



INSTITUT AFRICAIN D'INFORMATIQUE
Etablissement Inter-Etats d'Enseignement Supérieur
BP 2 263 Libreville - Gabon
Tel : +241 07505500 / www.iaisege.com



Fournisseur d'accès Internet et valeur ajoutée.
BP 12 155 Libreville – Gabon
Tel : +241 03082339 / www.ipi9.com

MEMOIRE DE FIN CYCLE

En vue de l'obtention du diplôme
D'Ingénieur de Conception en Informatique

THEME

**iPi9 Nexus : outil de Machine Learning
pour l'évaluation d'Entreprise.**

Réalisé et soutenu par :
GAN A BOL

Superviseur
M. Donaciен GUIFO
Enseignant permanent à IAI

Maitres de stage
M. Serge SEGbefia (Consultant iPi9)
M. Raz BIRAMAH DG iPi9

ANNEE ACADEMIQUE 2019-2020





THEME : iPi9 Nexus : outil de Machine Learning pour l'évaluation d'Entreprise.



DEDICACE

« Je dédie ce mémoire de fin de formation à tous les Chefs d'Etat des pays membres qui, par la Résolution N° 33/AEFT/Fort-Lamy du 29 janvier 1971, ont bien voulu créer l'Institut Africain d'Informatique (I.A.I) »

EPIGRAPHE

« La beauté lorsqu'elle est trouvée, est semblable au miroir de l'âme. Tout, mais alors tout devient clair et limpide, tout s'explique, tout s'illumine, tout paraît éternel, infini et réalisable... »

Gandhy.

REMERCIEMENTS

Par cette occasion unique, qui m'est donnée. J'exprime toute ma profonde et sincère gratitude à tous ceux et celles, qui ont pris part de près comme de loin, non pas seulement à l'aboutissement de ce mémoire, mais surtout à l'aboutissement de mes années de formation à l'IAI. Années de formation au cours desquelles, j'ai dû personnellement surmonter des expériences inédites au titre desquelles : une année blanche et cette pandémie mondiale provoquée par le coronavirus SARS-CoV-2 plus connue sous le nom de COVID-19.

Je tiens à exprimer toute ma reconnaissance tout d'abord à l'Éternel Dieu, le père tout puissant pour m'avoir sans cesse comblé et protégé durant ma formation.

J'adresse mes sincères remerciements au gouvernement de la République du Cameroun, pour la prise en charge des frais de scolarité durant tout le cycle et ses aides multiformes.

Mon profond respects et remerciements à L'Institut Africain d'Informatique (IAI), pour l'accueil, la formation et l'encadrement dont j'ai bénéficié.

Je remercie de façon toute particulière :

- Tout le personnel de la structure iPi9, pour l'accueil chaleureux et toute l'attention accordée à ma modeste personne, entre autres ;
 - M. Raz BIRAMAH, Directeur Général de iPi9 et maître de stage, pour toute la confiance dont il m'a gratifié ;
 - M. Serge SEGBEFIA, Consultant à iPi9 et maître de stage, pour tous ses multiples conseils, corrections et orientations.
- Le corps administratif, professoral et personnel d'appui de l'IAI pour l'encadrement, la formation et conseils qu'ils m'ont octroyés, notamment :
 - Monsieur Donacien GUIFO, pour avoir accepté superviser mes travaux durant ce stage ;
 - Monsieur BENADJINGAR Goldoum, Coordonnateur de la filière Ingénieur ;
 - Professeur Souleymane KOUSSOUBE et Dr. Roger NOUSSI pour le suivi continu.
- Toute la communauté étudiante de l'IAI qui pour moi est devenue une famille, notamment mes camarades de promotions ;
- La grande famille BOL A BEYECK, notamment mes très chers parents, mes précieux frères et sœurs, qui ont toujours été là pour moi, je vous suis redevable.

A tous ceux et celles dont le nom ne figure pas sur cette liste, je tiens à vous exprimer ma reconnaissance, gratitude car je vous suis à tous reconnaissant.

SOMMAIRE

DEDICACE.....	i
EPIGRAPHÉ.....	ii
REMERCIEMENTS	iii
SOMMAIRE	iv
AVANT-PROPOS.....	vii
RESUME.....	viii
ABSTRACT	ix
TABLE DES ILLUSTRATIONS.....	x
LISTE DES SIGLES ET ABREVIATIONS.....	xii
INTRODUCTION GENERALE.....	1
PREMIERE PARTIE : CONTEXTE GENERAL DE L'ETUDE	2
CHAPITRE I : PRESENTATION DE LA STRUCTURE D'ACCUEIL ET DU SUJET D'ETUDE... 3	3
Introduction.....	3
I.1 iPi9	3
I.1.1 Présentation générale.....	3
I.1.2 Les services de iPi9	4
I.1.3 Structure interne	5
I.2 Présentation du sujet.....	6
I.2.1 Contexte du stage	6
I.2.2 Problématique du sujet	6
I.2.3 Objectif du projet.....	7
I.2.4 Libellé du sujet	8
Conclusion	8
CHAPITRE II : CONCEPTS LIÉS A L'EVALUATION DE PROJET D'ENTREPRISE..... 9	9
Introduction.....	9
II.1 Principales notions	9
II.1.1 Qu'est-ce qu'une opportunité d'affaire ?.....	9
II.1.2 Qu'est-ce qu'un KPI ?	10
II.1.3 Qu'est-ce qu'une étude d'opportunité ou business case ?	12
Conclusion	15
CHAPITRE III : AUDIT & LIMITES DES SOLUTIONS EXISTANTES	16
Introduction.....	16
III.1 Solutions existantes	16
III.1.1 Créer Mon Business Plan	16
III.1.2 Moo.com.....	17
III.1.3 Les tests « Êtes-vous prêt à entreprendre ? ».	17
III.2 Limites des solutions existantes	18

Conclusion	19
DEUXIEME PARTIE : MACHINE LEARNING & SES ALGORITHMES.....	20
CHAPITRE IV : PRESENTATION DU MACHINE LEARNING.....	21
Introduction.....	21
IV.1 Histoire du Machine Learning.....	21
IV.2 Evolution du Machine Learning.....	22
IV.3 Types de problèmes en Machine Learning.....	24
IV.3.1 Supervised Learning (Apprentissage Supervisé).....	24
IV.3.2 Unsupervised Learning (Apprentissage non supervisée)	25
IV.3.3 LES ALGORITHMES DE ML	26
IV.4 Aborder un problème de Machine Learning.....	27
IV.4.1 Définition du problème.....	28
IV.4.2 Préparation des données	28
IV.4.3 Choix du bon algorithme	30
Conclusion	32
CHAPITRE V : Algorithme K-Nearest Neighbours	33
Introduction.....	33
V.1 Présentation et mode de fonctionnement de K-NN	33
V.2 Principe détaillé du fonctionnement de K-NN	33
V.3 Définition de la fonction de similarité dite de distance d de l'algorithme K-NN.....	34
V.4 Détermination de la plage de K	35
V.5 Construction de notre modèle.....	36
Conclusion	39
TROISIEME PARTIE : REALISATION DU SYSTEME	40
Chapitre VI : MÉTHODOLOGIE DE DÉVELOPPEMENT.....	41
Introduction.....	41
VI.1 Tour d'horizon des méthodes de développement	41
VI.2 Choix d'une méthode	42
VI.2.1 Présentation du langage UML	42
VI.2.2 2-TUP	43
Conclusion	44
Chapitre VII : MISE EN ŒUVRE DE LA SOLUTION.....	45
Introduction.....	45
VII.1 Itération 1 : Étude globale du système.....	45
VII.1.1 Étude préalable	45
VII.1.2 Capture des besoins fonctionnels	49
VII.1.5 Analyse	61

VII.1.6	Autres activités	62
VII.2	Itération 2 : Gestion profils entreprises	63
VII.2.1	Étude préalable	63
VII.2.2	Capture des besoins fonctionnels	64
VII.2.3	Capture des besoins techniques	65
VII.2.4	Analyse	65
VII.2.5	Conception préliminaire	67
VII.2.6	Conception détaillée	71
VII.2.7	Codage	74
	Conclusion	75
	QUATRIEME PARTIE : FINALISATION DE LA SOLUTION	76
	Chapitre VIII : DÉPLOIEMENT ET RÉSULTATS	77
	Introduction	77
VIII.1	Diagramme de déploiement	77
VIII.2	Résultats obtenus	78
	CHAPITRE IX : CONDUITE DE PROJET	80
	Introduction	80
IX.1	Organisation du projet	80
IX.1.1	Découpage du projet en phases & tâches	80
IX.1.2	Intervenants et déroulement du stage	81
IX.1.3	Diagramme de GANTT	82
IX.1.4	Estimation des charges et des coûts	84
IX.2	Bilan et perspectives	85
IX.2.1	Apports du stage et difficultés rencontrée	85
IX.2.2	Perspectives	86
	Conclusion	86
	CONCLUSION GÉNÉRALE	87
	RÉFÉRENCES BIBLIOGRAPHIQUES	i
	WEBOGRAPHIE	i
	ANNEXES	iii

AVANT-PROPOS

L’Institut Africain d’Informatique (**IAI**), est une École Inter-états d’Enseignement Supérieur qui compte actuellement onze (11) pays d’Afrique francophone, dont le siège est au Gabon (Libreville). Les États membres actuellement impliqués sont le Bénin, le Burkina Faso, le Cameroun, la République Centrafricaine, le Congo, la Côte d’Ivoire, le Gabon, le Niger, le Sénégal, le Tchad et le Togo. Il a été créé le 29 janvier 1971 à FORT-LAMY, ancien nom de la capitale de la République du Tchad. L’IAI comprend les cycles : Analyste Programmeur, Licence, MIAGE, Master et Ingénieur.

Dans le cursus de formation des Ingénieurs de Conception en Informatique, en fin de cycle, il intègre un stage de formation pratique d'une durée de cinq (05) mois en entreprise ou dans un laboratoire de recherche, et donne lieu à la rédaction puis à la soutenance d'un mémoire de fin d'études. Ce stage de fin d'étude vise à mettre les élèves-ingénieurs dans un environnement de conception d'applications informatiques ou dans un centre de Recherche et de Développement (laboratoire), afin de leur permettre une intégration en milieu professionnel et/ou scientifique.

Suivant les termes décrits dans le « Cahier des charges 3A Ingénieur », nous avons effectué notre stage au sein de la structure **iPi9**, Entreprise fournisseur d'accès Internet et des Services à valeur ajoutée du Gabon pour une durée de cinq (05) mois.

Le présent document fait office de l'aboutissement de trois (03) années de formation à l’Institut Africain d’Informatique à Libreville au Gabon et cinq (05) mois de stage pratique effectué au sein de **iPi9**. Il tient lieu de mémoire de fin de formation d’Ingénieur de Conception en Informatique. Le travail ainsi réalisé au sein de cette structure s’intitule :

« iPi9 Nexus outil de Machine Learning pour l'évaluation d'Entreprise. »

RESUME

Dans la pratique, l'évaluation d'une opportunité d'affaire (Entreprise), est soutenue par une proposition structurée : **le business case ou analyse de rentabilisation en français**. Cette dernière se trouve être en grande partie une étude de faisabilité économique. Si elle est concluante, elle est destinée à rassurer de la viabilité d'un investissement. Le business case est un document réalisé à la demande d'un investisseur auprès d'un expert en étude de projet, ce qui permet à l'investisseur de se rassurer de la viabilité de son investissement. Cependant, cette technique bien qu'efficace, repose sur le capital Homme. N'ayant pas fait l'objet d'automatisation, elle est fastidieuse, lente et laisse place à un risque d'erreur trop élevé lorsqu'il s'agira d'évaluer plusieurs opportunités d'affaire. Concevoir via la programmation classique, une application réalisant cette tâche serait trop difficile, car il va falloir gérer plusieurs conditions devant nous conduire vers un programme avec une complexité élevée. Par contre, en utilisant les techniques nouvelles de Machine Learning, résoudre ce problème en se basant sur un dataset assez fourni serait moins complexe et donc plus fiable.

Ainsi, à travers le présent mémoire, nous proposons une approche nouvelle pour résoudre ce problème. Notre objectif est de proposer **iPi9 Nexus** : une plateforme web tout en utilisant le paradigme objet couplée aux techniques de Machine Learning. Dès lors, notre solution permettra non seulement de mettre en relation les entrepreneurs locaux et les investisseurs du monde, mais surtout de pouvoir prédire avec un fort taux de réussite la viabilité ou non d'une opportunité d'affaire.

Nous avons particulièrement mis un accent sur la sécurisation des données collectées auprès des entrepreneurs car au fil de nos travaux nous nous sommes rendu compte de la sensibilité très forte de ces informations, ce qui pourrait justifier à juste titre la rareté de ces informations tant sur internet qu'auprès des organes en charge de la régulation de ce secteur. Pour la modélisation de notre solution nous utiliserons le langage de modélisation unifié (UML) et pour la programmation nous utiliserons le langage Python.

ABSTRACT

In practice, the assessment of a business opportunity (Company) is supported by a structured proposal: the business case or business case in French. The latter happens to be largely an economic feasibility study. If it is conclusive, it is intended to reassure the viability of an investment. The business case is a document produced at the request of an investor from a project study expert, which allows the investor to be reassured of the viability of his investment. However, this technique, although effective, relies on human capital. Not having been automated, it is tedious, slow and leaves room for too high a risk of error when it comes to evaluating multiple business opportunities. Designing via classical programming, an application performing this task would be too difficult, because it will be necessary to manage several conditions that should lead us to a program with a high complexity. On the other hand, by using new machine learning techniques, solving this problem based on a fairly extensive dataset would be less complex and therefore more reliable.

Thus, through this memo, we propose a new approach to solve this problem. Our goal is to offer iPi9 Nexus : a web platform while using the object paradigm coupled with Machine Learning techniques. Therefore, our solution will not only connect local entrepreneurs and investors from around the world, but above all to be able to predict with a high success rate the viability or otherwise of a business opportunity.

We have placed particular emphasis on securing the data collected from entrepreneurs because throughout our work we have realized the very high sensitivity of this information, which could rightly justify the scarcity of this information both on the internet. than with the bodies in charge of regulating this sector. For the modeling of our solution we will use the unified modeling language (UML) and for the programming we will use the Python language.

TABLE DES ILLUSTRATIONS

Liste des figures

Figure 1 iPi9 Puce 4G	4
Figure 2: Box internet iPi9	4
Figure 3: Box D-Link iPi9	4
Figure 4 logo iPi9	5
Figure 5 logo web7.....	5
Figure 6 Organigramme iPi9	5
Figure 7 Page d'accueil Créeer Mon Business Plan	16
Figure 8 Moo.com	17
Figure 9 Êtes-vous prêt à entreprendre ?	17
Figure 10 Point de rencontre Expertise métier, mathématique & informatique	23
Figure 11 Types de problèmes de ML	24
Figure 12 Illustration de la différence entre classification linéaire et régression linéaire.....	25
Figure 13 Résultat de clustering.....	26
Figure 14 Les algorithmes de Machine Learning.....	26
Figure 15 CRISP-DM.....	27
Figure 16 Process de choix d'une classe d'algorithmes	30
Figure 17 Ensemble d'activités de conception et production d'un modèle de ML.....	32
Figure 18 : Cycle en Y	43
Figure 19 Phases et activités du cycle de développement.....	44
Figure 20 Diagramme de contexte statique - Etude globale du système	47
Figure 21 Diagramme de contexte dynamique - Etude globale du système	47
Figure 22 Diagramme de cas d'utilisation – Etude Globale	53
Figure 23 Architecture MVT de la solution	60
Figure 24 Architecture retenue solution.....	60
Figure 25 Diagramme de classes - Etude globale du système	61
Figure 26 Diagramme d'activité - Etude globale du système.....	62
Figure 27 Diagramme de composants du système - Étude globale du système.....	62
Figure 28 Diagramme de cas d'utilisation - Gestion profils entreprises	64
Figure 29 Diagramme de classes - Gestion profils entreprises	65
Figure 30 Diagramme de séquence du cas d'utilisation « Éditer profil entreprise »	66
Figure 31 Diagramme de séquence du cas d'utilisation « Consulter profil entreprise »	66
Figure 32 Diagramme de séquence du cas d'utilisation « Désactiver profil entreprise »	67
Figure 33 DCP du cas d'utilisation « Editer profil entreprise ».....	68
Figure 34 DCP du cas d'utilisation « Consulter profil entreprise »	68
Figure 35 DSOP du cas d'utilisation « Créer profil Entreprise ».....	69
Figure 36 DSOP du cas d'utilisation « Consulter profil entreprise ».....	70
Figure 37 DSOP du cas d'utilisation « Désactiver profil entreprise »	70
Figure 38 Diagramme de classes détaillées (DCD)	71
Figure 39 DSOD du cas d'utilisation « Créer profil Entreprise »	72

Figure 40 DSD du cas d'utilisation « Consulter profil entreprise ».....	73
Figure 41 DSD du cas d'utilisation « Désactiver profil entreprise »	73
Figure 42 Implémentation de la classe « Entreprise ».....	74
Figure 43 Implémentation de la classe « Profil_Entreprise »	74
Figure 44 Implémentation des classes « Consulter & Message »	75
Figure 45 Diagramme de déploiement iPi9 Nexus	77
Figure 46 Page d'accueil de la plateforme iPi9 Nexus	78
Figure 47 Formulaire de connexion	78
Figure 48 Dashboard Entrepreneur	79
Figure 49 Dashboard Investisseur	79
Figure 50: SplashScreen iPi9TV	82
Figure 51: Page accueil iPi9TV	82
Figure 52: Espace de lecture de chaîne iPi9TV	82
Figure 53: Diagramme de Gantt prévisionnel	83
Figure 54: Diagramme de Gantt réel.....	84

Liste des tableaux

Tableau 1 Identification des messages Étude préalable	46
Tableau 2 Identification des cas d'utilisation - Etude globale du système	49
Tableau 3 Structuration en packages uses case - Etude globale du système.....	54
Tableau 4 Planification en itération - Etude globale du système	55
Tableau 5 Besoins techniques - Etude globale du système	58
Tableau 8 Identification des messages itération 2.....	64
Tableau 9 Identification des cas d'utilisation itération 2	64
Tableau 10: Phases et Tâches du Projet	81
Tableau 11: Coût logiciel du projet.....	85
Tableau 12: Coût matériel du projet.....	85
Tableau 13: Coût humain du projet.....	85

Liste des équations

Équation 1 Distance euclidienne	35
Équation 2 Distance Manhattan	35
Équation 3 Distance Hamming	35

Liste des annexes

Annexe 1: Questionnaire sur les informations des Entreprises.....	iii
---	-----

LISTE DES SIGLES ET ABBREVIATIONS

2TUP	2-Track Unified Process
4G	Quatrième Génération
CPU	Central Processing Unit
CRISP-DM	Cross Industry Standard Process for Data Mining
CSS	Cascading Style Sheets
DCD	Diagramme de classes détaillées
DCP	Diagramme de classes participantes
DSOD	Diagramme de séquences-objets détaillés
DSOP	Diagrammes de séquence-objet préliminaire
DSS	Diagrammes de séquence système
FCFA	Franc de la communauté financière africaine
HTML	Hypertext MakeUp Language
HTTP	HyperText Transfer Protocol
HTTPS	HyperText Transfer Protocol Secure
IAI	Institut Africain d'Informatique
iPi9	Innovation Performance Intégration dans les 9 Provinces
K-NN	K-nearest neighbors
KPI	Key Primary Indicator
ML	Machine Learning
MVC	Modèle Vue Control
MVT	Modèle Vue Template
OMT	Object Modeling
ORM	Object Relational Mapping
SGBD	Systèmes de Gestion de Bases de Données
SHA	Secure Hash Algorithm
SIM	Security Information Management System
SQL	Structured Query Language
SVC	Support Vector Clustering
SVM	Support Vector Machines
UML	Unified Modeling Language
WI-FY	Wireless Fidelity
WIMAX	Worldwide Interoperability for Microwave Access)

INTRODUCTION GENERALE

Nous vivons une ère où, la révolution informatique impacte considérablement tous les secteurs de la vie. L'émergence de nouvelles technologies a révolutionné notre quotidien et nous oblige à adopter des comportements nouveaux. Cela va bien au-delà de révolutionner le quotidien, ces nouvelles technologies repoussent aussi les barrières que connaissaient plusieurs travaux de recherches, de programmation et les défis de l'homme.

De ces technologies émergentes figure le data mining ou fouille de données, très souvent associée au Machine Learning. L'émergence du Machine Learning est venue redonner un nouvel élan à la façon de solutionner des problèmes. Associée au Big Data, il apporte une nouvelle vision et d'avantages de connaissances dans les activités de prédiction et de description.

Du fait de la crise économique sévissant dans plusieurs pays, la voie de l'entrepreneuriat est de plus en plus empruntée et les Etats injectent des sommes faramineuses dans ce secteur pour encourager la création des ressources. Malheureusement pour la grande majorité, l'accès au financement demeure un parcours de combattants et ce pour de multiples raisons.

Fort de ce constat, il est donc primordial de mettre en place un système permettant : aux entrepreneurs de publier leur entreprise (informations descriptives), afin non seulement d'avoir une visibilité, mais surtout de pouvoir lever des fonds et aux investisseurs d'évaluer une entreprise et d'avoir des recommandations.

Telles sont les préoccupations auxquelles nous avons été confrontées durant la période de stage que nous avons effectué au sein de la structure **iPi9**.

Le présent document présente le travail réalisé. Il est organisé en quatre (04) parties.

- Dans la première partie, nous présentons la structure d'accueil et le cadre général de l'étude. Il s'agit notamment de l'environnement du stage et des concepts clés du sujet.
- La deuxième partie sera principalement consacrée à la revue du Machine Learning et de ses algorithmes. Elle s'achèvera avec la construction du modèle prédictif.
- Dans la troisième partie nous présentons l'analyse des différentes méthodes d'analyse puis la conception et la mise en œuvre d'une interface applicative.
- La quatrième partie est dédiée à la réalisation et finalisation proprement dite du système et pour finir avec la présentation de la conduite de projet.

PREMIERE PARTIE : CONTEXTE GENERAL DE L'ETUDE

Dans cette partie de notre travail, nous situons le contexte dans lequel s'inscrit le sujet soumis à notre étude, la problématique qu'il soulève, son intérêt et les objectifs visés. Nous présenterons dans un premier temps la structure iPi9, qui nous a chaleureusement accueillis. Ensuite nous allons présenter les concepts clés nécessaires à la compréhension du sujet au moyen d'une revue bibliographique. Nous aborderons dans cet ordre, les notions d'opportunité d'affaire ; la création d'entreprise ; les KPI ; le business case. Et pour finir nous effectuerons une présentation de l'existant, qui nous permettra de donner les raisons nous ayant conduit vers l'adoption de l'outil Machine Learning.

Chapitres

CHAPITRE I : PRESENTATION DE LA STRUCTURE D'ACCUEIL ET DU SUJET D'ETUDE.....
CHAPITRE II : EVALUATION DE PROJET D'ENTREPRISE
CHAPITRE III : AUDIT & LIMITES DES SOLUTIONS EXISTANTES

CHAPITRE I : PRESENTATION DE LA STRUCTURE D'ACCUEIL ET DU SUJET D'ETUDE

Introduction

Dans ce chapitre, nous présentons la structure d'accueil : **iPi9** (Innovation Performance Intégration dans les 9 Provinces), ainsi que le sujet sur lequel nous avons travaillé.

I.1 iPi9

I.1.1 Présentation générale

iPi9 est une entreprise fournisseur d'accès Internet et des Services à valeur ajoutée. Crée en 2010 dans le but premier de fournir l'accès internet pour tous à toutes les couches sociales. iPi9 s'appuie sur des formules et du matériel novateur qui lui ont permis de chambouler l'environnement Internet au Gabon.

Avec pour mission de créer au Gabon, une « **révolution internet** » comparable à celle observée dans le secteur de la téléphonie mobile en démocratisant son accès à toutes les couches de la population, au moyen d'une offre :

- ✓ Compatible avec tous les besoins ;
- ✓ Compatible avec tous les budgets ;
- ✓ Compatible avec toutes les envies.

Dès octobre 2012, iPi9 a réussi une percée sans précédent sur le marché internet grand public au Gabon. Après 19 mois d'exploitation de la technologie WIMAX, elle comptait déjà quelques **36 000** clients sur environ **50 000 internautes** que comptabilisait le Gabon en 2010 répartis auprès de **06 opérateurs** majeurs à savoir **Gabon Télécom, Wifly, TLDC, Internet Gabon et SOLSI**, dépassant ainsi tous les opérateurs historiques de la place avec une croissance de l'ordre de **1 500** clients par mois.

L'identité de la structure est principalement reconnaissable via les concepts clés suivants :

- **Un internet sur mesure**
 - ✓ Du temps de connexion **prépayé** sur la base de **cartes de recharge** (vouchers) à partir de **1 000 FCFA** ;
 - ✓ Un Internet sans engagement ;
 - ✓ Pas de souscription obligatoire ;
 - ✓ Pas de montant forfaitaire à payer mensuellement.

- **Un internet sans frais d'installation**

- ✓ Seul l'achat du modem suffit : on branche et ça marche (100% Plug & Play) ;
- ✓ Pas besoin de recourir à des techniciens pour son installation.

A ce jour iPi9, conserve sa vision qui est de : Maintenir son **leadership** de l'internet grand public au Gabon en **améliorant** constamment sa **qualité de service** à l'aide des **technologies** les plus performantes.

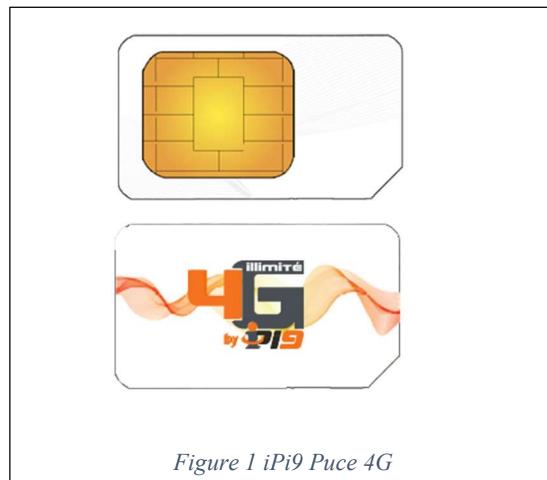
I.1.2 Les services de iPi9

Comme mentionné plus haut, iPi9 est une entreprise spécialisée dans la fourniture d'accès Internet et des Services à valeur ajoutée. A ce jour, ses services phares sont les suivants.

I.1.2.1 iPi9 accès à internet

Ce service est rendu possible grâce aux produits :

➤ **La SIM 4G**



➤ **Routeur 4G Mobile Wi-Fi**



I.1.2.2 iPi9 TV



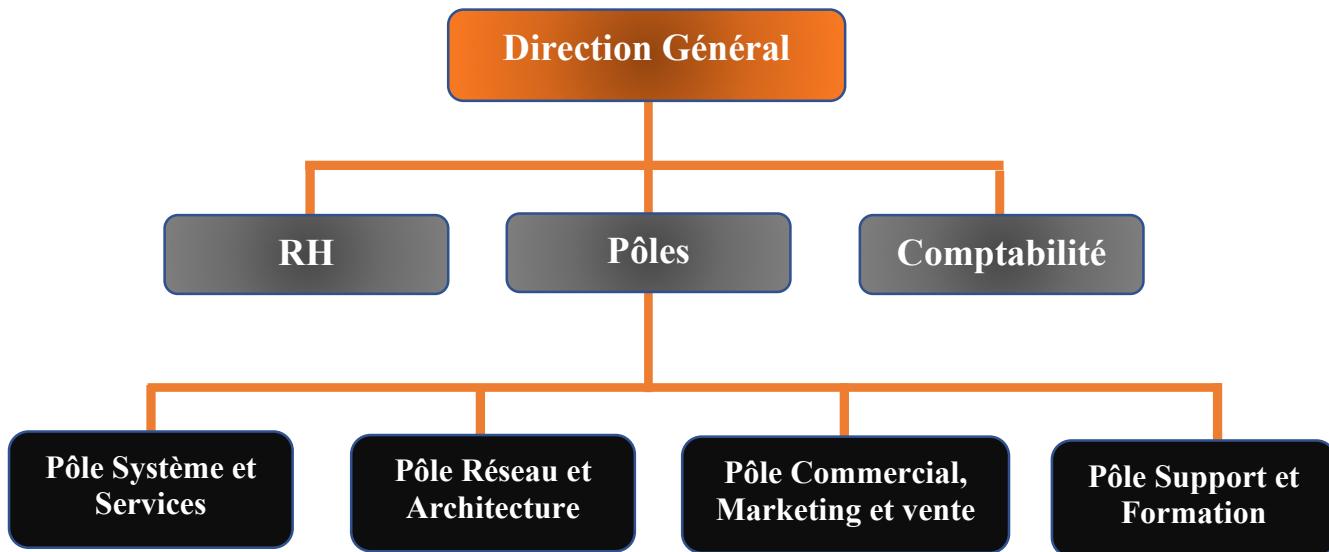
Service mettant à disposition une application mobile moderne. Offrant à tout internaute, pas uniquement aux abonnés iPi9, un accès à différentes chaînes de télévision en direct ou en différé, d'écouter la radio, de regarder des vidéos à la demande.

