. L'une des techniques d'apprentissage partiellement supervisé consiste à déduire les étiquettes des exemples sans étiquette, puis à entraîner le modèle avec les étiquettes déduites afin de créer un nouveau modèle. L'apprentissage partiellement supervisé peut être utile si les étiquettes sont coûteuses, mais que les exemples sans étiquette abondent.

I.3.5 Apprentissage par renforcement

L'apprentissage par renforcement consiste, pour un agent autonome (robot, etc.), à apprendre les actions à prendre, à partir d'expériences, de façon à optimiser une récompense quantitative au cours du temps. L'agent est plongé au sein d'un environnement, et prend ses décisions en fonction de son état courant. En retour, l'environnement procure à l'agent une récompense, qui peut être positive ou négative. L'agent cherche, au travers d'expériences itérées, un comportement décisionnel (appelé stratégie ou politique, et qui est une fonction associant à l'état courant l'action à exécuter) optimal, en ce sens qu'il maximise la somme des récompenses au cours du temps.

L'apprentissage par renforcement est utilisé dans plusieurs applications : robotique, gestion de ressources, vol d'hélicoptères, chimie. Cette méthode a été appliquée avec

Chapitre I : Apprentissage Automatique et Algorithmes de Classification

succès à des problèmes variés, tels que le contrôle robotique, le pendule inversé, la planification de tâches, les télécommunications, le backgammon et les échecs.

I.3.6 Apprentissage par transfert

Capacité d'un système à appliquer des connaissances ou compétences apprises au cours de tâches antérieures sur de nouvelles tâches comportant des similitudes. L'apprentissage se fait donc en fonction d'un apprentissage effectué auparavant.

Utilise une simulation du cerveau humain, connue sous le nom de réseaux neuronaux. Ces réseaux neuronaux sont constitués de neurones qui sont similaires à l'unité fondamentale du cerveau humain. Les neurones constituent un modèle de réseau neuronal et ce domaine d'étude est appelé deep learning.

Le résultat final d'un réseau neuronal est appelé modèle de deep learning. La plupart du temps, dans le deep learning, on utilise des données non structurées à partir desquelles le modèle de deep learning extrait des caractéristiques par lui-même grâce à un entraînement répété sur les données. Ces modèles conçus pour un ensemble particulier de données peuvent être utilisés comme point de départ pour le développement d'un autre modèle avec un ensemble différent de données et de caractéristiques, ce que l'on appelle transfer learning (apprentissage par transfert). En termes simples, l'apprentissage par transfert est une méthode populaire où un modèle développé pour une tâche particulière est à nouveau utilisé comme point de départ pour développer un modèle pour une autre tâche.

Pour comprendre la notion de base de l'apprentissage par transfert, considérons qu'un modèle X est entraîné avec succès à effectuer la tâche A avec le modèle M1. Si la taille de l'ensemble de données pour la tâche B est trop petite, empêchant le modèle Y de s'entraîner efficacement ou provoquant un surajustement des données, nous pouvons utiliser une partie du modèle M1 comme base pour construire le modèle Y afin d'effectuer la tâche B.

I.4 Conclusion

Dans cette partie, nous avons fait une étude sur la notion d'apprentissage automatique et nous avons survolé quelques-uns des algorithmes utilisés aussi bien en mode : supervisé, semi-supervisé, ou non supervisé. Dans le prochain chapitre, nous nous allons présenter la conception de notre solution.

Chapitre II : Conception et modélisation de la solution proposée

Chapitre II : Conception et modélisation de la solution proposée

II.1 Introduction

Dans ce chapitre nous allons voir la description des étapes de la réalisation de notre système d'Extraction des Caractéristiques basé sur l'Apprentissage Automatique dans les Hadiths.

II.2 Architecture globale

Afin de choisir les meilleurs algorithmes pour notre contribution, nous avons d'abord effectué des recherches sur les différentes sectes dans différentes ressources afin de pouvoir réaliser un Dataset exploitable.

Le schéma qui suit (figure 4), représente l'architecture sommaire de notre solution.

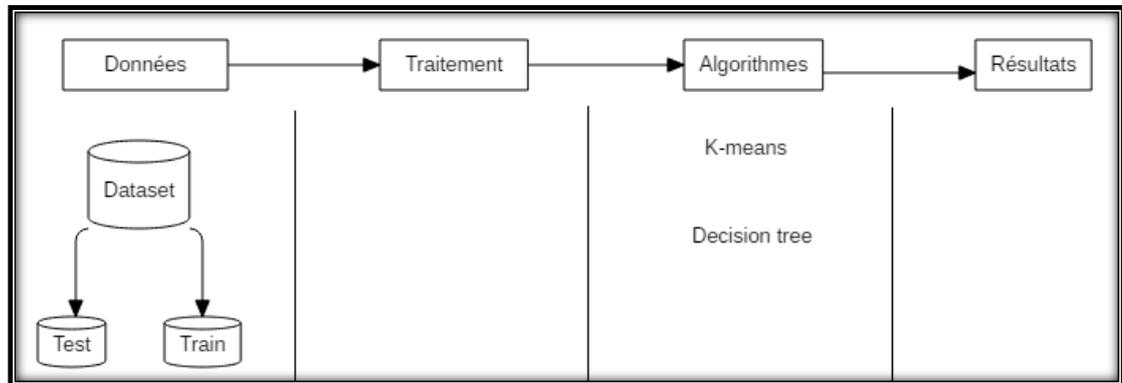


Figure 4 : Architecture globale du système

Dans les sections suivantes, nous allons décrire chacun des composants de cette architecture.

II.3 Description des données

Dans cette sous-section, nous allons nous intéresser aux données d'entrée du système.

II.3.1 Données en entrée

Avant de nous focaliser sur le livre de la différence entre les sectes (الفرق بين الفرق) (للعالم عبد القاهر بن طاهر بن محمد البغدادي الإسفرائيني التميمي الملل و) qui a été écrit par Chahristani (النحل للعالم ابن ابي بكر الشهرستاني) et qui est considéré comme une référence dans le domaine de la description des différentes sectes de l'Islam, mais à cause de son volume important, nous avons dans un premier temps essayer de nous intéresser à un seule tome de son œuvre afin de pouvoir limiter le travail. Par ailleurs, nous avons eu

Chapitre II : Conception et modélisation de la solution proposée

recours à différentes sources web afin de pouvoir avoir une vision claire de l'état actuel des choses, mais c'était très dur et très confus sans la présence d'un expert à nos côtés et si ce n'est 'intervention de notre promoteur, Monsieur Cherif Zahar, nous n'aurions jamais pu trouver une issue.

La figure qui suit représente un récapitulatif des principaux courants de l'Islam :

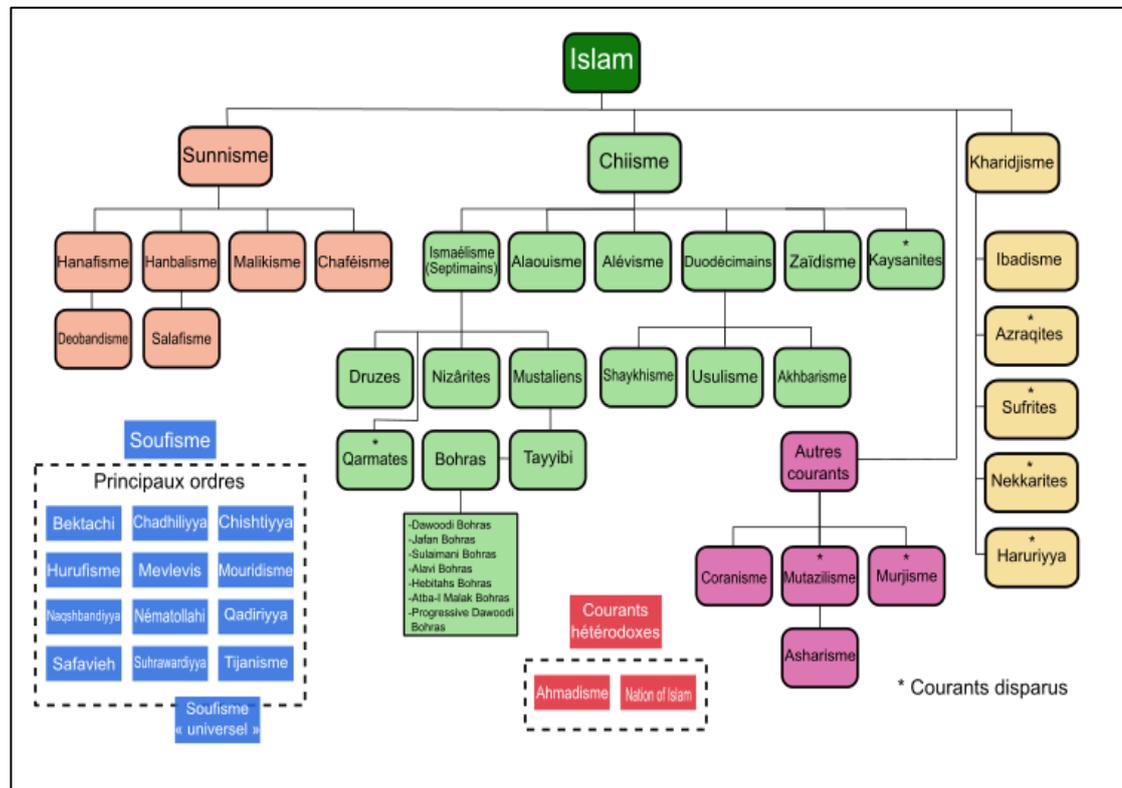


Figure 5: Les principaux courants de l'Islam.

Avant de détailler les différentes étapes de l'architecture globale montrée dans la figure 4, il est à noter que jusqu'au jour d'aujourd'hui, les tests n'ont pas pu être effectués et toute la partie qui concerne le Traitement Automatique de la Langue et l'Apprentissage Automatique et toutes les expérimentations derrière est restée au stade théorique.

II.3.2 Source de données

Pour effectuer les expérimentations et évaluations nécessaires, nous avons d'abord pensé à la meilleure représentation des données. Dans un premier temps, une lecture approfondie de l'œuvre « الملل و النحل للعالم ابن ابي بكر الشهرستاني » nous a permis de faire une première ébauche du corpus. Suivant, la répartition suivante :

الفرق	العلماء والكتب	الخصائص	فرق ثانوية
-------	----------------	---------	------------

Chapitre II : Conception et modélisation de la solution proposée

Et voici une capture d'écran comprenant deux lignes du dataset initial :

المجموع = 80	الفرق	الغناء والكتب	الخصائص	فرق تاتوية
1	أهل السنة والجماعة	أبو حنيفة الإمام مالك الشافعي أحمد بن حنبل الشيخان البخاري ومسلم أصحاب السنن التسلي أبو داود، والترمذي ابن ماجه الإمام النووي ابن قيم الجوزية ابن تيمية		كَلْبِيَّة
1	فرقة الأشعرية	أبي الحسن الأشعري البيهقي الباقلاني القشيري الجويي الغزالي الفخر الرازي النووي	لا يكفرون أحداً من أهل القبلة بتقديم النص على العقل جعلوا العقل متخلاً في فهم النص تأويل بعض صفات الله عز وجل يعتمد على مخالفة عواطف البرء استخدموا الحجج العقلية الجمع بين العاطفة والعقل (يعتمد على مخالفة عواطف البرء)	

Figure 6: Capture d'écran du premier dataset.

L'idée de départ était de collecter toutes les caractéristiques et détails des différentes sectes présentes dans le premier tome du livre ainsi que des informations sur quelques-uns des savants qui s'y sont intéressés afin de pouvoir recueillir certains de leurs travaux intéressants dans le domaine et constituer ainsi une collection de document sur laquelle nous pourrions faire nos tests.

L'idée était de permettre à un utilisateur de saisir le nom d'une secte ou ses caractéristiques et d'avoir en retour les informations sur la secte la plus proche de sa recherche et le cas échéant, les quelques documents qui pourraient lui donner plus d'information sur le sujet. Mais le travail était laborieux et en l'absence d'un spécialiste en Religion musulmane, l'idée est bien vite devenue une voie sans issue.

Par la suite, le livre de la différence entre les sectes (الفرق بين الفرق للعالم عبد القاهر) (بن طاهر بن محمد البغدادي الإسفرائيني التميمي) a été retenu car il offrait une meilleure prise en considération de notre cas d'étude en offrant une version simplifiée et facilement assimilable pour une éventuelle collecte d'information.

Ainsi, la codification suivante a été adoptée :

- Chaque secte pouvait comportait une à plusieurs sous-sectes (ou sectes filles) qui se partageait les mêmes croyances.
- Chaque secte (respectivement secte fille) possédait un certain nombre de caractéristiques desquels, nous nous sommes intéressés uniquement aux quinze les plus importants et qui sont représentés dans le tableau suivant :

Tableau 2: Caractéristiques retenues pour les sectes musulmanes.

Chapitre II : Conception et modélisation de la solution proposée

d1
الخلافة في قریش
الخلافة في جميع المسلمين
الإمامة في علي وبعض ذريته من بعده
d2
الخلافة شورى
الخلافة تنصيب
يجوز الاستخلاف كما تجوز الشورى
الخلافة انتزاع مع اقتدار
d3
الإيمان اعتقاد بالجنان
الإيمان اعتقاد بالجنان وقول باللسان
الإيمان اعتقاد بالجنان وقول باللسان وعمل بالجوارح
d4
تفويض معاني صفات الله
صفات الله كصفات المخلوقين
إنكار السماء والصفات جميعا
إنكار الصفات وإثبات الأسماء
إنكار بعض الصفات وإثبات بعضها وإثبات الأسماء
صفات الله وأسماءه كما جاءت في القرآن والسنة من غير تحريف ولا تعطيل ولا تكييف ولا تمثيل
d5
الإمامة ركن من أركان الدين
الخلافة واجب شرعي
d6
خلافة الخلفاء الثلاثة صحيحة
خلافتهم باطلة
خلافة كل الخلفاء الذين تمت لهم البيعة صحيحة
خلافة الأولين صحيحة فقط دون الثالث والرابع
d7
العاصي كافر
العاصي مؤمن كامل الإيمان
العاصي لا مؤمن ولا كافر وهو مخلد في النار
العاصي مسلم وهو تحت مشيئة الله إن مات من غير توبة
d8
القرآن مخلوق
القرآن كلام الله غير مخلوق
d9

Chapitre II : Conception et modélisation de la solution proposée

الله بائن من خلقه
الله يحل في بعض خلقه
d10
ليس في الوجود خالق ومخلوق بل الكل هو الله
ثانوية الوجود أي هنالك خالق ومخلوق
d11
الإنسان مكلف من البلوغ إلى الموت
سقوط التكاليف ببلوغ درجة معينة
d12
الخروج على أئمة الجور غير جائز
الخروج على أئمة الجور واجب
d13
الجنة والنار مهلوقتان الآن
الجنة والنار غير مخلوقتان الآن
d14
الجنة والنار لا تفنيان ولا تبديدان
الجنة لا تفنى والنار تفنى
كلاهما يفنى
d15
رؤية الله في الآخرة حاصلة للمؤمنين
رؤية الله في الآخرة غير ممكنة

L'idée était de recueillir toutes les informations qui pouvait être recueillie à partir du livre et ce, en prenant en considération les caractéristiques ci-dessus. En assignant à chacune des caractéristique un poids afin de constituer la matrice allant être exploitée par les algorithmes de classification.

Le résultat de cette collecte est disponible en deux formats :

1. « SQL » pour que les données puissent être analysées et manipulées via la plateforme.
2. « CSV » pour pouvoir effectuer les tests d'apprentissage automatique.

Les deux figures qui suivent représentent un extrait de ces deux formats :

Chapitre II : Conception et modélisation de la solution proposée

<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	idcaracteristique	categorieCaracteristique	Poids
<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	1	الخلافة والإمامة	0.3
<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	2	الخلافة والشورى	0.45
<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	3	الإيمان	1
<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	4	صفات الله وأسماءه	1
<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	5	الإمامة في الدين	0.23
<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	6	خلافة الخلفاء	0.82
<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	7	حكم العاصي	0.56
<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	8	قضية خلق القرآن الكريم	1
<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	9	الله واحد	1
<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	10	الله هو الخالق	1
<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	11	حكم التكليف التكليف	0.22
<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	12	الخروج على أئمة الجور	0.98
<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	13	خلق الجنة والنار	0.88
<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	14	فناء الجنة والنار	0.46
<input type="checkbox"/>	 Éditer	 Copier	 Supprimer	15	رؤية الله في الآخرة	0.76

Figure 7: liste des caractéristiques dans notre base de données.

Chapitre II : Conception et modélisation de la solution proposée

1	sec.D1,D2,D3,D4,D7,D8,D9,D10,D11,D12,D14,D15
2	1,0,0,0,0.96,0.8,0,0,6,0.8,0,0,0,0
3	2,0,0,0,0,0.8,0,0,6,0.8,0,0,0,0
4	3,0,0,0,0,0,0,0.6,0.8,0,0,0,0
5	4,0,0,1,0.96,0,0,0,6,0.8,0,0,0,1,0
6	5,0,0,0,0.96,0.8,0,0,6,0.8,0,0,0,0
7	6,0,0,1,0,0,0,0.6,0.8,0.135,0,0,0
8	7,0,0,1,0,0,0,0.6,0.8,0.135,0,0,0
9	8,0,0,1,0.96,0,0,0,6,0.8,0.135,0,0,0
10	9,0,0,1,0.96,0,0,0,6,0.8,0,0,0,0
11	10,0,0,1,0,0,0.23,0.6,0.8,0,0,0,0
12	11,0,0,1,0.96,0.8,0,0,6,0.8,0,0,0,0
13	12,0,0,1,0.96,0.8,0,0,6,0.8,0,0,0,0
14	13,0,0,0,0,0,0.23,0.6,0.8,0,0,0,0
15	14,0,0,0,0,0,0.23,0.6,0.8,0,0,0,0
16	15,0,0,0,0,0,0.23,0.6,0.8,0,0.5,0,0
17	16,0,0,1,0.96,0.8,0,0,6,0.8,0,0,0,1,0.33
18	17,0,0,0,0.96,0.8,0,0,6,0.8,0,0,0,0.33
19	18,0.7,0,0,0.96,0.8,0,0,0,0,0,0.33
20	19,0,0.8,1,0.96,0.8,0,0,6,0,0.135,0,0,0
21	23,0.7,0,0,0,0.8,0,0,0,0,0,0,0
22	24,0,0.8,0,0,0.8,0,0,0,0,0,0,0
23	25,0,0.8,0,0,0.8,0,0,0,0,0,0,0
24	26,0.7,0,0,0,0.8,0,0,0,0,0,0,0
25	27,0.7,0,0,0,0,0,0,0,0,0,0,0
26	28,0.7,0,0,0,0,0,0,0,0,0,0,0
27	29,0.7,0,0,0,0,0,0,0,0,0,0,0
28	30,0,0.8,0,0,0,0,0,0,0,0,0,0
29	31,0.7,0,0,0,0,0,0,0,0,0,0,0
30	32,0.7,0,0,0,0,0,0,0,0,0,0,0
31	33,0.7,0,0,0.96,0,0,0,0.135,0,0,0
32	34,0,0,0,0.8,0,0,0,0,0,0,0,0
33	35,0,0,0,0,0,0,0,0,0,0,0,0
34	36,0,0,0,0.8,0,0,0,0,0,0,0,0
35	37,0,0,0.96,0.8,0,0,0,0,0,0,0,0
36	38,0,0,1,0.8,0.23,0.6,0.8,0.135,0.5,0.1,0
37	39,0,0,1,0,0.8,0.23,0.6,0.8,0.135,0.5,0.1,0.33

Figure 8: Quelques lignes de notre fichier Csv.

II.4 Expression des besoins

II.4.1 Identification des acteurs

La plateforme devra pouvoir être utilisée par trois types d'utilisateurs :

1. Un administrateur :

Les tâches de ce dernier peuvent inclure, entre autres :

- Gestion des comptes utilisateurs :
 - Validation des comptes (Vérification de la crédibilité des experts).
 - Ajouts, visualisation, suppression, ou modification des informations de comptes.
- Opérations de visualisation :
 - Visualiser le nombre des sectes et des sous-secte et aussi pour les caractéristique et les utilisateurs.

Chapitre II : Conception et modélisation de la solution proposée

2. Des experts en religion musulmane :

Les tâches de ces derniers peuvent inclure, entre autres :

- Gestion de leur propre compte utilisateur :
 - Demande de création de compte en saisissant les informations personnelles et en fournissant de pièces justificatives (facultatives).
 - Ajouts, visualisation, suppression, ou modification des informations de comptes.
- Alimentation de la base des connaissances :
 - Ajouter des sectes et de leurs spécificités (ajout, modification, suppression...)
 - Ajouter des ressources pour la classification
 - Ajouter des ressources par secte
 - Ajouter des Eulma et autres détails par secte.
 - Envoi de messages ou de commentaires à l'administrateur ou experts.
- Opérations de visualisation :
 - Même tâches que celles spécifiées pour l'administrateur.

3. Des utilisateurs lambdas :

Ces derniers n'ont pas besoin de compte et leurs tâches se limitent aux opérations de visualisation vues précédemment. Ces derniers peuvent éventuellement utiliser le formulaire de contact pour envoyer un message à l'administrateur.

II.4.2 Diagrammes des cas d'utilisation

II.4.2.1 Diagramme de cas d'utilisation global

La figure ci-dessous (figure 9) représente le diagramme de cas d'utilisation global de notre application.

Chapitre II : Conception et modélisation de la solution proposée

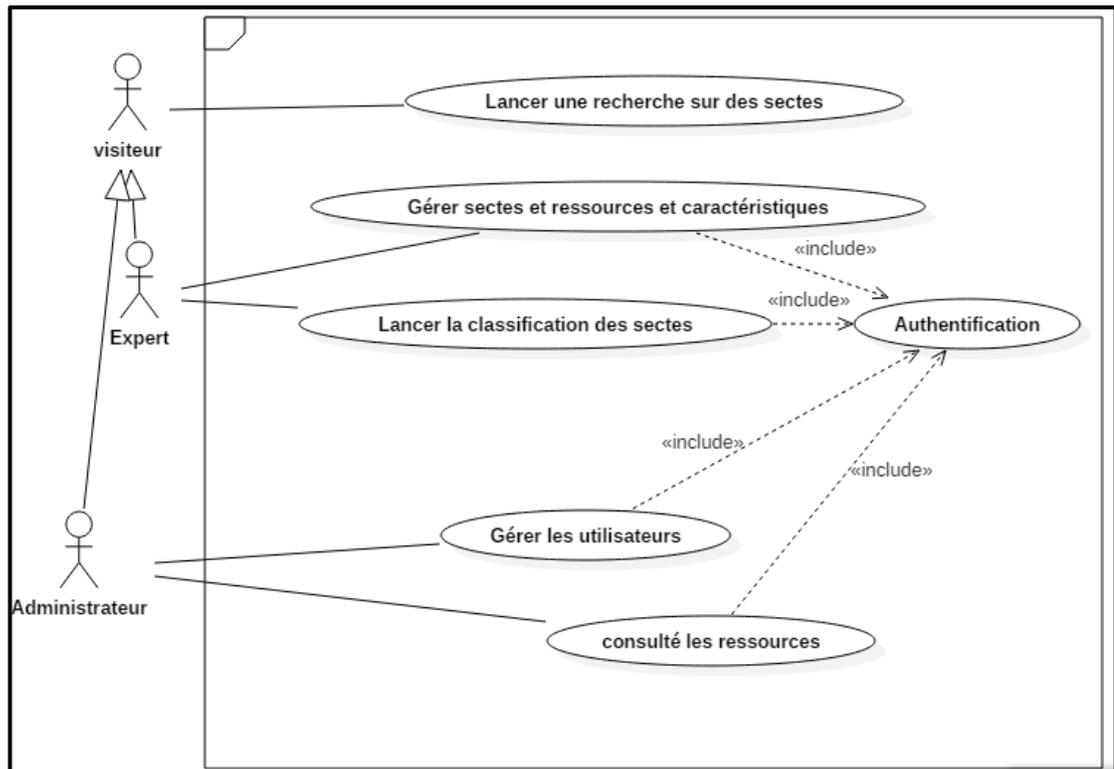


Figure 9 : Diagramme de cas d'utilisation global.

Le tableau ci-dessous décrit ce diagramme de cas d'utilisation :

Tableau 3: Description du diagramme des cas d'utilisation global.

Cas d'utilisation	Authentification
Acteur	Expert.
But	Accéder à la page appropriée.
Description	L'accès à la page approprié doit passer par le formulaire qui contient les renseignements adéquats.
Pré-condition	Les informations doivent être les mêmes de celles il a fourni lors de l'inscription (le nom d'utilisateur et le mot de passe)
Post-condition	L'utilisateur peut effectuer les tâches qui lui sont permises.
Exception	Annulation ; Si l'utilisateur (expert ou administrateur) tape un nom d'utilisateur ou un mot de passe qui ne convient pas, le système affiche un message d'erreur.
Cas d'utilisation	Lancer une recherche sur des sectes
Acteur	Expert.
But	Faire des recherches.
Description	Pour faire une recherche il faut accéder à la page de la recherche et taper les informations dans la barre de la recherche.
Pré-condition	La barre de la recherche n'est pas vide (il faut écrire quelques chose).

Chapitre II : Conception et modélisation de la solution proposée

Post-condition	Pas de post-condition.
Exception	Pas d'exception.
Cas d'utilisation	Lancer la classification des sectes
Acteur	Expert.
But	Appliquer les différents algorithmes de classification sur notre base de données (Dataset).
Description	Pour faire lancer un algorithme de classification il faut choisir le type de l'algorithme KNN, Naïve Bayes ...
Pré-condition	Les informations à traité sont préparées pour appliquer l'algorithme.
Post-condition	Expert peut effectuer les tâches qui lui sont permises.
Exception	Pas d'exception.

II.4.2.2 Diagramme de cas d'utilisation « Gérer secte »

La figure ci-dessous représente le diagramme de cas d'utilisation « Gestion des sectes ».

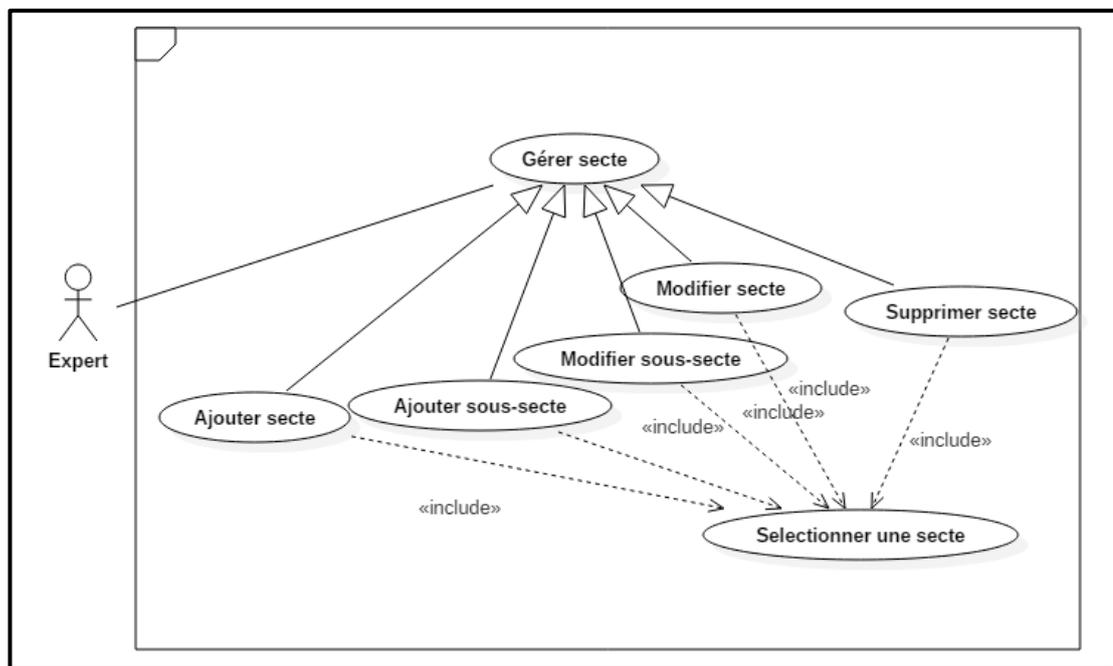


Figure 10 : Diagramme de cas d'utilisation « Gestion des sectes ».

Tableau 4: Description du diagramme de cas d'utilisation "Gestion des sectes".

Cas d'utilisation	Gérer secte
	Ajouter secte
Acteur	Expert.
But	Ajouter secte.
Description	L'ajout d'une secte doit passer par un formulaire qui contient les renseignements adéquats.
Pré-condition	La secte, n'existe pas dans la base de données.

Chapitre II : Conception et modélisation de la solution proposée

Post-condition	Expert peut effectuer les tâches qui lui sont permises.
Exception	Annulation ; si l'expert ajoute une secte qui existe, Le système affiche un message « secte existe Déjà »
<i>Ajouter sous-secte</i>	
Acteur	Expert.
But	Ajouter sous-secte.
Description	L'ajout d'une sous-secte doit passer par un formulaire qui contient les renseignements adéquats.
Pré-condition	La sous-secte, n'existe pas dans la base de données.
Post-condition	Expert peut effectuer les tâches qui lui sont permises.
Exception	Annulation ; si l'expert ajoute une sous-secte qui existe, Le système affiche un message « secte existe Déjà »
<i>Supprimer secte</i>	
Acteur	Expert.
But	Supprimer secte
Description	La suppression d'une secte se trouve dans l'interface de secte insérer par expert.
Pré-condition	La secte existe dans la base de données et ajoutée par l'expert.
Post-condition	Expert peut effectuer les tâches qui lui sont permises.
Exception	Annulation ; si l'expert supprime une secte qui a un ou plusieurs sous-sectes, Le système affiche un message « secte a des sous-sectes »
<i>Supprimer sous-secte</i>	
Acteur	Expert.
But	Supprimer sous-secte
Description	La suppression d'une sous-secte se trouve dans l'interface de secte insérées par expert.
Pré-condition	La secte existe dans la base de données et ajoutée par l'expert.
Post-condition	Expert peut effectuer les tâches qui lui sont permises.
Exception	Pas d'exception.
<i>modifier secte</i>	
Acteur	Expert.
But	Modifier secte
Description	La modification d'une secte doit passer par un formulaire qui contient les renseignements adéquats.
Pré-condition	Expert sélectionne une secte existe déjà dans la base de données.
Post-condition	Pas de post-condition.
Exception	Pas d'exception
<i>modifier sous-secte</i>	
Acteur	Expert.
But	Modifier sous-secte
Description	La modification d'une sous-secte doit passer par un formulaire qui contient les renseignements adéquats.
Pré-condition	Expert sélectionne une sous-secte existe déjà dans la base

Chapitre II : Conception et modélisation de la solution proposée

	de données.
Post-condition	Pas de post-condition.
Exception	Pas d'exception.

II.4.2.3 Diagramme de cas d'utilisation « Gérer ressources »

La figure ci-dessous représente le diagramme de cas d'utilisation de « Gestion des ressources »

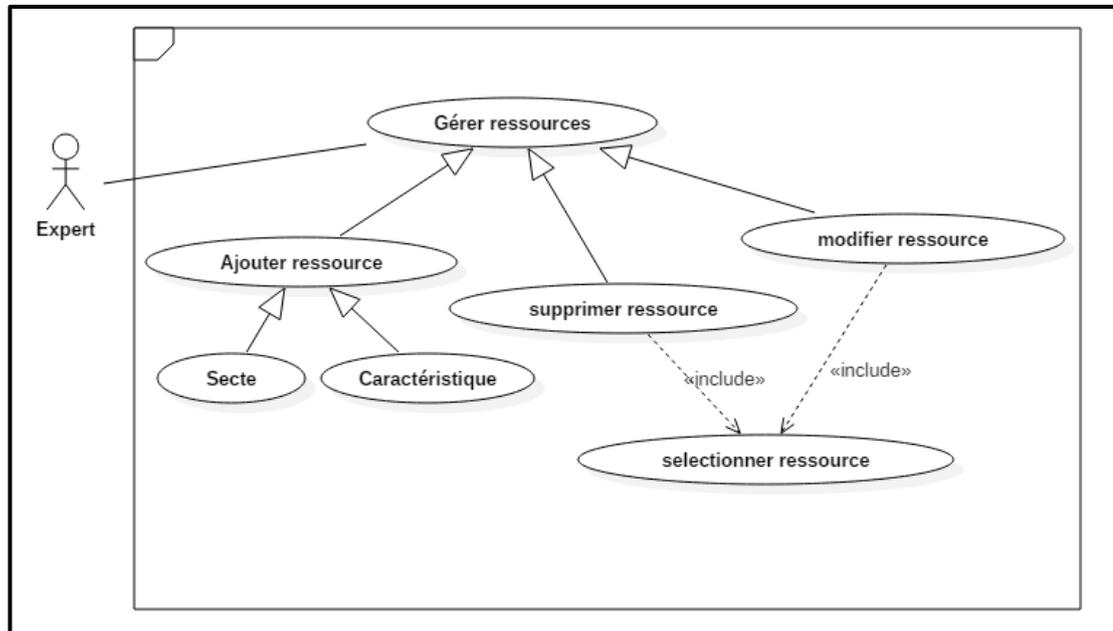


Figure 11 : Diagramme de cas d'utilisation « Gestion ressources ».

Le tableau ci-dessous décrit ce diagramme :

Tableau 5: Description du cas d'utilisation "Gestion des ressources".

Cas d'utilisation	Gérer ressources
Ajouter caractéristique	
Acteur	Expert.
But	Ajouter caractéristique.
Description	L'ajout d'une caractéristique doit passer par un formulaire qui contient les renseignements adéquats.
Pré-condition	La caractéristique n'existe pas dans la base de données.
Post-condition	Expert peut effectuer les tâches qui lui sont permises.
Exception	Annulation ; si l'expert ajoute caractéristique qui existe, Le système affiche un message « caractéristique existe Déjà »
Supprimer caractéristique	
Acteur	Expert.
But	Supprimer caractéristique
Description	La suppression d'une caractéristique se trouve dans

Chapitre II : Conception et modélisation de la solution proposée

	l'interface de caractéristique insérées par expert.
Pré-condition	Caractéristique, existe dans la base de données et ajoutée par l'expert.
Post-condition	Expert peut effectuer les tâches qui lui sont permises.
Exception	Pas d'exception.
<i>modifier caractéristique</i>	
Acteur	Expert.
But	Modifier une caractéristique
Description	La modification d'une caractéristique doit passer par un formulaire qui contient les renseignements adéquats.
Pré-condition	Expert sélectionne une caractéristique existe déjà dans la base de données.
Post-condition	Pas de post-condition.
Exception	Pas d'exception.

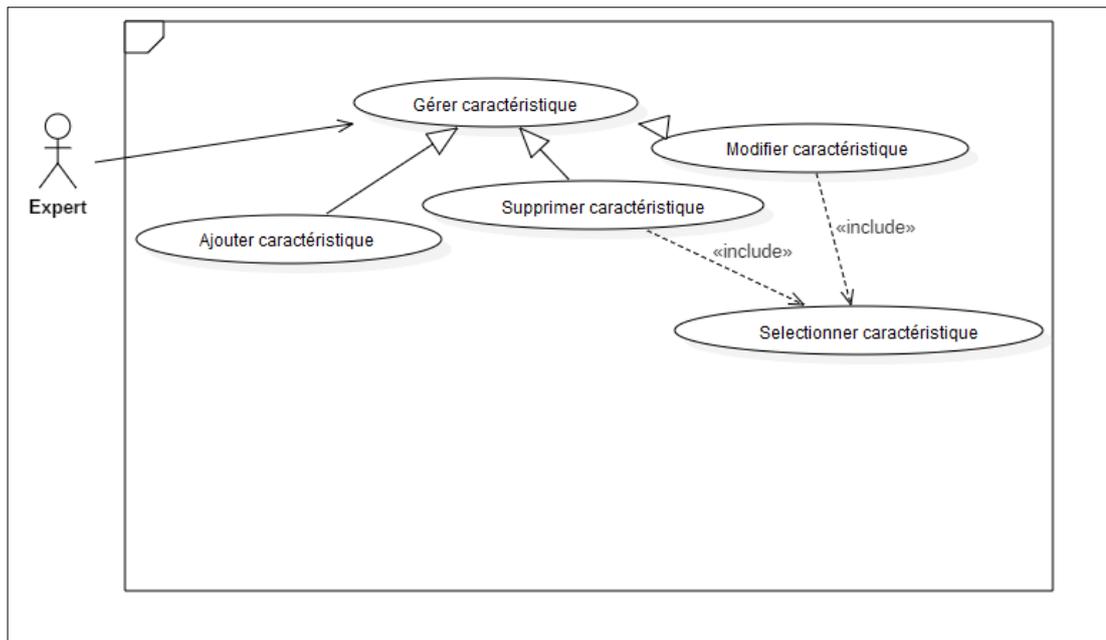


Figure 12 : Diagramme de cas d'utilisation « Gestion caractéristique ».