I.1.2.3 Web7



C'est un service de conception de site web personnalisé livré en 7 jours (ouvrés) chrono. Il est hébergé au Gabon en www.votrenomdedomaine.ga, livré clé en main avec une fonctionnalité de mise à jour en toute autonomie pour son propriétaire (via une interface d'administration), ainsi qu'avec des adresses email au format votreaddress@votredomaine.ga et plein d'autres services et fonctionnalités associés.

I.1.3 Structure interne



C'est au sein du **Pôle Système et Services** que nous avons été affecté pour effectuer notre stage.

I.2 Présentation du sujet

I.2.1 Contexte du stage

Depuis quelques années déjà, il est clairement admis que l'Afrique est et sera le lieu des plus importants taux de croissance dans les années à venir. Cette statistique positive est basée sur le fait que les économies africaines sont de plus en plus stables, et donc propices à la création d'entreprises utiles à l'écosystème africain et à ses communautés. Un tel environnement devient alors idéal pour les investisseurs. Cependant, ces derniers ont besoin d'évaluer les opportunités disponibles pour en ressortir celles qui présenteraient un fort potentiel de profits à court, moyen ou long terme.

Ce stage s'intègre dans une réflexion globale visant à rendre l'environnement des affaires africain attractif pour les investisseurs de toute origine tout en étant hautement profitable pour les entrepreneurs africains dont les solutions aideront à régler les problématiques des communautés autochtones.

I.2.2 Problématique du sujet

Évaluer une opportunité d'affaires est une étape importante avant d'investir du temps, de l'argent et d'autres ressources dans un projet.

A tout financeur potentiel, différentes approches sont proposées jusqu'à présent pour évaluer la viabilité d'un projet d'entreprise et de faire des recommandations sur les projets à fort potentiel lorsqu'il en existe plusieurs. Chacune de ces approches dépendent principalement de l'œil bien formé ou expérimenté de la personne qui l'utilise. On peut par exemple analyser le marché ou la situation industrielle de l'opportunité d'affaires en se posant comme questions:

- **Existe-t-il un marché pour l'idée d'entreprise ?**
- **Quelle est la taille dudit marché ?**
- **Quel est le taux de croissance de l'industrie ?**

Or, Évaluer une opportunité d'affaires nécessite en réalité de prendre en considération plusieurs autres paramètres tels que les indicateurs de performance financiers, commerciaux, de productivité, organisationnels... afin d'avoir une estimation objective. En effet, aborder le problème sous un angle trop restreint ou au contraire trop large pourrait conduire vers une mauvaise évaluation. Comme elles sont par ailleurs essentiellement manuelles, ces approches rendent la tâche encore plus ardue et rallonge la période d'évaluation.

De ce qui précède il se dégage donc les questions suivantes :

- ◆ Quelles sont les meilleurs critères à prendre en compte pour évaluer une opportunité d'affaire donnée ?
- ◆ Quel est le processus de sélection le plus approprié d'un projet du point de vue d'un investisseur ?
- ◆ Est-ce possible d'automatiser ce processus ? Et si oui, comment le faire en réduisant grandement la complexité du programme ?

I.2.3 Objectif du projet

Ce projet a pour but de concevoir et réaliser une plateforme web de Machine Learning (ML) qui permettra d'une part, d'effectuer des prédictions de viabilité d'un projet à court, moyen ou long terme sur la base d'un ensemble de données récoltées et, d'autre part, de proposer des recommandations sur les opportunités d'entreprises qui, sur la base de ces prédictions, présenteraient un potentiel intéressant en termes de retour sur investissement.

Le résultat final attendu est un espace de mise en relation des investisseurs, en quête d'opportunités de placements, avec les entrepreneurs locaux, à la recherche de partenaires pour faire évoluer leur entreprise.

La plateforme devra permettre aux entrepreneurs locaux de présenter leur entreprise, leur niveau d'avancement, les évolutions prévues et un grand nombre d'autres informations pertinentes sur ladite entreprise.

Toutes ces données sur les entreprises après avoir été soigneusement identifiées, feront l'objet :

- D'une récolte (en cas d'absence de données) ;
- D'un traitement et de mise à disposition dans un format facilement exploitable ;
- D'un stockage sécurisé.

L'atteinte des objectifs visés par ce stage, nous obligera à concevoir et réaliser :

- Un modèle de Machine Learning (ML), qui permettra d'étudier un projet d'entreprise, à court terme prédire l'état de l'entreprise et à long terme lui attribuer un taux de viabilité ;
- Une plateforme web pour la mise en relation des entrepreneurs locaux et de potentiels investisseurs. Cette plateforme web devra intégrer le modèle de ML construit.

Le système sera ainsi en mesure d'analyser un projet entreprise, comparer divers projets et sélectionner ceux qui auront rempli des critères préalablement fournis.

I.2.4 Libellé du sujet

iPi9 Nexus : outil de Machine Learning pour l'évaluation d'Entreprise.

Conclusion

Ce chapitre nous a permis de présenter la structure au sein de laquelle nous avons effectué notre stage et le travail qui nous a été confié tout en analysant en profondeur son contexte. Dans le chapitre qui suit, nous présenterons les concepts clés indispensables à la compréhension du sujet.

CHAPITRE II : CONCEPTS LIÉS A L'EVALUATION DE PROJET D'ENTREPRISE

Introduction

Dans ce chapitre, il sera principalement question pour nous de passer en revue les notions clés nécessaires à la compréhension du processus d'évaluation d'une opportunité d'affaire. Nous aborderons dans cet ordre, les notions d'opportunité d'affaire ; la création d'entreprise ; les KPI ; le business case.

II.1 Principales notions

II.1.1 Qu'est-ce qu'une opportunité d'affaire ?

II.1.1.1 Définition

Dans notre contexte, elle peut être vue comme une entreprise ou organisation ou encore une unité institutionnelle, mue par un projet décliné en stratégies, en politiques et en plans d'action. Son but est de produire et de fournir des biens et/ou des services à destination d'un ensemble de clients ou d'usagers. Si elle est acquise par un investisseur, l'opportunité d'affaire lui permettra de réaliser des bénéfices à court, moyen ou long terme.

II.1.1.2 La création d'entreprise

La réussite de la création d'une entreprise exige une démarche rigoureuse et consciente, elle suppose de la part du futur entrepreneur des qualités et des capacités entrepreneuriales pour conduire de manière progressive et efficiente les différentes étapes du processus de création qui sont :

- Chercher une bonne idée de création d'entreprise,
- Vérifier si cette idée constitue réellement une opportunité exploitable,
- Transformer cette opportunité en une entreprise viable qui, dès son démarrage maximise son potentiel de rentabilité et de croissance.

Aucun promoteur aussi doué soit-il, ne peut espérer réussir s'il ne propose pas un produit ou un service capable de répondre à un besoin insatisfait ou mal satisfait aussi, s'il ne peut vendre ce produit ou ce service à un prix acceptable pour ses clients, et rentable pour sa future entreprise.

Afin de s'assurer une vue globale sur le long terme et avoir un moyen de pilotage, de suivi et de contrôle de la réalisation de son projet, l'entrepreneur doit mettre sur pied un plan d'affaire : business plan. Il couvrira les aspects les plus importants de l'étude du projet, ainsi que ses différentes phases de réalisation. Le business plan se décline ainsi qu'il suit :

- La description de l'activité, du produit ou service, des marchés visés, des facteurs-clés de réussite ;
- L'étude de marché et marketing ;
- La structure juridique et le statut de l'entreprise ;
- La technologie, la production et le fonctionnement ;
- La direction de l'entreprise et la propriété ;
- Le personnel ;
- Les investisseurs ;
- Les fonds (description et leurs provenances) ;
- Le bilan des risques et opportunités ;
- Les projections financières triennales de chiffre d'affaires et de résultats ;
- Et une conclusion qui projette l'entreprise à moyens terme en soulignant les principaux besoins de financement, les opportunités de sortie pour les futurs investisseurs.

II.1.2 Qu'est-ce qu'un KPI ?

II.1.2.1 Définition

Un indicateur clé de performance (ICP) ou KPI en anglais, est une mesure ou un ensemble de mesures braquées sur un aspect critique de la performance globale de l'organisation.

Les indicateurs de performance, qu'ils soient qualitatifs ou quantitatifs, sont multiples et variés.

II.1.2.2 Les indicateurs de performance financier

Permettent d'analyser la santé réelle d'une entreprise.

On peut citer.

- Chiffre d'affaires annuelle ;
- Chiffre d'affaires mensuel ;
- Dépenses mensuelles ;

- Trésorerie disponible (somme nette en compte en banque) ;
- Le taux de rentabilité (ratio permettant de mesurer l'impact des capitaux investis sur les profits générés) ;
- Le taux de rendement des actifs (ratio entre le résultat net et l'actif mobilisé dans l'activité) ;
- Les besoins en fonds de roulement (BFR) : montant nécessaire pour couvrir toutes les dépenses et décaissements liés à l'activité ;
- Le cycle d'exploitation des produits (temps nécessaire à l'entreprise pour transformer des marchandises achetées et l'écoulement sur le marché. En d'autres termes, il s'agit de la période entre l'achat des matières premières et le paiement de la facture par le client.) ;
- Les délais de règlement clients et fournisseurs.

II.1.2.3 Les indicateurs de performance commerciaux

Permettent de mesurer les actions commerciales réalisées par l'entreprise.

- Taux de transformation d'opportunités, ou taux de conversion prospect/client : mesure le nombre de ventes par rapport au nombre de rendez-vous prospects ;
- Taux de fidélisation : concerne les clients qui ont acheté vos produits ou services au moins deux fois ;
- Taux d'efficacité : le nombre de rendez-vous pris ou de ventes conclues par le nombre d'appels ;
- Nombre de ventes conclues ;
- Nombre de ventes en cours ;
- Nombre de ventes perdues ;
- Nombre de nouveaux clients ;
- Valeur des ventes ;
- Panier moyen : le chiffre d'affaires rapporté au nombre de ventes ;
- Nombre de clients payants (permet d'évaluer l'action commerciale) ;
- Connaissance de la concurrence ;
- Connaissance du parcours client.

II.1.2.4 Les indicateurs de performance organisationnelle

Ils mesurent l'impact organisationnel global de l'entreprise.

- Taux d'absentéisme du personnel ;
- Coût d'intégration de nouveaux employés ;
- Niveau de performance ;
- Taux de roulement : c'est le ratio du nombre d'employés qui ont quitté volontairement ou non l'entreprise sur l'effectif total d'employés ;
- Taux d'accident.

II.1.2.5 Les indicateurs de performance de productivité

- Les coûts de production ;
- Capacité de production ;
- Nombre de produits ;
- Nombre de sites ;
- Revenu par produit/site ;
- Nombre de concurrents.

II.1.3 Qu'est-ce qu'une étude d'opportunité ou business case ?

Dans la pratique, l'évaluation d'une opportunité d'affaire est soutenue par une proposition structurée : le business case. Cette dernière qui se trouve être une étude de faisabilité économique. Si elle est concluante, est destinée à rassurer de la viabilité d'un investissement. Si dans la plupart des cas il servira de base pour autoriser un projet et évaluer l'atteinte des objectifs, il permet surtout de valider le bien-fondé du projet.

Une étude d'opportunité doit contenir les étapes suivantes :

Étape 1 : Le besoin de l'organisation

Le business case est souvent précédé d'une évaluation des besoins, dont le but est de comprendre les objectifs, les points à traiter et les opportunités à saisir. Cette rubrique doit porter au minimum sur les points suivants :

- Déterminer les raisons qui poussent à l'action ;

Par exemple une nouvelle contrainte réglementaire, une évolution des comportements de consommation des clients, une initiative de la concurrence et le besoin de ne pas se laisser

dépasser par cette concurrence, un changement dans le climat économique, la mise en œuvre d'une nouvelle orientation stratégique de votre société etc.

- Documenter le problème à traiter ou l'opportunité à saisir ;
- Identifier les parties prenantes ;

Quels sont les services et les équipes à impliquer dans ce projet. Il est important de préciser s'il faut faire appel à des ressources externes et si oui lesquelles, précisez pour les ressources internes à quels services de l'entreprise elles appartiennent.

- Identifier le périmètre.

Ici on décrit précisément le périmètre de l'étude d'opportunité en termes de marché, de structures, d'équipes impliquées, des processus impliqués et des fonctions touchées.

Étape 2 : Analyse de la situation

Cette analyse globale de la situation doit porter, mais sans s'y limiter, aux axes suivants :

- Identifier les stratégies et les objectifs de l'organisation ;
- Identifier la ou les causes originales du problème ou les principaux aspects de l'opportunité ;
- Évaluer la capacité de l'organisation à supporter le projet ;
- Identifier les risques (aspects organisationnels, fonctionnels, techniques, humains) et à évaluer comment les contourner ou les mitiger ;
- Identifier les facteurs clés de succès ;
- Identifier les critères d'évaluation utilisés pour analyser les différentes options et scénarios.

Exemples de catégories de critères utilisées pour analyser une situation :

- **Requis** : Ce critère doit être obligatoirement satisfait pour traiter le problème ou l'opportunité ;
- **Souhaité**. Il est souhaité de remplir ce critère pour traiter le problème ou l'opportunité ;
- **Facultatifs**. Ce critère n'est pas essentiel. Lorsqu'il est rempli, il devient un élément qui permet de faire la différence entre les autres plans d'action.

Étape 3 : État futur

L'état désiré par l'organisation décrit l'état qui devrait résulter si le besoin est satisfait à la fin du projet. Il doit inclure des résultats mesurables qui peuvent être utilisés pour évaluer la pertinence de l'étude d'opportunité et ou de la solution proposée.

A cette étape il faut s'assurer que les objectifs du projet sont pilotés par des indicateurs SMART (Spécifique, Mesurable, Ambitieux, Réaliste, Temporel).

Étape 4 : Options et scénarios

Identification d'un ensemble d'options ou Scénarios à prendre en compte pour traiter le problème ou l'opportunité. Les options représentent tous les choix qui s'offrent à l'organisation pour gérer la situation.

À titre d'exemple, une note d'opportunité peut présenter les trois options suivantes :

- **Option 1 : Statu quo** ou accepter de maintenir la situation courante, en choisissant cette option, le projet n'est pas autorisé ;
- **Option 2 :** Faire le minimum pour traiter le problème ou l'opportunité ;
- **Option 3 :** Faire plus que le minimum pour traiter le problème ou l'opportunité. Cette option répond à l'ensemble minimal des critères ainsi qu'à d'autres critères documentés.

Pour chaque scénario, on décrit dans les grandes lignes les ressources nécessaires (humaines, techniques et financières) et les coûts.

Pour chaque scénario, on précise comment les objectifs du projet peuvent être atteints en termes d'organisation, de ressources techniques et humaines.

Étape 5 : Recommandation

Énoncer l'option recommandée par l'auteur du business case pour autoriser et suivre le projet.

Cet énoncé doit inclure au minimum les points suivants :

- Résultats d'analyse de l'option recommandée ;
- Contraintes, hypothèses et risques liées à cette option ;
- Indicateurs de mesures de réussite ;
- Plan de mise en œuvre à haut niveau (délais, acteurs, rôles les responsabilités).

Étape 6 : Évaluation

Énoncer le plan de mesure des bénéfices offerts par le projet. Il devrait inclure tous les aspects opérationnels et indicateurs de mesure de la réussite de l'option recommandée après sa mise en œuvre.

Exemple d'indicateurs de mesure des bénéfices :

- Réduction des dépenses (personnel, fournitures, ...);
- Augmentation de la part de marché ;
- Augmentation du périmètre fonctionnel de compétences d'un service ;
- Image de marque plus respectueuse de l'environnement ;
- Image de marque plus innovante ;
- Meilleur taux de satisfaction des clients ;
- Réduction du nombre de tâches opérationnelles pour un même processus ;

Etc.

Proposer des indicateurs concrets pour mesurer le retour sur investissement du projet :

- Évolution de la part de marché ou part de marché si nouveau produit ou service ;
- Chiffre d'affaires ;
- Montant des frais généraux de la société ;
- Évolution du trafic sur le site internet ;
- Évolution du nombre d'inscrits sur le site ;
- Évolution du nombre de clients ;
- Nombre d'utilisateurs pour une application interne

Etc.

Conclusion

Après cette immersion dans les généralités de l'évaluation de projets d'entreprise, nous avons pu cerner d'avantage notre sujet.

Dans le chapitre suivant, nous allons faire un tour d'horizon rapide des solutions existantes et dégager leurs éventuelles limites, et finir par un listing des raisons qui nous conduisent à utiliser le Machine Learning comme outil dans notre solution.

CHAPITRE III : AUDIT & LIMITES DES SOLUTIONS EXISTANTES

Introduction

Dans ce chapitre, nous allons procéder séquentiellement à l'étude de quelques solutions existantes, par la suite lever leurs limites. A la fin, nous allons dans une certaine mesure dégager quelques pistes de raisons nous conduisant à utiliser le Machine Learning pour automatiser la tâche d'évaluation d'Entreprise.

III.1 Solutions existantes

Avant de rentrer de plein fouet dans le développement de notre solution, il est nécessaire de faire un tour d'horizon des solutions existantes en la matière pour en dégager les qualités et particulièrement les limites. Cela nous permettra de dégager des approches nouvelles d'automatisation.

III.1.1 Créez Mon Business Plan

The screenshot shows the homepage of 'Créez Mon Business Plan'. At the top, there's a navigation bar with 'CRÉER MON BUSINESS PLAN' (with a green house icon), 'BUSINESS PLAN', 'EVALUATION DE PROJET' (highlighted in green), 'ACTUALITÉS', and 'CONNECTEZ-VOUS'. Below the navigation is a large image of hands typing on a laptop keyboard. Overlaid on this image is a white box containing the text 'ÉVALUATION DE PROJET' and a paragraph explaining the software's purpose: 'Ce logiciel d'évaluation est gratuit et ne nécessite pas d'installation. Il vous permet de mesurer le potentiel de réussite de votre projet. Il vous suffit de répondre aux questions pour obtenir en quelques minutes les forces et les faiblesses mais aussi les pistes d'amélioration qui vous permettront de sécuriser votre projet.' At the bottom of this box is a green button labeled 'DÉBUTER LE TEST'.

Déterminez le potentiel de réussite de votre projet

Débutez la réalisation de votre Business Plan en évaluant vos forces et vos faiblesses

This screenshot shows a section of the website encouraging users to test the service for free. It features a green box with the text 'TESTEZ GRATUITEMENT VOTRE PROJET DE CRÉATION' and '55 612 ÉVALUATIONS réalisées ces 12 derniers mois'. To the right, there are two columns: one for 'Facile' (with a checkmark icon) and one for 'Rapide' (with a rocket icon). Both columns contain text describing the user experience: 'Vous êtes guidé tout au long d'un formulaire abordant les différents aspects de votre projet : le créateur, le projet, le produit ou service, les marchés cibles, etc.' and 'Quelques minutes vous suffiront pour répondre à cette série de questions. Nous avons conçu le questionnaire pour aller droit au but'.

Figure 7 Page d'accueil Créez Mon Business Plan

Créez Mon Business Plan est une solution simple permettant de créer un business plan gratuitement en ligne. Mais le lien du site <https://www.creer-mon-business-plan.fr/evaluation-de-projet/> propose également un outil pour évaluer un projet de création d'entreprise. Il se charge d'analyser la pertinence de l'idée à travers une série de questions

portant sur les thématiques suivantes : le créateur d'entreprise ; la localisation ; le projet ; le produit ou service ; le marché et ses acteurs.

III.1.2 Moo.com

Comment évaluer votre idée commerciale

Si vous envisagez de créer votre propre entreprise, vous devrez tout d'abord examiner attentivement votre idée. S'agit-il d'une idée géniale ou votre projet risque-t-il d'échouer dès le début ? L'équipe MOO.COM a créé ce tableau interactif pour vous aider à vous poser les bonnes questions !

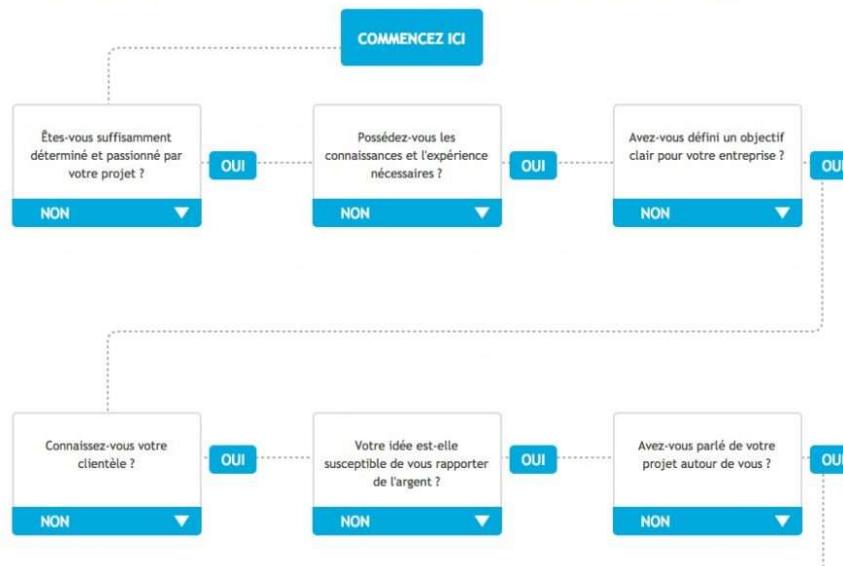


Figure 8 Moo.com

Sur Moo.com : <https://www.moo.com/fr/startup-business-toolkit/flowchart/>, on y retrouve un tableau interactif très pratique pour évaluer une idée de projet.

III.1.3 Les tests « Étes-vous prêt à entreprendre ? »

ÊTES-VOUS PRÊT À ENTREPRENDRE ?

Avez-vous le profil d'un entrepreneur ? Des tests sont à votre disposition pour le savoir.

MOTS-CLÉS : # COMPÉTENCES # TEST # ENTREPRENEUR # PROFIL



Découvrez vos traits de caractère et vos ressources

Deux tests sont disponibles en ligne. Evaluatez vos traits de caractère et vos ressources.

Avez-vous les traits de caractères d'un entrepreneur ?

Avez-vous les ressources internes pour entreprendre ?

Figure 9 Étes-vous prêt à entreprendre ?

La Chambre de Commerce et de l'Industrie de France a mis en place trois tests sur le site : <https://www.cci.fr/web/creation-d-entreprise/etes-vous-un-entrepreneur> à l'attention des entrepreneurs en herbe :

- Le premier s'intéresse aux déterminations de l'entrepreneur ;
- Le deuxième sur les traits de caractères ;
- Et le dernier vise à évaluer vos aptitudes et vos ressources internes.

Car pour eux, les propres qualités de l'entrepreneur, traits de caractères et motivations sont au moins aussi importants que le projet d'entreprise en lui-même.

III.2 Limites des solutions existantes

Les solutions ci-hauts listées existantes pour une évaluation d'une idée d'entreprise et spécialement dédiées au porteur de projet (entrepreneur) et non à l'investisseur, ont pour point commun une logique d'évaluation axée sur 02 points à savoir :

- ✓ Les points forts et faibles du porteur de projet (l'entrepreneur) ;
- ✓ L'idée projet.

S'il est vrai que l'entrepreneur au regard de ses point forts, est un élément déterminant dans la réussite d'un projet, et qu'aucun entrepreneur, aussi doué soit-il, ne peut espérer réussir s'il ne propose pas un produit ou un service capable de répondre à un besoin insatisfait ou mal satisfait. Mais pour un investisseur ou un analyste financier chargé d'étudier la viabilité d'une idée entreprise, il existe des paramètres quantifiables et mesurables aussi importants à prendre en compte que les points forts du porteur de projet et l'idée projet tels : les indicateurs de performances financiers, commerciaux, organisationnelles... qui ne sont pas exploités dans les solutions existantes présentées.

Cela vient justifier notre postulat présenté plus haut selon lequel : l'évaluation proprement dite de l'entreprise est toujours jusqu'à présent une tâche manuelle.

A ce jour, avec les techniques de programmation nouvelles, l'automatisation d'une telle tâche peut être faite soit par :

- 1) La conception via les paradigmes de programmation traditionnelle d'une application où il va falloir prendre en considération chaque paramètre, ce qui obligera à gérer un grand nombre de cas. Cette approche complexe tant dans le codage, que dans la lourdeur de l'application est non seulement difficilement implémentable, mais en plus elle ne présente aucune garantie d'aboutir à des prédictions (résultats) fiables.

- 2) Une seconde approche plus prometteuse de nos jours, qui s'inscrit à l'intersection de l'expertise métier, l'expertise mathématique et l'expertise informatique est la Data Science et plus précisément le Machine Learning.

Rappelons que le Machine Learning peut être utilisé dans la résolution des problèmes :

- ✓ Que l'on ne sait pas résoudre ;
- ✓ Que l'on sait résoudre, mais dont on ne sait pas formaliser en termes d'algorithmes ;
- ✓ Que l'on sait résoudre, mais avec des procédures trop gourmandes en ressources informatiques.

En bref, le ML est utilisé quand les données sont abondantes (relativement), mais les connaissances peu accessibles ou peu développées.

Ainsi, il sera question de transmettre à l'ordinateur, via l'apprentissage d'un algorithme de Machine Learning, la capacité humaine à évaluer une Entreprise donnée en se basant sur un grand jeu de données. A ce jour, le Machine Learning continue de faire ses preuves dans la résolution de plusieurs problématiques complexes, il est non seulement moderne mais surtout offre des garanties certaines tant sur la rapidité d'exécution de l'application, que sur la fiabilité d'obtention de résultats. Car en observant un grand volume de données, l'ordinateur pourra dégager la logique sous-jacente et ainsi sera en mesure de réaliser une prédiction lorsqu'on lui donnera une nouvelle donnée (ne faisant pas partie du grand volume de données de départ).

Logiquement, après tout ce qui précède, on utilisera la deuxième approche pour essayer de solutionner notre problématique.

Conclusion

Par ce chapitre dédié à l'audit des solutions existantes, nous mettons un terme à la première partie de notre travail. La deuxième partie portera sur l'état de l'art du Machine Learning outil retenu pour résoudre notre problématique. Nous finirons cette section avec à la mise en place de notre modèle prédictif via l'environnement de programmation Jupyter.

DEUXIEME PARTIE : MACHINE LEARNING & SES ALGORITHMES

Dans cette partie, nous allons mettre en avant le Machine Learning qui comme le disait Arthur Samuel est le champ d'étude qui donne aux ordinateurs la capacité d'apprendre sans être explicitement programmés. Ici, nous allons réaliser une analyse, conception et construction d'un modèle prédictif, sur lequel on s'appuiera pour l'évaluation d'opportunité d'affaire. Nous présenterons toutes les étapes nécessaires à la mise en place d'un modèle avec des taux de prédiction efficaces.

Chapitres

CHAPITRE IV : PRESENTATION DU MACHINE LEARNING
CHAPITRE V : Algorithme K-Nearest Neighbours.....

CHAPITRE IV : PRESENTATION DU MACHINE LEARNING

Introduction

L'apprentissage automatique (ou Machine Learning), branche de l'intelligence artificielle, connaît une forte expansion ces dernières décennies grâce au Big Data (Gandomi and Haider, 2015). Car notons-le en 2017, une étude montre que 90% des données existantes ont été générées sur les 02 dernières années, comme illustrations de nos jours :

- Chaque seconde environ 5 900 tweets sont expédiés sur le site de micro-blogging Twitter. Cela représente 504 millions de tweets par jour ou 184 milliards par an (<https://www.planetoscope.com>).
- Le nombre de messages quotidiens échangés sur WhatsApp est de 60 milliards dont 760 millions de vidéos, 3,3 milliards de photos et 80 millions de GIFs. Cela représente 21 900 milliards de messages WhatsApp par année ou l'équivalent de 694 000 messages par seconde.
- 350 millions de photos sont envoyés sur le réseau social le plus célèbre au monde Facebook chaque jour ...