II.4.2.4 Diagramme de cas d'utilisation « Gestion des utilisateurs »

La figure ci-dessous présente le diagramme de cas d'utilisation de « Gestion des utilisateurs ».

Chapitre II : Conception et modélisation de la solution proposée

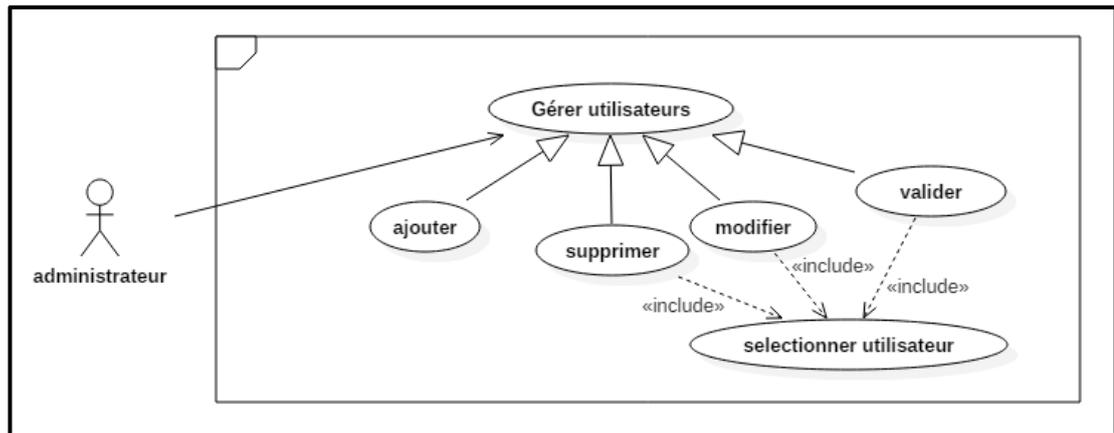


Figure 13 : Diagramme de cas d'utilisation « Gérer utilisateurs »

Le tableau ci-dessous décrit ce diagramme :

Tableau 6: Description du diagramme de cas d'utilisation "Gestion des utilisateurs".

Cas d'utilisation	Gérer utilisateurs
Ajouter	
Acteur	administrateur.
But	Ajouter caractéristique.
Description	L'ajout d'une caractéristique doit passer par un formulaire qui contient les renseignements adéquats.
Pré-condition	La caractéristique n'existe pas dans la base de données.
Post-condition	Expert peut effectuer les tâches qui lui sont permises.
Exception	Annulation ; si l'expert ajoute caractéristique qui existe, Le système affiche un message « caractéristique existe Déjà »
Supprimer	
Acteur	administrateur.
But	Supprimer caractéristique.
Description	La suppression d'une caractéristique se trouve dans l'interface de caractéristique insérées par expert.
Pré-condition	Caractéristique, existe dans la base de données et ajoutée par l'expert.
Post-condition	Expert peut effectuer les tâches qui lui sont permises.
Exception	Pas d'exception.
Modifier	
Acteur	Administrateur
But	Modifier une caractéristique
Description	La modification d'une caractéristique doit passer par un formulaire qui contient les renseignements adéquats.
Pré-condition	Expert sélectionne une caractéristique existe déjà dans la base de données.
Post-condition	Pas de post-condition.
Exception	Pas d'exception.

II.5 Analyse des besoins

Dans cette section, nous analysons les différents besoins de notre application à travers les diagrammes de séquences.

II.5.1 Diagramme de séquence « Authentification »

Le scénario du diagramme de séquence « Authentification » est le suivant :

1. Le formulaire d'authentification est affiché à l'utilisateur.
2. L'utilisateur saisi ces informations personnelles.
3. Le système vérifie les informations avec la base de données.
4. Une vérification est effectuée au niveau de la base de données.
5. Si les informations sont correctes, l'utilisateur est orienté vers son espace personnel.
6. Si les informations sont erronées, l'utilisateur est redirigé vers l'interface d'accueil.

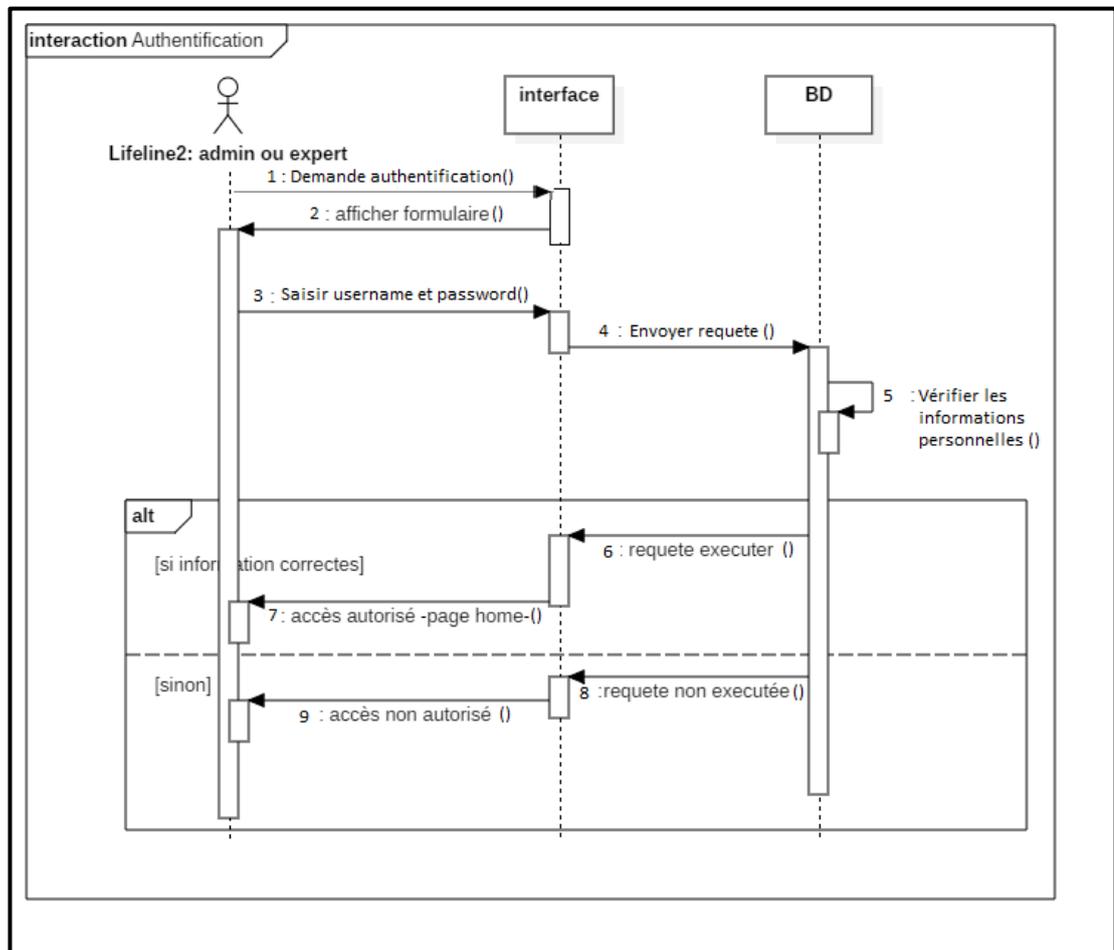


Figure 14 : Diagramme de séquences « Authentification ».

Chapitre II : Conception et modélisation de la solution proposée

II.5.2 Diagramme de séquence « Création compte pour un expert »

Le scénario du diagramme de « Création de comptes » est le suivant :

1. Formulaire d'inscription est affiché pour le demandeur.
2. L'utilisateur saisi et remplit les champs avec les informations demandées.
3. Vérification des informations saisies comme l'email et le mot de passe.
4. Si le nom d'utilisateur existe ou est déjà utilisé, alors un message d'erreur s'affiche pour demander de le changer.
5. Si les mots de passes saisis ne sont pas identiques, le système affiche un message d'erreur pour indiquer ce problème.
6. Si toutes les informations saisies sont correctes alors le système confirme à l'utilisateur que le compte a été créé avec succès et qu'il faut attendre la validation de l'administrateur pour commencer à l'utiliser.
7. Si c'est l'administrateur qui crée le compte, dans ce cas le compte est automatiquement validé.

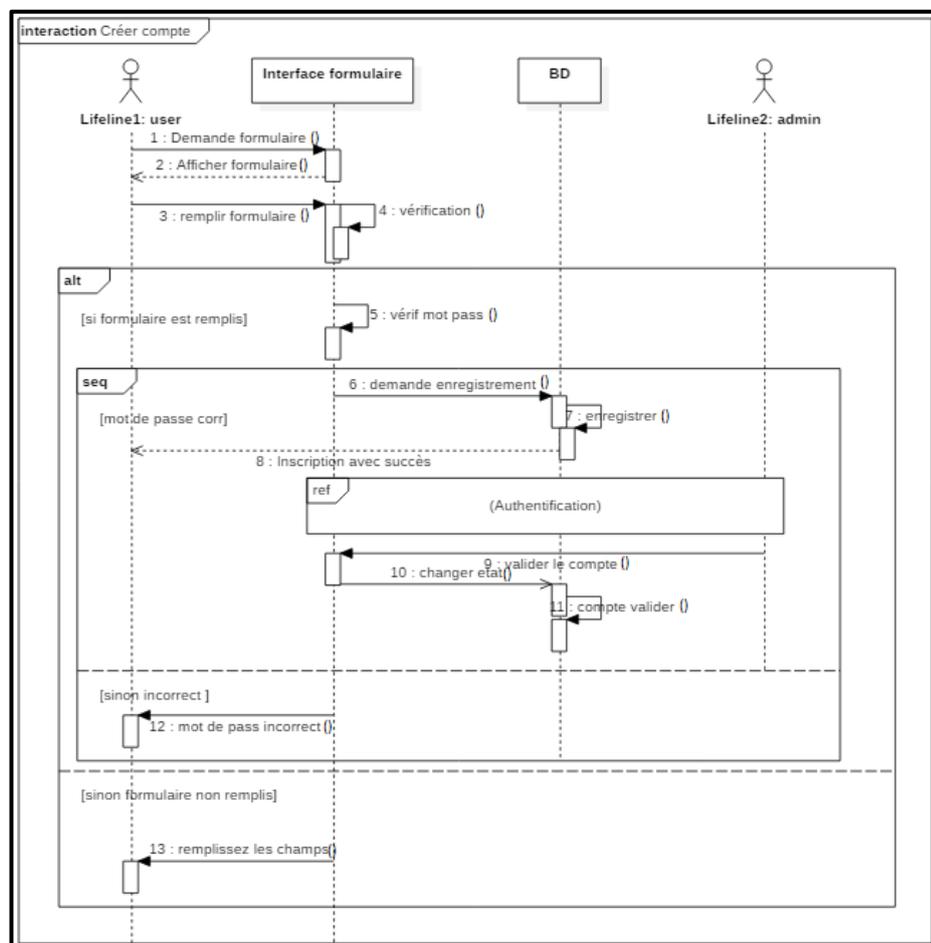


Figure 15 : Diagramme de séquences « Création d'un compte ».

Chapitre II : Conception et modélisation de la solution proposée

II.5.3 Diagramme de séquence gestion des ressources

Ci-dessous nous présentons les différents diagrammes de gestion des ressources :

II.5.3.1 Diagramme de séquence ajouter une ressource

Le scénario du diagramme « Ajouter une ressource » est le suivant :

1. L'expert sélectionne l'option ajouter une ressource.
2. Formulaire d'ajout est affiché pour l'expert.
3. L'expert saisi les informations de ressource.
4. Le système vérifie les champs s'ils ont été remplis correctement.
5. Le système vérifie les informations avec la base de données s'il n'existe pas de doublons.
6. Une vérification est effectuée au niveau de la base de données.
7. Si le résultat de la vérification indique que les ressources n'existent pas alors les ressources seront ajoutées à la BD.
8. Si les ressources existent alors le système affiche un message pour indiquer qu'ils existent déjà dans la BD.

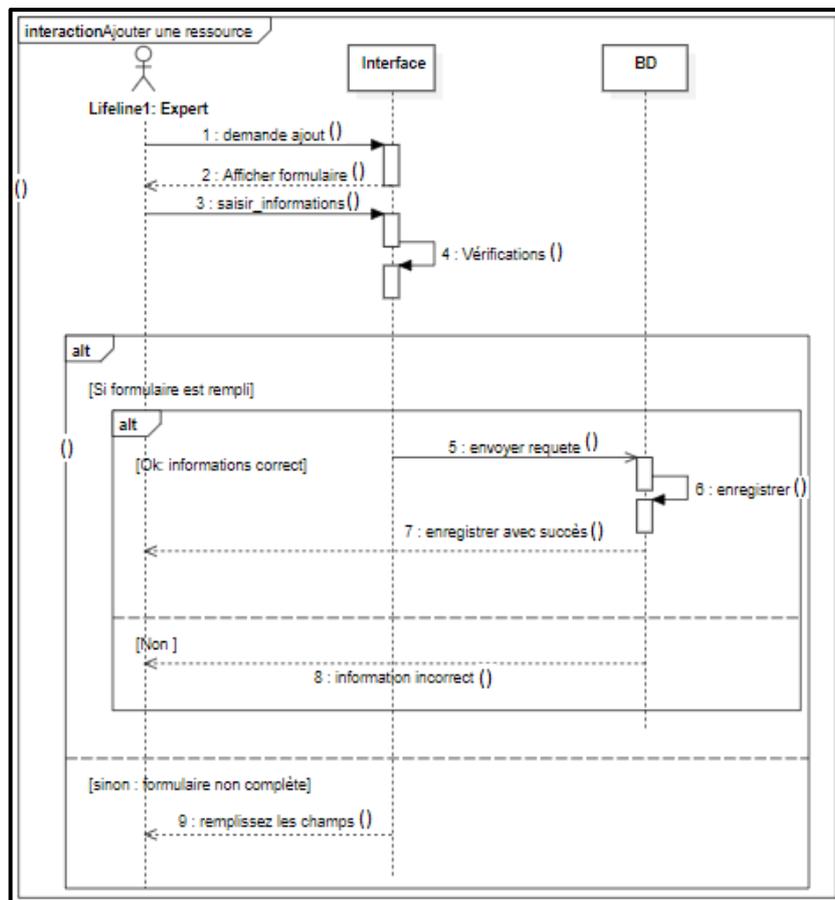


Figure 16 : Diagramme de séquence ajouter ressource.

Chapitre II : Conception et modélisation de la solution proposée

II.5.3.2 Diagramme de séquence supprimer une ressource

Le scénario du diagramme de « Suppression d'une ressource » est le suivant :

1. L'expert saisit les informations de la ressource.
2. Le système vérifie les informations avec la base de données sur l'existence de la ressource.
3. Une vérification est effectuée au niveau de la base de données.
4. Si la ressource existe alors l'expert peut effectuer la suppression.
5. Sinon le système affiche que la ressource n'existe pas dans la BD.

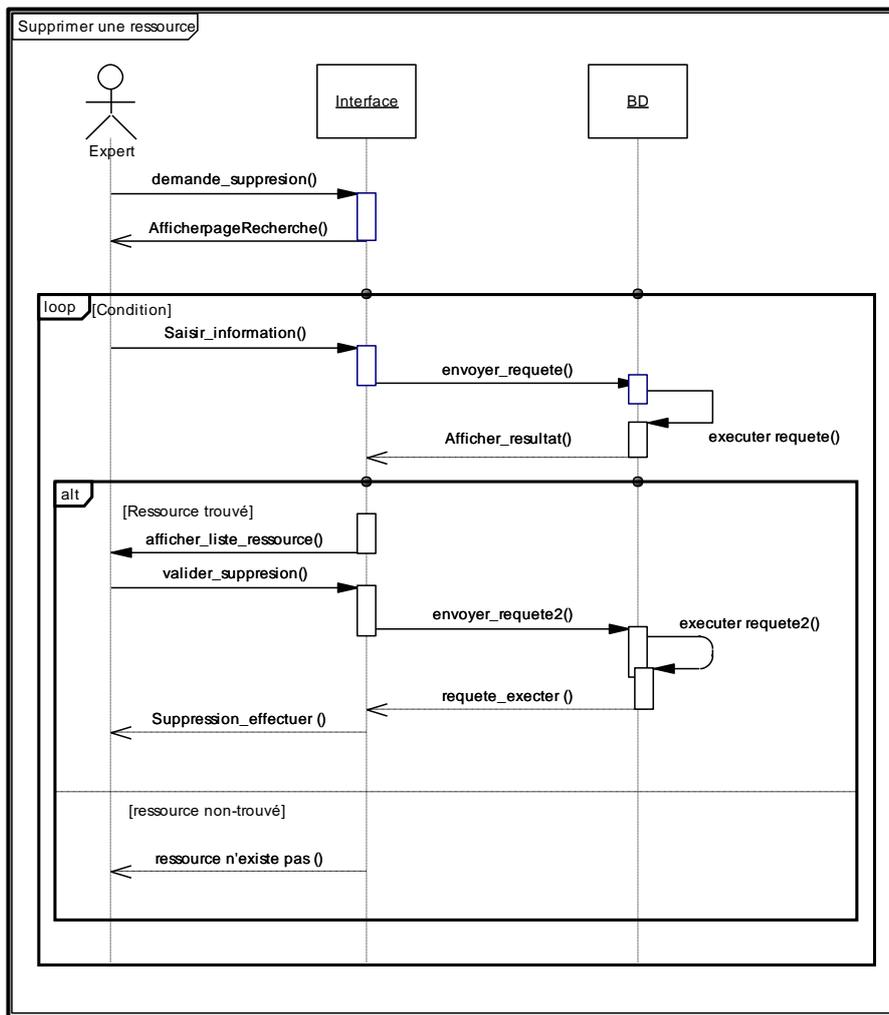


Figure 17: Diagramme de séquence supprimer une ressource

II.5.3.3 Diagramme de séquence modifier une ressource

Le scénario du diagramme de « Modifier une ressource » est le suivant :

6. L'expert saisit les informations de ressource.
7. Le système vérifie l'existence des informations avec la base de données.
8. Une vérification est effectuée au niveau de la base de données.
9. Si la ressource existe alors l'expert peut saisir les nouvelles informations.

Chapitre II : Conception et modélisation de la solution proposée

10. Les informations doivent être correctes et complètes sinon le système affiche un message d'erreur.
11. Sinon (n'existe pas) le système affiche que la ressource n'existe pas dans la BD.

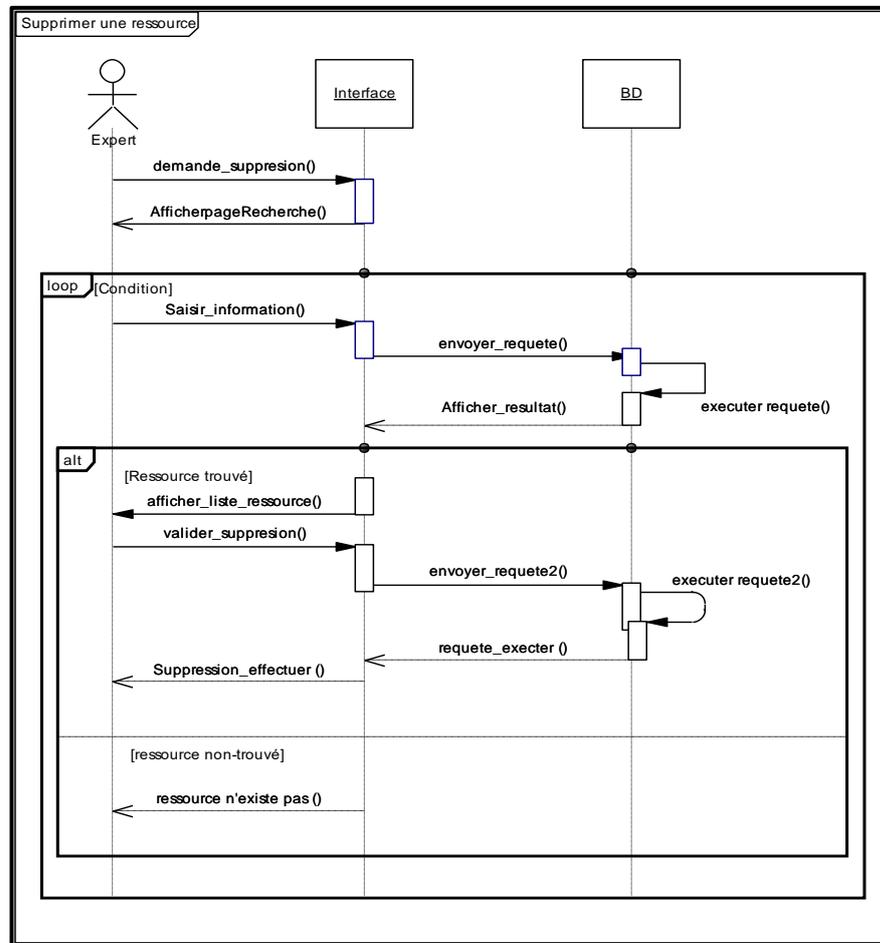


Figure 18 : Diagramme de séquence « Modifier une ressource »

II.5.4 Diagramme de séquence gestion des sectes

Les diagrammes ci-dessous correspondent aux différentes actions sur les sectes qui sont permises par la plateforme.

II.5.4.1 Ajouter secte

Le scénario du diagramme d' « Ajout d'une secte » est le suivant :

1. Le formulaire d'ajout est affiché pour l'expert.
2. L'expert saisi les informations de secte.
3. Le système vérifie les informations avec la base de données en cas de doublons.

Chapitre II : Conception et modélisation de la solution proposée

4. Une vérification est effectuée dans la base de données.
5. Si le résultat de la vérification indique que la secte n'existe pas alors elle sera ajoutée (insérée) dans la table correspondante.
6. Si la secte existe déjà, alors le système affiche un message pour l'indiquer à l'utilisateur.

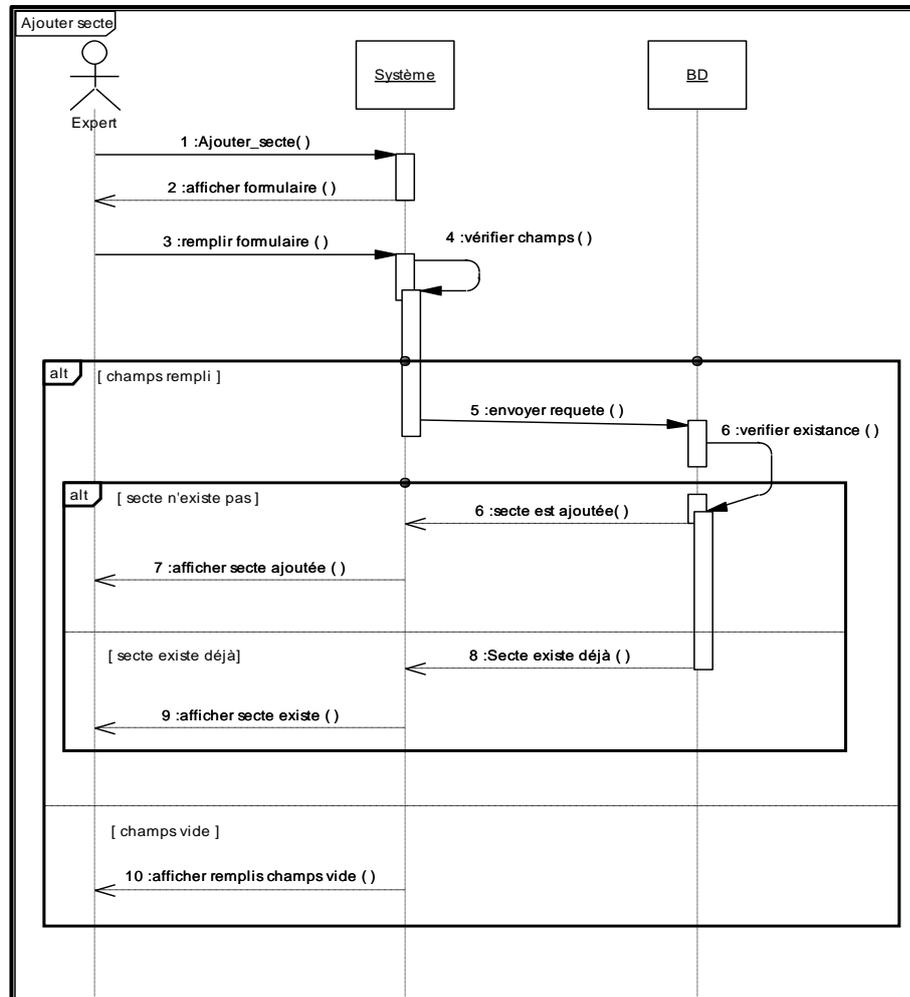


Figure 19 : Diagramme de séquence « Ajouter une secte ».

II.5.4.2 Modifier secte

Le scénario du diagramme de « Modification des sectes » est le suivant :

1. L'expert demande modification d'une secte.
2. Le système affiche le formulaire correspondant.
3. L'expert saisie les informations de la secte allant être modifiée.
4. Le système vérifie les informations avec la base de données pour éviter les doublons.
5. Une vérification est effectuée au niveau de la base de données.
6. Si la secte existe alors l'expert peut saisir les nouvelles informations.

Chapitre II : Conception et modélisation de la solution proposée

7. Les informations doivent être correctes et complètes sinon le système affiche un message d'erreur.
8. Sinon (n'existe pas) le système affiche que la secte n'existe pas dans la BD.

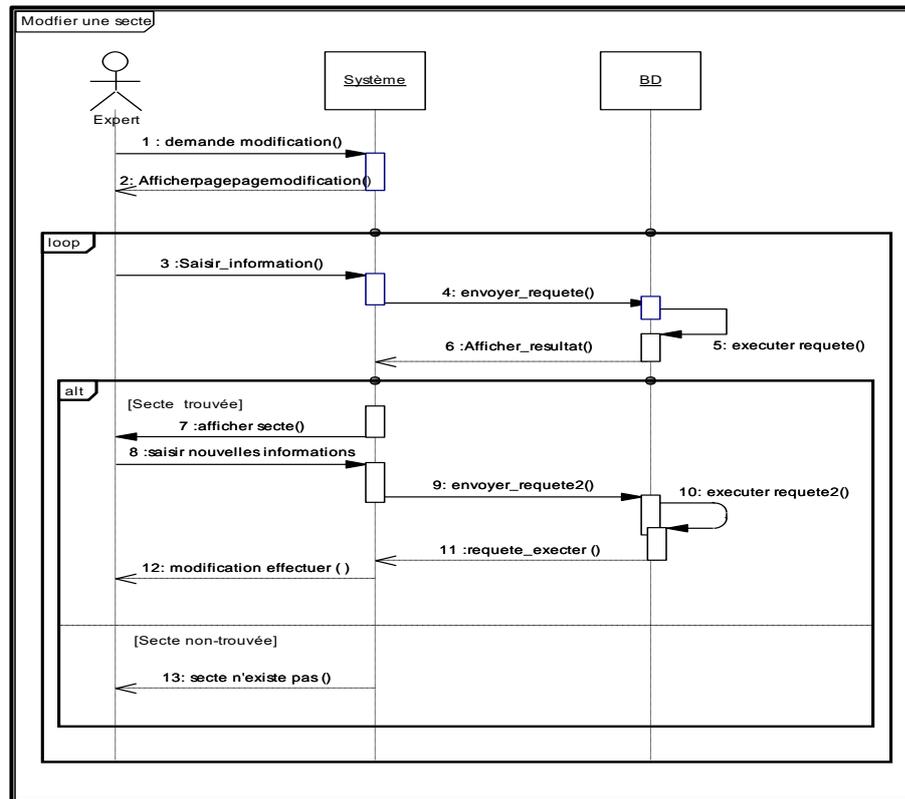


Figure 20: Diagramme de séquence « Modifier une secte ».

II.5.4.3 Supprimer secte

Le scénario du diagramme de « Suppression d'une secte » est le suivant :

1. L'expert demande la suppression d'une secte.
2. Le système affiche page de suppression (formulaire).
3. L'expert saisie les informations de la secte.
4. Le système effectue une vérification sur les données saisies.
5. Si les informations sont complètes alors le système envoie une requête pour récupérer les informations de la base de données.
6. Si la secte est trouvée alors le système affiche le résultat et demande à l'expert la confirmation de la suppression.
7. Une fois la confirmation de la suppression est faite, le système envoie une requête pour supprimer la secte de la BD.
8. Afficher le message « suppression effectuée » si la requête est exécutée correctement.

Chapitre II : Conception et modélisation de la solution proposée

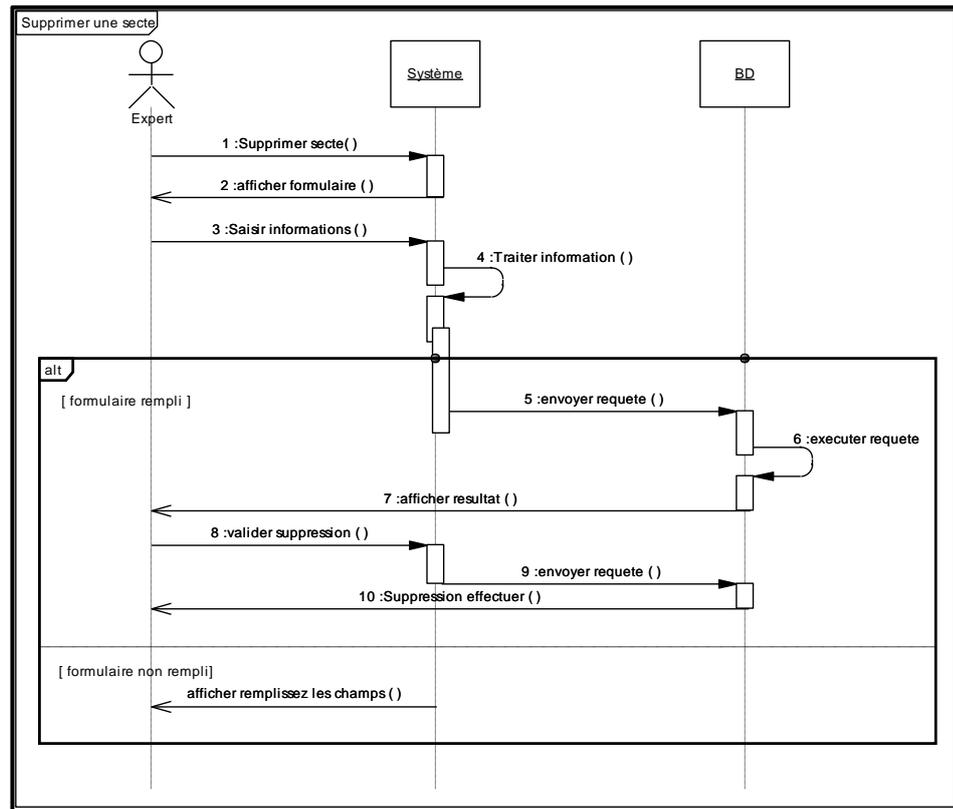


Figure 21 : Diagramme de séquence « supprimer une secte ».

II.6 Conception de l'application

Nos informations peuvent être organisées en un ensemble de tables d'une base de données :

- Une table « Utilisateur » qui contient les informations personnelles, les informations d'identification ainsi que le type de l'utilisateur.
- Une table « ressource » qui contient les informations des ressources ajoutées par les utilisateurs.
- Une table « Secte » qui contient les sectes principales et leurs informations.
- Une table « Sous-secte » qui contiennent du détail plus fin (avec comme clé étrangère, la secte mère)
- Une table « caractéristiques » qui résume les caractéristiques retenues pour les tests.

II.6.1 Diagramme de classes

Le diagramme ci-dessous représente le diagramme de classes de notre application (figure 8).

Chapitre II : Conception et modélisation de la solution proposée

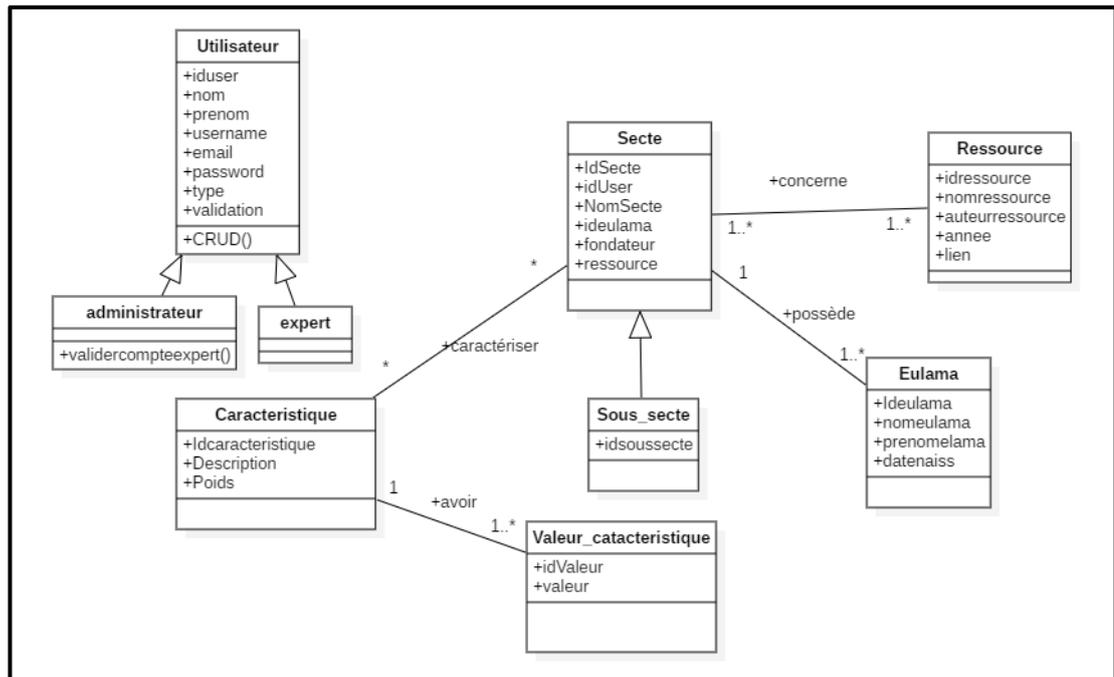


Figure 22: Diagramme de classes de l'application.

Le tableau ci-dessous décrit brièvement ce diagramme :

Tableau 7: Description du diagramme de classes.

Classe	Rôle	attributs
Utilisateur	Cette classe regroupe tous les utilisateurs de notre plateforme	-idUser : entier -nom : chaîne de caractères -prénom : chaîne de caractères -username : chaîne de caractères -email : chaîne de caractères -password : chaîne de caractères -type : caractère('A', 'S') -validation : entier(0,1)
Secte	C'est la classe mère de sous secte. Elle regroupe tous les sectes de notre plateforme.	-idSecte : entier -idUser : entier -nomSecte : chaîne de caractères -eulama : chaîne de caractères -fondeur : chaîne de caractères -ressource : entier -caracteristique : chaîne de caractères

Chapitre II : Conception et modélisation de la solution proposée

Sous-secte	Cette classe regroupe les sectes dérivé d'une secte mère elle prend les mêmes caractéristique de la classe mère et ajoute leurs nouvelles caractéristiques	-Idsoussecte : entier -Sectemère : entier -caractéristique : chaîne de caractère
Caractéristique	Décrire principale caractéristique ou la secte est arrivé.	-Idcaractéristique : chaîne de caractère -valeur : chaîne de caractère -Poids : float
Valeur_caractéristique	Décrire les questions sur lesquelles où les nations diffèrent par rapport aux principale caractéristique	-Idvaleur : entier -idcaractéristique : entier -valeur : chaîne de caractère -Poids : entier
Ressource	Cette classe contient les listes des références pour chaque secte.	-Idressource : entier -Nomressource : chaîne de caractère -Auteur : chaîne de caractère -Année : date -Lien : chaîne de caractère

Ci-après, nous présentons le schéma relationnel de notre diagramme de classes qui se présente comme suit :

- Utilisateur (**iduser**, nom, prenom, username, email, password, type, validation).
- Secte (**idsecte**, iduser, nomsecte, NomSecte, ideulama, fondateur).
- Concerne(**#idsecte** , **#idressource**).
- caractériser (**#idCaractéristique**, **#idSecte**).
- Valeur_caractéristique(**idValeur**, **idcaractéristique**, valeur).
- Caractéristique (**idCaractéristique**, description , poids, # idValeur).
- Manipuler (**#idUtilisateur**, **#idSecte**).
- Ressource(**idressource** , nomressource , auteur , annee , lien, #idSecte).
- Eulama(**ideulama**, nomeulama, prenomeulama, datenaiss, #IdSecte)
Sous-secte(**idsoussecte**, idsecte , iduser , nomsecte , NomSecte , ideulama, fondateur).

II.7 Conclusion

Nous avons présenté dans ce chapitre, la conception de notre application web à travers l'étude des besoins et leur analyse. Dans le prochain chapitre, nous allons nous intéresser au développement de cette dernière en présentant les outils utilisés et les interfaces de l'application.