Cette masse d'information vient alimenter le flot d'informations ("big data") publiée par l'humanité chaque jour sur internet. Appliquée dans des domaines variés comme le marketing digital, la biologie, l'économie ou l'industrie automobile, l'analyse prédictive de cette masse d'informations représente un enjeu stratégique considérable.

IV.1 Histoire du Machine Learning

C'est dans les années 80 que le Machine Learning est né, il sera dans ces années fortement liés à l'analyse de données et aux algorithmes de décision. Mais la notion d'apprentissage automatique elle, remonte à bien plus avant, autour des années 1667, où elle fera d'abord référence à la compréhension de la pensée humaine, étudiée par Descartes puis par Leibniz dans son ouvrage « *De Arte combinatoria* ». Le philosophe tente dans cet ouvrage de définir les raisonnements les plus simples de la pensée à l'aide d'un alphabet qui permettront, une fois combinés, de formuler des pensées très complexes. C'est en 1854 que ces travaux seront formalisés par G. Boole dans son ouvrage « *An investigation of the laws of thought, on which are founded the mathematical of logic and Probability* ». Il faudra attendre 1956 avec la conférence de Dartmouth (USA), organisée par M. Minsky et J. McCarthy, pour que

l’Intelligence Artificielle connaisse un tournant majeur. Pour beaucoup, cette conférence est considérée comme l’acte de naissance de l’IA où un principe très proche de celui évoqué par Leibniz est formulé : « *chaque aspect de l'apprentissage ou toute autre caractéristique de l'intelligence peut être si précisément décrit qu'une machine peut être conçue pour le simuler* ». A cette date, la puissance de calcul des ordinateurs n’étant pas encore assez développée, c’est dans les années 80 que l’IA se développe avec les lois de Gordon E. Moore qui prévoient l’évolution chaque 18 mois de la puissance de calcul des ordinateurs.

John Joseph Hopfield développera en 1982 un nouveau type de réseau neuronal baptisé réseau de Hopfield. Ce dernier sera largement utilisé et repris. Comme date majeure dans la naissance du ML, on peut citer 1984 avec la création des arbres à décisions par Leo Breiman, Jerome Friedman, Richard Olshen et Edward Carroll Stone, plus tard en 2001, Breiman et A. Culter vont étendre ce concept avec les forêts aléatoires plus connues en ML sous le nom de random forests.

Dans l’Histoire du ML, en termes d’applications marquantes on peut citer :

- En 1997, l’ordinateur IBM Deep Blue, qui va battre pour la toute première fois le champion du monde du Jeu d’échec Garry Kasparov ;
- En 2014, c’est un chatbot russe « Eugene Goostman » qui va devenir le tout premier « robot » à passer le fameux test de Turing.

IV.2 Evolution du Machine Learning

Le Machine Learning est un domaine issu de nombreuses disciplines comme les statistiques, l’optimisation, l’algorithme ou le traitement du signal. Domaine en pleine mutation, il occupe de plus en plus notre société via les solutions et applications que nous manipulons au quotidien. Notons qu’il est utilisé depuis plusieurs années dans : la reconnaissance automatique de caractères ; les filtres anti-spam ; la protection de fraudes bancaires ; recommandation de livres, films ou autres produits sans oublier l’identification des visages dans les viseurs de nos caméras.

Les années à venir offriront encore de beaux jours au ML, car de plus en plus l’on observe une explosion de données, l’augmentation fulgurante des puissances de calcul de nos ordinateurs et l’optimisation des techniques mathématiques. Facteurs qui permettront au ML d’améliorer les secteurs tels : la sécurité routière (aussi avec les voitures autonomes) ; les réponses d’urgences aux catastrophes naturelles, le développement de nouveaux médicaments.

L'apprentissage est une question qui continue de fasciner les informaticiens et mathématiciens, tout comme les neurologues, pédagogues, philosophes ou artistes. Une définition de l'apprentissage s'appliquant au programme informatique tel le robot et aussi à un animal et voir même l'être humain est celle de Fabien Benureau (2015) « *l'apprentissage est la modification du comportement sur la base d'une expérience* ».

Le ML est devenu l'outil qu'utilise la **Data Science**, qui est ce champ d'étude s'inscrivant à l'intersection de l'expertise métier, expertise mathématique et statistiques, l'expertise informatique particulièrement programmation, comme l'illustre la figure ci-dessous, dans l'optique d'extraire de la connaissance à partir des données.

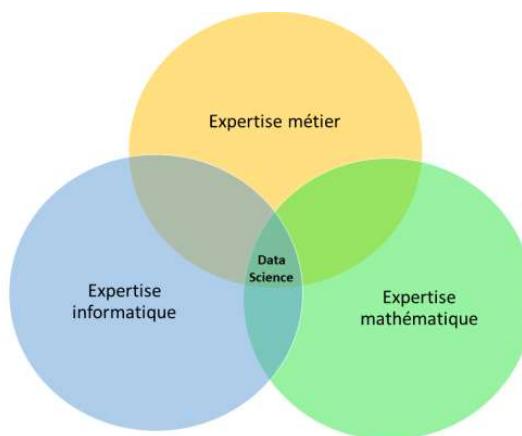


Figure 10 Point de rencontre Expertise métier, mathématique & informatique

La Data Science nous permet de combiner les notions clés du domaine d'étude, avec celles de la précision des formules mathématiques tout en s'arrogeant la puissance de calcul de plus en plus grandissante des ordinateurs.

C'est donc fort de cela que le machine Learning outil de la Data Science, traite des sujets complexes où la programmation traditionnelle trouve ses limites. En réalité, construire un programme qui conduit une voiture serait très complexe, voire même impossible en utilisant la programmation classique. Cela étant principalement dû aux infinis cas possibles à traiter. Avec le Machine Learning, au lieu de spécifier des instructions pour dire à l'ordinateur quoi faire à chaque cas, l'ordinateur apprendra par lui-même comment conduire simplement en se basant sur des expérimentations.

En fonction des données d'expérimentation que prendra l'algorithme d'apprentissage en entrée, il déduira par lui-même une hypothèse de fonctionnement (modèle) qui permettra d'expliquer la cause du phénomène étudié. Par la suite, il utilisera cette dernière pour de nouveaux cas, et affinera son expérience au fil du temps.

IV.3 Types de problèmes en Machine Learning

Le Machine Learning est un domaine assez vaste, nous allons dresser dans cette section une liste des plus grandes classes de problématiques auxquels il peut être appliqué. On distingue 03 types de problématiques en Machine Learning comme l'illustre la figure ci-dessous. Mais pour les besoins de notre rapport, nous allons nous attarder sur les 02 types les plus utilisés.

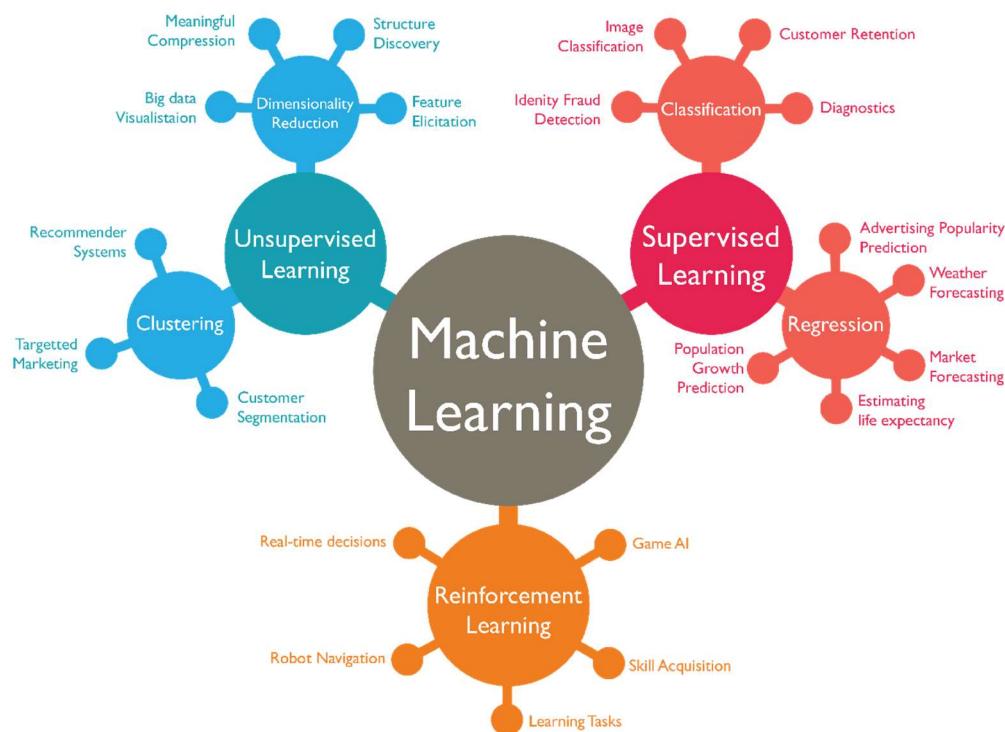


Figure 11 Types de problèmes de ML

IV.3.1 Supervised Learning (Apprentissage Supervisé)

C'est le type le plus récurrent. Il s'agit de fournir aux algorithmes d'apprentissage un jeu de données d'apprentissage (**Training Set**), pouvant être représenté sous forme d'un couple (**X**, **Y**) avec :

- X qui seront les variables prédictives (caractéristiques ou **features** en anglais) ;
- Et Y le résultat de l'observation.

En se basant sur le Training Set donné, l'algorithme va trouver une fonction mathématique qui permet de transformer dans le meilleur des cas, X vers Y. En d'autres termes, on peut dire que l'algorithme va trouver une fonction mathématique F tel que : **F(X) \approx Y**.

L'apprentissage supervisé est le concept derrière plusieurs applications sympas de nos jours :

- Reconnaissance faciale de nos photos par les smartphones,

- Filtres anti-spam des emails ...

Les caractéristiques X peuvent être des valeurs numériques, alphanumériques, des images... et quant à la **variable prédictive Y**, elle peut être de deux catégories :

- **Variable discrète** : Dans ce cas, la variable à prédire peut prendre une valeur d'un ensemble fini de valeurs définies dans une **classe**.

Par exemple, pour prédire si un mail est SPAM ou non, la variable Y peut prendre deux valeurs possibles : $Y \in \{\text{SPAM}, \text{NON SPAM}\}$

- **Variable continue** : La variable Y peut prendre n'importe quelle valeur. Pour illustrer cette notion, avec un algorithme qui prend en entrée des caractéristiques d'un véhicule, et tentera de prédire le prix du véhicule (la variable Y).

Notons que la catégorie de la variable prédictive Y, fait décliner l'apprentissage supervisé en deux sous catégories :

- ✓ **La classification** (lorsque la variable Y est une valeur discrète) ;
- ✓ **La régression** (lorsque la variable Y est une valeur continue).

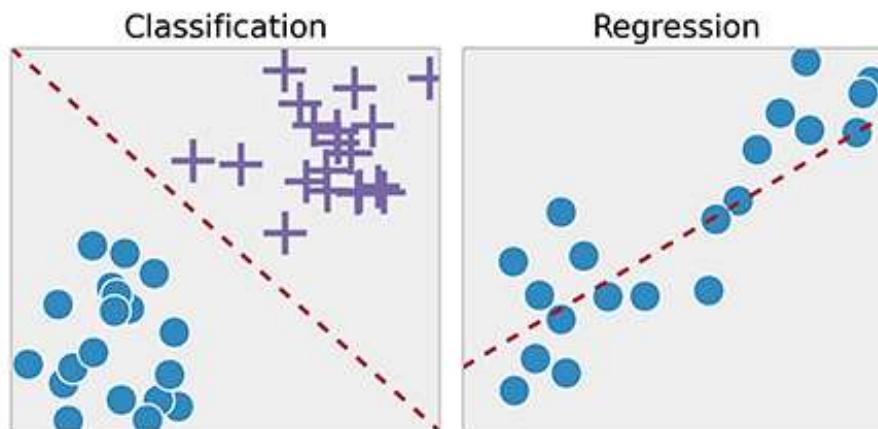


Figure 12 Illustration de la différence entre classification linéaire et régression linéaire

IV.3.2 Unsupervised Learning (Apprentissage non supervisé)

Dans ce type, on va donner à l'algorithme d'apprentissage des données éventuellement non structurées. Nous entendons par là sans label i-e pas de variable à prédire Y. Et on laisse l'algorithme trouver une sorte de structure ou patterns dans les données. Cela peut être des regroupements de données (Clustering).

Un exemple :

En disposant des données d'achats des internautes dans un site e-commerce, l'algorithme de clustering va trouver les produits qui se vendent le mieux ensemble.

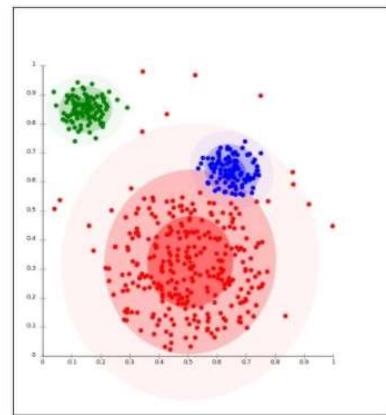


Figure 13 Résultat de clustering

IV.3.3 LES ALGORITHMES DE ML

Pour faire simple, le Machine Learning peut être vu comme un ensemble de méthodes algorithmiques permettant, entre autres, de produire des modèles prédictifs comme nous l'avons souligné plus haut, ou détecter des patterns dans les données.

La réalisation d'un modèle (vu ici comme une représentation de la réalité), prédictif en Machine Learning est rendu possible grâce à l'existence de nombreux algorithmes qui, pour la plupart connaissent déjà une implémentation dans des librairies. Ces algorithmes pour être utilisables vont se voir passer un jeu de données qui au préalable devrons être traitées afin d'être entraînés et produire un modèle qui sera capable de produire des résultats.

Les 02 illustrations suivantes présentent de manière organisée quelques-uns de ces algorithmes suivant le problème.

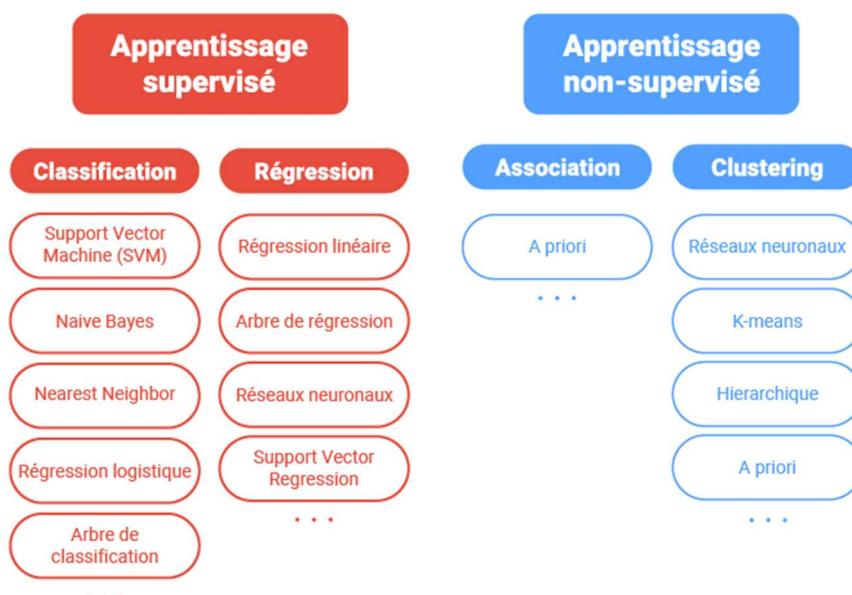
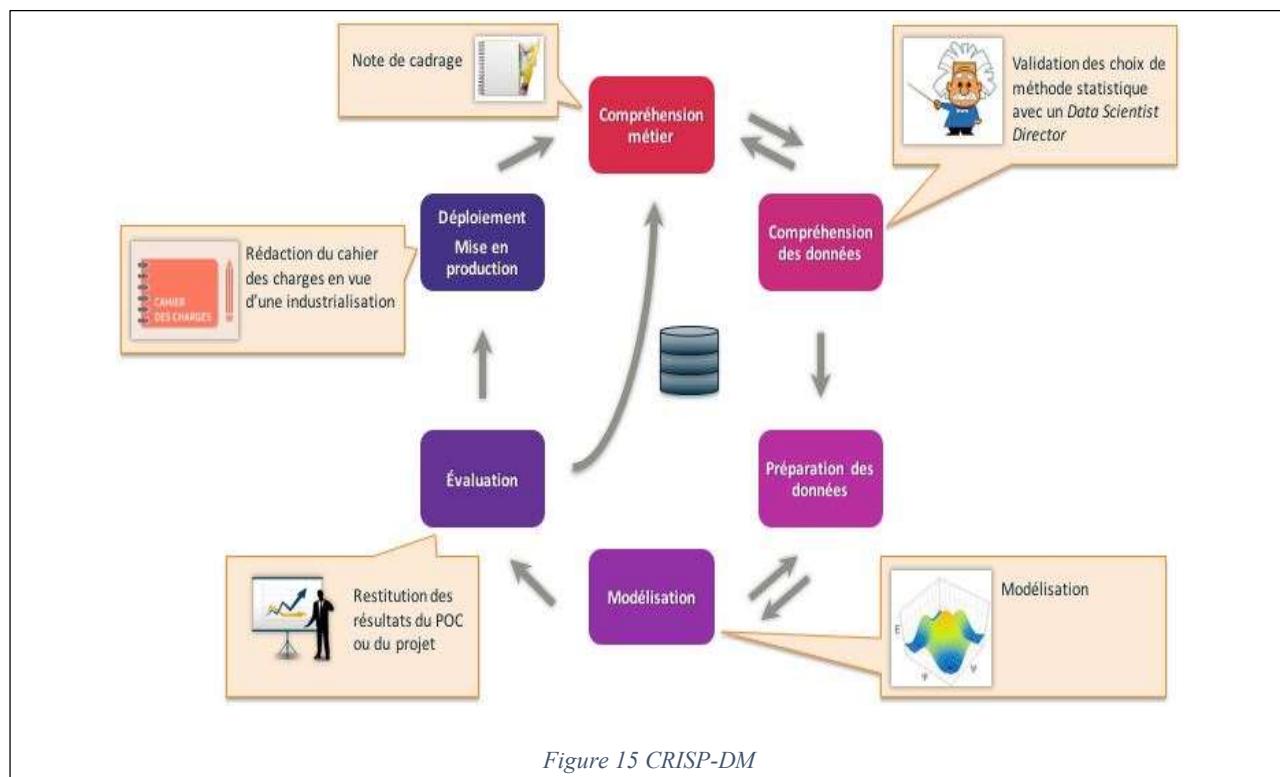


Figure 14 Les algorithmes de Machine Learning

IV.4 Aborder un problème de Machine Learning

Parce qu'un problème de Machine Learning, de par sa définition est souvent complexe à résoudre, avoir une méthodologie rigoureuse garantie qu'aucun aspect du problème ne sera oublié tout au fil de sa résolution, car la problématique complexe de départ se verra découper en étapes plus petites devant ainsi nous faciliter la compréhension et donc la résolution.

La méthode **CRISP** initialement connue comme **CRISP-DM** développée par IBM dans les années 60 pour réaliser les projets Datamining, reste aujourd'hui la seule méthode utilisable efficacement pour tous les projets Data Science et donc par ricochet en Machine Learning. Cette dernière se décompose en 06 étapes comme le montre la figure ci-dessous :



Nous pouvons résumer la démarche de la méthode CRISP-DM en 05 grandes articulations à savoir :

- 1) Définition du problème ;
- 2) Préparation des données ;
- 3) Choix des bons algorithmes ;
- 4) Optimisation des résultats ;
- 5) Mise en production.

Dans la suite de notre travail, quelques-uns de ces points seront présentés tout en prenant comme cas pratique la résolution de notre problème et les autres points seront abordées dans le chapitre suivant.

IV.4.1 Définition du problème

Cette étape est nécessaire et primordiale lorsque le problème à résoudre n'est pas encore clair, tant dans les esprits que sur papier.

Pour clarifier le problème à résoudre, il est recommandé de commencer par une description du problème. On commence par une description informelle du problème dans le but de formuler en une phrase facilement compréhensible le problème.

Comme exemple pratique : *Nous voulons une plateforme capable de prédire la viabilité d'une entreprise donnée.*

Ayant alors notre description informelle, on a dès à présent des idées plus claires et donc on peut à présent formuler une description plus formelle du problème.

Pour le faire nous allons nous baser sur une définition de Tom Mitchell en 1997 qui définissait un programme de Machine Learning comme suit : « *A computer program is said to learn if its performance at a task T, as measured by a performance P, improves with experience E* »

Pour notre définition formelle, il va falloir définir le triptyque **T** (tâche), **P** (Performance) et **E** (Expérience).

- **Tâche (T)** : Prédire viabilité d'une entreprise.
- **Performance (P)** : La précision de la prédition faite par l'algorithme (à quel point elle est proche de la réalité).
- **Expérience (E)** : Les descriptions des entreprises suivies de leur car leur état viable ou non viable.

IV.4.2 Préparation des données

Sans données, il est difficile voir impossible de solver un problème de Machine Learning. C'est montrer à quel point les données sont au centre des travaux du data scientist qui voudra en extraire de la connaissance.

Dans cette étape, nous allons nous atteler à :

➤ **Recenser le type des données utiles pour la résolution du problème**

Dans la définition d'une Entreprise, nous avons identifié 02 parties à savoir les informations d'identification de l'Entreprise telles : le numéro de registre du commerce, la dénomination, le secteur d'activité ... et les kpi (mesures clés ou indicateurs de performance), pour l'évaluation de sa santé financière.

Nous prévoyons collecter auprès des entreprises ces informations. La première partie (identification de l'Entreprise), sera réservée pour la présentation de l'entreprise et la seconde (kpi), sera utilisée pour son évaluation proprement dite.

➤ **Définir les sources de données**

Le succès de notre projet dépend grandement de la **quantité et la qualité de données**. Pour les données devant constituer le dataset nous allons dans un premier temps soumettre des formulaires de questions à différents responsables et Chefs d'Entreprise ceci dans l'optique d'avoir des données reflétant la réalité. Les données récupérées depuis la base de données des entreprises enregistrées via la plateforme serviront à affiner notre modèle tout en respectant le principe de la séparation des données d'entraînement et de test.

➤ **Définir le degré de difficulté pour collecter ces données sur les entreprises**

Malheureusement pour nous, les données de l'Afrique et donc du Gabon dans ce secteur sont encore éparses et presque inexistantes. Chaque entreprise évolue en vase clos et l'organe en charge ne dispose pas à ce jour d'une base de données centralisant l'ensemble de ses informations.

Pour collecter ces données, faut l'avouer sensibles sur les entreprises, nous avons avec l'aide d'un Expert en Analyse et montage de projets, mis sur pied 02 questionnaires : un questionnaire qui sera soumis aux entrepreneurs locaux (**ceux ayant réussis et aussi ceux n'existant plus à ce jour**), et le deuxième lui sera soumis aux investisseurs. Nous prévoyons réaliser ses questionnaires en format physique (papiers) et en ligne (via un formulaire **google form**) cf. Annexes 1.

➤ **Définir les traitements préparatoires à réaliser sur ces données.**

Après la phase dite d'administration de formulaires, nous allons passer à la phase de **sécurisation, traitement et sauvegarde** de ses données en base de données (sur un

serveur), et dans un format .csv (Comma-separated values), qui facilitera leur exploitation dans les langages R ou python lors de la construction du modèle prédictif.

IV.4.3 Choix du bon algorithme

Gardons à l'esprit qu'à chaque type de problème de Machine Learning (classification, régression, clustering...), on peut avoir plusieurs algorithmes candidats pour le résoudre. Mais les facteurs entrant en jeu pour choisir un algorithme et non l'autre des algorithmes candidates peuvent être nombreux, notamment :

- Le type de la variable target (Y) (discrète ou continue) ;
- Le nombre de caractéristiques (features) ;
- La quantité de données que l'on dispose...

Le schéma suivant nous aiguillé dans le bon choix d'un algorithme pour résoudre notre problème.

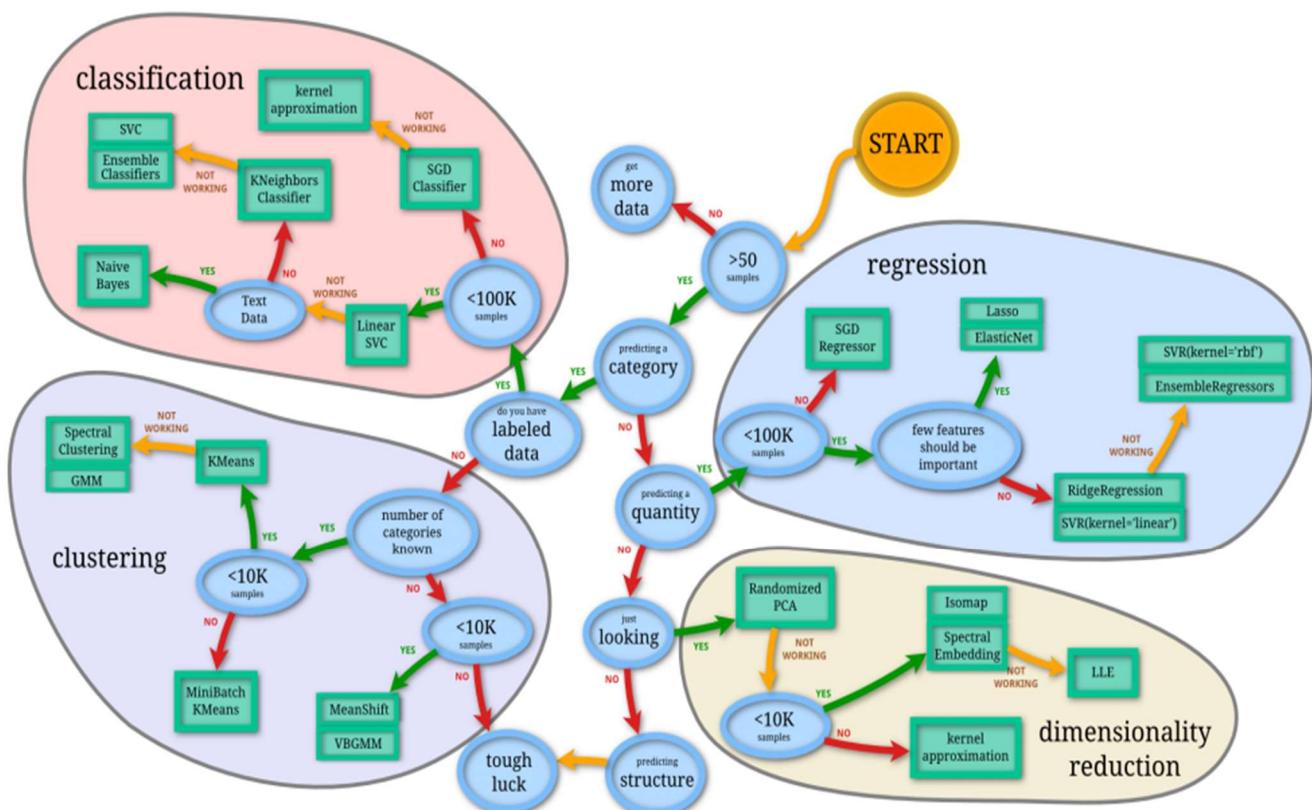


Figure 16 Process de choix d'une classe d'algorithmes

Ainsi, partant du principe que :

- ✓ Notre datasets de départ sera obligatoirement constitué de plus de **50** jeux de données sur les entreprises locaux (pour un début) entreprises florescentes ou ayant

fermées car n'ayant pas survécus à ce jour. Notons-le, en deçà de 50 exemples notre modèle prédictif ne sera pas acceptable du fait de la faiblesse du jeu de données.

Pour entraîner notre modèle nous prévoyons avoir dans notre dataset de départ au moins 50 observations sur les entreprises viables et au moins 50 sur ceux non viables.

- ✓ Nous prévoyons labéliser nos données avec l'aide de notre expert en analyse et montage de projet, durant la phase de traitement des données collectées, suivant la classe {Entreprise à fort potentiel, Entreprise à faible potentiel}. Cette phase est déterminante c'est durant celle-ci qui nous allons nous assurer qu'elles sont consistantes, sans valeurs aberrantes ni manquantes. La labélisation du jeu de données étant possible, nous prévoyons y procéder afin que l'apprentissage supervisé de notre algorithme puisse être possible.
- ✓ L'objectif pour le modèle que nous désirons construire sera de prédire la viabilité d'une opportunité d'affaire grâce au modèle sous-jacent construit.

Au vu de ce qui précède et dans une certaine mesure, il sera difficile pour nous de construire un datasets dans un premier temps avec plus de 100 000 exemples d'entreprises. Suivant le schéma présenté ci-dessus, nous dégagons logiquement comme algorithmes candidates : **le linear SVC et le KNeighnors Classifier**.