Chapitre III: Tests et implémentation

Chapitre III: Tests et Implémentation

III.1 Tests

III.1.1 Introduction

Cette partie représente toutes les phases des tests qui ont été effectués. Nous avons testés différents classificateurs tels que K-means algorithme non-supervisé et Decision tree (Arbre de décision) pour un algorithme supervisé. Ces derniers ont déjà fait l'objet d'une présentation dans le chapitre précédent.

III.1.2 Jeux de données

Afin de tester les classificateurs nous avons utilisées un Dataset personnel qui contient des sectes, leurs caractéristiques et un poids de chaque caractéristique selon son importance dans notre jeu de classification.

Notre Dataset est composé de :

- 100 lignes représentant les sectes que nous avons récupérées.
- 10 colonnes représentant les caractéristiques possibles de la secte.
- Chaque caractéristique possède un poids entre 0 et 1 tels que 1 représente la caractéristique la plus importante (el-immane).

III.1.2.1 Résultats de l'algorithme de classification K-means

III.1.2.1.1 Représentation des données

Le Dataset que nous avons construit ne nous permet pas d'appliquer les différents algorithmes que nous avons déjà choisi, donc nous avons fait des changements sur le Dataset initial pour qu'on puisse faire la classification selon les deux algorithmes K-means et l'arbre de décision en suivant les étapes suivantes :

1. Attribuer un poids entre 0 et 1 pour les caractéristiques des sectes la caractéristique la plus importantes possèdera un poids plus élevé (c'est-à-dire proche de 1).
2. Convertir les noms des sectes (Firaq) en numéro de 1 à 100 (100 sectes dans notre cas).
3. Les sectes sont placées horizontalement (en lignes).
4. Les colonnes représentent des caractéristiques (D1, D2, D3...)
5. Les caractéristiques qui ne figuraient dans aucune secte de notre collecte ont été supprimées.
6. Les cellules de notre fichier de tests correspondent aux poids des caractéristiques.

III.1.2.1.2 Résultats

Chapitre III: Tests et Implémentation

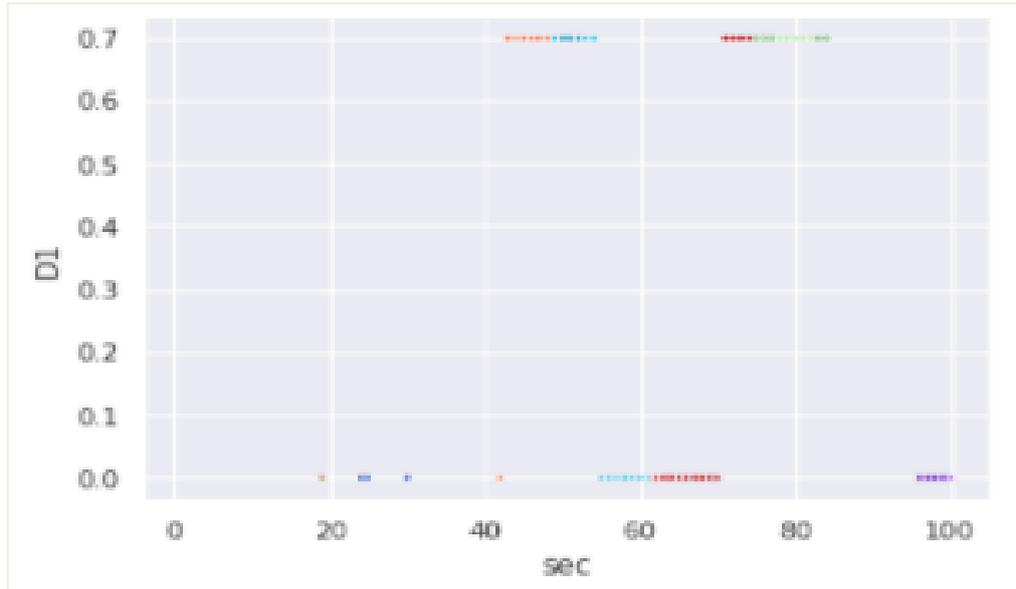


Figure 23: résultat 1 de l'algorithme K-means.

La figure 23 montre comment les sectes elles sont positionnées avant l'application de l'algorithme dans des groupes avec des différentes couleurs.

secte	D1	D2	D3	D4	D7	D8	D9	D10	D11	D12	D14	D15
1	0	0	0	0.96	0.8	0	0.6	0.8	0	0	0	0
2	0	0	0	0	0.8	0	0.6	0.8	0	0	0	0
3	0	0	0	0	0	0	0.6	0.8	0	0	0	0
4	0	0	1	0.96	0	0	0.6	0.8	0	0	0.1	0
5	0	0	0	0.96	0.8	0	0.6	0.8	0	0	0	0
6	0	0	1	0	0	0	0.6	0.8	0.135	0	0	0
7	0	0	1	0	0	0	0.6	0.8	0.135	0	0	0
8	0	0	1	0.96	0	0	0.6	0.8	0.135	0	0	0
9	0	0	1	0.96	0	0	0.6	0.8	0	0	0	0
10	0	0	1	0	0	0.23	0.6	0.8	0	0	0	0
11	0	0	1	0.96	0.8	0	0.6	0.8	0	0	0	0
12	0	0	1	0.96	0.8	0	0.6	0.8	0	0	0	0
13	0	0	0	0	0	0.23	0.6	0.8	0	0	0	0
14	0	0	0	0	0	0.23	0.6	0.8	0	0	0	0
15	0	0	0	0	0	0.23	0.6	0.8	0	0.5	0	0
16	0	0	1	0.96	0.8	0	0.6	0.8	0	0	0.1	0.33
17	0	0	0	0.96	0.8	0	0.6	0.8	0	0	0	0.33
18	0.7	0	0	0.96	0.8	0	0	0	0	0	0	0.33
19	0	0.8	1	0.96	0.8	0	0.6	0	0.135	0	0	0
23	0.7	0	0	0	0.8	0	0	0	0	0	0	0
24	0	0.8	0	0	0.8	0	0	0	0	0	0	0

Chapitre III: Tests et Implémentation

25	0	0.8	0	0	0.8	0	0	0	0	0	0	0
26	0.7	0	0	0	0.8	0	0	0	0	0	0	0
27	0.7	0	0	0	0	0	0	0	0	0	0	0

Tableau 8: Quelques lignes de dataset avant la classification.

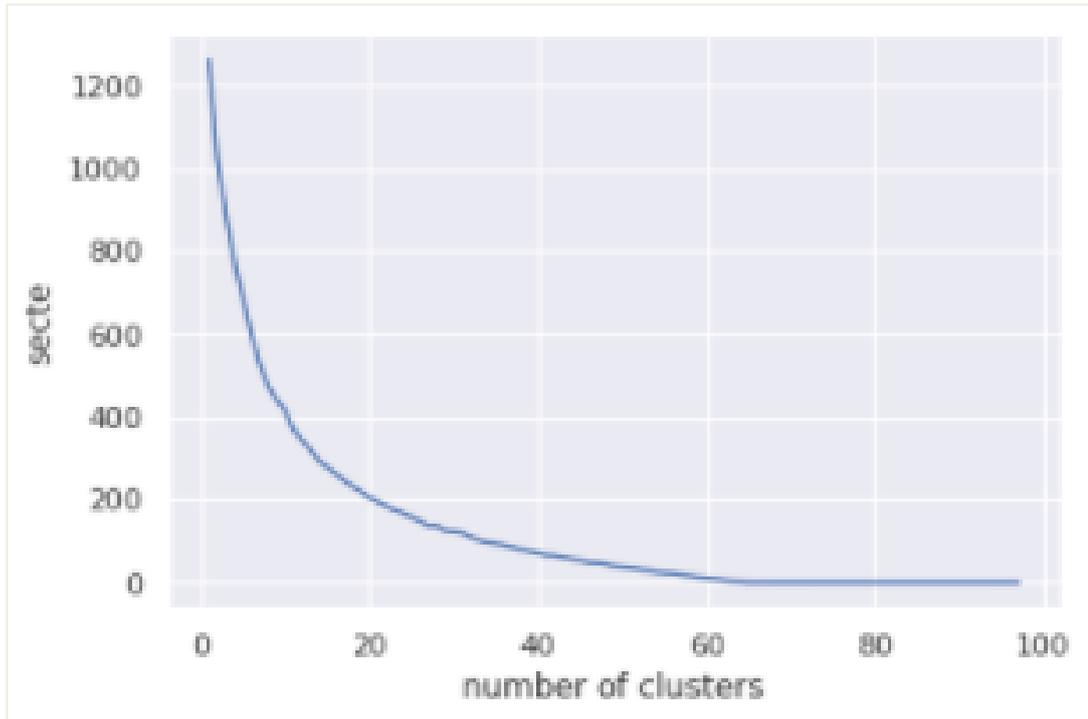


Figure 24: Graphe nombre de sectes par rapport aux clusters.

Le graphe de la figure 24 nous aide à déterminer le nombre de secte que l'on veut extraire autrement dit que le nombre de cluster représente le nombre des sectes qui ont des caractéristiques similaires donc elles deviennent une secte mère.

Chapitre III: Tests et Implémentation

	sec	D1	D2	D3	D4	D7	D8	D9	D10	D11	D12	D14	D15	cluster_pred
0	1	0.0	0.0	0	0.96	0.8	0.00	0.6	0.8	0.000	0.0	0.0	0.00	5
1	2	0.0	0.0	0	0.00	0.8	0.00	0.6	0.8	0.000	0.0	0.0	0.00	5
2	3	0.0	0.0	0	0.00	0.0	0.00	0.6	0.8	0.000	0.0	0.0	0.00	8
3	4	0.0	0.0	1	0.96	0.0	0.00	0.6	0.8	0.000	0.0	0.1	0.00	5
4	5	0.0	0.0	0	0.96	0.8	0.00	0.6	0.8	0.000	0.0	0.0	0.00	5
5	6	0.0	0.0	1	0.00	0.0	0.00	0.6	0.8	0.135	0.0	0.0	0.00	8
6	7	0.0	0.0	1	0.00	0.0	0.00	0.6	0.8	0.135	0.0	0.0	0.00	8
7	8	0.0	0.0	1	0.96	0.0	0.00	0.6	0.8	0.135	0.0	0.0	0.00	5
8	9	0.0	0.0	1	0.96	0.0	0.00	0.6	0.8	0.000	0.0	0.0	0.00	5
9	10	0.0	0.0	1	0.00	0.0	0.23	0.6	0.8	0.000	0.0	0.0	0.00	8
10	11	0.0	0.0	1	0.96	0.8	0.00	0.6	0.8	0.000	0.0	0.0	0.00	5
11	12	0.0	0.0	1	0.96	0.8	0.00	0.6	0.8	0.000	0.0	0.0	0.00	5
12	13	0.0	0.0	0	0.00	0.0	0.23	0.6	0.8	0.000	0.0	0.0	0.00	8
13	14	0.0	0.0	0	0.00	0.0	0.23	0.6	0.8	0.000	0.0	0.0	0.00	8
14	15	0.0	0.0	0	0.00	0.0	0.23	0.6	0.8	0.000	0.5	0.0	0.00	8
15	16	0.0	0.0	1	0.96	0.8	0.00	0.6	0.8	0.000	0.0	0.1	0.33	5
16	17	0.0	0.0	0	0.96	0.8	0.00	0.6	0.8	0.000	0.0	0.0	0.33	5
17	18	0.7	0.0	0	0.96	0.8	0.00	0.0	0.0	0.000	0.0	0.0	0.33	1
18	19	0.0	0.8	1	0.96	0.8	0.00	0.6	0.0	0.135	0.0	0.0	0.00	5
19	23	0.7	0.0	0	0.00	0.8	0.00	0.0	0.0	0.000	0.0	0.0	0.00	1
20	24	0.0	0.8	0	0.00	0.8	0.00	0.0	0.0	0.000	0.0	0.0	0.00	1
21	25	0.0	0.8	0	0.00	0.8	0.00	0.0	0.0	0.000	0.0	0.0	0.00	1
22	26	0.7	0.0	0	0.00	0.8	0.00	0.0	0.0	0.000	0.0	0.0	0.00	1
23	27	0.7	0.0	0	0.00	0.0	0.00	0.0	0.0	0.000	0.0	0.0	0.00	1
24	28	0.7	0.0	0	0.00	0.0	0.00	0.0	0.0	0.000	0.0	0.0	0.00	1
25	29	0.7	0.0	0	0.00	0.0	0.00	0.0	0.0	0.000	0.0	0.0	0.00	1
26	30	0.0	0.8	0	0.00	0.0	0.00	0.0	0.0	0.000	0.0	0.0	0.00	1
27	31	0.7	0.0	0	0.00	0.0	0.00	0.0	0.0	0.000	0.0	0.0	0.00	1
28	32	0.7	0.0	0	0.00	0.0	0.00	0.0	0.0	0.000	0.0	0.0	0.00	1
29	33	0.7	0.0	0	0.96	0.0	0.00	0.0	0.0	0.135	0.0	0.0	0.00	1
30	34	0.0	0.0	0	0.00	0.8	0.00	0.0	0.0	0.000	0.0	0.0	0.00	1
31	35	0.0	0.0	0	0.00	0.0	0.00	0.0	0.0	0.000	0.0	0.0	0.00	1
32	36	0.0	0.0	0	0.00	0.8	0.00	0.0	0.0	0.000	0.0	0.0	0.00	1
33	37	0.0	0.0	0	0.96	0.8	0.00	0.0	0.0	0.000	0.0	0.0	0.00	1

Figure 25: Quelques lignes de Dataset après la classification.

Après la classification, la figure 25 montre l'affectation de chaque secte dans un cluster la colonne « cluster_pred » montre le numéro de cluster.

Chapitre III: Tests et Implémentation

III.1.2.1.3 Pseudo code pour K-means

```
# -*- coding: utf-8 -*-
# Commented out IPython magic to ensure Python compatibility.
# %matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns; sns.set() # for plot styling
import numpy as np

from sklearn.datasets.samples_generator import make_blobs
X, y_true = make_blobs(centers=4,n_features=15,cluster_std=0.60, random_state=20)

plt.scatter(X[:, 0], X[:, 1],c='black', s=20 );

#2ème exemple sur un fichier CSV
import numpy as np
import pandas as pd
import seaborn as sns
sns.set()
from sklearn.cluster import KMeans

data=pd.read_csv('secte1.csv')

plt.scatter(data['sec'],data['D1'],data['D2'],data['D14'] )
plt.xlabel('D1')
plt.ylabel("D2")
plt.show

#Section Features
x=data.copy()
p=x
#Clustering
kmeans = KMeans()
print(p)
kmeans.fit(p)

#clustering result
clusters=x.copy()
clusters['cluster_pred']=kmeans.fit_predict(p)
#affichage plot
plt.scatter(clusters['sec'],clusters['D1'],clusters['D2'],c=clusters ['cluster_pred'],cmap='rainbow')
plt.xlabel('sec')
plt.ylabel('D1')
plt.show()
cluster_new.head()
#7. Standardizing the variables
from sklearn import preprocessing
x_scaled=preprocessing.scale(p)
x_scaled
print(p)
#8. The Elbow Method
secte=[]
for i in range(1,98):
    kmeans=KMeans(i)
    kmeans.fit(x_scaled)
    secte.append(kmeans.inertia_)
secte
```

Chapitre III: Tests et Implémentation

```
#9. Visualizing the Elbow Method
plt.plot(range(1,98),secte)
plt.xlabel('number of clusters')
plt.ylabel('secte')
plt.show()

kmeans_new=KMeans(16)
kmeans.fit(x_scaled)
cluster_new=x.copy()
cluster_new['cluster_pred']=kmeans_new.fit_predict(x_scaled)
#cluster_new

plt.scatter(cluster_new['sec'], cluster_new['D1'], c=cluster_new['cluster_pred'],cmap='rainbow')
plt.xlabel('sec')
plt.ylabel('D1')
plt.show()
cluster_new.head(100)
```

III.1.2.2 Résultats de l'algorithme arbre de décision

Dans cette section, nous allons présenter les résultats obtenus avec l'algorithme arbre de décision.

III.1.2.2.1 Paramétrage de l'algorithme

Le paramétrage de l'algorithme est le suivant :

- Il faut indiquer le nom du fichier correspondant au Dataset avec extension CSV et lire les données dans notre cas le Dataset ne doit pas contenir d'entête car cette dernière sera définie dans d'autres instructions après la lecture.
- Il faut obligatoirement choisir une variable cible pour que l'algorithme fasse les traitements par rapport à cette variable.
- Dans notre cas nous avons utilisé la colonne D3 c'était la colonne correspondant à la caractéristique el-imman « الإيمان ».
- Pour que l'affichage de l'arbre ne soit pas très grand il faut spécifier la profondeur selon le code suivant.

```
# Create Decision Tree classifier object
clf = DecisionTreeClassifier(criterion="entropy", max_depth=40)
```

Figure 26: Code pour spécifier le profondeur de l'arbre dans l'algorithme DT.

III.1.2.2.2 Résultat

Nous avons utilisé la mesure d'évaluation accuracy qui représente :

- **Accuracy** : il montre à quel fréquence le classificateur prédit les valeurs correctes et peut être calculé comme suit :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

- TP : True positive (vrai positif) les bons clusters correctement prédits.
- TN : True négative (vrai négatif) les mauvais clusters correctement prédits.
- FP : False positive (faux positif) les bons clusters non correctement prédits.
- FN : False négative (faux négatif) les mauvais clusters non correctement prédits.

Pour la caractéristique D3, on obtient une accuracy égale à 1. Et nous n'avons pas obtenu un résultat pour ensemble des clusters comme la figure 30.

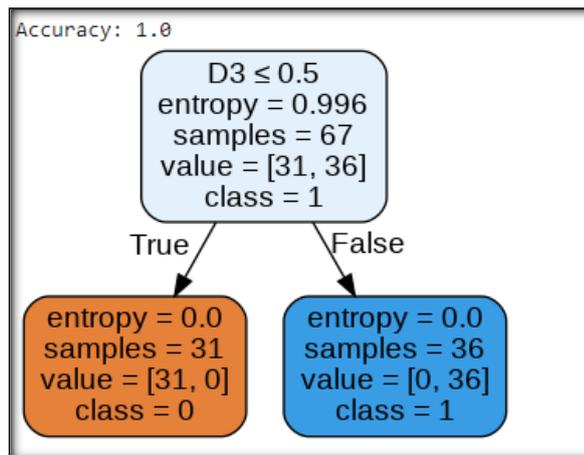


Figure 27: résultat de l'arbre et valeur d'accuracy par rapport à D3.

Pour une autre colonne, Le résultat de l'accuracy est égale 0.3.

- **Entropy** est une mesure d'information qui indique le désordre des caractéristiques avec la cible. Semblable à l'indice de Gini, la division optimale est choisie par la caractéristique avec moins d'entropie. Il obtient sa valeur maximale lorsque la probabilité des deux classes est la même et un

Chapitre III: Tests et Implémentation

nœud est pur lorsque l'entropie a sa valeur minimale. L'entropie est calculé comme suit :

$$\sum P_j \cdot \log_2 \cdot P_j \quad (6)$$

Les valeurs minimale et maximale de l'entropie :

$$\text{Entropy}_{\min} = -1 \cdot \log_2(1) = 0$$

$$\text{Entropy}_{\max} = -0.5 \cdot \log_2(0.5) - 0.5 \cdot \log_2(0.5) = 1$$

- **Samples** : sont les classes dans la figure 27 on trouve 31 samples(classes) de nœud gauche et 36 samples dans le nœud à droite (classes).

```
DecisionTreeClassifier(ccp_alpha=0.0, class_weight=None, criterion='entropy',
max_depth=40, max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, presort='deprecated',
random_state=None, splitter='best')
Accuracy: 0.3
```

Figure 28: résultat de précision de l'arbre de décision par rapport à une autre caractéristique.

Chapitre III: Tests et Implémentation

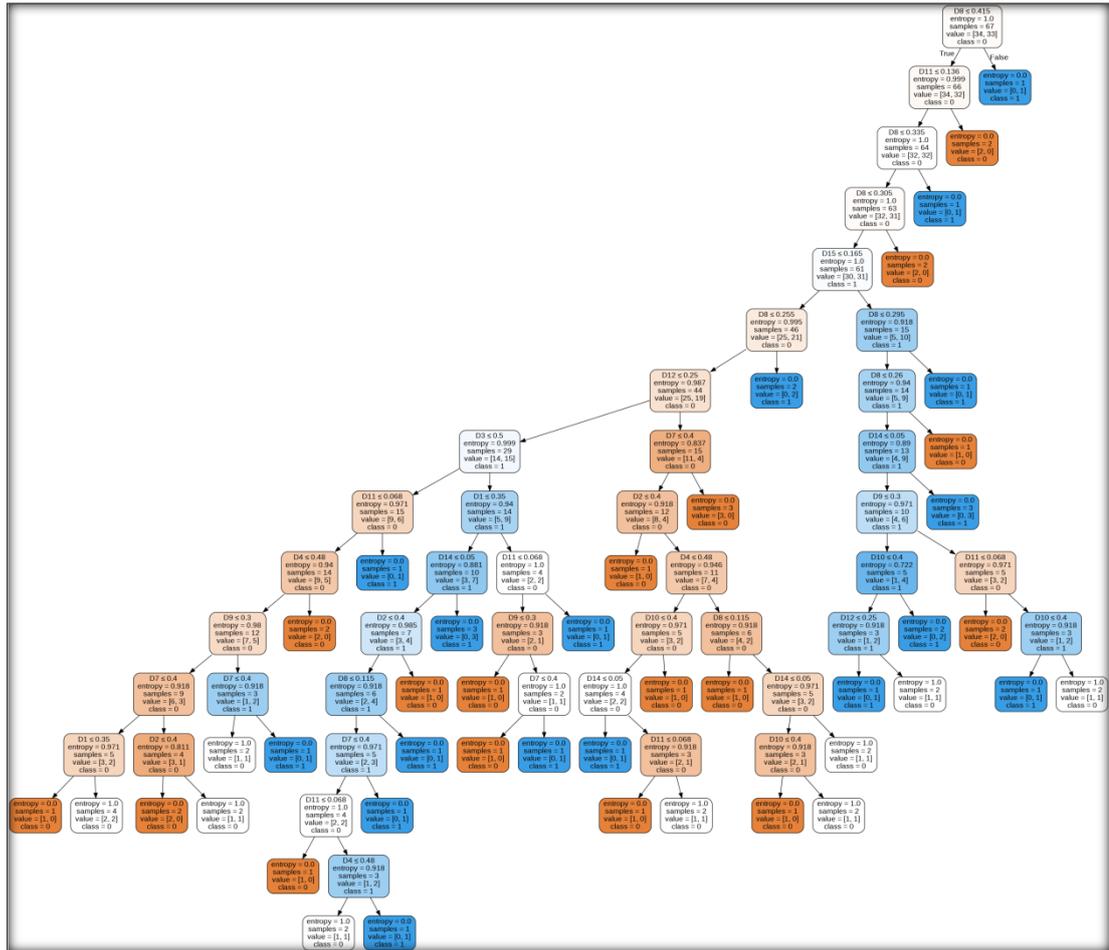


Figure 29: Arbre de notre Dataset par rapport à une autre caractéristique.

L'arbre montre les différents chemins pour obtenir le meilleur regroupement des sectes dans des clusters. Cet arbre avec une entrée de profondeur égale à 40 on a obtenu 16 lignes comme profondeur maximum.

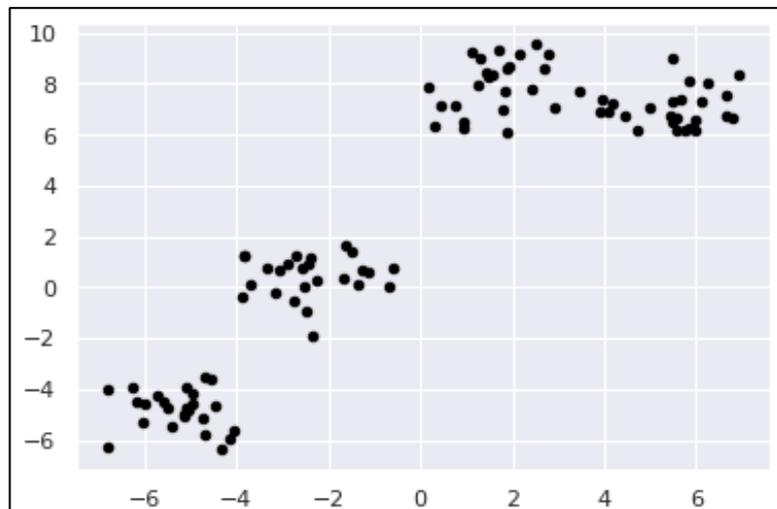


Figure 30: résultat de classification par rapport au DT algorithm.

Chapitre III: Tests et Implémentation

La figure 30 montre ensemble des clusters que sont les grandes sectes est les points en noir représente les sous-secte des clusters.

III.1.2.2.3 Pseudo code pour l'arbre de décision

```
# Load libraries
import pandas as pd
from sklearn.tree import DecisionTreeClassifier # Import Decision Tree Classifier
from sklearn.model_selection import train_test_split # Import train_test_split
function
from sklearn import metrics #Import scikit-learn metrics module for accuracy
calculation
from sklearn.externals.six import StringIO
from IPython.display import Image
from sklearn.tree import export_graphviz
import pydotplus

import matplotlib.pyplot as plt
import seaborn as sns; sns.set() # for plot styling

from sklearn.datasets.samples_generator import make_blobs
X, y_true = make_blobs( centers=4,n_features=6,cluster_std=0.90,
random_state=20)

plt.scatter(X[:, 0], X[:, 1],c='black', s=20 );
col_names = ['sec', 'D1', 'D2', 'D3', 'D4', 'D7', 'D8', 'D9', 'D10', 'D11', 'D12', 'D14',
'D15','D16']

# load dataset
pima = pd.read_csv("secte1 - Copie.csv", header=None, names=col_names)

#split dataset in features and target variable
feature_cols = [ 'D1', 'D2', 'D3', 'D4', 'D7', 'D8', 'D9', 'D10', 'D11', 'D12', 'D14',
'D15']
X = pima[feature_cols] # Features
y = pima.D16 # Target variable
X.head(10)
# Split dataset into training set and test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=1) # 70% training and 30% test

# Create Decision Tree classifier object
clf = DecisionTreeClassifier(criterion="entropy", max_depth=40)

# Train Decision Tree Classifier
clf = clf.fit(X_train,y_train)
print(clf)
#Predict the response for test dataset
```

Chapitre III: Tests et Implémentation

```
y_pred = clf.predict(X_test)

# Model Accuracy, how often is the classifier correct?
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))

dot_data = StringIO()

export_graphviz(clf, out_file=dot_data,
                filled=True, rounded=True,
                special_characters=True, feature_names =
feature_cols,class_names=['0','1'])
graph = pydotplus.graph_from_dot_data(dot_data.getvalue())

graph.write_png('secte.png')
Image(graph.create_png())
```

III.1.3 Discussion

Après plusieurs tests, nous avons bien remarqué que les deux algorithmes donnent des résultats, mais les résultats de l'algorithme de K-means nous a donné des résultats plus cohérentes tels que le nombre des sectes dans chaque cluster (groupe ou classe) et aussi on peut avec l'algorithme de k-means déduire le nombre des sectes voulus comme 73 sectes et classification se fait automatiquement dans l'autre côté on ne peut pas spécifier un nombre exacte dans l'algorithme de l'arbre de décision.

Selon les tests, nous avons remarqué que les fonctionnalités de l'algorithme de K-means nous conviennent mieux que l'arbre de décision. D'où le meilleur choix c'était le K-means.

III.1.4 La différence entre K-means et arbre de décision

Le tableau suivant récapitule les différences entre les deux algorithmes choisis :

<i>Décision Tree</i>	<i>K -means</i>
Supervisé	non supervisé
pour la classification	pour le clustering
prédit une classe pour un vecteur d'entrée donné. Les attributs peuvent être numériques ou nominaux.	l'algorithme s'appuie uniquement sur la dynamique des caractéristiques indépendantes pour faire des inférences sur des données invisibles
La formation implique la construction itérative d'un arbre (considérant l'algorithme ID-3)	peut créer les informations de cluster pour les nœuds voisins
Temps de test faible puisqu'il s'agit d'une	crée les classes représentées par les

Chapitre III: Tests et Implémentation

traversée vers le bas d'un arbre	étiquettes centroïde et classe des échantillons appartenant à chaque classe.
Pas besoin de choisir une métrique de distance puisque les scissions se produiront en fonction des valeurs inhérentes à chaque caractéristique.	Lorsqu'il s'agit d'un grand ensemble de données, l'application d'une technique de dendrogramme fera planter l'ordinateur en raison de la charge de calcul et des limites de mémoire dynamique.
Un arbre existant ne peut pas être mis à jour un nouvel arbre entier doit être créé	Il s'adapte aux divers changements de vos données. En cas de souci, l'ajustement du segment de cluster permettra d'apporter rapidement des modifications nécessaires à l'algorithme.
- Pas besoin de stocker les données d'entraînement - ne stockez que le modèle d'arbre	Il convient à un grand nombre d'ensembles de données et est calculé beaucoup plus rapidement que le plus petit. Il peut également produire des clusters plus élevées.

Tableau 9: La différence entre K-mens et Arbre décision.

III.2 Implémentation

Dans cette section, nous allons décrire les moyens matériels et logiciels utilisés dans le développement de notre application. Ayant comme objectif d'aboutir à un produit final, exploitable par les utilisateurs. Nous allons présenter les différentes interfaces de notre application mais nous commençons d'abord par décrire l'environnement matériel et logiciel utilisé.

III.2.1 Environnement de développement

Cette partie consiste à présenter et spécifier les différents outils utilisés pour développer notre plateforme.

III.2.2 Environnement Matériel

Nous avons utilisé comme environnement matériel un ordinateur avec les caractéristiques suivantes :

➤ **PC SONY VAIO Fit 15e**

- ✓ *Processor : Intel(R) Core(TM) i3-3217U CPU @ 1.80GHz
1.80GHz*
- ✓ *Mémoire vive : 4 Go.*
- ✓ *Disque dur : 500 Go.*
- ✓ *Taille d'écran : 15,6 pouces.*

Chapitre III: Tests et Implémentation

- ✓ *Type de système : Windows 10, système d'exploitation x64 bits, processeur x64.*

III.2.3 Environnement et langages de programmation utilisés

III.2.3.1 Langages de programmation

➤ PHP

(Hypertext Preprocessor), plus connu sous son sigle PHP (sigle autoréférentiel), est un langage de programmation libre²⁴, 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.

➤ Javascript

JavaScript est un langage de programmation de scripts principalement employé dans les pages web interactives et à ce titre est une partie essentielle des applications web. Avec les technologies HTML et CSS, JavaScript est parfois considéré comme l'une des technologies cœur du World Wide Web³. Une grande majorité des sites web l'utilisent⁴, et la majorité des navigateurs web disposent d'un moteur JavaScript⁵ dédié pour l'interpréter, indépendamment des considérations de sécurité qui peuvent se poser le cas échéant.

III.2.3.2 Modules externe utilisé

➤ *Flask :*

Flask est un Framework d'application Web. Il est conçu pour rendre la mise en route rapide et facile, avec la possibilité de s'adapter à des applications complexes. Il a commencé comme un simple wrapper autour de Werkzeug et Jinja et est devenu l'un des frameworks d'application Web Python les plus populaires [54]. Il est utilisé pour faire un lien entre le code python et l'interface de notre application.

III.2.3.3 Environnement de développement

➤ Sublime text

Sublime Text¹ est un éditeur de texte générique (en anglais) codé en C ++ et Python, disponible sur Windows, Mac et Linux. Pour rappel. Sublime Texte prend en charge plus de 44 langages de programmation majeurs.

➤ Spyder²

Spyder est un environnement scientifique puissant écrit en Python, pour Python, Il présente une combinaison unique des fonctionnalités avancées d'édition, d'analyse,

¹ <https://www.sublimetext.com/3>

² <https://www.spyder-ide.org/>

Chapitre III: Tests et Implémentation

de débogage et de profilage d'un outil de développement complet avec l'exploration de données, l'exécution interactive, l'inspection approfondie et les superbes capacités de visualisation d'un package scientifique.

➤ *Google Colab*³

Google Colaboratory ou Colab, un outil Google simple et gratuit qui permet d'améliorer vos compétences de codage en langage de programmation Python, et d'utiliser un environnement de développement (Jupyter Notebook) qui ne nécessite aucune configuration. Il a été utilisé pour exécuter les commandes d'entraînement.

III.2.4 Réalisation du projet

Dans ce qui suit, nous allons présenter quelques interfaces de l'application :

III.2.5 Schéma d'accessibilité de la plateforme

Tout d'abord, nous allons présenter le schéma d'accessibilité de notre plateforme avec les trois espaces « Administrateur », « Expert », « Espace tiers »:

³ <https://colab.research.google.com/>

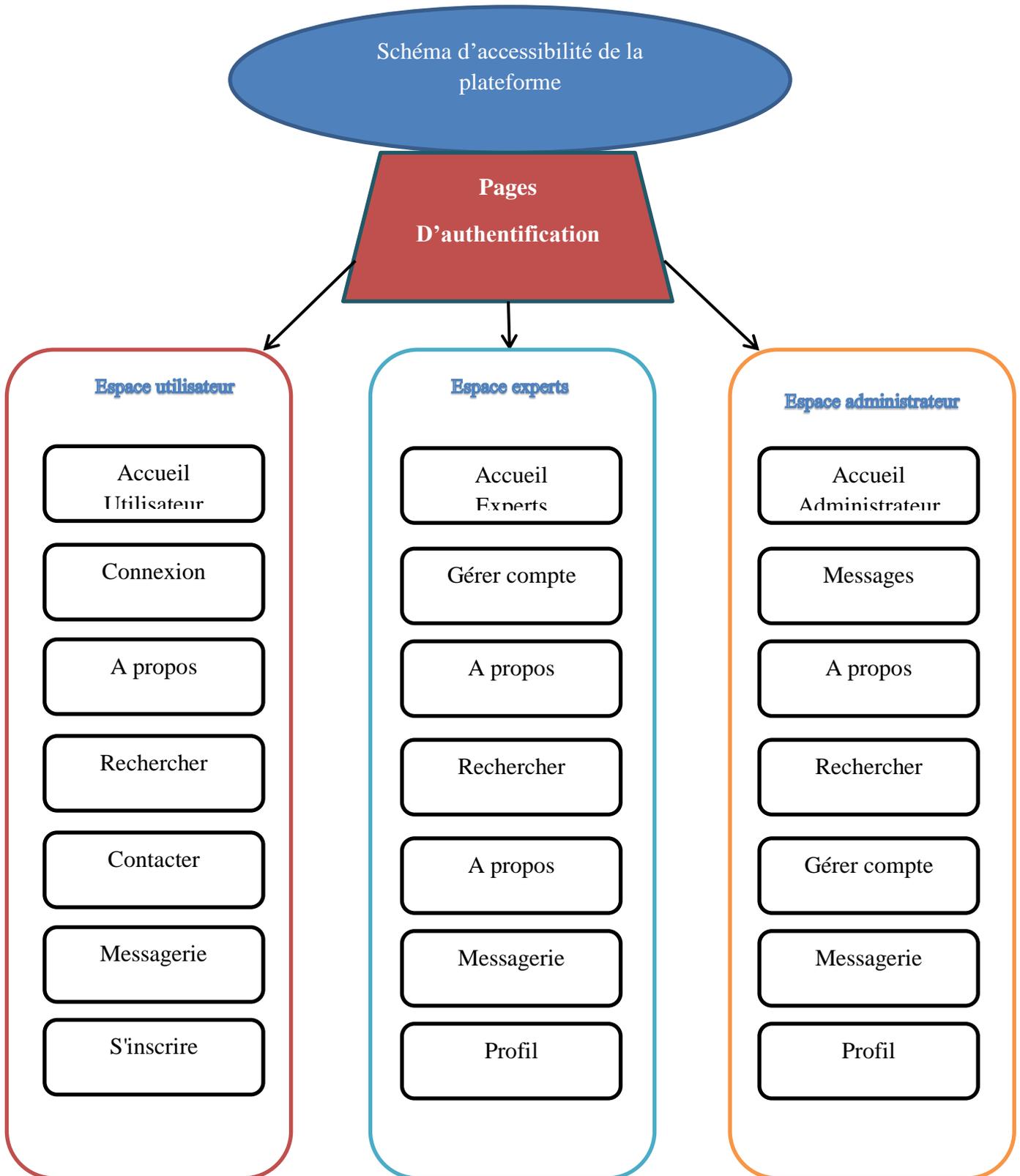


Figure 31 : Schéma d'accessibilité

Chapitre III: Tests et Implémentation

III.2.6 Les interfaces

L'accès à notre Système passe obligatoirement par une page d'authentification (voir Figure 14) donc il faut avoir un compte pour pouvoir accéder au Système. L'utilisateur s'identifie en saisissant, un login et un mot de passe qui lui sont propres plus le type de son compte. Si l'authentification est réussie, le système affiche la Session correspondante à l'utilisateur selon le type de son compte (Administrateur, Expert). Dans ce qui suit, nous allons présenter les interfaces correspondantes à chaque type de compte.