➤ **le linear SVC**

Linear SVC pour **Support Vector Classifier** en anglais et **Classification Vectorielle de Support** en français, est une méthode d'apprentissage supervisé. C'est un algorithme faisant partir des classes qu'offre **SVM (Support Vector Machines)**, qui peuvent être utilisés pour les problèmes de classification, régression et la détection des valeurs aberrantes. Les SVM sont très efficaces dans les espaces de grandes dimensions et sont généralement utilisés dans les problèmes de classification. Les SVM sont populaires et économies en mémoire car ils utilisent un sous-ensemble de points d'apprentissage dans la fonction de décision.

➤ **KNeighnors Classifier**

L'algorithme K-NN (K-Nearest Neighbors) est une méthode d'apprentissage supervisé. Il peut être implémenté aussi bien pour la résolution de problèmes de régression que pour la classification. Le mode de fonctionnement de K-NN peut être assimilé au principe suivant « *essaie de m'indiquer quels sont tes plus proches voisins, et je te prédirai qui tu es* ».

Nous allons travailler dans la suite avec le **KNeighbors Classifier**, car non seulement notre dataset de départ n'est pas encore assez fourni en données et de plus, plus tard nous prévoyons prédire avec une précision forte le pourcentage de viabilité en valeur numérique.

Le schéma ci-dessous résume à suffisance toutes les activités que nous avons pratiquées pour rendre le modèle utilisable dans le projet :

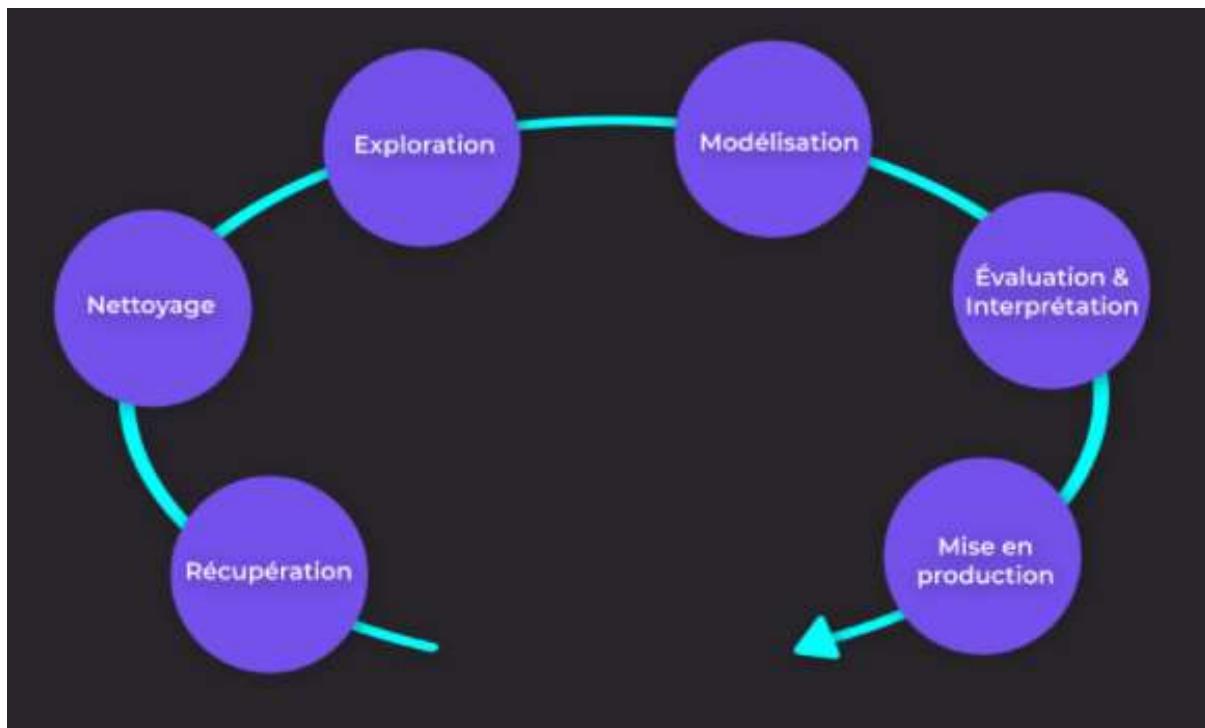


Figure 17 Ensemble d'activités de conception et production d'un modèle de ML

Conclusion

Ce chapitre nous a permis de présenter le Machine Learning, les grands types de problèmes pouvant être résolus par le ML, les algorithmes du ML suivant une hiérarchie bien établie et surtout les étapes de résolution d'un problème de ML et aussi les étapes de construction d'un modèle de ML. Dans le chapitre suivant nous présenterons l'algorithme K-NN (K Nearest Neighbors), retenu pour résoudre notre problème.

CHAPITRE V : Algorithme K-Nearest Neighbours

Introduction

Ce chapitre sera principalement axé sur la présentation de l'algorithme K-Nearest Neighbors (**K-NN**). Nous allons voir le mode fonctionnement de cet algorithme, et construire notre modèle prédictif.

V.1 Présentation et mode de fonctionnement de K-NN

K-NN pour **K-Nearest Neighbours** en anglais et pour **méthode des k plus proches voisins** en français. C'est un algorithme standard de classification, pouvant tout aussi être utilisé pour des problèmes de régression. Comme nous l'avons souligné plus haut, il fait partir de la classe des algorithmes dits supervisés.

Son mode de fonctionnement peut être décrit comme suit : En se basant sur tout le jeu de données étiquetées (d'où l'apprentissage supervisé), K-NN contrairement aux autres algorithmes de machine learning, en réalité ne produit pas à proprement parler un modèle à partir d'un Training set, mais chercher à estimer (prédirer) la classe (valeur) d'une nouvelle donnée en regardant quelle est la classe le plus proche possible des k données retenues comme voisines de la nouvelle donnée à évaluer. C'est ce qui explique son nom de K-Nearest Neighbours.

V.2 Principe détaillé du fonctionnement de K-NN.

Comme mentionné, l'algorithme K-NN utilise tout le jeu de données de départ. Pour une observation nouvelle, ne faisant pas parti du dataset, dont il est question de réaliser une prédiction, K-NN cherche les **K** instances (exemples) du dataset les plus proches (voisins) de notre observation. Par la suite, pour les **K** voisins retenus, l'algorithme évaluera chacune leur variable de sortie (output ou encore target variable), y pour essayer de calculer la valeur de la variable target y de l'observation qu'on souhaite prédire.

Notons généralement que :

- Si K-NN est utilisé pour la régression, c'est la moyenne (ou la médiane) des variables **y** des **K** plus proches observations qui servira pour la prédiction.
- Si K-NN est utilisé pour la classification, c'est le mode des variables **y** des **K** plus proches observations qui servira pour la prédiction.

Le principe de fonctionnement ci-dessus nous permet de schématiser l'algorithmique en pseudo-code suivant :

Algorithme K-NN

Déclaration des variables :

- Un ensemble de données **D**.
- Une fonction de définition distance **d**.
- Un nombre entier **K**

Début

Pour une nouvelle observation **X** dont on veut prédire sa variable de sortie **y** Faire :

- 1) Calculer toutes les distances de cette observation **X** avec les autres observations du jeu de données **D**.
- 2) Retenir les **K** observations du jeu de données **D** les plus proches de **X** en utilisant la fonction de calcul de distance **d**.
- 3) Prendre les valeurs de **y** des **K** observations retenues :
- 4) Si on effectue une régression, calculer la moyenne (ou la médiane) de **y** retenues
- 5) Si on effectue une classification, calculer le mode de **y** retenues
- 6) Retourner la valeur calculée dans l'étape 3 comme étant la valeur qui a été prédite par KNN pour l'observation **X**.

Fin Algorithm

V.3 Définition de la fonction de similarité dite de distance ‘d’ de l'algorithme K-NN

Comme présenté dans le pseudo-code algorithmique, il faut mettre sur pied une fonction de calcul de distance entre deux observations. Plus deux observations sont proches l'une de l'autre, plus ils sont similaires et vice versa. Cette fonction sera appliquée sur tous les exemples du jeu de données avec la nouvelle donnée à prédire.

En mathématique, il existe plusieurs fonctions de calcul de distance, on peut citer la distance :

Euclidienne ; de Manhattan ; de Minkowski ; de Jaccard ; de Hamming...

Dans la pratique, il est recommandé de choisir la fonction de calcul de distance en fonction des types de données (features) qu'on manipule. Ainsi pour :

- ✓ Les données quantitatives exemple : poids, salaires, taille, montant de panier électronique etc... et toutes (les inputs) du même type, la **distance euclidienne** sera très bien adaptée.
- ✓ Alors que la distance de Manhattan, sera un bon candidat quand les données (input variables) ne sont pas tous du même type (exemple : âge, sexe, longueur, poids etc...).

Note : Dans la pratique, il est utile de savoir que des librairies de Machine Learning comme Scikit Learn, effectuent ces calculs en interne. Il suffit juste d'indiquer la mesure de distance qu'on souhaite utiliser. Néanmoins, ci-dessous nous présentons les formules mathématiques afférentes à quelques distances énumérées ci-haut.

- **La distance euclidienne** (calcule la racine carrée de la somme des différences carrées entre les coordonnées de deux points)

$$D_e(x, y) = \sqrt{\sum_{j=1}^n (x_j - y_j)^2}$$

Équation 1 Distance euclidienne

- **Distance Manhattan** (calcule la somme des valeurs absolues des différences entre les coordonnées de deux points)

$$D_m(x, y) = \sum_{i=1}^k |x_i - y_i|$$

Équation 2 Distance Manhattan

- **Distance Hamming** (la distance entre deux points donnés est la différence maximale entre leurs coordonnées sur une dimension)

$$D_h(x, y) = \sum_{i=1}^k |x_i - y_i|$$

Équation 3 Distance Hamming

V.4 Détermination de la plage de K

Une question déterminante dans la réalisation d'une bonne prédiction en utilisant l'algorithme K-NN est le choix idéal de la valeur K à utiliser pour effectuer une prédiction. En règle générale, ce choix varie en fonction du jeu de données, ce qui est sûr :

- Si on arrête un nombre K petit i-e **moins on utilisera de voisins** : plus on tend vers un sous apprentissage (underfitting) du modèle.
- A contrario, **plus on utilise de voisins** (un nombre K grand) : plus on risque tendre vers surapprentissage de notre modèle et avoir de fausses prédictions.
- Si on détermine K nombre de voisins avec **K=N**, où N est le nombre d'observations (taille de notre dataset), on risque d'avoir du overfitting i-e du surapprentissage.

Afin de réussir la détermination de notre K, nous allons procéder par un tracé de la relation entre les valeurs de K et la précision de test correspondante. Ainsi, on pourra observer avec exactitude dans quelle plage tirer notre K et déterminer par la même occasion la précision de la prédiction. Cela sera rendu possible en utilisant les bibliothèques sklearn et ainsi que matplotlib.

V.5 Construction de notre modèle

#DECLARATION DES LIRAIRIES

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
import joblib
```

#CONSTANTES

```
np.random.seed(1)
n = 50
```

#INITIALISATION DES OBSERVATIONS D'ENPREPRISES VIABLES

```
df = pd.DataFrame({
    'chiffre_affaire': np.random.randint(10000000, 200000000, n),
    'nombre_client_payant': np.random.randint(100000, 1000000, n),
    'tresorerie_disponible': np.random.randint(10000000, 20000000, n),
    'depense_annuelle': np.random.randint(20000000, 100000000, n),
    'target': np.full(n, 1)
})
```

#VERIFICATION & IDENTIFICATION DES ENTREPRISES NON VIABLES

```
df['target'][((df.chiffre_affaire      <      df.depense_annuelle)      |      (df.tresorerie_disponible      <
(df.depense_annuelle/12)))] = 0
```

#INITIALISATION DES OBSERVATIONS D'ENPREPRISES NON VIABLES

```
df2 = pd.DataFrame({
    'chiffre_affaire': np.random.randint(20000000, 40000000, n),
    'nombre_client_payant': np.random.randint(10000, 100000, n),
    'tresorerie_disponible': np.random.randint(10000000, 20000000, n),
    'depense_annuelle': np.random.randint(20000000, 100000000, n),
```

```
'target': np.full(n, 0)
})

#VERIFICATION & IDENTIFICATION DES ENTREPRISES VIABLES
df2['target'][df2.chiffre_affaire > df2.depense_annuelle] = 1

#CREATION D'UN DATAFRAME & SHUFFLE
dataframe = df.append(df2, ignore_index=True)
dataframe = dataframe.sample(frac=1).reset_index(drop=True)
```

dataframe

	chiffre_affaire	nombre_client_payant	tresorerie_disponible	depense_annuelle	target
0	190483109	469137	12652309	39138701	1
1	150980108	230091	11344203	60021204	1
2	10491263	578234	14544775	35033131	0
3	23419401	757204	17411789	69876614	0
4	103440847	895984	13246792	43463629	1
...
95	28442573	94198	13546005	69513169	0
96	21654927	56163	12404223	62243888	0
97	26630809	67071	14921308	58279669	0
98	20816949	11200	18674281	43467404	0
99	24042452	92237	14699373	92573261	0

100 rows × 5 columns

```
#PROCESSUS DE SEPARATION DES DONNEES EN POUR ENTRAINNEMENT & TESTS
X = dataframe.iloc[:, :4]
y = dataframe.target
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.30)
```

```
#DATA SCALING « MISE A ECHELLE »
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
scaler.fit(X_train)
X_train = scaler.transform(X_train)
X_test = scaler.transform(X_test)
```

```
print(X_train.shape)
print(X_test.shape)
```

```
(70, 4)
(30, 4)
```

#EVALUATION DU TAUX DE FIABILITE DES PREDICTIONS SUIVANT LES VALEURS DE K

```
from sklearn.neighbors import KNeighborsClassifier
```

```
from sklearn import metrics
```

```
# Nous allons l'exécuter pour k = 1 à 25 et enregistrerons la précision des tests, les tracerons, montrerons la matrice de confusion et le rapport de classification:
```

```
range_k = range(1,25)
```

```
scores = {}
```

```
scores_list = []
```

```
for k in range_k:
```

```
    classifier = KNeighborsClassifier(n_neighbors=k)
```

```
    classifier.fit(X_train, y_train)
```

```
    y_pred = classifier.predict(X_test)
```

```
    scores[k] = metrics.accuracy_score(y_test,y_pred)
```

```
    scores_list.append(metrics.accuracy_score(y_test,y_pred))
```

```
result = metrics.confusion_matrix(y_test, y_pred)
```

```
print("Confusion Matrix:")
```

```
print(result)
```

```
result1 = metrics.classification_report(y_test, y_pred)
```

```
print("Classification Report:")
```

```
print (result1)
```

```
Confusion Matrix:
```

```
[[20  1]
 [ 1  8]]
```

```
Classification Report:
```

	precision	recall	f1-score	support
0	0.95	0.95	0.95	21
1	0.89	0.89	0.89	9
accuracy			0.93	30
macro avg	0.92	0.92	0.92	30
weighted avg	0.93	0.93	0.93	30

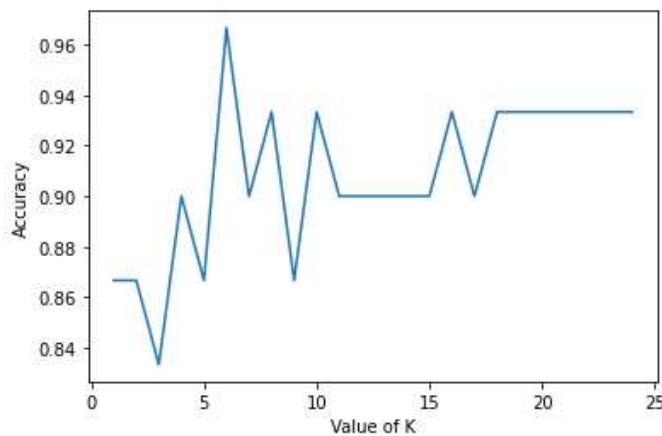
```
%matplotlib inline
```

```
import matplotlib.pyplot as plt
```

```
plt.plot(range_k,scores_list)
```

```
plt.xlabel("Value of K")
```

```
plt.ylabel("Accuracy")
```



Pour notre modèle, nous pouvons choisir la valeur optimale de K (toute valeur entre 19 et 25, car la précision est la plus élevée pour cette plage)

```
classifier = KNeighborsClassifier(n_neighbors = 20)
classifier.fit(X_train, y_train)
```

#SAUVEGARDE DU MODELE POUR UNE PERSISTANCE
joblib.dump(classifier, 'entreprise_prediction.joblib')

#MINI TEST

```
classes = {0:'Non viable', 1:'Viable'}
```

```
x_new = [[60000000, 40000, 28000000, 32000000],[20000000, 1000, 8000000, 92000000]]
```