D'abord l'interface de connexion (authentification) :

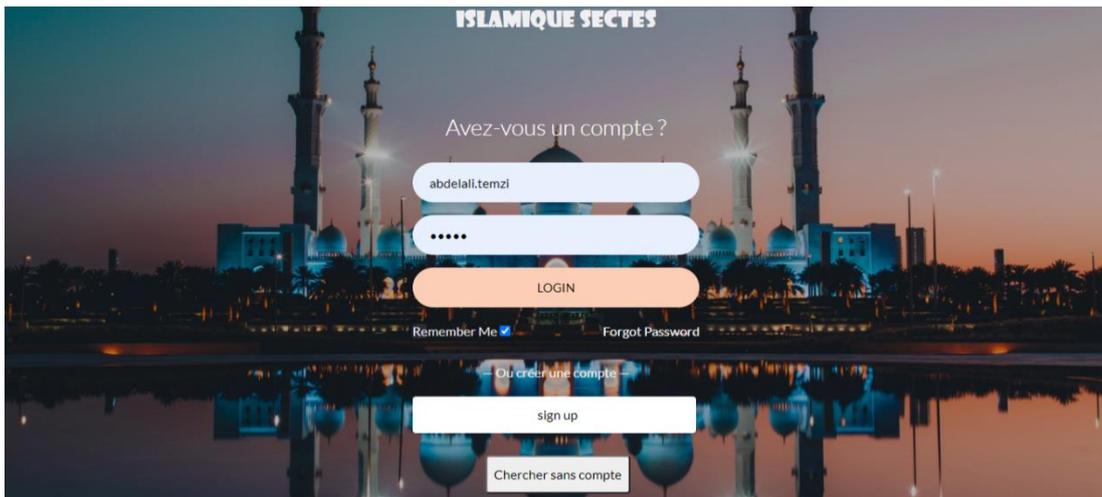


Figure 32: Page d'authentification.

III.2.6.1 Session Administrateur

❖ Page d'accueil de l'Administrateur

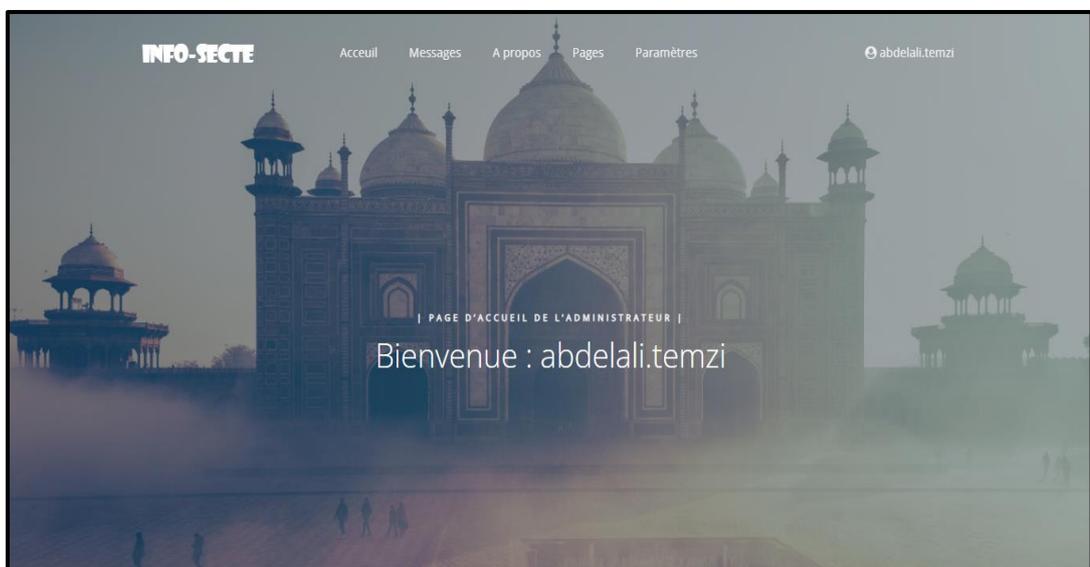
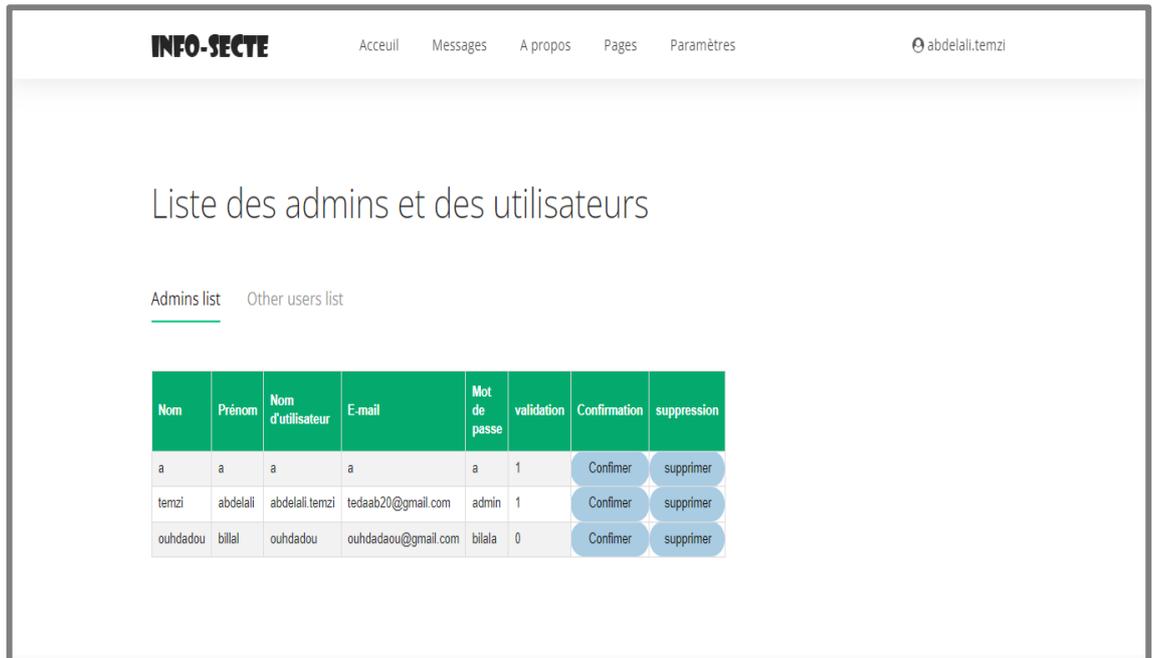


Figure 33 : Administrateur page d'accueil 1

Chapitre III: Tests et Implémentation

❖ Page de validation et de suppression des utilisateurs

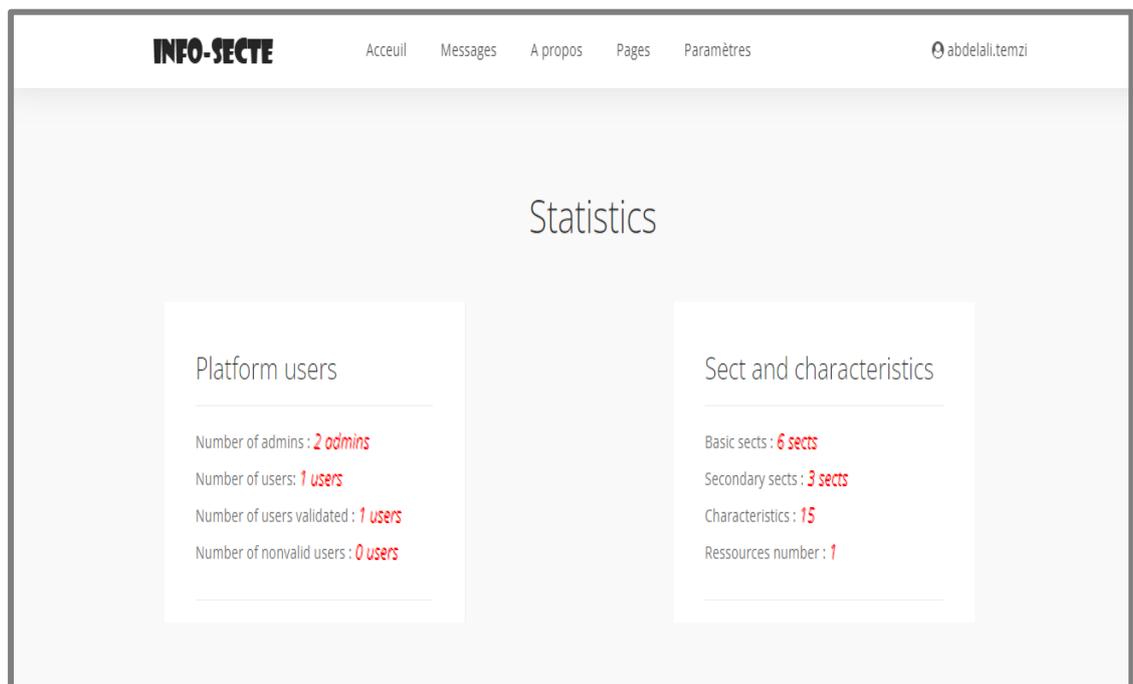


The screenshot shows the 'INFO-SECTE' administrator interface. At the top, there is a navigation menu with links for 'Accueil', 'Messages', 'A propos', 'Pages', and 'Paramètres', along with a user profile icon for 'abdelali.temzi'. The main heading is 'Liste des admins et des utilisateurs'. Below this, there are two tabs: 'Admins list' (which is active) and 'Other users list'. A table displays the user list with columns for 'Nom', 'Prénom', 'Nom d'utilisateur', 'E-mail', 'Mot de passe', 'validation', 'Confirmation', and 'suppression'. The table contains three rows of user data.

Nom	Prénom	Nom d'utilisateur	E-mail	Mot de passe	validation	Confirmation	suppression
a	a	a	a	a	1	Confirmer	supprimer
temzi	abdelali	abdelali.temzi	tedaab20@gmail.com	admin	1	Confirmer	supprimer
ouhdadou	billal	ouhdadou	ouhdadaou@gmail.com	bilala	0	Confirmer	supprimer

Figure 34 : Administrateur page d'accueil 2.

❖ Page des statistiques



The screenshot shows the 'INFO-SECTE' administrator interface for the 'Statistics' page. The navigation menu and user profile are consistent with the previous screenshot. The main heading is 'Statistics'. There are two summary cards displayed:

- Platform users:**
 - Number of admins: **2 admins**
 - Number of users: **1 users**
 - Number of users validated: **1 users**
 - Number of nonvalid users: **0 users**
- Sect and characteristics:**
 - Basic sects: **6 sects**
 - Secondary sects: **3 sects**
 - Characteristics: **15**
 - Ressources number: **1**

Figure 35 : Administrateur page d'accueil 3 -statistiques-.

Chapitre III: Tests et Implémentation

❖ Ajouter un administrateur

The screenshot shows the 'Ajouter un administrateur' page of the INFO-SECTE application. The page has a header with the logo 'INFO-SECTE' and navigation links: 'Accueil', 'Messages', 'A propos', 'Pages', and 'Paramètres'. The user 'abdelali.temzi' is logged in. Below the header, there are two tabs: 'Ajouter admin' (active) and 'Ajouter un utilisateur'. The main content area features a green header 'Ajouter un administrateur' and a form with the following fields: 'Nom *' (Firstname), 'Prénom *' (Lastname), 'Nom d'utilisateur:' (Username), 'E-mail*' (E-mail), and 'Mot de passe*' (Password). A 'valider' button is located at the bottom of the form.

Figure 36: Administrateur page 4 ajouter un administrateur.

❖ Envoyer message

The screenshot shows the 'Contactez-nous' page of the INFO-SECTE application. The page has the same header as Figure 36. The main content area features the heading 'Contactez-nous' and a form with the following fields: 'Nom et prénom', 'Address e-mail', and 'votre message'. A dark 'Envoyer' button is located to the right of the 'Address e-mail' field. At the bottom of the page, there is a copyright notice: 'Copyright © 2021 Abdelalitemzi - Design: link' and social media icons for Facebook, Twitter, and Instagram.

Figure 37 : Page Messagerie

Chapitre III: Tests et Implémentation

III.2.6.2 Session expert

❖ Page d'accueil

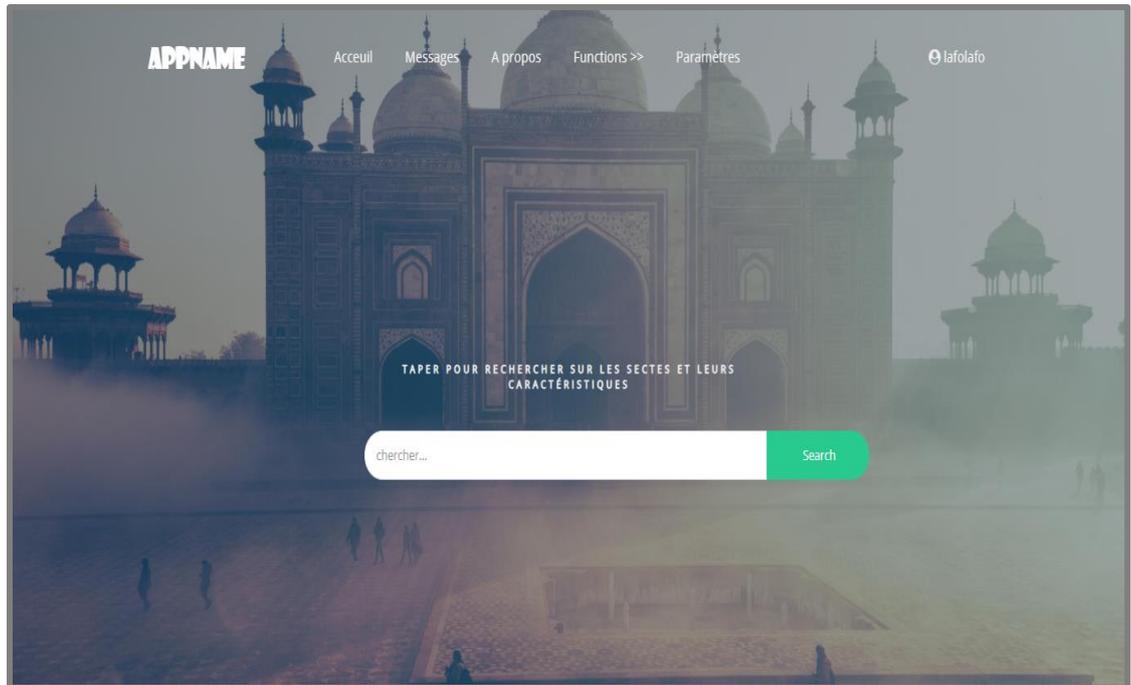


Figure 38 : page d'accueil expert 1.

❖ Taches de l'expert

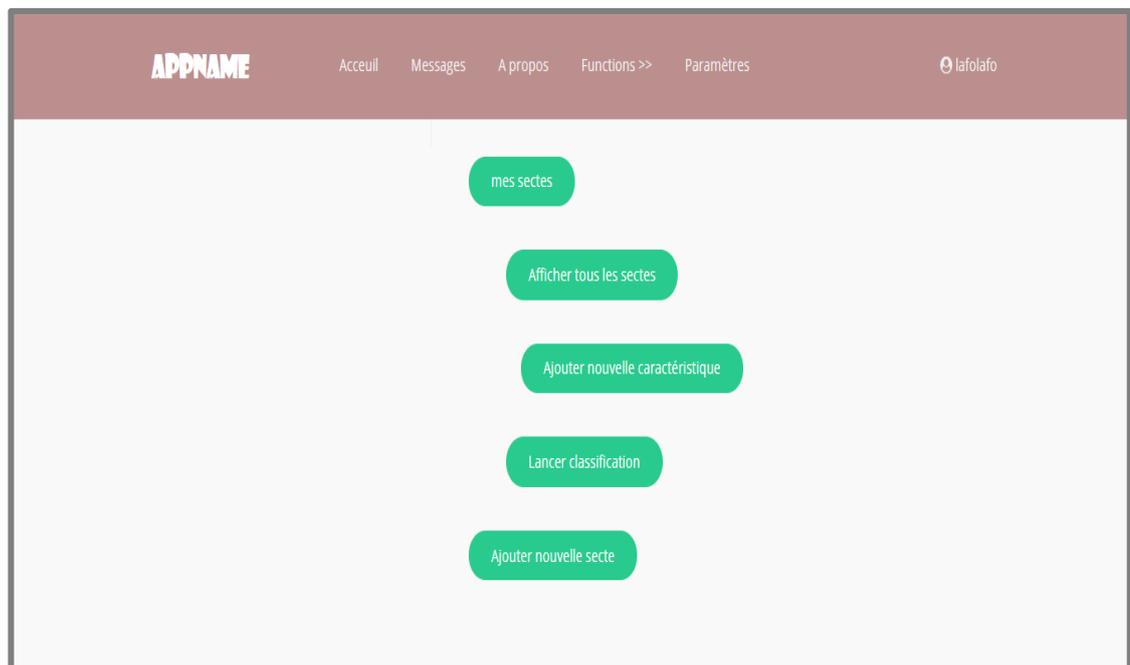


Figure 39: taches que l'expert peut les faire.

❖ Informations sur les sectes



Figure 40: Quelques statistiques que l'expert peut consultées.

❖ Page Messagerie

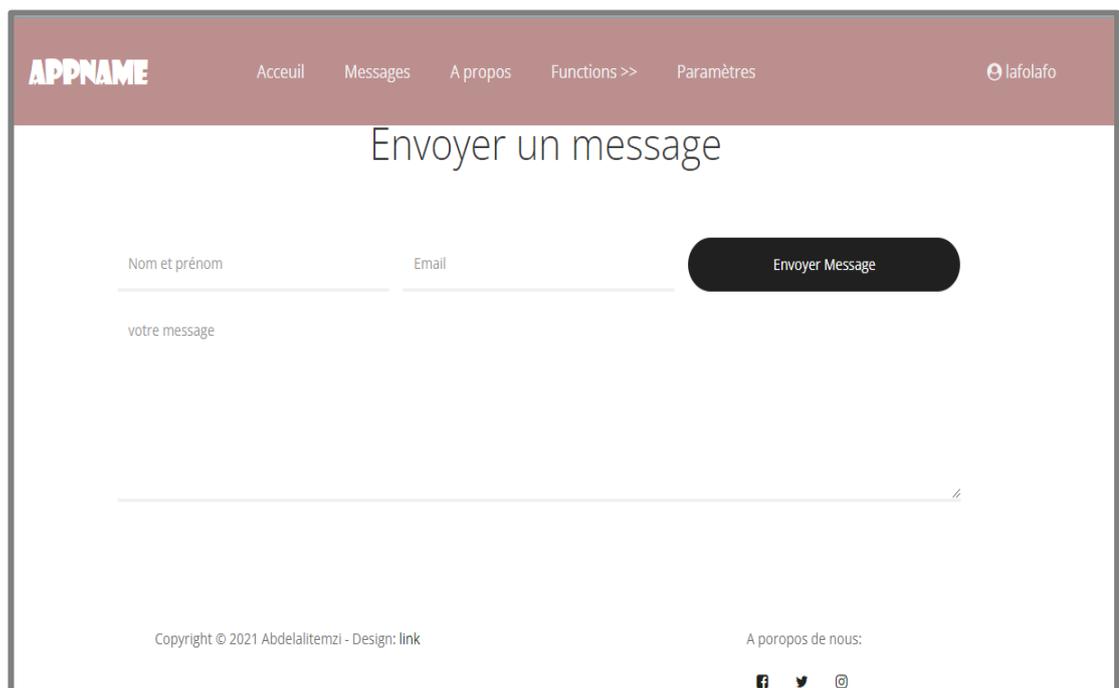
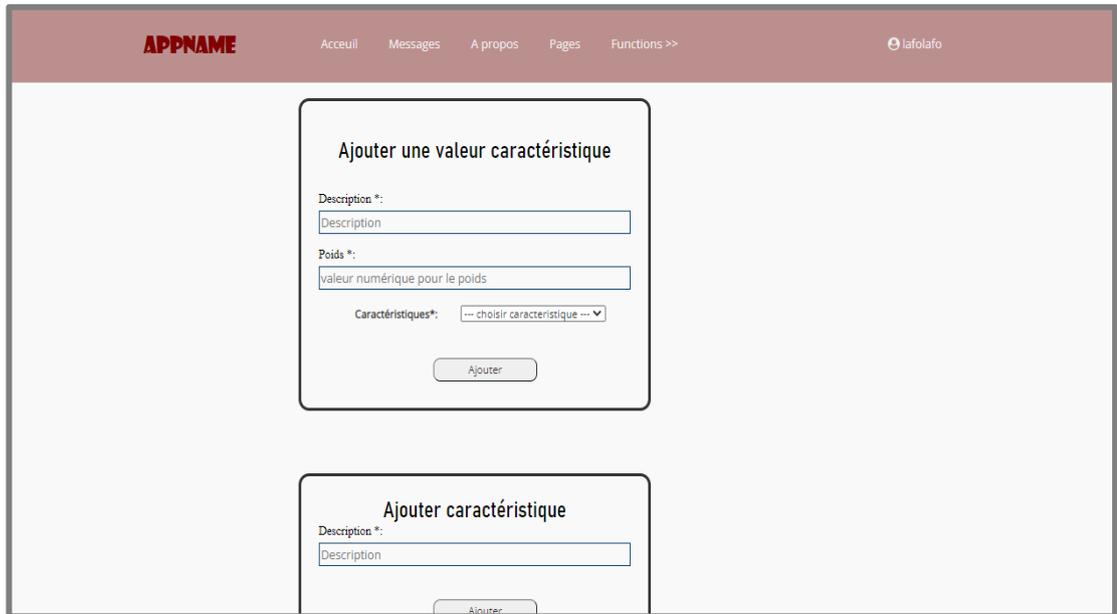


Figure 41 : interface pour envoyer un message.

Chapitre III: Tests et Implémentation

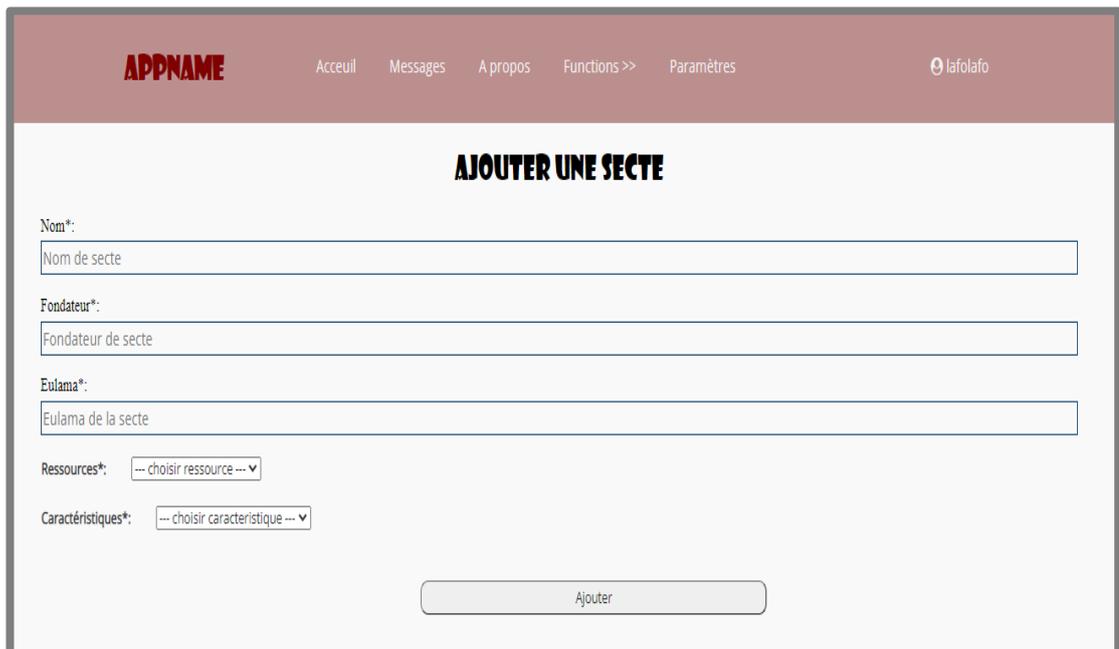
❖ Page ajouter des caractéristique et valeurs pour les caractéristiques



The screenshot shows two forms on a web page. The top form is titled 'Ajouter une valeur caractéristique' and contains the following fields: 'Description *:' with a text input containing 'Description'; 'Poids *:' with a text input containing 'valeur numérique pour le poids'; and 'Caractéristiques*:' with a dropdown menu showing '--- choisir caractéristique ---'. Below these fields is an 'Ajouter' button. The bottom form is titled 'Ajouter caractéristique' and contains the field 'Description *:' with a text input containing 'Description'. Below this field is an 'Ajouter' button.

Figure 42 : ajouter des caractéristiques et valeur pour des caractéristiques

❖ Ajouter une secte



The screenshot shows a form titled 'AJOUTER UNE SECTE'. The form contains the following fields: 'Nom*:' with a text input containing 'Nom de secte'; 'Fondateur*:' with a text input containing 'Fondateur de secte'; 'Eulama*:' with a text input containing 'Eulama de la secte'; 'Ressources*:' with a dropdown menu showing '--- choisir ressource ---'; and 'Caractéristiques*:' with a dropdown menu showing '--- choisir caractéristique ---'. Below these fields is an 'Ajouter' button.

Figure 43: Ajouter une nouvelle secte.

❖ La liste de toutes les sectes qui ont été ajoutées par l'expert de ce compte

Chapitre III: Tests et Implémentation

APPNAME Acceuil Messages A propos Pages Paramètres 				
Liste de sectes que vous avez déjà ajouté				
nom secte	Fondateur	Eulama de secte	Ressource	Caractéristique
القدرية	1	1	الفرق بين الفرق	قصيدة خلق القرآن الكريم
القدرية	1	1	الفرق بين الفرق	الخروج على أئمة الجور
الجهية	--	الجهيم بن صفوان الترمذي	الفرق بين الفرق	فناء الجنة والنار
الجهية	--	الجهيم بن صفوان الترمذي	الفرق بين الفرق	رؤية الله في الآخرة
معتزلة	السيبيل إلى معرفة الحق	واصل بن عطاء	الفرق بين الفرق	صفات الله وأسماءه
معتزلة	السيبيل إلى معرفة الحق	واصل بن عطاء	الفرق بين الفرق	الإمامة في الدين
معتزلة	السيبيل إلى معرفة الحق	واصل بن عطاء	الفرق بين الفرق	خلافة الخلفاء

Figure 44: Afficher tous les sectes ajoutées par l'expert

- ❖ La liste de toutes les sectes de la base de données avec l'expert qui les a ajouté

APPNAME Acceuil Messages A propos Pages >> Paramètres 						
Liste de toutes les sectes seulement les sectes qui ont été ajouté par notre experts						
ID secte	nom secte	fondateur	Eulama	Ressource	Caractéristique	Ajouté par -id-
0	السنة	amkzjdh	mlazlmazraz	razi001	11	
1	aaa	aa	aa	الفرق بين الفرق	12	1
2	azdazd	azdazd	azdazd	الفرق بين الفرق	11	1
6	a	az	ódz	الفرق بين الفرق	13	1
7	dazdazd	azdazdaz	dazdazda	الفرق بين الفرق	7	1
10	temzi	abdelali	temzi	الفرق بين الفرق	14	54696

Figure 45: La liste de toutes les sectes de la base de données

Chapitre III: Tests et Implémentation

III.2.6.3 Session visiteur

Dans notre cas on a fait seulement une interface pour les visiteurs avec une fonctionnalité de la recherche sur les informations des sectes.

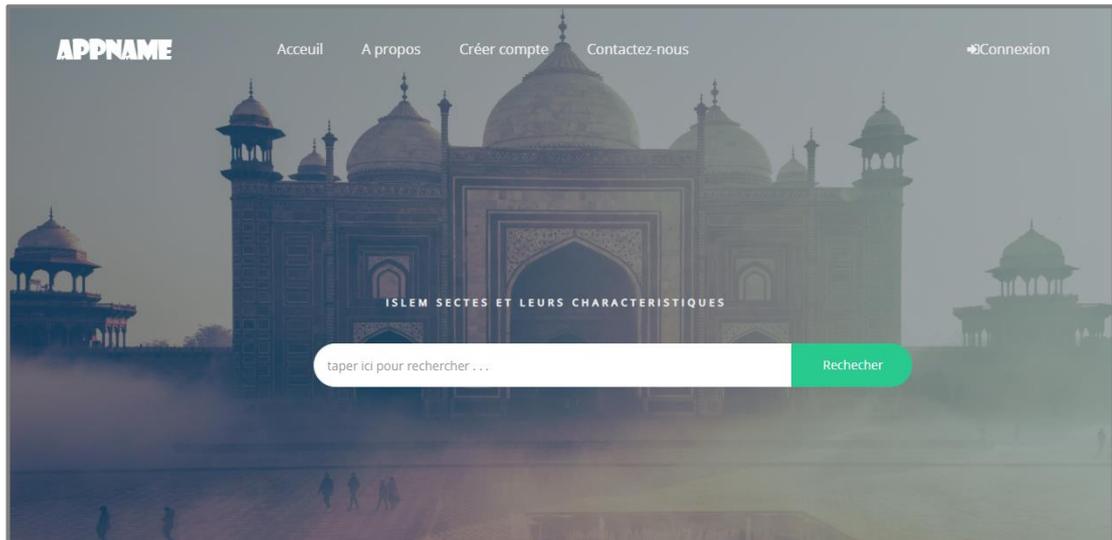


Figure 46: Interface 1 de la recherche de visiteur.



Nom de secte	Fondateur	Eulama	Ressources	Caractéristique
القدرية	كتاب المصنف	عولان المشفي	من كتاب الرازي	يتألمون قدر الله تعالى

Figure 47: Interface de l'affichage les résultats de la recherche.

III.3 Conclusion

Dans ce chapitre, nous avons défini l'Architecture matérielle et technique du système. Par ailleurs, nous avons aussi présenté les supports logiciels utilisés et l'environnement de développement (langage de programmation, serveurs,...) choisis sur lequel repose notre application.

Enfin, nous avons présenté la vue réelle de notre application, et des principaux modules qui la font fonctionner en toute simplicité, rapidité, grâce à une interface intuitive et adaptative (responsive), simple respectant les concepts fondamentaux et les contraintes actuelles dans le monde de la programmation Web.

Conclusion générale

Conclusion générale

III.1 Synthèse

Dans un hadith authentique, le Messager d'Allah a dit : « La communauté juive s'est scindée en soixante et onze ou soixante-douze factions, toutes iront en enfer. La communauté chrétienne s'est également scindée en soixante et onze ou soixante-douze factions, toutes iront en enfer. Ma communauté-ci se divisera en soixante-treize factions, toutes iront en enfer sauf une : celle qui a la même religion que moi et mes compagnons avons aujourd'hui. »

Ce hadith très connu a fait l'objet d'études de savants religieux pour essayer de déterminer les sectes vouées au feu de l'enfer d'entre celles se prévalant de l'islam. Toutefois, ces mêmes oulémas ont buté sur un problème : trouver les traits distinctifs qui permettent de retrouver soixante-treize parties. En fait, certaines parties se sont ramifiées au point que le nombre total des parties et de leurs ramifications diverses dépasse le nombre signalé. Assurément, il y a une erreur dans la façon d'approcher le problème : soit on considère, à partir d'un point donné, que les ramifications forment la même classe que la classe initiale, et dans ce cas, on ne retrouve pas le nombre de 73, soit on garde les ramifications comme des parties distinctes et on trouve plus de 73 !

L'idée était d'essayer diverses approches prenant en considération les choses les plus fondamentales sur lesquelles il y a eu des divergences et essayer de parvenir à une approche qui donne le nombre de soixante-treize recherché même s'il faut pour cela faire des expérimentations en évinçant tel caractère distinctif ou tel autre.

Bien que pour l'informaticien, le problème soit un problème de classification, il n'en demeure pas moins que les retombées peuvent être bénéfiques : est-ce que la partie à laquelle on appartient est la bonne ou bien sommes-nous dans l'erreur.

Cette tâche aurait pu être résolue en combinant des techniques de récupération d'informations et d'apprentissage automatique. Tel qu'il existe deux approches impliquées dans le traitement de la classification de texte : la première approche est liée à l'extraction des termes caractéristiques qui sont reconnus comme des mots-clés efficaces dans la phase d'apprentissage qui distinguent chaque catégorie de l'autre, et

Conclusion générale

la seconde approche concerne la classification réelle du document à l'aide de ces termes caractéristiques dans la phase de test

Ainsi, les principaux objectifs derrière ce projet étaient les suivants :

- Trouver les fameux termes caractéristiques qui permettraient de dresser les traits distinctifs des différentes communautés et construire une base de connaissances pouvant servir dans la classification.
- Tester différents algorithmes de classification des textes religieux en différentes catégories prédéfinie.
- Déterminer les performances et l'efficacité du modèle de catégorisation et essayer d'obtenir une relation ou une corrélation entre les classifications obtenues.

III.2 Perspectives

Au terme de ce travail, nous avons effectué quelques tests de classification avec les algorithmes d'Apprentissage Automatique: k-means et arbre de décision et nous avons pu tirer quelques conclusions et vu un ensemble de clusters représentant une éventuelle classification des sectes et un éventuel nombre de ces dernières. Nous estimons toutefois que nous avons été un peu loin d'atteindre les objectifs escomptés à cause de la non disponibilité des données et la non complétude et la confirmation des informations sur les caractéristiques des sectes et bien que l'application de gestion des sectes, offre quelques-unes des fonctionnalités souhaitées afin d'alimenter les caractéristiques des sectes de manière manuelle, beaucoup de tâches restent en suspens. Ci-dessous, nous allons en citer les plus importantes à notre connaissance :

- Envisager la mise en place d'une interface de navigation interactive dans l'arborescence des sectes afin de permettre de passer en revue les différentes sectes avec la possibilité de s'attarder sur une secte, une ressource, ou une information en particulier.
- Envisager le recours aux techniques de Traitement Automatique de la Langue afin de pouvoir faire une extraction automatique et non manuelle des caractéristiques et ce, à partir des livres sélectionnés pour la classification ou les ressources des différents savants (eulamas) par sectes.

Conclusion générale

- Prévoir une interface de recherche d'information à travers laquelle l'utilisateur sera en mesure de satisfaire son besoin en Information dans le domaine en saisissant : soit le nom d'une secte, le nom d'une sous secte, le nom d'un savant, le nom d'une ressource, quelques caractéristiques d'une secte.

Références bibliographiques

Bibliographie

- [1] Christopher D. Manning, Hinrich Schütze, Foundations of Statistical Natural Language Processing, MIT Press (1999), (ISBN 978-0-262-13360-9), p. xxxi.
- [2] Hausmane Issarane, Consulté le 17 juillet 2021, <https://analyticsinsights.io/5-apprentissage-supervise/>.
- [3] Amine, S., Dartigues-Pallez, C., & Gaetan, R. (2020). CLASSIFICATION SUPERVISÉE DE DONNÉES PÉDAGOGIQUES POUR LA RÉUSSITE DANS L'ENSEIGNEMENT SUPÉRIEUR (Doctoral dissertation, I3S, Université Côte d'Azur).
- [4] Rakotomalala, R. (2005). Arbres de décision. Revue Modulad, 33, 163-187.
- [5] SIGISMONDI, C. (2011). Docteur en Sciences (Doctoral dissertation, UNIVERSITÉ de Nice-Sophia Antipolis).
- [6] J. B. MacQueen (1967). « Some Methods for classification and Analysis of Multivariate Observations [archive] » dans Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability 1: 281–297 p. Consulté le 25 juillet 2009.
- [7] S. Kapil, M. Chawla and M. D. Ansari, "On K-means data clustering algorithm with genetic algorithm," 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2016, pp. 202-206, doi: 10.1109/PDGC.2016.7913145.
- [8] En ligne, Consulté le 17 juillet 2021, <https://dataanalyticspost.com/Lexique/reduction-de-dimensionnalite/>
- [9] A.Kassambara, (15/10/2017), Consulté le 01 aout 2021, Méthodes des Composantes Principales dans R: Guide Pratique, <http://www.sthda.com/>
- [10] Taj al-Din Abu al-Fath Muhammad ibn Abd al-Karim ash-Shahrastani(1978) .Al-Milal wa al-Nihal [En Ligne]. Al-Halabi et Partners Est.
- [11] Abou-Mansur 'Abd-al-Kahir ibn-Tahir al-BaghdadiSeelye(1919). Al-Fark Bain al-Firak [En Ligne]. Maison des Livres Scientifiques Egypt.