```
y_predict = classifier.predict(x_new)
```

```
print(classes[y_predict[0]])
print(classes[y_predict[1]])
```

```
Viable
Non viable
```

Conclusion

Au terme de ce chapitre, exclusivement dédié à la présentation de l'algorithme K-NN qui est l'algorithme retenu pour l'apprentissage, nous mettons un terme à cette deuxième partie. Ce chapitre nous a permis d'atteindre un certain nombre d'objectifs à nous fixés dans notre cahier de charges de départ à savoir : la collecte de données sur les entreprises, l'exploration et traitement de ces données afin de construire notre modèle de ML. La partie suivante portera sur la réalisation de notre système.

TROISIEME PARTIE : REALISATION DU SYSTEME

Afin d'aboutir à une solution finale robuste et de qualité, nous allons faire du benchmark des méthodologies de développement existantes et d'en choisir une. Cette partie présentera la méthode de développement que nous allons utiliser pour parvenir à une solution ainsi que les différentes activités menées tout au long du processus d'élaboration (analyse, conception) de notre plateforme.

Chapitres

Chapitre VI : MÉTHODOLOGIE DE DÉVELOPPEMENT
Chapitre VII : MISE EN ŒUVRE DE LA SOLUTION

Chapitre VI : MÉTHODOLOGIE DE DÉVELOPPEMENT

Introduction

“A method is a set of principles for selecting and applying technics and tools in order to construct an efficient artefact”. Une méthodologie de développement se présente comme un cadre utilisé pour structurer, planifier et contrôler le développement d'une application dans le but de livrer un produit final dans les temps, avec une forte qualité et en minimisant les coûts. Dans ce chapitre nous allons procéder à la revue des méthodes de développement existantes en vue d'en choisir une qui nous servira de garde-fou tout au long de notre travail.

VI.1 Tour d'horizon des méthodes de développement

Avant de rentrer de plein fouet dans la présentation des différentes méthodes, nous allons donner une définition et lister les composants d'une méthode de développement.

De manière simpliste, une méthode de développement de systèmes logiciels est un ensemble d'étapes à exécuter suivant une chronologie précise dans l'optique d'analyser – concevoir – implémenter des systèmes logiciels.

Lorsque l'objet se situe dans la discipline informatique, une méthode de développement est composée de 03 éléments à savoir :

- **Des langages** : systèmes de communication permettant de spécifier, créer et documenter les éléments d'un système logiciel.
- **D'un processus** : ensemble structuré d'activités et agencées dans un ordre bien défini, nécessaires pour développer un produit logiciel.
- **Des modèles** : ensembles de concepts et de règles à appliquer.

Il existe une pléthore de méthodes de développement des systèmes d'information, ne pouvant pas tous les lister, nous allons présenter quelques-unes d'entre elles :

- **Les méthodes cartésiennes ou fonctionnelles** : L'idée principale qui se dégage est une vue fonctionnelle du système ou organisation à étudier.
- **Les méthodes systémiques** : Elles voudraient que le système étudié soit pris comme un artefact d'entités (processus) qui mis en relation communiquent autant en interne

qu'avec son environnement direct. Une méthode systémique = la modélisation des données + modélisation des traitements.

- **Les méthodes orientées-objet** : définissent un système informatique comme un ensemble d'entités appelées objets qui s'échangent des messages de manière cohérente. Chaque objet est identifiable et possède des propriétés clairement connues qui peuvent être soient des attributs ou des fonctions
- **Les méthodes agiles** : Elles basent son essence sur 02 qualificatifs principalement : itératif et incrémental. Pour elles, le processus de développement repose sur des activités à appliquer à chaque étape qui compose la méthode agile.
- **Les méthodes formelles** : Elles visent à exploiter les techniques d'inférences et les règles d'équivalences issues des mathématiques discrètes et de la logique. C'est une des raisons pour lesquelles le langage utilisé dans ces méthodes est textuel.

VI.2 Choix d'une méthode

Après une revue des méthodes de développement des systèmes d'information et eu égard de notre cahier de charges établi, nous estimons être judicieux de préconiser un processus de développement qui nous offrira une assise solide dès l'étape de détermination des besoins jusqu'à la validation finale en passant tour à tour par la conception et le codage final. C'est à juste titre que nous avons opté pour une démarche basée sur le processus **2-TUP**. Elle exploite le langage UML, standard incontournable de la modélisation objet.

VI.2.1 Présentation du langage UML

UML pour Unified Modeling Language, de manière brève se définit comme un langage de spécification, visualisation, construction et aussi de documentation des éléments d'un système logiciel conçu.

UML est une fusion des concepts clés des méthodes phares OMT, BOOCH et OOSE...

A travers ses 13 diagrammes, UML permet ainsi la modélisation d'un système depuis la spécification des besoins jusqu'à l'implémentation dans une machine. UML nous oblige à observer un système d'information selon deux aspects complémentaires : la structuration statique et celle dynamique ou comportementale. Dans chacune d'elle on y retrouve des diagrammes spécifiques.

Notre choix s'est porté sur UML car :

- ✓ La syntaxe de UML définit l'ensemble des concepts de modélisation UML, leurs attributs et leurs relations, ainsi que les règles permettant d'associer ces concepts afin de créer des modèles UML partiels ou complets.
- ✓ Il fournit une explication détaillée de la sémantique de chaque concept de modélisation. La sémantique définit, d'une façon indépendante de la technologie, comment les concepts UML doivent être mis en œuvre par les ordinateurs.
- ✓ Il spécifie des éléments de notation lisibles par l'homme pour représenter chaque concept de modélisation, ainsi que les règles pour les combiner au sein d'une grande variété de diagrammes correspondant à différents aspects des systèmes modélisés.

VII.2.2 2-TUP

2-TUP (2 Track Unified Process), appelé aussi modèle de développement en **Y**, est un processus de développement qui implémente le processus **UP** (Unified Process) qui fait de lui un processus unifié.

Le processus de développement 2-TUP couvre via ses 03 branches (voir figure ci-dessous) :

- L'étude des besoins fonctionnels ;
- L'étude des besoins techniques ;
- Le tout est unifié dans la réalisation.

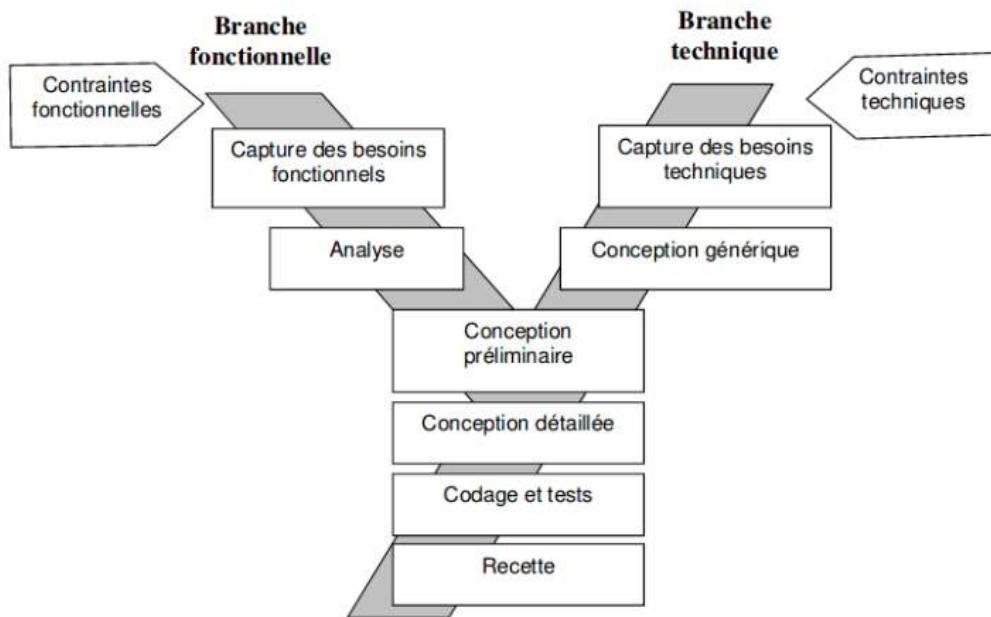


Figure 18 : Cycle en Y

Les principales caractéristiques de tout processus unifié sont : basé sur UML, itératif et incrémental, conduit par les cas d'utilisation, orienté par les risques, centré sur l'architecture du système.

Le cycle de développement de UP organise la mise en œuvre d'un système logiciel en un ensemble sous-projets afin de faciliter le développement. Chaque sous-projet a un cycle de vie en quatre phases :

- **Création** : Dans cette phase, il sera principalement question pour l'équipe en charge de collecter, analyser les besoins tant fonctionnels que techniques et procéder à une évaluation des risques.
- **Élaboration** : Ici, une étude approfondie du rendu de la phase précédente est faite.
- **Construction** : C'est ici que la conception et codage des fonctions du sous projet est réalisé de manière indépendante de l'ensemble.
- **Transition** : Une intégration du produit logiciel construit à l'étape précédente est faite à la solution finale, s'en suit des tests d'unification avant déploiement du produit final.

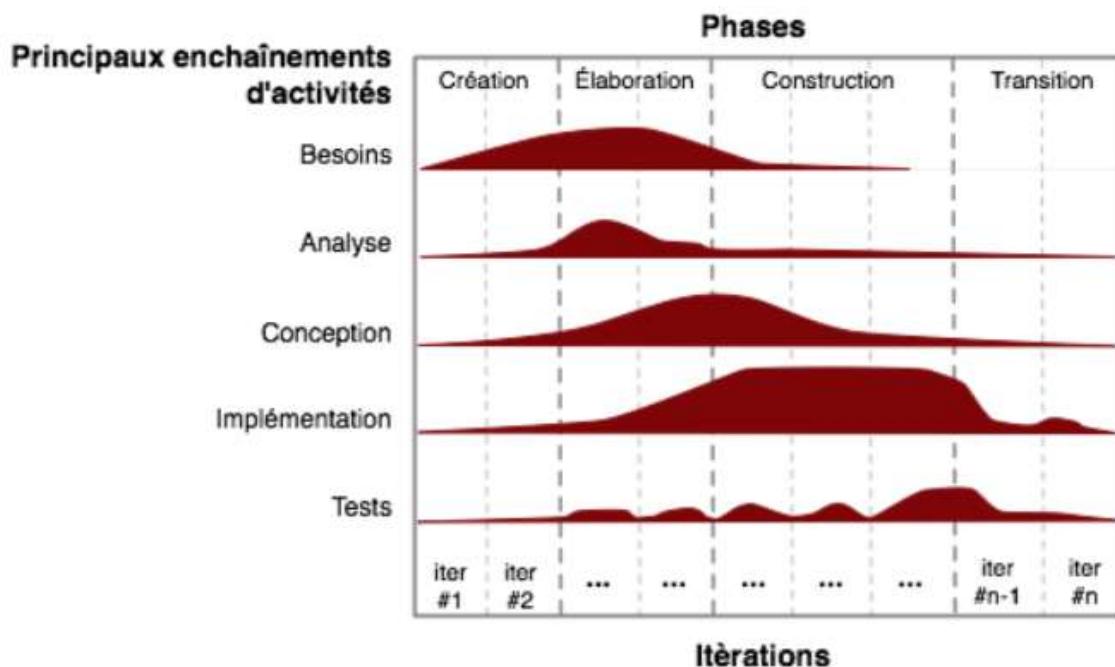


Figure 19 Phases et activités du cycle de développement

Conclusion

Ce chapitre nous a permis de réaliser un tour d'horizon rapide des méthodes de développement. Ceci, dans le but de dégager celle qui convient le mieux à notre solution logicielle. Ce qui nous a permis de choisir le processus unifié basé sur le langage UML et 2-TUP. Dans le chapitre suivant, nous aborderons l'analyse et la conception avec les moyens offerts par ledite processus.

Chapitre VII : MISE EN ŒUVRE DE LA SOLUTION

Introduction

Dans ce chapitre, nous allons détailler la démarche ; l'organisation nécessaire pour construire notre solution en respectant rigoureusement les jalons de la méthode choisie. En effet, la démarche 2-TUP recommande la construction de notre système en une succession de sous-projets. Chaque sous-projet aura un cycle de vie en 4 phases : création, élaboration, construction et transition. Pour les besoins de notre mémoire, nous allons présenter la mise en œuvre de 02 sous-projets ; le premier sous-projet étant **la définition globale de l'architecture de notre système et ses fonctionnalités**.

VII.1 Itération 1 : Étude globale du système

VII.1.1 Étude préalable

VII.1.1.1 Projet à réaliser

Ce projet consiste en une étude préliminaire des fonctionnalités de notre système.

VII.1.1.2 Recueil des besoins fonctionnels

Il s'agit des grandes fonctionnalités du système. Ces besoins fonctionnels expriment les principaux processus métiers que doit traiter le système :

1. Enregistrement des informations sur les entreprises ;
2. Consultation des profils d'informations des entreprises ;
3. Gestion des comptes investisseurs ;
4. Gestion des comptes entrepreneurs ;
5. Gestion des levées de fonds ;
6. Élaboration des statistiques.

VII.1.1.3 Recueil des besoins opérationnels

Il s'agira pour nous ici d'expliciter les volumes et la définition de la sécurité à implémenter afin d'assurer un fonctionnement optimal.

- ✓ Les échanges entre les acteurs et le système doivent être sécurisés ;
- ✓ Les informations d'une entreprise ne doivent qu'être accessibles et consultables que par un investisseur identifié ;

- ✓ La création d'un compte utilisateur se fait en fournissant obligatoirement une adresse email ou un numéro de téléphone valide ;
- ✓ Journalisation des actions effectuées dans le système ;
- ✓ Le site web doit être responsive.

VII.1.1.4 Identification des acteurs

Dans ce paragraphe, nous allons recenser les acteurs du système et préciser les prérogatives de chacun. Notons qu'un **acteur** est le rôle qu'un utilisateur ou une entité externe peut jouer en interagissant avec le système. Ceci dit, les entités revendiquant ces qualifications dans le cadre de notre projet sont les suivantes :

- **Le visiteur** est toute personne qui se rend sur la plateforme et effectue certaines actions sans être connecté à un compte préalablement créé par ce dernier.
- **L'entrepreneur ou responsable d'entreprise** est toute personne désirant publier sur la plateforme des informations sur l'entreprise dont il a la charge et qui aimerait nouer de partenariats avec de potentiels investisseurs pour l'évolution de son entreprise.
- **L'investisseur** est une personne physique ou morale qui alloue une part de ses avoirs (argent) personnels dans une opportunité d'affaire dans l'attente d'un retour sur investissement.
- **L'administrateur** qui est chargé de gérer les ressources du système ainsi que les comptes utilisateurs.

VII.1.1.5 Identification des messages

Un message représente la spécification d'une communication unidirectionnelle entre les objets, il transporte de l'information avec l'intention de déclencher une activité.

Le système reçoit les messages suivants	Le système envoie les messages suivants
<ul style="list-style-type: none"> • Création et modification de compte entrepreneur/investisseur • Demande de connexion • Soumission des informations • Demande liste des entreprises • Enregistrement levée de fonds et ou des sessions liées à une levée. • Statistiques d'entreprises 	<ul style="list-style-type: none"> • Informations sur le compte • Informations sur l'entreprise • Historique des levées de fonds et ou des Sessions liées à une levée • Liste des entreprises avec prédiction • Liste des messages • Statistiques

Tableau 1 Identification des messages Étude préalable

Note : Ces informations pourront être complétées lorsque nous aurons mené une analyse approfondie durant la phase d'analyse telle que prescrite par 2-TUP.

VII.1.1.6 Modélisation du contexte

La modélisation du contexte se fait à l'aide des diagrammes de contexte. Ils servent à délimiter le contour du système en cours d'étude, c'est-à-dire définir clairement ses frontières et les acteurs avec lesquels il communique. Il en existe deux types : Diagramme de contexte statique et diagramme de contexte dynamique.

1) Diagramme de contexte statique

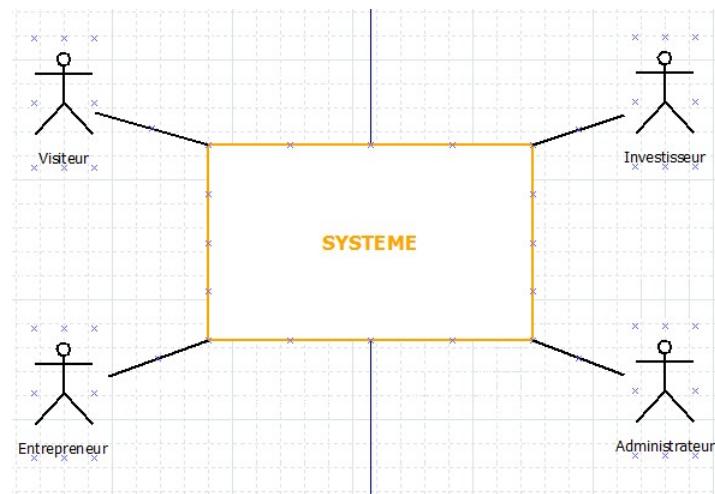


Figure 20 Diagramme de contexte statique - Etude globale du système

2) Diagramme de contexte dynamique

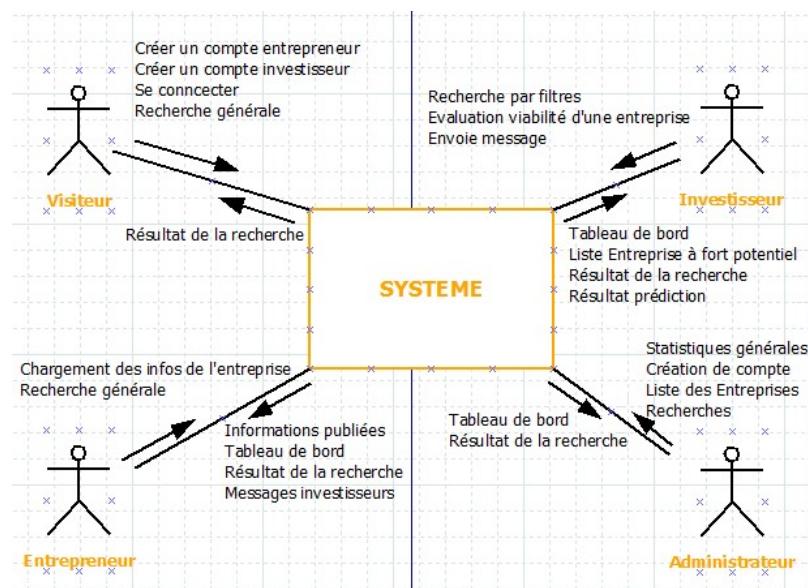


Figure 21 Diagramme de contexte dynamique - Etude globale du système

VII.1.1.7 Définition des choix techniques

1) Architecture matérielle

Notre solution est pensée avant tout pour fonctionner dans les environnements réseaux de type TCP/IP basé sur ETHERNET ou WIFI.

L'architecture de configuration matérielle est constituée d'un :

- Poste de travail ou tout autre terminal capable de se connecter via un navigateur à la plateforme. Ce poste n'aura pas besoin de ressources particulières car aucun traitement ne sera effectué côté client.
- Serveur web pour l'hébergement des composants de notre plateforme. Ce serveur devra supporter Django ; PHP ; Python principalement.

2) Langage de programmation

Le langage de programmation choisi doit :

- ✓ Supporter l'orientée-objet ;
- ✓ Avoir un typage assez flexible afin d'éviter les erreurs et de faciliter le débogage ;
- ✓ Être multi-plateforme, portable et supporter la programmation concurrente ;
- ✓ Possédant un riche panel de bibliothèque implémentant les fonctions mathématiques ;
- ✓ Possédant un riche panel de bibliothèque implémentant les algorithmes d'apprentissage ;
- ✓ Être un langage pour lequel des compétences et composantes peuvent être facilement trouvées.

3) Architecture logicielle

L'architecture choisie devra :

- ✓ Faciliter le découplage des fonctionnalités du système, car c'est via cette architecture que l'on identifie les composants et que l'on sait comment ces derniers communiquent ;
- ✓ Permettre le découpage du système en composants par une séparation claire des préoccupations (présentation, logique applicative, logique métier, accès aux données).

Les données de notre système seront sauvegardées dans un SGBDR (système de gestion de bases de données relationnelles), car ce dernier offre un support fiable de transactions. Notre système sera déployé au travers d'un serveur d'applications. Nous aurons besoin d'un framework qui facilite le travail de développement en offrant un ensemble de composants pour

gérer des problématiques telles que l'accès aux données, l'authentification, le développement des composants graphiques, la journalisation, etc.

4) Outils de travail

Pour le développement notre système, nous aurons besoin d'un :

- Environnement de développement intégré pour faciliter l'édition de code, le déploiement, l'écriture de tests ;
- Modeleur UML ;
- Outil de gestion de version de code source ;
- Outil de traitement de texte pour l'édition de nos rapports ;
- Outil d'édition et de présentation de diapositives.

VII.1.2 Capture des besoins fonctionnels

VII.1.2.1 Identification des cas d'utilisation

Les cas d'utilisations décrivent les fonctionnalités fournies par le système à un acteur. Ils permettent de mettre en évidence les relations fonctionnelles entre les acteurs et le système étudié. Dans le cadre de notre travail, nous avons pu identifier les cas d'utilisation listés dans le tableau ci-après :

N°	Cas d'utilisation	Acteur(s)	
01	S'authentifier	Entrepreneur,	Investisseur,
02	Modifier compte	Entrepreneur,	Investisseur,
03	Créer compte entrepreneur	Visiteur	
04	Créer compte investisseur	Visiteur	
05	Editer profil entreprise	Entrepreneur	
06	Enregistrer levée de fonds	Entrepreneur	
07	Enregistrer session de levée de fonds	Entrepreneur	
08	Visualiser statistiques consultation	Entrepreneur	
09	Consulter messages	Entrepreneur, Investisseur	
10	Envoyer message	Investisseur, Administrateur	
11	Consulter profil entreprise	Investisseur, Administrateur	
12	Désactiver compte	Administrateur	
13	Visualiser les statistiques globales	Administrateur	

Tableau 2 Identification des cas d'utilisation - Etude globale du système

VII.1.2.2 Description textuelle des cas d'utilisation

Nous allons décrire quelques cas d'utilisation.

■ Cas d'utilisation « S'authentifier »

- ◆ **Sommaire d'identification**
 - **Titre :** S'authentifier
 - **Résumé :** Ce cas d'utilisation permet à un utilisateur d'ouvrir une session sur le système.
 - **Acteurs :** Entrepreneur, Investisseur, Administrateur.
 - **Date de création :** 06-10-2020
 - **Version :** 1.1 | Responsable : Gan à Bol
- ◆ **Description des enchaînements**
 - **Préconditions**
 - Le système est accessible.
 - **Postconditions**
 - L'événement est enregistré dans le journal du système.
- ◆ **Scénario nominal**
 1. L'utilisateur clique sur le bouton de connexion.
 2. Le système affiche le formulaire d'identification.
 3. L'utilisateur remplit le formulaire avec son email et son mot de passe, puis valide.
 4. Le système vérifie qu'il existe un utilisateur dans la base de données dont l'email et le mot de passe correspondent à ceux renseignés dans le formulaire.
 5. Le système signale que l'opération s'est déroulée avec succès.
- ◆ **Scénario alternatif**

A1 : Le système ne trouve pas d'utilisateur correspondant aux identifiants saisis.

L'enchaînement A1 démarre au point 4 du scénario nominal.

1. Le système renvoie un message d'erreur et signale à l'utilisateur d'essayer à nouveau.
2. Retour au point 2 du scénario nominal

■ Cas d'utilisation « Créer compte entrepreneur »

- ◆ **Sommaire d'identification**
 - **Titre :** Créer compte entrepreneur
 - **Résumé :** Ce cas d'utilisation permet à un visiteur de s'inscrire en tant qu'entrepreneur sur la plateforme.
 - **Acteur :** Visiteur.
 - **Date de création :** 06-10-2020
 - **Version :** 1.1 | Responsable : Gan à Bol
- ◆ **Description des enchaînements**
 - **Préconditions**

- Le système est accessible.
- **Postconditions**
 - Le nouvel entrepreneur est enregistré en base de données.
 - L'événement est enregistré dans le journal du système.
- ◆ **Scénario nominal**
 1. Le visiteur clique sur le bouton **s'inscrire** du formulaire de connexion.
 2. Le système affiche une fenêtre modale avec 02 boutons.
 3. Le visiteur clique sur le bouton « **En tant qu'Entrepreneur** »
 4. Le système affiche le formulaire d'inscription.
 5. L'utilisateur remplit le formulaire avec les informations suivantes : noms, prénoms, email, mot de passe, numéro de téléphone et puis, valide.
 6. Le système vérifie que les informations soumises sont valides : noms, email, mot de passe et numéro de téléphone sont obligatoires.
 7. Le système renvoie une notification de succès.
- ◆ **Scénario alternatif**

A2 : Les informations renseignées par le visiteur ne sont pas valides.
L'enchaînement A2 démarre au point 6 du scénario nominal

 1. Le système renvoie un message d'erreur et signale à l'utilisateur d'essayer à nouveau.
 2. Retour au point 4 du scénario nominal

■ Cas d'utilisation « Éditer profil entreprise »

- ◆ **Sommaire d'identification**
 - **Titre** : Editer profil entreprise.
 - **Résumé** : Ce cas d'utilisation permet à un entrepreneur de renseigner les informations de profil de son entreprise tels que les **kpi** financiers, commerciaux, organisationnels ...
 - **Acteur** : Entrepreneur.
 - **Date de création** : 07-10-2020
 - **Version** : 1.1 | Responsable : Gan à Bol
- ◆ **Description des enchaînements**
 - **Préconditions**
 - Le système est accessible.
 - L'utilisateur est connecté.
 - **Postconditions**
 - Les informations de profile de l'entreprise sont enregistrées en base de données.
 - L'événement est enregistré dans le journal du système.
- ◆ **Scénario nominal**
 1. L'utilisateur clique sur le bouton de « **profil d'entreprise** ».
 2. Le système affiche le formulaire **profil d'entreprise**.

3. L'utilisateur remplit le formulaire avec les informations nécessaires et les soumet.
4. Le système vérifie que les informations soumises sont valides.
5. Le système renvoie une notification de succès.

◆ Scénario alternatif

A5 : Les informations renseignées par l'utilisateur ne sont pas valides.
L'enchaînement A5 démarre au point 4 du scénario nominal.

1. Le système renvoie un message d'erreur et signale à l'utilisateur d'essayer à nouveau.
2. Retour au point 2 du scénario nominal

■ Cas d'utilisation « Consulter profil entreprise »

◆ Sommaire d'identification

- **Titre :** Consulter profil entreprise.
- **Résumé :** Ce cas d'utilisation permet à un investisseur de visualiser le profil d'une entreprise.
- **Acteurs :** Investisseur, Administrateur.
- **Date de création :** 25-02-2021
- **Version :** 1.2 | Responsable : Gan à Bol

◆ Description des enchaînements

- **Préconditions**
 - Le système est accessible.
 - L'utilisateur est connecté.
 - L'évaluation des Entreprises générée.
- **Postconditions**
 - L'événement est enregistré dans le journal du système.
 - Une alerte (message) de consultation de profil est enregistrée et envoyée à l'entrepreneur.

◆ Scénario nominal

1. L'utilisateur clique sur une entreprise dans la liste.
2. Le système affiche le profil d'une entreprise.

VII.1.2.3 Diagramme des cas d'utilisation

Le diagramme de cas d'utilisation permet de représenter visuellement des fonctions ou services rendus par un système à un acteur bien défini. Ceci indépendamment de son fonctionnement interne. Des cas d'utilisation et des acteurs identifiés plus haut, nous avons construit le tout premier diagramme de UML suivant :

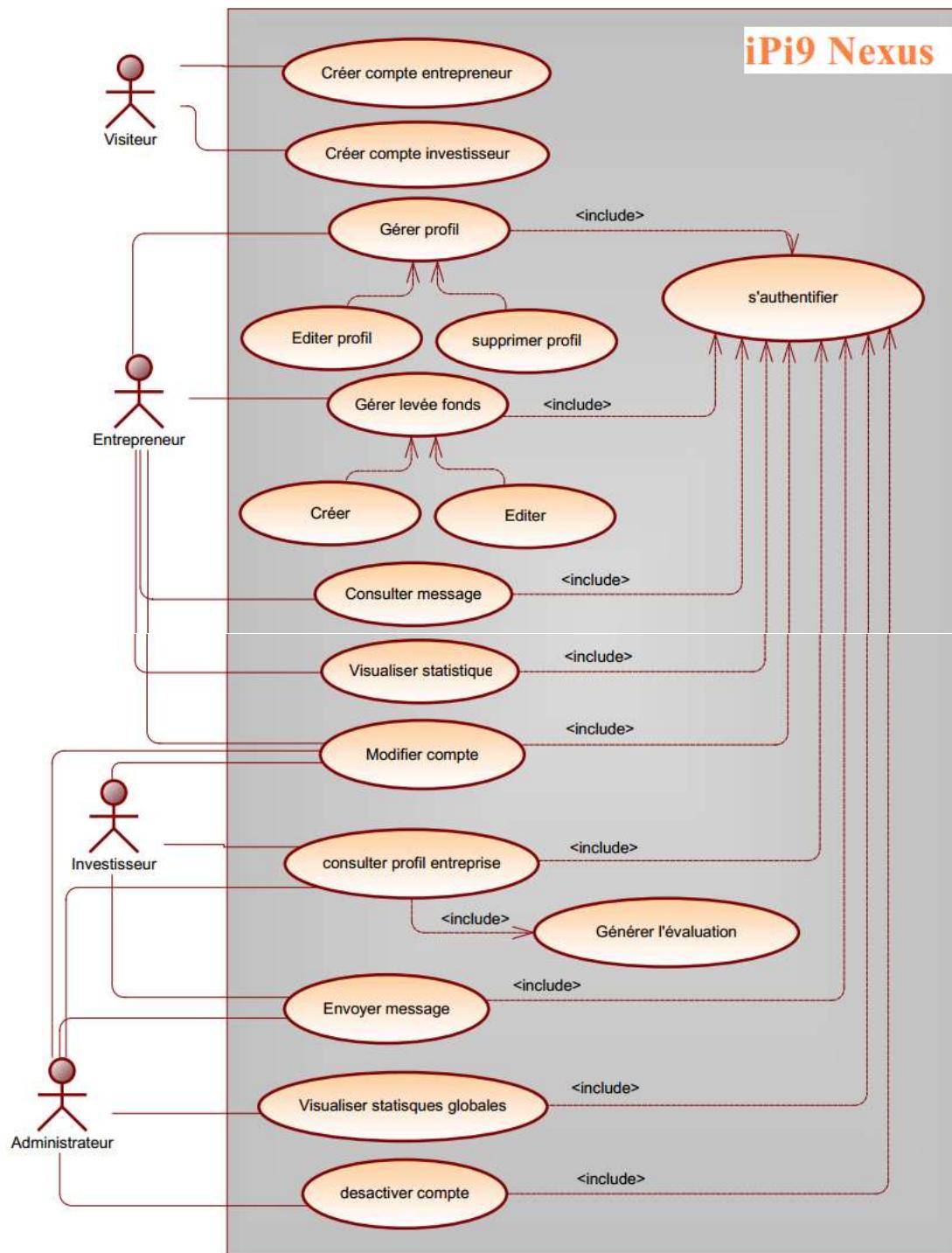


Figure 22 Diagramme de cas d'utilisation – Etude Globale

VII.1.2.4 Structuration des cas d'utilisation

Un package est un regroupement logique de uses cas à forte cohérence interne et faible couplage externe. Le découpage en packages de notre sujet a donné le résultat suivant :

N°	Cas d'utilisation	Packages
01	S'authentifier	Connexion
02	Créer compte entrepreneur	
03	Créer compte investisseur	
04	Modifier compte	Gestion des comptes
05	Désactiver compte	
06	Visualiser les statistiques globales	
07	Editer profil entreprise	
08	Visualiser statistiques consultation	Gestion profil
09	Consulter profil entreprise	
10	Enregistrer levée de fonds	Gestion des levées de fonds
11	Enregistrer session de levée de fonds	
12	Envoyer message	Messagerie
13	Consulter messages	

Tableau 3 Structuration en packages uses case - Etude globale du système

VII.1.2.5 Planification du travail en itérations

2-TUP étant une démarche itérative et incrémentale, notre planification sera faite sur la base des cas d'utilisation identifiés et décrits. La planification des cas d'utilisation est exposée aux risques suivants :

- ❖ **Information** : ce risque traduit la difficulté à comprendre les vocabulaires métiers du domaine. Pour chaque itération comportant ce type de risque, les concepts métiers méconnus ou mal appréhendés doivent être clarifiés dans le moindre détail.
- ❖ **Conception** : pour ce type de risque, une bonne réflexion sur les aspects conceptuels concernant la réalisation des cas d'utilisation s'impose. C'est le risque le plus important car la conception constitue le fondement de l'implémentation.
- ❖ **Implémentation** : les techniques, concepts et outils de programmation choisis doivent être maîtrisés et approfondis afin de satisfaire tous les cas d'utilisation constituant l'itération.

En outre, les différents cas d'utilisation seront classés en tenant compte des deux facteurs suivants :

- **La priorité fonctionnelle**, déterminée avec les valeurs : haut, moyen, bas ;

- **Les risques techniques**, estimés par le chef de projet avec les valeurs : haut, moyen, bas.

Pour cela, le chef de projet doit prendre en compte de façon combinée la priorité fonctionnelle et l'estimation du risque selon la logique suivante :

- Si la priorité est haute et le risque également, il faut planifier le cas d'utilisation dans les premières itérations.
- Si la priorité est basse et le risque également, on peut reporter le cas d'utilisation à une des toutes dernières itérations.
- Les choses se compliquent lorsque les deux critères sont antagonistes. Le chef de projet doit alors décider en pesant le pour et le contre.

Pour l'élaboration de notre système, nous proposons la planification en itérations suivantes :

Nº	Cas d'utilisation	Priorité fonctionnelle	Risque technique	Niveau de risque	Itération
01	S'authentifier	Haute	Bas	Implémentation	1
02	Créer compte entrepreneur	Haute	Bas	Information, conception	1
03	Créer compte investisseur	Haute	Bas	Information, conception	1
04	Modifier compte	Basse	Bas	Implémentation	5
05	Editer profil entreprise	Haute	Moyen	Information	2
06	Visualiser statistiques consultation	Basse	Bas	Implémentation	5
07	Consulter messages	Moyen	Moyen	Implémentation	3
08	Enregistrer levée de fonds	Haute	Moyen	Information, conception	2
09	Enregistrer session de levée de fonds	Haute	Moyen	Information, conception	2
10	Envoyer message	Basse	Bas	Implémentation	4
11	Consulter profil entreprise	Haute	Bas	Implémentation	2
12	Désactiver compte	Moyen	Bas	Implémentation	4
13	Visualiser les statistiques globales	Basse	Bas	Implémentation	5

Tableau 4 Planification en itération - Etude globale du système

VII.1.3 Capture des besoins techniques

Il s'agit des contraintes techniques qui caractérisent le système. Ce sont des besoins en matière de performance, de disponibilité, d'utilisabilité, de type de matériels, type

d'implémentation (langage de programmation, SGBD, système d'Exploitation...). Parmi les contraintes techniques, nous avons les contraintes suivantes :

- ◆ Le stockage et la manipulation fiable des données. Pour cela nous allons faire usage d'un **SGBDR** supportant les transactions. Pour notre solution, nous avons fait le choix de MySQL.
- ◆ Utilisation du paradigme objet : pour la mise en œuvre de ce paradigme, nous avons opté pour le langage Python (v.3.9.0). Il est multiparadigme, multi-plateforme, portable et supporte la programmation concurrente. HTML (v.5), CSS (v.3) et JavaScript seront aussi utilisés.
- ◆ La manipulation des modèles de données selon les préceptes de l'orientée-objet. Pour cela nous mettrons en œuvre une couche de mappage. La solution adoptée est l'ORM de Django.
- ◆ Sécurité de communication : notre solution sera déployée en intégrant les chiffrements des communications par le protocole SSL.
- ◆ Prototypage rapide de fonctionnalités et fiable : pour la mise en œuvre de cette contrainte, nous ferons usage du framework python Django.
- ◆ L'architecture matérielle répondra aux exigences suivantes :
 - Le poste de travail doit être doté de capacités (connectique réseau, CPU, Disque dur ...), pour faire tourner un système d'exploitation moderne (doté d'un environnement graphique et d'une pile réseau TCP/IP) comme Linux, Windows NT, UNIX/BSD, etc. Ce poste de travail devra être aussi doté d'un navigateur récent supportant les technologies HTML 5 et CSS 3. Pour répondre aux exigences de performances, le poste doit être multiprocesseurs avec minimum 2 Ghz de fréquence, 2 Go de mémoire vive, disposé d'une connexion à Internet d'au moins 100 ko/s de bande passante, des connectiques WIFI 802.11 a/b/g, 50 Go minimum de disque dur.
 - Le serveur sur lequel sera hébergé notre plateforme doit être une machine multiprocesseurs possédant : au minimum 2 Ghz de fréquence, 4 Go de mémoire vive, disposé d'une connexion à Internet d'au moins 512 ko/s de bande passante, 100 Go minimum de disque dur.

Le reste de nos besoins techniques est résumé dans le tableau ci-dessous.

Catégorie	Nom	Description	Logo
Gestion du Versionning	Git	Crée par Linus Torvald, Git est un logiciel permettant une gestion efficace de versions d'un projet (ensemble de fichiers) ou d'un fichier depuis sa création. Les nouvelles versions créées ne suppriment pas les anciennes, ce qui permettra de restaurer une version ultérieure si besoin est.	
Modeleur UML	DIA V 0.97.2-2	Crée à base du langage de programmation C, Dia est un logiciel gratuit et open source spécialisé dans la création de diagrammes dans plusieurs secteurs polyvalents et donc UML.	
SGBD	Power AMC 15	Conçu et mise en ligne par la société SAP, PowerAMC est un logiciel payant spécialisé dans la conception de modèles informatiques et dans des paradigmes multiples. Il favorise le passage du niveau logique au niveau physique directement implémentable dans une base de données.	
Serveur Web	MySQL v5.7.28	MySQL est la base de données open source la plus populaire au monde. Il est réputé pour être un serveur de base de données SQL rapide, stable, multiutilisateurs et multi-thread.	
	Wamp server	Wamp est un cadre web intégré de mise en production pour système d'exploitation windows. Sa version linux est XAMP. Avec Wamp il est possible de faire fonctionner en local des scripts PHP. Il comprenant trois serveurs, un interpréteur de script, ainsi que phpMyAdmin pour l'administration Web des bases MySQL	

Environnements de développement	PyCharm 2020.2.1	PyCharm est un éditeur de code intégré utilisé pour programmer en Python. Comme tout éditeur évolué, il offre des outils d'analyse de code et un débogueur graphique. Il permet également la gestion des tests unitaires, l'intégration de logiciel de gestion de versions, et supporte le développement web avec Django.	
Framework	Django 3.1.7	Django est un framework web open source de Python. Les concepteurs de ce framework avaient en tête d'offrir un cadre de développement python qui soit à la fois complet, puissant et qui, une fois pris en main par un développeur web, ce dernier verrai son rythme de travail aisément.	
Machin Learning	ANACONDA 3	Anaconda est un tout en un, libre et open source. Où on a la possibilité via son interface d'accéder à un ensemble d'outils, logiciels et langages tel R ou Python. Très orienté pour la réalisation d'applications dédiées à la science des données et à l'apprentissage automatique.	
	JUPYTER V 0.32.0	Jupyter Notebook de son nom complet est une produit logiciel type Web (s'exécutant dans un navigateur), open source qui permet la gestion de documents contenant du code, des équations, des graphiques et du texte narratif en direct.	
	SCIKIT LEARN V0.23.2	Scikit-learn est une librairie (regroupement de code exécutable), libre écrit en Python destinée à l'apprentissage automatique.	

Tableau 5 Besoins techniques - Etude globale du système

Notre plateforme de développement est bâtie sur une machine HP Pro Book, dotée de 8 Go de RAM, de 1 Tera de disque dur et d'un microprocesseur Intel core i5 64 bits 8^{ème} génération. Cette machine tourne sous un système Windows Professionnel 10.

VII.1.4 Conception générique

L'objectif de cette partie est de décrire l'architecture logique de notre système. L'architecture logicielle décrit la manière dont seront agencés les différents composants qui compose l'application et comment ils interagissent entre eux. Le modèle d'architecture décrit comment un système informatique doit être conçu de manière à répondre aux spécifications. Nous avons opté pour une architecture en 3-tiers associée à une architecture MVC via sa variante MVT implémentée dans Django.

VII.1.4.1 Le MVT

MVT (Modèle-Vue-Template) est un modèle de conception plus connu sous le terme de design pattern qui favorise une bonne organisation de son code source. Pour accélérer le processus de développement et surtout mieux maîtriser de bout en bout son code, le développeur a besoin d'un patron de conception qui lui permettra de mieux : anticiper des problématiques et améliorer la lisibilité du code. Avec MVT, ce dernier pourra savoir quels fichiers créer, quand le créer, mais surtout définir son rôle.

MVT sépare la logique du code en trois parties :

- **Un modèle** : C'est la partie du framework qui à l'exclusivité d'interagir avec la base de données. Lorsqu'une requête arrive, c'est au modèle de l'interpréter, de l'exécuter et de renvoyer une réponse en respectant le formalisme définit. Le modèle s'appuie sur un **ORM** (Object Relational Mapping, ou Mapping objet-relationnel en français).
- **Une vue** : C'est un peu comme le cœur, le pivot du MVT, car la vue est le point d'entrée de toutes les demandes utilisateurs et c'est le point de sortie de tout traitement réalisé par le système. Implémentant la logique métier, avec majoritairement des fichiers python (.py), c'est la vue qui sait à qui attribuer la charge d'exécution de la demande (Modèle ou Template).
- **Un template** : C'est la couche de présentation constituée principalement de fichiers HTML, qui gère complètement la partie de l'interface utilisateur. Concrètement, le template peut recevoir des objets Python, interpréter des variables et les afficher.

La figure suivante présente le rôle de chaque élément :

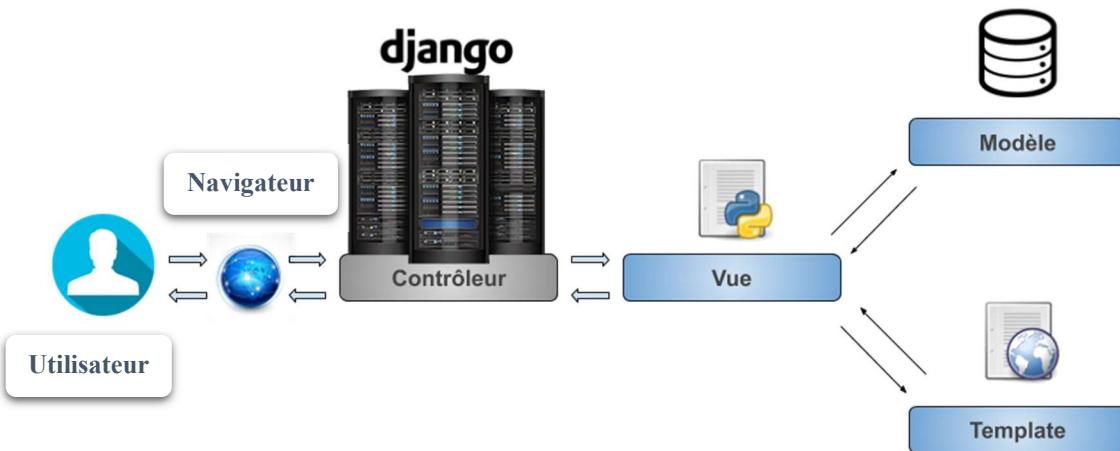


Figure 23 Architecture MVT de la solution

VII.1.4.2 Présentation de notre architecture

L’architecture client-serveur à trois niveaux est l’architecture technique correspondant au mieux à notre projet. Sa description est la suivante :

- Le client (premier tiers) est le client léger HTML, CSS, JavaScript s’exécutant dans un navigateur web. En effet, le client web enverra des requêtes HTTPS au serveur et recevra des fichiers HTML, CSS et JavaScript de ce dernier en guise de réponse.
- Le second tiers qui constitue le cœur de notre système, implémente la logique applicative et métier. Il est basé sur une architecture MVT implanté par le framework Django présentée plus haut. Il implémente également les objets métiers et la couche d’accès aux données.
- Le troisième tiers est le serveur de données MySQL.

Le schéma ci-dessous résume ce qui a été dit précédemment :

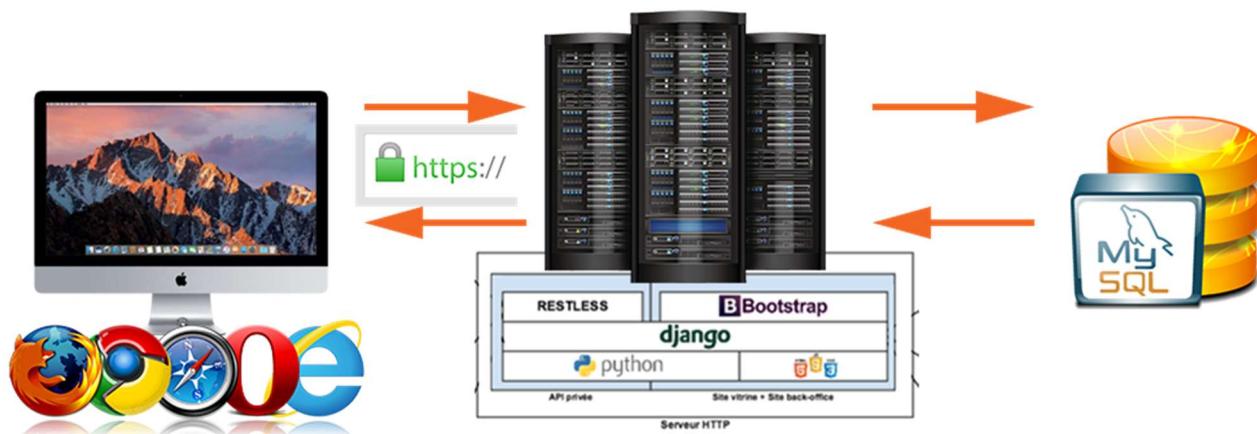


Figure 24 Architecture retenue solution

VII.1.5 Analyse

VII.1.5.1 Classes candidates

Il s'agit d'identifier les principales classes métiers de notre système. Ces classes métiers sont les suivantes :

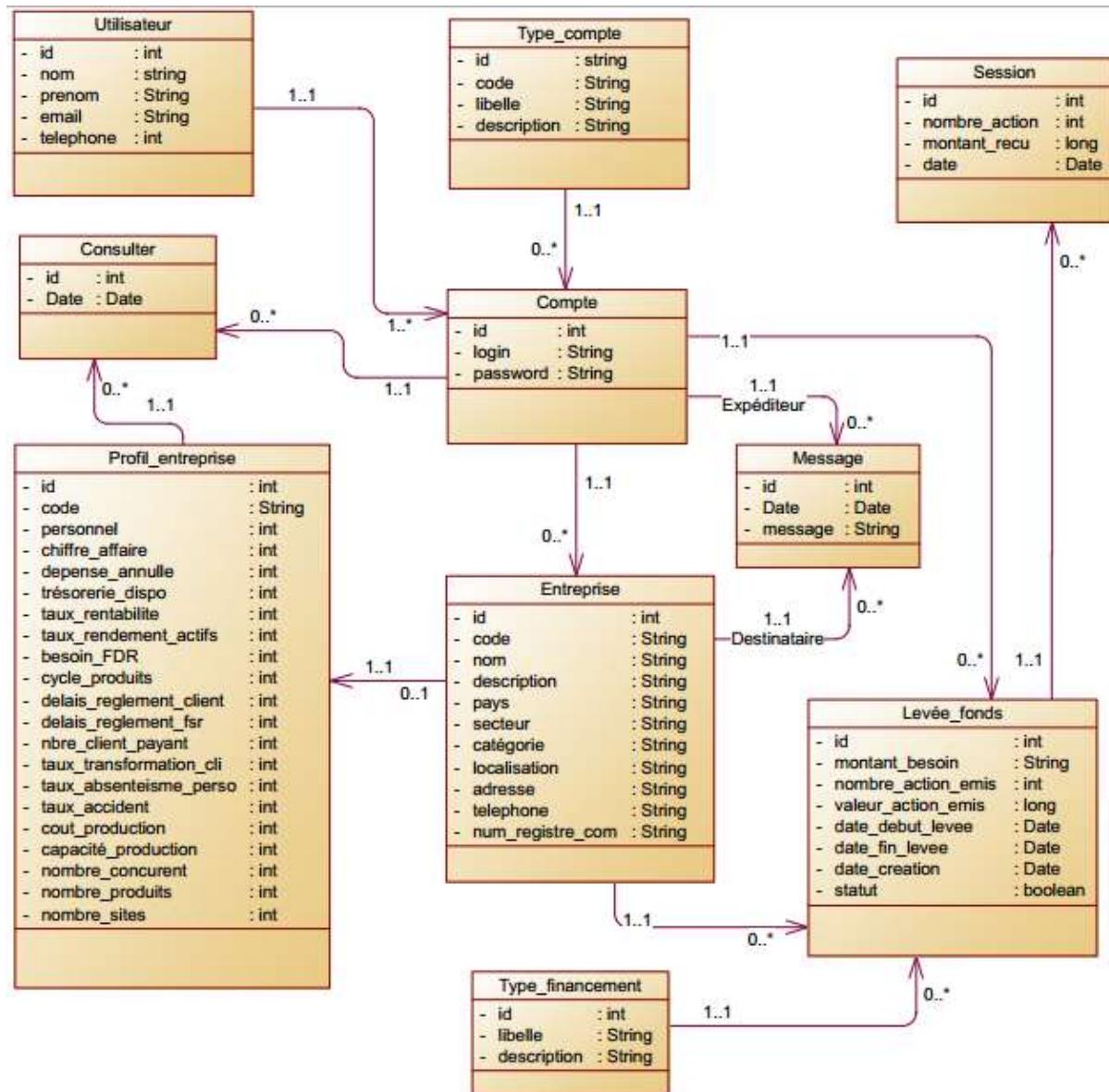


Figure 25 Diagramme de classes - Etude globale du système

VII.1.5.2 Diagramme d'activités de navigation

Un diagramme d'activités représente graphiquement la suite d'étapes qui constituent un processus, le comportement d'une méthode ou le déroulement d'un cas d'utilisation et les contraintes de séquencement. Il permet de mettre l'accent sur les traitements. Ci-dessous, le diagramme d'activités de navigation global de notre système.

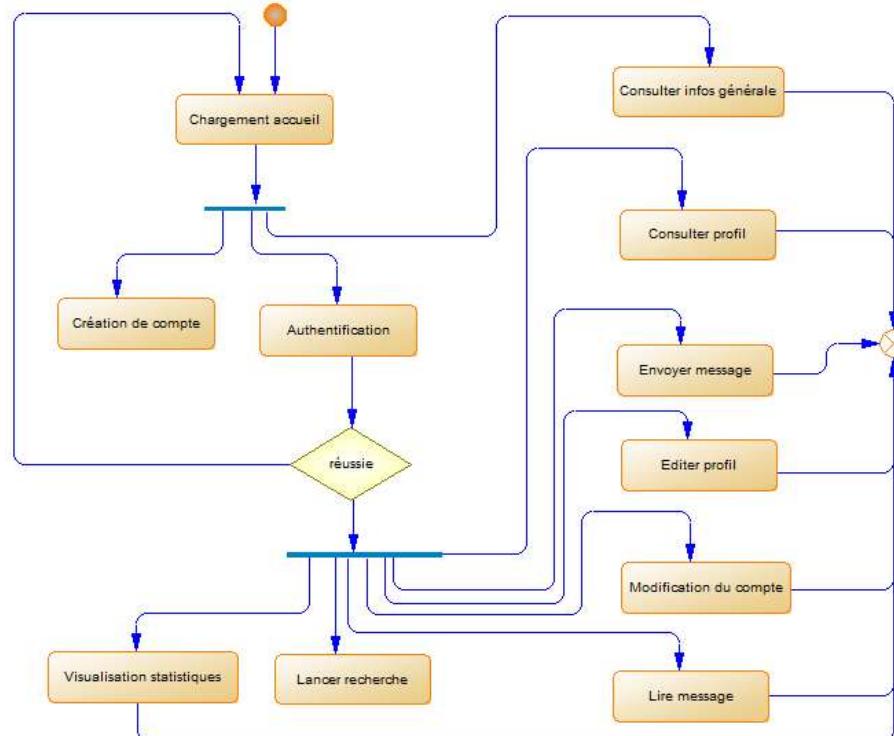


Figure 26 Diagramme d'activité - Etude globale du système

VII.1.5.3 Diagramme de composants

Ce diagramme nous permettra de mettre en évidence l'organisation du système avec un accent sur les différents nœuds qui vont le constituer.

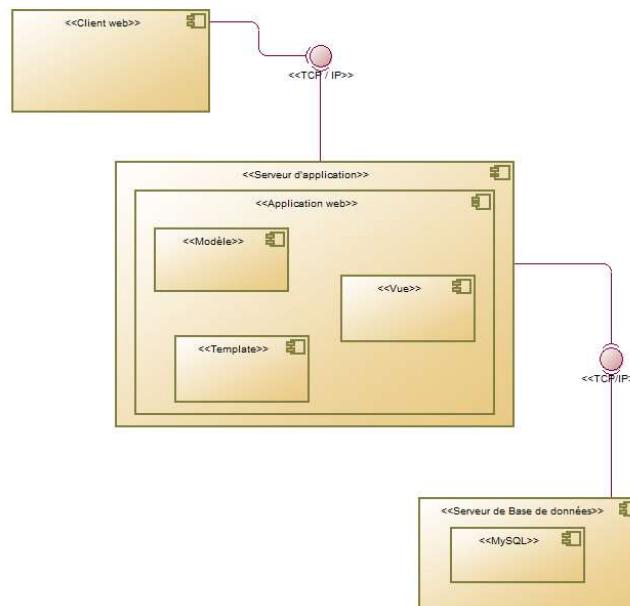


Figure 27 Diagramme de composants du système - Étude globale du système

VII.1.6 Autres activités

Dans l'optique de respecter la démarche incrémentale préconisée par les processus unifiés, nous ne mettons pas un accent sur les activités de conception préliminaire, conception détaillée, codage, tests et recette dans ce sous-projet.

VII.2 Itération 2 : Gestion profils entreprises

VII.2.1 Étude préalable

VII.2.1.1 Projet à réaliser

Le but de ce projet consiste à mettre en œuvre l'ensemble des fonctionnalités liées à la gestion et manipulations des données liées au profil de l'entreprise.

VII.2.1.2 Recueil des besoins fonctionnels

Les principaux besoins fonctionnels sont :

- Enregistrement des informations d'identification de l'entreprise ;
- Enregistrement des KPI de l'entreprise ;
- Enregistrement des consultations de profil d'entreprise ;
- Consulter profil entreprise ;
- Visualiser statistiques de consultation ;

VII.2.1.3 Recueil des besoins opérationnels

- L'édition du profil de l'entreprise ne sera faite que par un utilisateur identifié et connecté comme entrepreneur sur la plateforme ;
- Chaque profil d'entreprise devra être unique ;
- Toute consultation du profil de l'entreprise par un investisseur devra faire l'objet d'une notification à son propriétaire ;
- L'envoi des informations au serveur doit être sécurisé ;
- Seuls les investisseurs enregistrés peuvent consulter la liste des entreprises enregistrées.

VII.2.1.4 Identification des acteurs

- ✓ **Entrepreneur** : seul acteur capable d'éditer le profil de l'entreprise et visualiser les statistiques de consultation de son profil ;
- ✓ **Investisseur, Administrateur** : peuvent consulter les profils des entreprises enregistrées.
- ✓ **Administrateur** : il peut lister, désactiver un profil d'entreprise ne répondant pas aux normes et ou mal renseigné.

L'administrateur, entrepreneur et investisseur, doivent s'authentifier avant de réaliser les besoins listés ci-dessus.

VII.2.1.5 Identification des messages

Le système émet les messages suivants	Le système reçoit les messages suivants
<ul style="list-style-type: none"> • Profil entreprise • Prédiction de viabilité de l'Ese. • Alerte de consultation de profil 	<ul style="list-style-type: none"> • Enregistrement profil entreprise • Modification profil entreprise • Demande profil entreprise • Désactivation profil entreprise

Tableau 6 Identification des messages itération 2

VII.2.2 Capture des besoins fonctionnels

VII.2.2.1 Identification des cas d'utilisation

Cas d'utilisation	Acteur (s)	Interface d'accès
Editer profil entreprise	Entrepreneur	Application web
Visualiser statistiques de consultation	Entrepreneur	Application web
Consulter profil entreprise	Investisseur - Administrateur	Application web

Tableau 7 Identification des cas d'utilisation itération 2

VII.2.2.2 Diagramme des cas d'utilisation

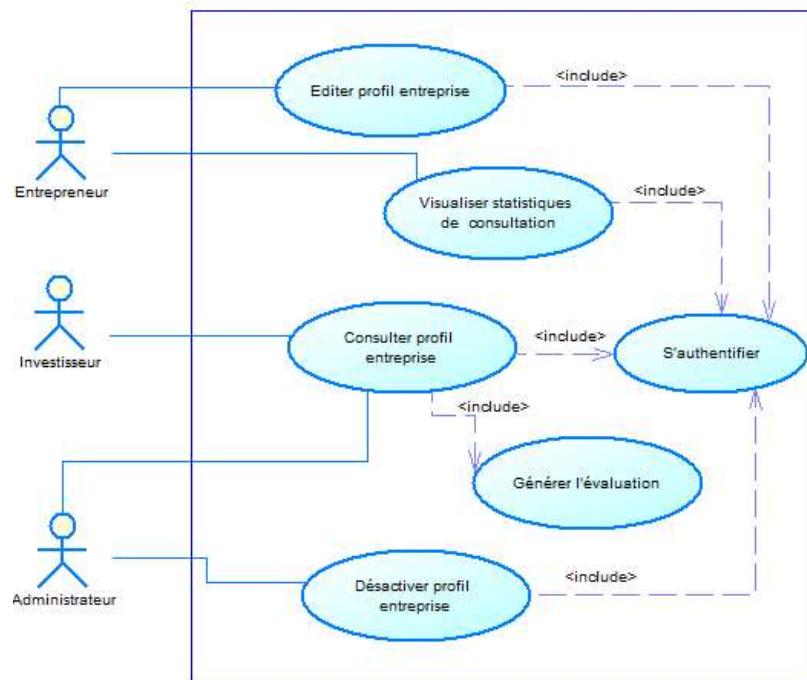


Figure 28 Diagramme de cas d'utilisation - Gestion profils entreprises

VII.2.3 Capture des besoins techniques

Les principaux besoins techniques sont :

VII.2.3.1 L'unicité de l'entreprise.

Les informations d'identification d'une entreprise devant être unique dans la plateforme, nous prévoyons garantir ce besoin technique en assurant l'unicité de : la dénomination ; le numéro de téléphone ; et le numéro de la fiche unique (numéro de registre du commerce), pour chaque entreprise enregistrée.

VII.2.3.2 Protection des informations sur les entreprises.

Il sera primordial pour nous de mettre en place une sécurité pendant l'édition, l'envoie et l'utilisation des données des entreprises, afin d'avoir une confiance des utilisateurs. Pour ce faire, la protection des informations envoyées au serveur sera garantie par le protocole HTTPS et l'accès à ces données sera réservé exclusivement aux personnes autorisées.

VII.2.4 Analyse

VII.2.4.1 Diagramme de classes d'analyse

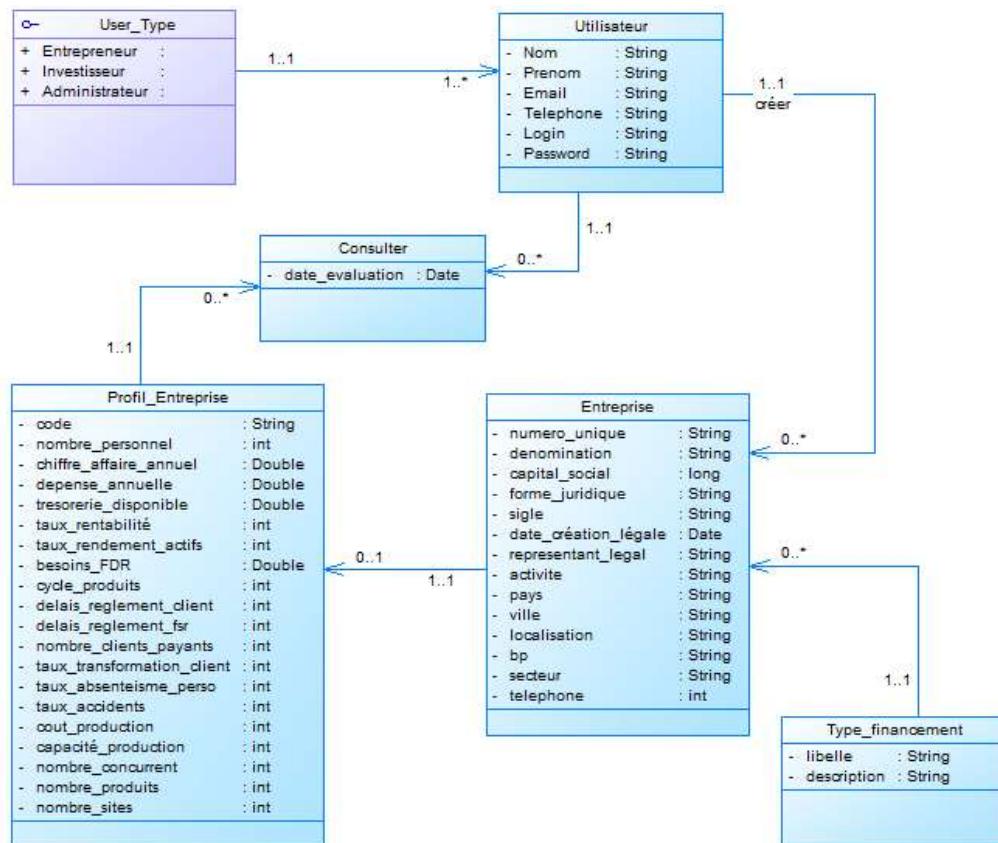


Figure 29 Diagramme de classes - Gestion profils entreprises

VII.3.4.2 Diagrammes de séquence système (DSS)

Le diagramme de séquences permet de montrer les interactions entre les acteurs et le système dans le cadre d'un scénario d'un diagramme des cas d'utilisation, tout en mettant l'accent sur l'aspect temporel.

1) Diagramme de séquence du cas d'utilisation « Éditer profil entreprise »

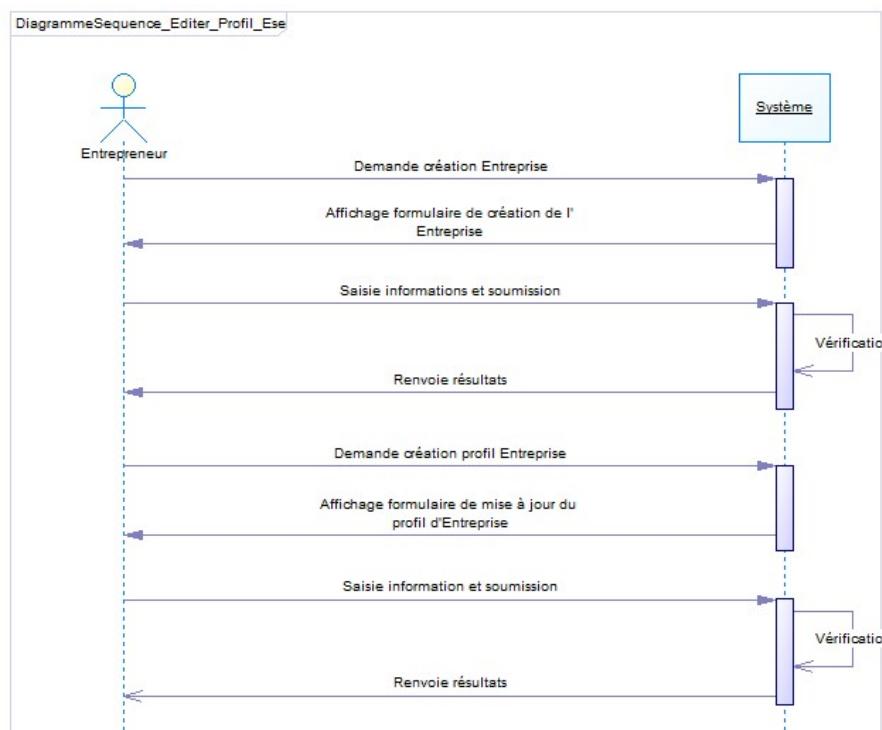


Figure 30 Diagramme de séquence du cas d'utilisation « Éditer profil entreprise »

2) Diagramme de séquence du cas d'utilisation « Consulter profil entreprise »

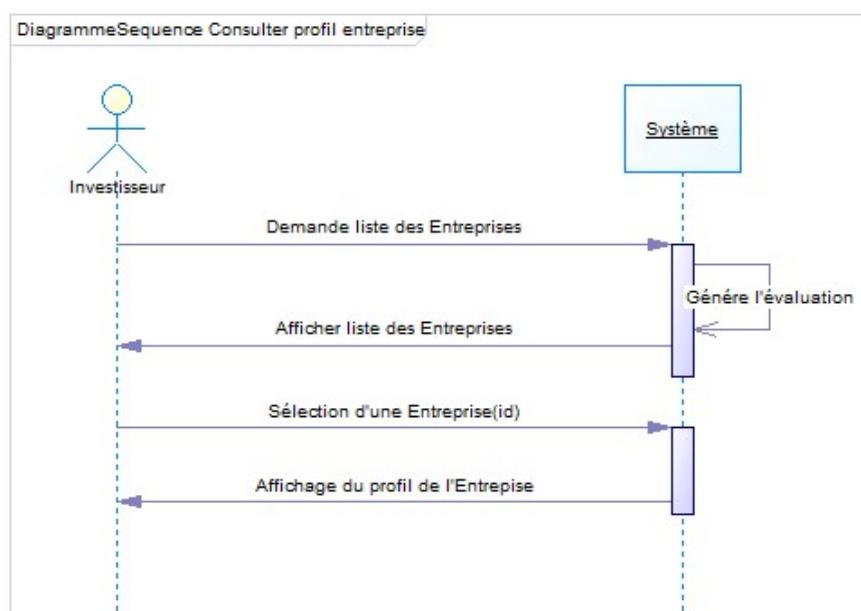


Figure 31 Diagramme de séquence du cas d'utilisation « Consulter profil entreprise »

3) Diagramme de séquence du cas d'utilisation « Désactiver profil entreprise »

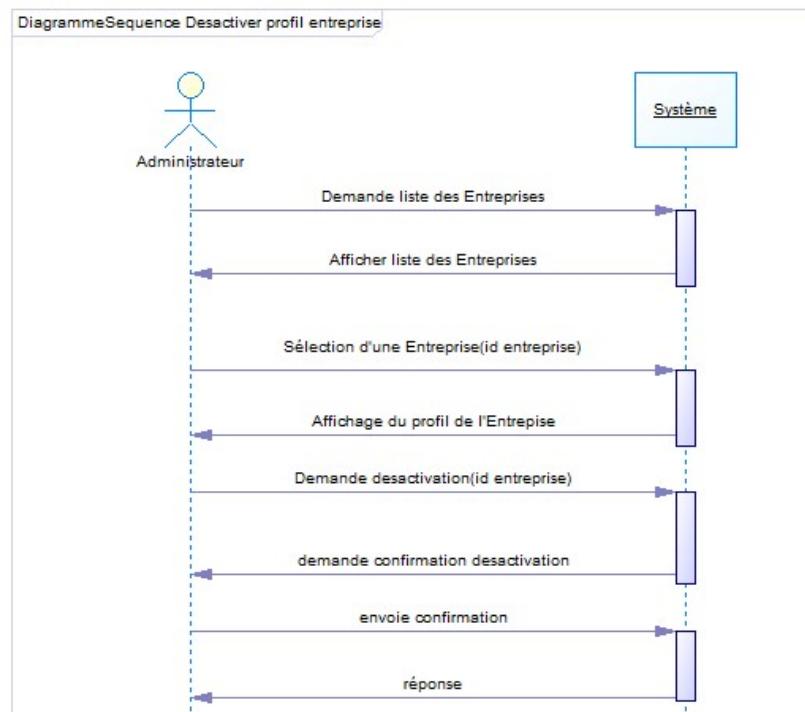


Figure 32 Diagramme de séquence du cas d'utilisation « Désactiver profil entreprise »

VII.2.5 Conception préliminaire

Il est question de décrire le système vu comme une boîte blanche en le décomposant en ses différents composants de dialogues, **contrôles et entités**. De plus la description doit de se faire de façon générique en étant le plus éloigné possible des mécanismes liés à la plateforme d'exécution ou de programmation.

VII.2.5.1 Diagramme de classes participantes (DCP)

Le diagramme des classes participantes modélise trois types de classe d'analyse :

- ◆ Les classes de type dialogue « **boundary** » :
Se sont typiquement les écrans proposés à l'utilisateur. On y retrouve par exemple : les formulaires de saisie, les résultats de recherche...
- ◆ Les classes de type métier « **entity** » :
Elles représentent les règles métiers et proviennent directement du modèle du domaine mais sont confirmées et complétées pour chaque cas d'utilisation.
- ◆ Les classes de type contrôle « **control** » :
Elles sont les plaques tournantes de l'application et font la transition entre les dialogues et les classes métiers, en permettant aux écrans de manipuler des informations détenues par un ou plusieurs objets métiers.

Nous allons identifier les entités, les dialogues et les contrôles, puis réaliser les diagrammes des classes participantes (DCP) des cas d'utilisation suivants.

➤ **DCP du cas d'utilisation « Éditer profil entreprise »**

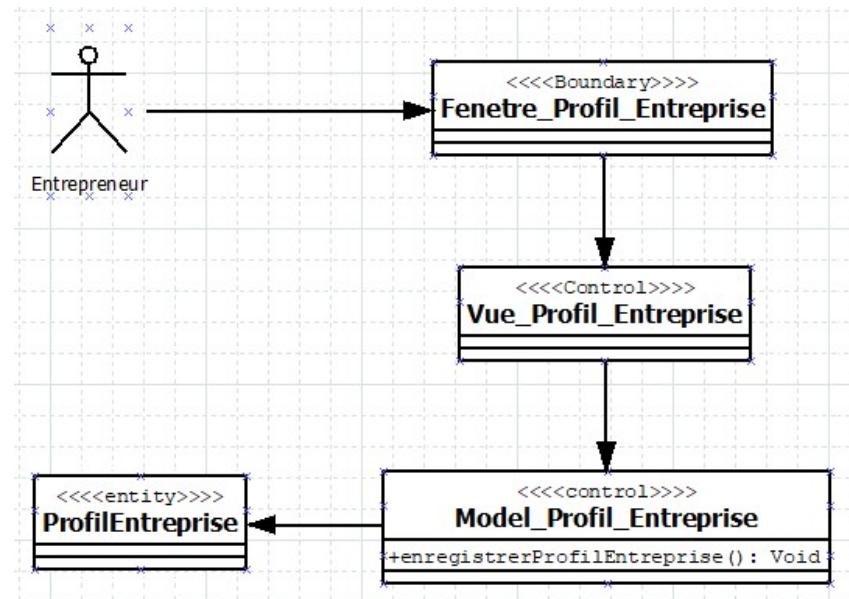


Figure 33 DCP du cas d'utilisation « Editer profil entreprise »

➤ **DCP du cas d'utilisation « Consulter profil entreprise »**

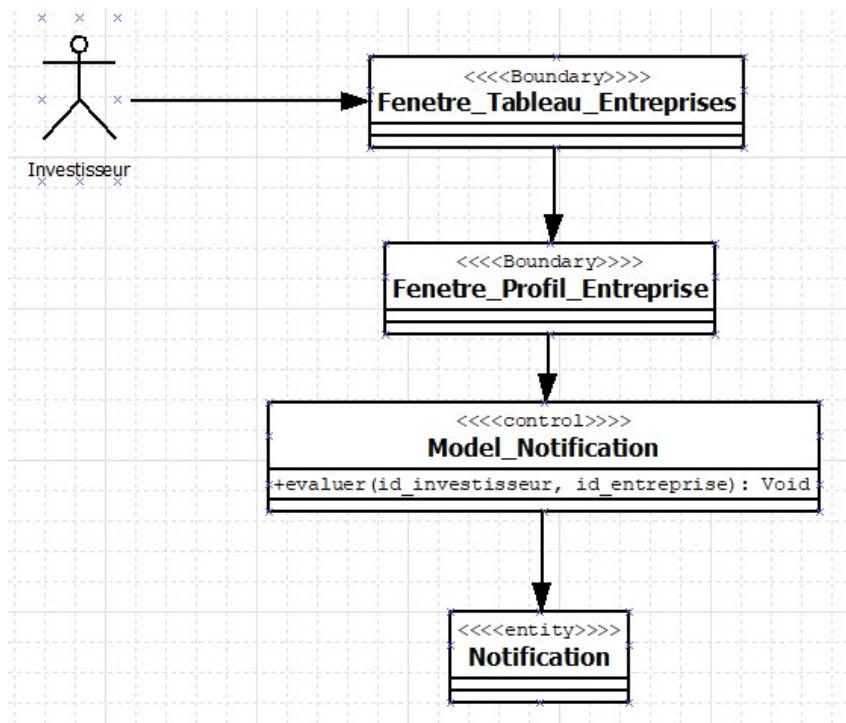


Figure 34 DCP du cas d'utilisation « Consulter profil entreprise »

VII.3.5.2 Diagrammes de séquence-objet préliminaire (DSOP)

Dans cette section, à la différence de la section précédente où le diagramme de séquence d'un scénario représente une séquence d'interactions entre le système et ses acteurs. Le système était alors considéré comme une boîte noire. A présent, nous allons remplacer le système par une collaboration d'objets intervenants dans chaque scénario.

➤ DSOP du cas d'utilisation « Créer profil Entreprise »

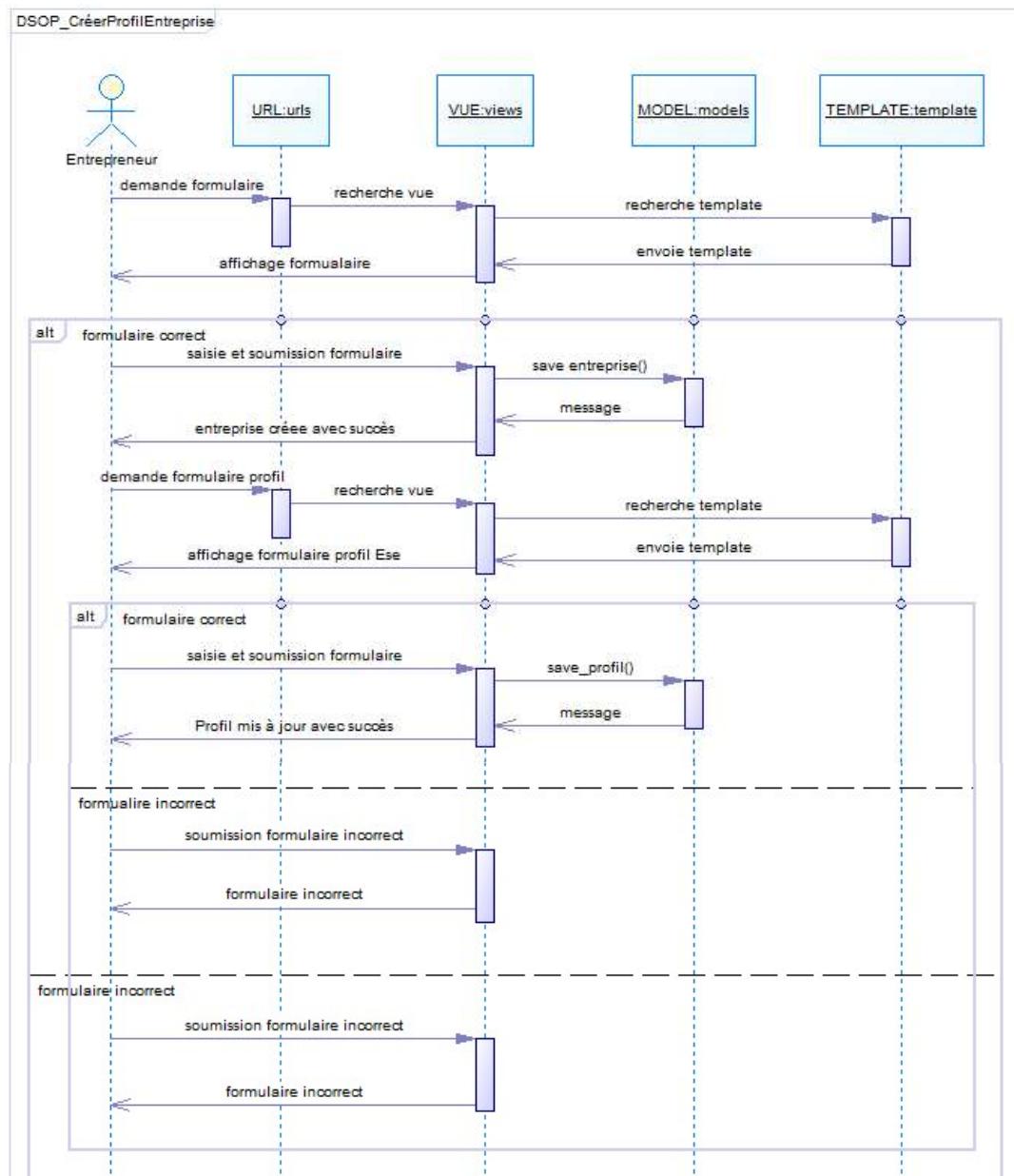


Figure 35 DSOP du cas d'utilisation « Créer profil Entreprise »

➤ **DSOP du cas d'utilisation « Consulter profil entreprise »**

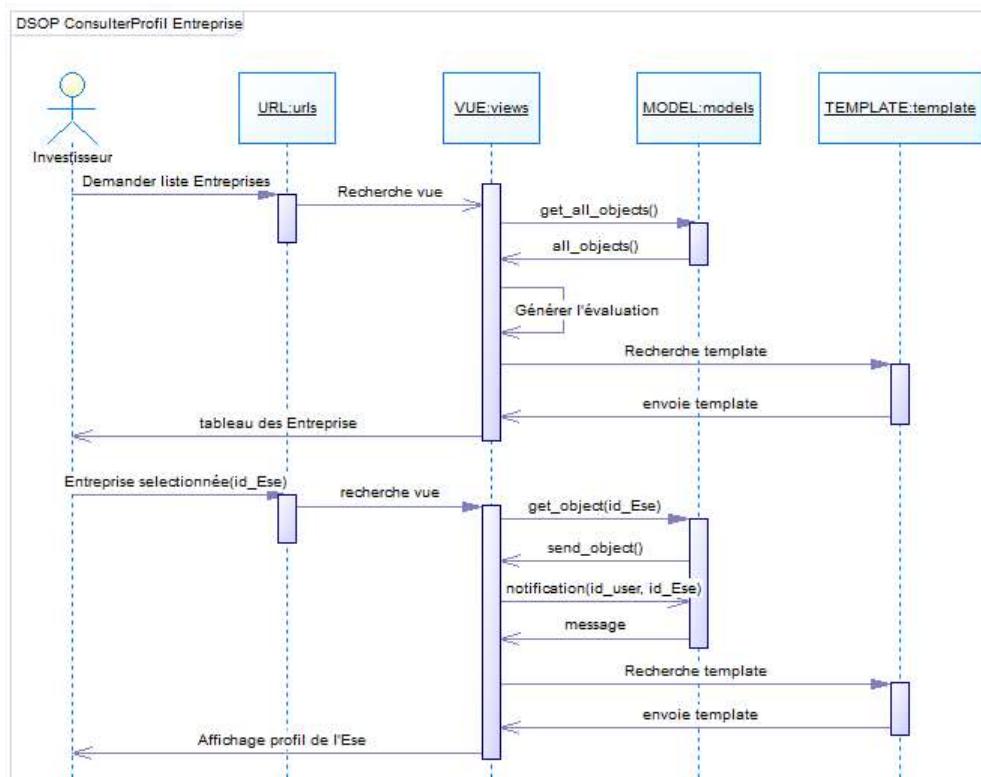


Figure 36 DSOP du cas d'utilisation « Consulter profil entreprise »

➤ **DSOP du cas d'utilisation « Désactiver profil entreprise »**

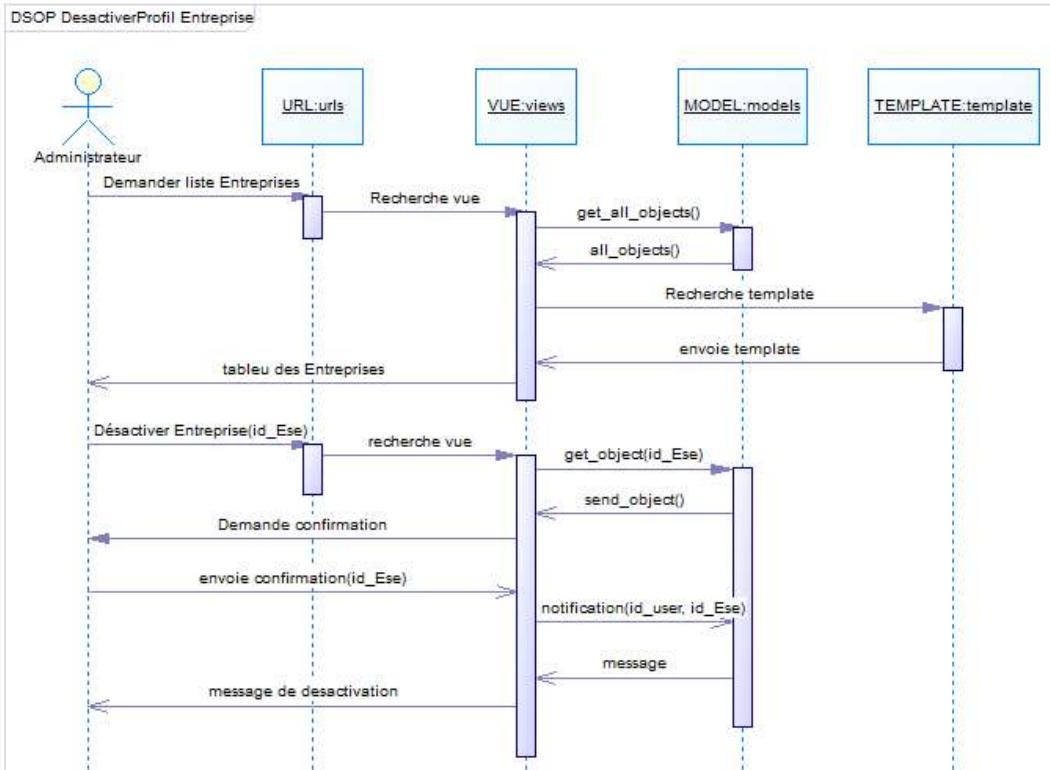


Figure 37 DSOP du cas d'utilisation « Désactiver profil entreprise »

VII.2.6 Conception détaillée

Nous allons nous atteler à construire et à documenter de manière plus précise les classes, les interfaces, les tables et les méthodes qui constitueront le code de la solution. S'appuyant fortement à la fois sur les Frameworks techniques (à utiliser dans la phase de codage), et les regroupements propres au métier, la conception détaillée de cette itération nous permettra d'avoir une idée précise et claire pour la fabrication et l'assemblage des sous-systèmes de configuration logicielle car toutes les boîtes noires existantes jusqu'ici seront détaillées.

VII.2.6.1 Diagramme de classes détaillées (DCD)

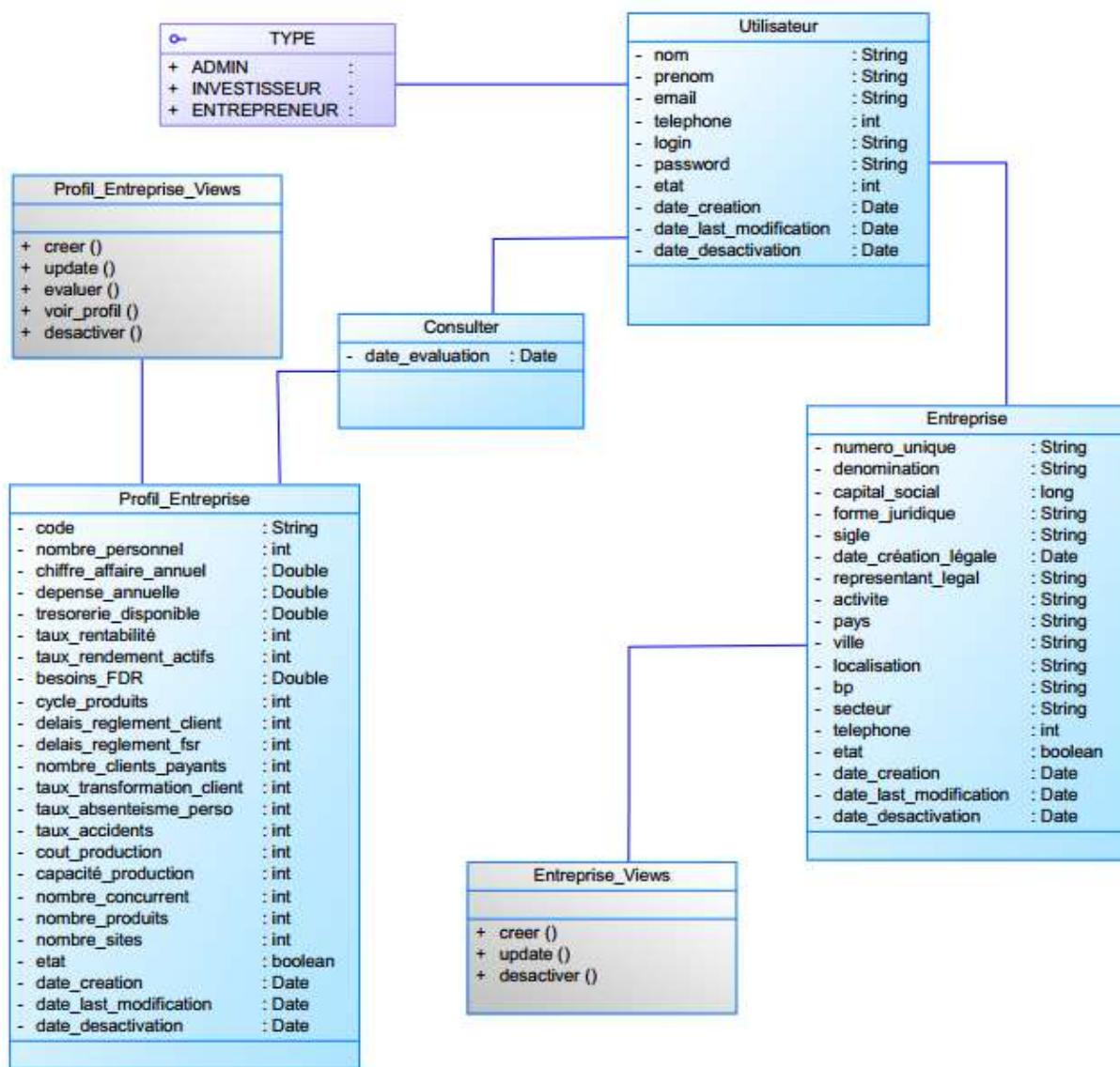


Figure 38 Diagramme de classes détaillées (DCD)

VII.3.6.2 Diagramme de séquences-objets détaillés (DSOD)

A ce niveau d'abstraction, nous allons faire intervenir le 3^{ème} tiers de notre architecture 3-tiers qui est la base de données.

➤ DSOD du cas d'utilisation « Créer profil Entreprise »

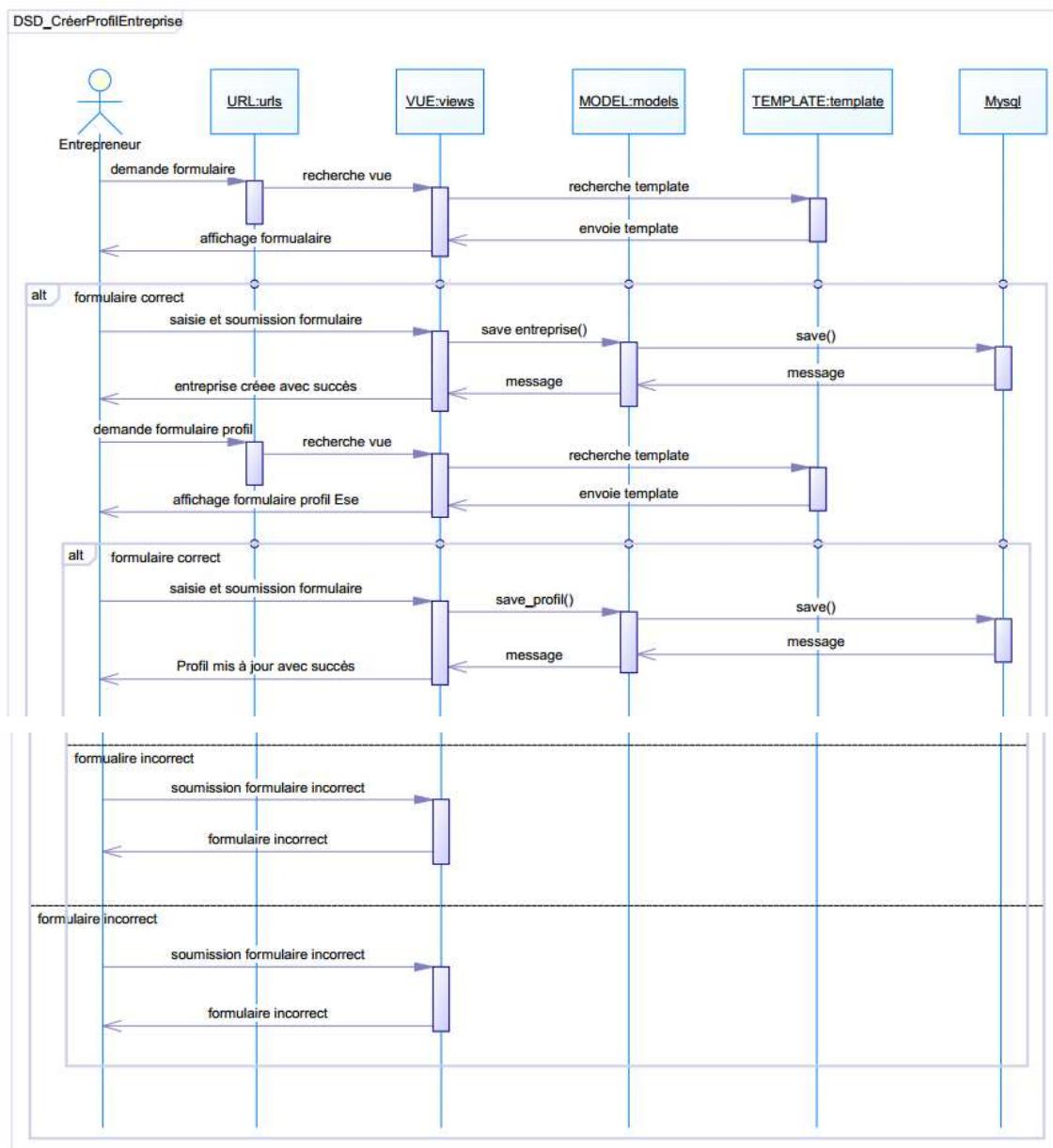


Figure 39 DSOD du cas d'utilisation « Créer profil Entreprise »

➤ **DSD du cas d'utilisation « Consulter profil entreprise »**

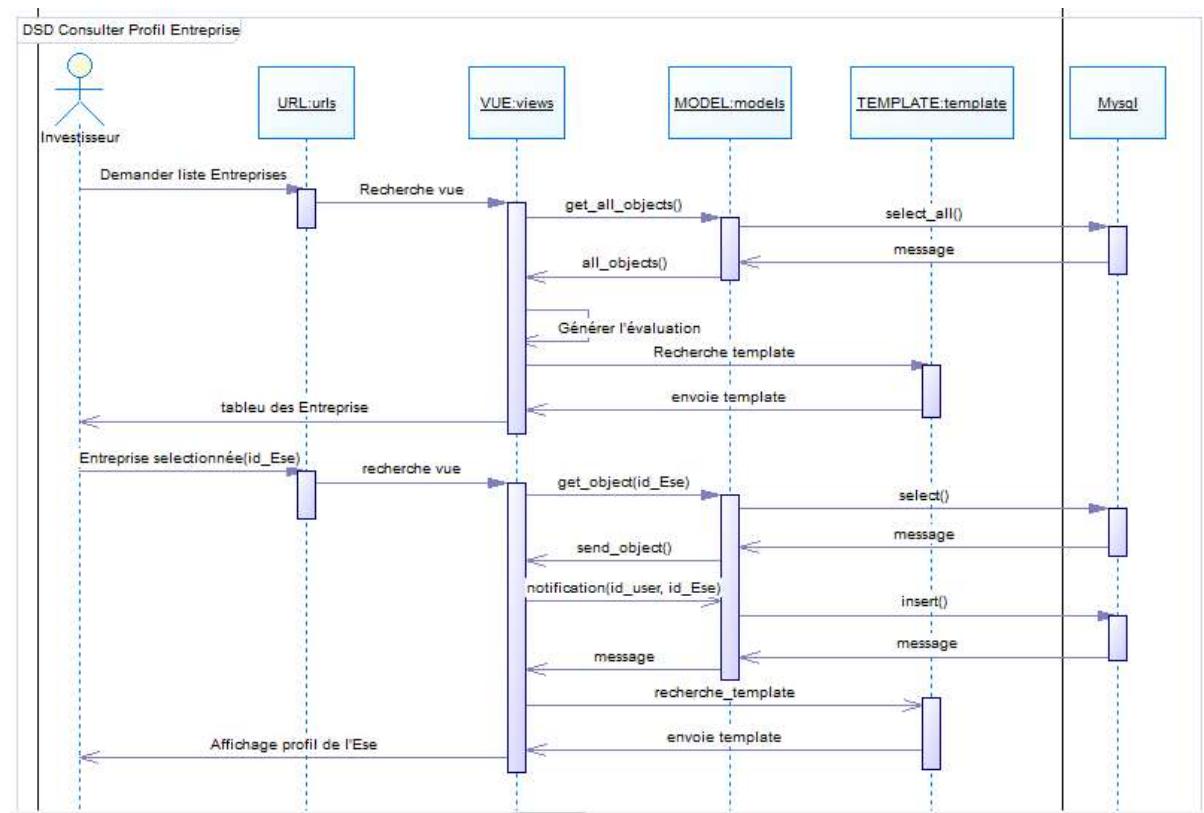


Figure 40 DSD du cas d'utilisation « Consulter profil entreprise »

➤ **DSD du cas d'utilisation « Désactiver profil entreprise »**

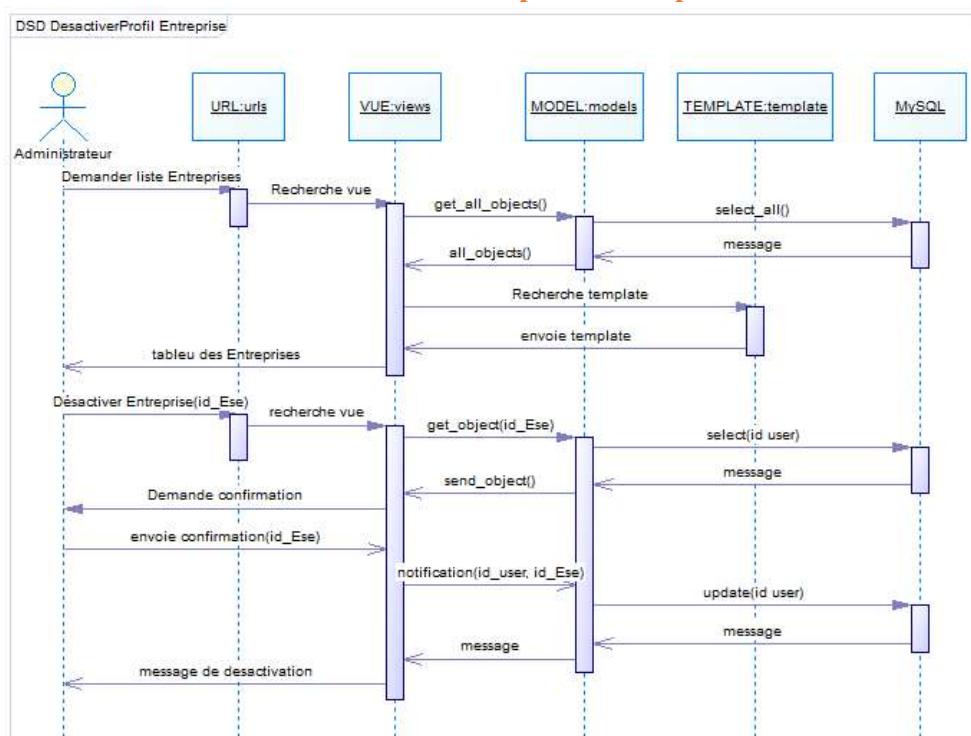


Figure 41 DSD du cas d'utilisation « Désactiver profil entreprise »

VII.2.7 Codage

Nous présentons les captures de l'implémentation de certaines classes.

```

class Entreprise(models.Model):
    entrepreneur = models.ForeignKey(Utilisateur, null=True, on_delete=models.CASCADE)
    code = models.CharField(max_length=10, null=False, blank=False)
    denomination = models.CharField(max_length=100, null=False, blank=False, unique=True)
    description = models.CharField(max_length=2000, null=False, blank=False)
    capital_social = models.BigIntegerField(blank=False, null=False)
    forme_juridique = models.CharField(max_length=2000, null=False, blank=False)
    date_creation_juridique = models.CharField(max_length=100, null=True, blank=True)
    manager = models.CharField(max_length=2000, null=False, blank=False)
    pays = models.CharField(max_length=100, null=True, blank=True, default='Gabon')
    secteur = models.CharField(max_length=100, null=False, blank=False)

    CATEGORIES_TYPES = (
        ("Microentreprise", "Petite & moyenne entreprise"),
        ("Entreprise de taille intermédiaire", "Entreprise de taille intermédiaire"),
        ("Grande Entreprise", "Grande Entreprise"),
    )
    categorie = models.CharField(max_length=50, choices=CATEGORIES_TYPES, blank=True, default="Microentreprise")
    localisation = models.CharField(max_length=100, null=False, blank=False)
    adresse = models.CharField(max_length=100, null=False, blank=False)
    telephone = models.CharField(max_length=100, null=False, blank=False, unique=True)
    numero_registre_commerce = models.CharField(max_length=100, null=False, blank=False)

    date_creation = models.DateTimeField(default=timezone.now(), blank=True)
    date_last_modification = models.DateField(null=True, blank=True)
    date_desactivation = models.DateField(null=True, blank=True)
    statut = models.BooleanField(default=False)

Entrepri

```

Figure 42 Implémentation de la classe « Entreprise »

```

class Profil_entreprise(models.Model):
    entreprise = models.ForeignKey(Entreprise, null=True, on_delete=models.CASCADE)
    code = models.CharField(max_length=10, null=False, blank=False)
    nombre_personnel = models.IntegerField(blank=True, null=True)
    chiffre_affaire = models.BigIntegerField(blank=True, null=True)
    depense_annuelle = models.IntegerField(blank=True, null=True)
    tresorerie_disponible = models.IntegerField(blank=True, null=True)
    taux_rentabilite = models.IntegerField(blank=True, null=True)
    taux rendement_actifs = models.IntegerField(blank=True, null=True)
    besoin_FDR = models.IntegerField(blank=True, null=True)
    cycle_produits = models.IntegerField(blank=True, null=True)
    delais_reglement_aclient = models.IntegerField(blank=True, null=True)
    delais_reglement_fournisseur = models.IntegerField(blank=True, null=True)
    nombre_client_payant = models.IntegerField(blank=True, null=True)
    taux_transformation_client = models.IntegerField(blank=True, null=True)
    taux_absenteisme_personnel = models.IntegerField(blank=True, null=True)
    taux_accident = models.IntegerField(blank=True, null=True)
    cout_production = models.IntegerField(blank=True, null=True)
    nombre_concurrent = models.IntegerField(blank=True, null=True)
    nombre_produits = models.IntegerField(blank=True, null=True)
    nombre_sites = models.IntegerField(blank=True, null=True)

    date_creation = models.DateTimeField(default=timezone.now(), blank=True)
    date_last_modification = models.DateField(null=True, blank=True)
    date_desactivation = models.DateField(null=True, blank=True)
    statut = models.BooleanField(default=True)

Entrepri

```

Figure 43 Implémentation de la classe « Profil_Entreprise »

```
class Consulter(models.Model):
    investisseur = models.ForeignKey(Utilisateur, null=True, on_delete=models.CASCADE)
    profil_entreprise = models.ForeignKey(Profil_entreprise, null=True, on_delete=models.CASCADE)
    date = models.DateTimeField(default=timezone.now(), blank=True)

class Message(models.Model):
    investisseur = models.ForeignKey(Utilisateur, null=True, on_delete=models.CASCADE)
    profil_entreprise = models.ForeignKey(Profil_entreprise, null=True, on_delete=models.CASCADE)
    message = models.TextField(null=False, blank=False)
    date_envoie = models.DateTimeField(default=timezone.now(), blank=True)
    date_suppression = models.DateField(null=True, blank=True)
    statut = models.BooleanField(default=True)
```

Figure 44 Implémentation des classes « Consulter & Message »

Conclusion

Avec ce chapitre, nous mettons un terme à cette partie dans laquelle nous avons choisi la méthode de développement et procéder à l'analyse des besoins et à la mise au point des différents modèles de notre système. La partie suivante sera consacrée à la présentation de la solution obtenue et le suivi du projet.

QUATRIEME PARTIE : FINALISATION DE LA SOLUTION

*Cette partie sera principalement réservée à la présentation dans son entiereté de **iPi9 Nexus**, notre solution ainsi que les éléments nécessaires à son déploiement. Nous terminerons cette partie avec la présentation de la conduite de projet.*

Chapitres

CHAPITRE VIII : DÉPLOIEMENT ET RÉSULTATS
CHAPITRE IX : CONDUITE DE PROJET

Chapitre VIII : DÉPLOIEMENT ET RÉSULTATS

Introduction

Dans ce chapitre, il sera essentiellement question de présenter les résultats de nos travaux. Il s'agira ici du rendu des différentes intégrations du produit logiciel construit avec le modèle prédictif développé.

VIII.1 Diagramme de déploiement

Il décrit la disposition physique des ressources matérielles qui composent le système et montre la répartition des composants sur ces matériels. Chaque ressource étant matérialisée par un nœud, le diagramme de déploiement précise comment les composants sont répartis sur les nœuds et quelles sont les connexions entre les composants ou les nœuds.

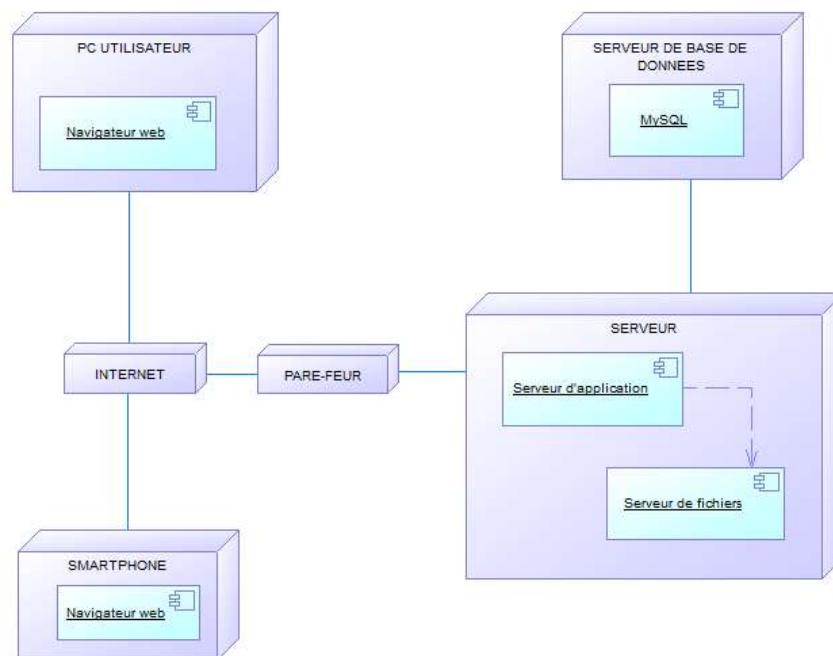


Figure 45 Diagramme de déploiement iPi9 Nexus

VIII.2 Résultats obtenus



Figure 46 Page d'accueil de la plateforme iPi9 Nexus

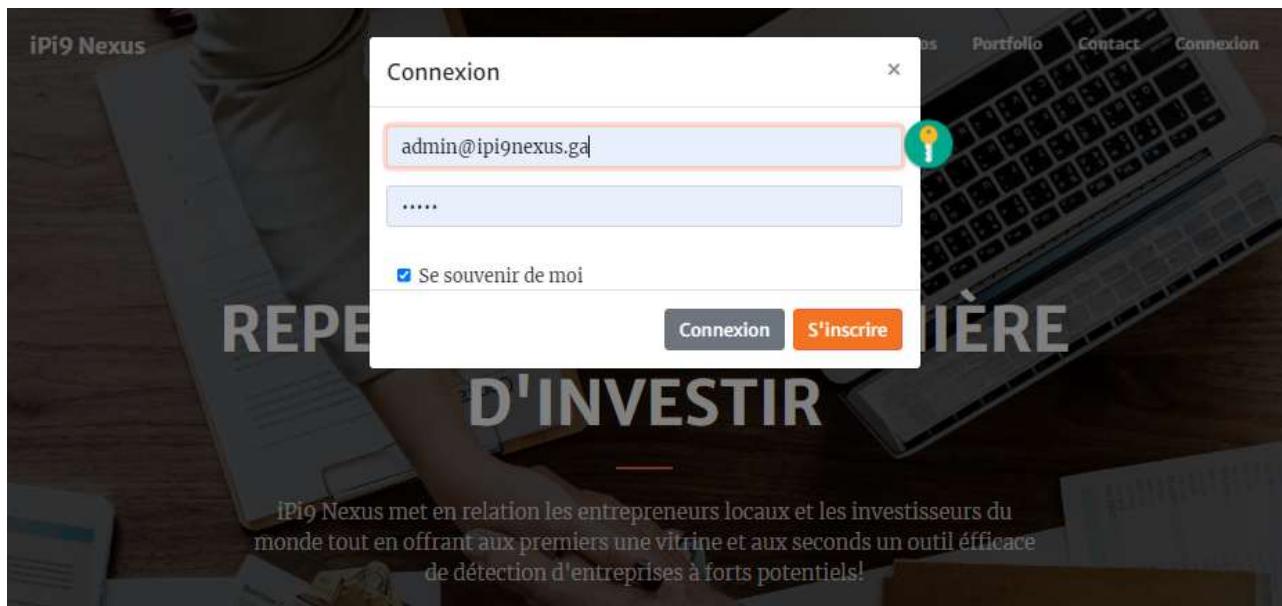


Figure 47 Formulaire de connexion

iPi9 Nexus

DASHBOARD

Accueil

MENU

- Mon Compte
- Leyée de fonds
- Messagerie
- Statistiques

Nexus

Connecté en tant que:
Abol (Entrepreneur)

Rechercher...

ESPACE ENTREPRENEUR

Home > Dashboard

Profil de l'entreprise
Enregistrer ou consulter les informations relatives à votre entreprise.
[Voir les Détails](#)

Messages (2)
Consulter les messages reçus.
[Voir les Détails](#)

Statistiques
Visualiser une courbe graphique du rapport de global de votre entreprise.
[Voir les Détails](#)

Alertes (48)
Toutes les informations relatives aux consultations de votre profil entreprise.
[Voir les Détails](#)

Copyright © iPi9 Nexus - IAI 2020

Politique de confidentialité - Conditions d'utilisation

Figure 48 Dashboard Entrepreneur

DASHBOARD

Accueil

MENU

- Mon Compte
- Messagerie
- Entreprises

Nexus

Connecté en tant que:
Danger (Investisseur)

Rechercher...

ESPACE INVESTISSEUR

Home > Dashboard

Informations disponibles

Les entreprises enregistrées à ce jour.

Profils entreprises
2
[Voir les Détails](#)

Pays de soumission
1
[Voir les Détails](#)

Secteurs d'activités
2
[Voir les Détails](#)

Levée de fonds en cours
1
[Voir les Détails](#)

Copyright © iPi9 Nexus - IAI 2020

Politique de confidentialité - Conditions d'utilisation

Figure 49 Dashboard Investisseur

CHAPITRE IX : CONDUITE DE PROJET

Introduction

Ce chapitre sera principalement dédié à la présentation de l'organisation mise en place pour mener à terme notre projet de stage. Notamment, son découpage en phases et tâches, les différents intervenants au projet, sa planification dans le temps et le rendu de l'estimation des charges globales du projet.

IX.1 Organisation du projet

IX.1.1 Découpage du projet en phases & taches

Découper un projet en phases, permet de pouvoir mieux le conduire à terme en respectant les impératifs de qualité, de coût et de délai. Chaque phase est découpée en tâches et accompagnée d'un bilan de fin d'étape destinée à la validation de la phase écoulée avant de passer à la phase suivante. L'ensemble des phases et des tâches de notre projet et le découpage effectué sont présentés dans le tableau ci-dessous.

Phases	Tâches
Etude des concepts du domaine auxquels se rapporte le sujet.	1- Revue littéraire sur la Data Science 2- Revue littéraire sur le Machine Learning et ses algorithmes 3- Revue littéraire sur les KPI d'une Entreprise 4- Revue littéraire sur l'évaluation d'Entreprise 5- Apprentissage du langage Python et Django 6- Prise en main des outils de Machine Learning
Etude du projet	1- Recensement des attentes du projet 2- Définition des objectifs du projet
Etat de l'art sur l'évaluation d'Entreprise	Etude du Business Case et de certaines solutions existantes.
Construction d'un modèle prédictif de ML	1- Construction du dataset 2- Construction du modèle 3- Evaluation du modèle

Analyse et conception d'une plateforme web	1- Choix méthodologique
	2- Etude préalable
	3- Analyse des besoins & fonctionnalités
	4- Découpage en itérations
Réalisation	1- Modélisation métier des itérations
	2- Implémentation
	3- Tests unitaires
Test Global	1- Intégration du modèle prédictif
	2- Mise en œuvre des tests
	3- Interprétation des résultats
Finalisation	1- Bilan & perspectives
	2- Finalisation du mémoire
	3- Présentation du projet à l'équipe « DSS iPi9 »
	4- Présentation du projet aux Maitres de stages

Tableau 8: Phases et Tâches du Projet

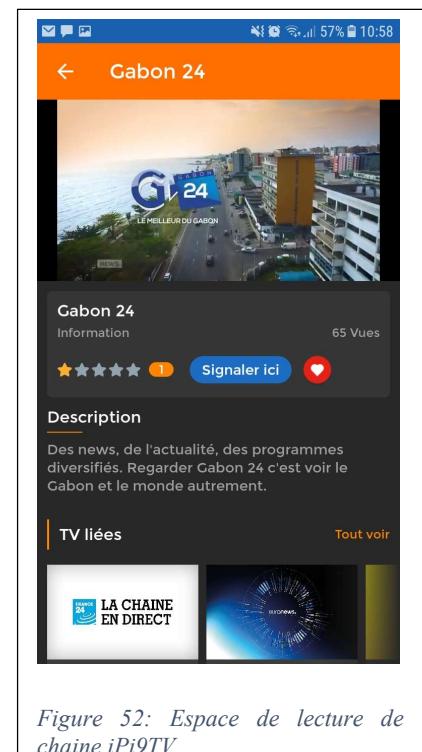
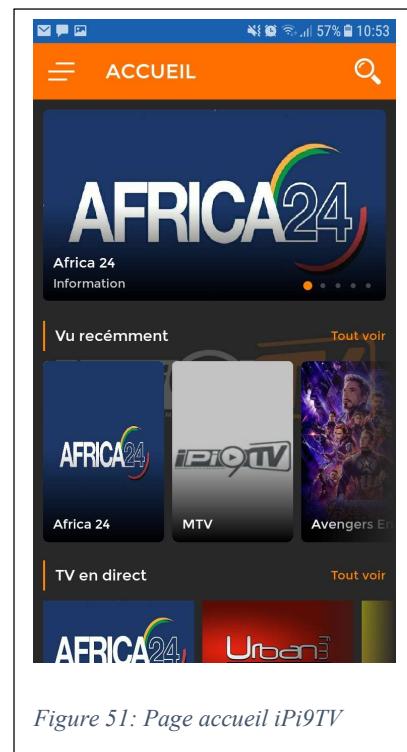
IX.1.2 Intervenants et déroulement du stage

Durant toute la période de notre stage, nous avons travaillé avec plusieurs personnes et solliciter l'appréciation et encadrement académique de différents enseignants, leurs interventions se fessaient sur différents domaines relatifs au sujet de stage. Les principaux intervenants sont de fait :

- Messieurs Serge SEGBEFIA Consultant iPi9 & Raz BIRAMAH Directeur Général iPi9, en qualité de Maitres de stage. Ils ont supervisé l'avancement du projet et organisé les sessions de travail.
- M. Donacien GUIFO, enseignant permanent à l'IAI, en qualité de Superviseur académique.
- Professeur Souleymane KOUSSOUBE, enseignant permanent à l'IAI, qui a suivi la modélisation.
- Dr. Roger NOUSSI, enseignant permanent à l'IAI, qui a suivi l'orientation globale du sujet.
- M. Maman MATY enseignant permanent à l'IAI, qui nous a aidé durant la prise en main des outils d'analyse de données et de Data Science.

- M. Nobert FOUDA, Expert en Analyse & Montage de projet, dont la mission a été de nous apporter d'amples explications sur les concepts d'évaluation d'Entreprise.

Nous avons par ailleurs au cours de ce stage travaillé sur le projet **iPi9 TV**, donc l'objectif est de fournir aux internautes un accès illimité aux chaînes TV et contenu VOD locaux via une application mobile. Nous avons principalement intervenu sur la conception et réalisation d'une nouvelle application mobile.



IX.1.3 Diagramme de GANTT

Le diagramme de Gantt nous permet de modéliser la planification des différentes tâches nécessaires à la réalisation de notre projet. Il permet ainsi de définir une feuille de route pour l'avancement de notre projet et une bonne communication entre les différents intervenants. Les diagrammes de Gantt prévisionnel et effectif sont présentés par la suite.

IX.1.3.1 Diagramme de Gantt prévisionnel du projet

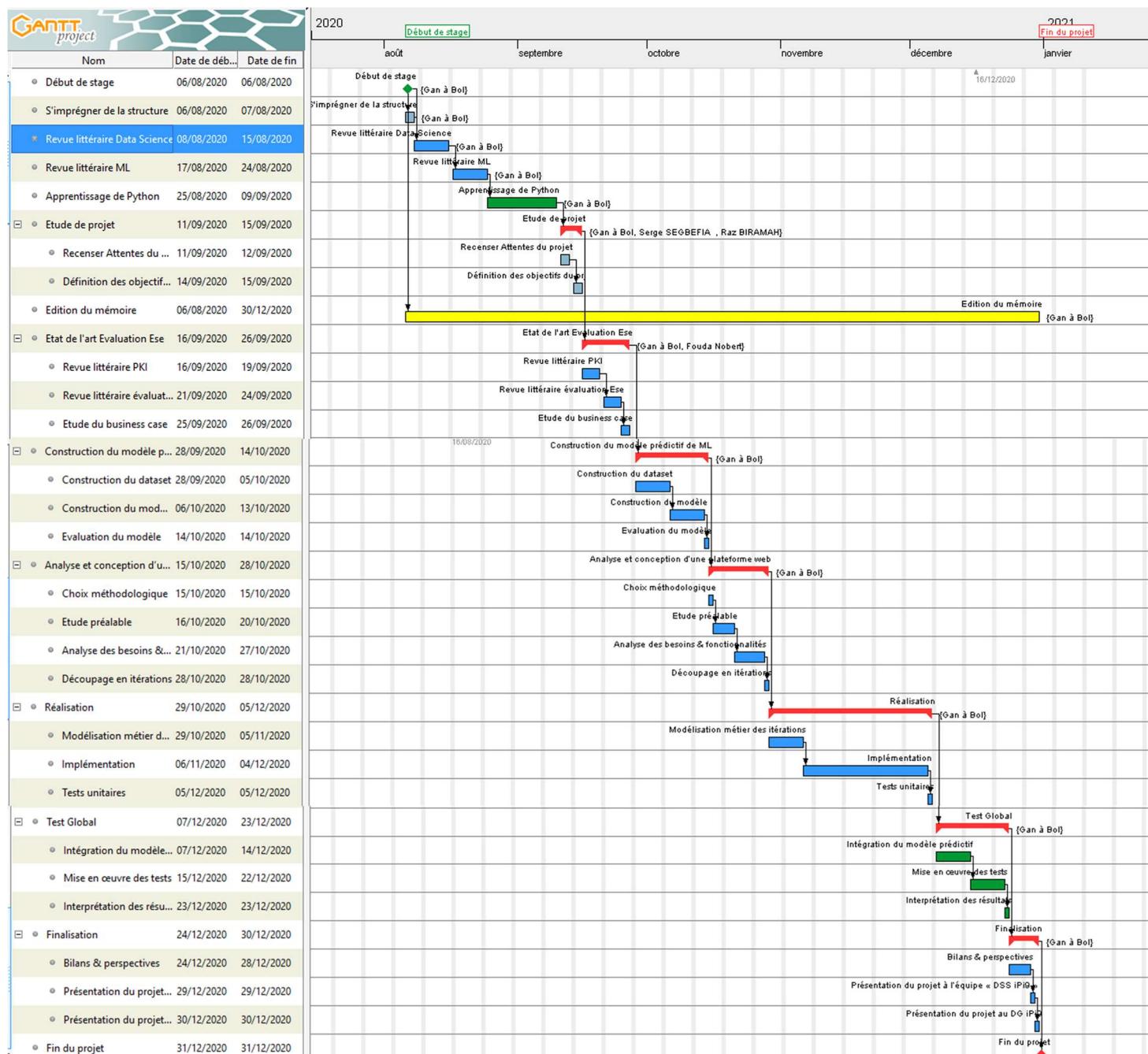


Figure 53: Diagramme de Gantt prévisionnel

IX.1.3.2 Diagramme de Gantt réel du projet

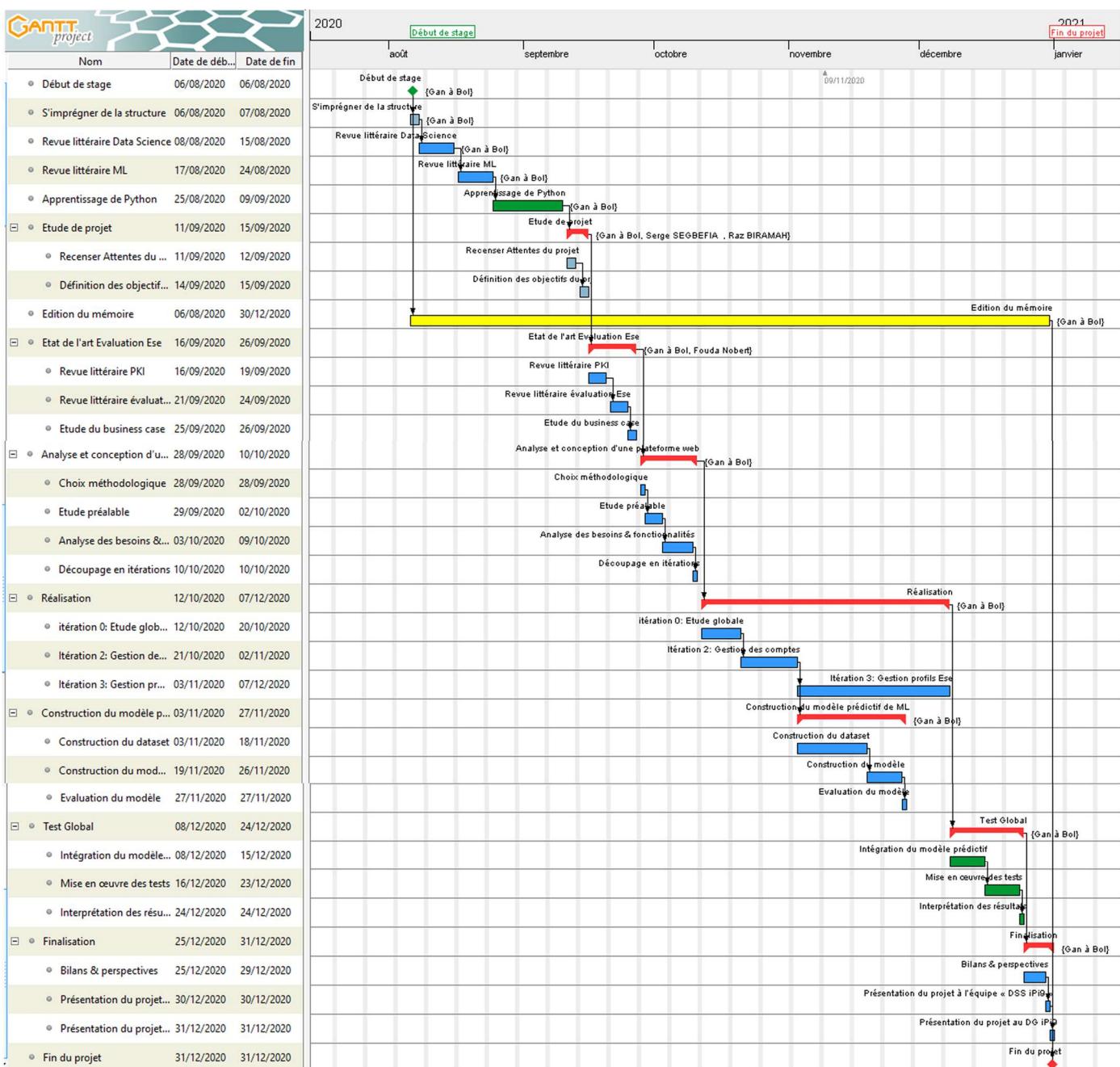


Figure 54: Diagramme de Gantt réel

IX.1.4 Estimation des charges et des coûts

Pour la réalisation de ce projet, des moyens ont été mis en œuvre, les tableaux ci-dessous en donnent le récapitulatif.

Rubrique	Quantité	Indication	Prix Unitaire	Prix total
Windows 10 OEM pro	01		209€	209 €
PowerAMC	01		5671 €	5671 €
Concepteur Studio Ese				
Anaconda	01	Gratuit	0 €	0 €
PyCharm Community Edition	01	Gratuit	0 €	0 €
Django	01	Gratuit	0 €	0 €
Total				5 880 €

Tableau 9: Coût logiciel du projet

Rubrique	Quantité	Indication	Prix Unitaire	Prix total
HP Probook 450 G6 i5-8265U	01		858 €	858 €
Total				858 €

Tableau 10: Coût matériel du projet

Rubrique	Durée du projet	Nombre de participant	Coût
Estimation de l'ensemble des activités	140 jours	01	140 jours-homme

Tableau 11: Coût humain du projet

IX.2 Bilan et perspectives

IX.2.1 Apports du stage et difficultés rencontrée

IX.2.1.1 Apports

Notre stage au sein de la structure **iPi9** a été positif à plus d'un titre. Car il nous a permis de travailler dans un environnement de télécommunication (fourniture d'accès à internet) et développement d'applications. **iPi9** est une structure adaptée pour asseoir les connaissances apprises pendant la formation, aussi il offre un cadre favorable pour effectuer un travail à la hauteur du titre d'ingénieur auquel nous prétendons.

Durant ces 05 mois de stage nous avons eu l'occasion de :

- ✓ Nous familiariser avec le monde professionnel ;
- ✓ Explorer l'univers nouveau de la Data Science et du Machine Learning ;
- ✓ Explorer l'univers du développement d'application web et mobile ainsi des bots ;
- ✓ Assister à la gestion des projets d'envergure et de nous rapprocher des experts du métier ;
- ✓ Dans le volet technique travailler auprès de personnes expérimentées nous a permis d'améliorer nos connaissances sur le langage de programmation python et la manipulation de bien d'autres outils et de techniques de programmation.

IX.2.1.2 Difficultés

Après notre phase chaleureuse d'insertion dans la structure, nous avons dès le début fait face une difficulté, ceci particulièrement dû au fait que le domaine dans lequel s'inscrit notre travail est nouveau et encore en pleine expansion. Grâce aux conseils, orientations et mise à notre disposition de documents venant de nos encadreurs, nous avons pu réussir notre immersion dans la science de Data Science et Machine Learning.

Une autre entrave à notre projet et non des moindres à laquelle nous avons fait face est la difficulté à collecter des informations auprès des Entreprises après mise en ligne et soumission des questionnaires, information plus que importantes pour la réussite de notre travail.

IX.2.2 Perspectives

En perspectives à court termes il sera question pour nous d'approfondir l'apprentissage de notre modèle pour garantir un taux de fiabilité encore plus élevé de prédiction. Mais aussi pouvoir amener notre modèle à donner des prédictions plus précises sur le taux (valeur numérique), de viabilité des Entreprises. Nous prévoyons tout aussi d'intégrer un module d'envoi de rapport complet d'évaluation de la viabilité d'une entreprise donnée à la demande d'un investisseur.

Conclusion

Dans ce chapitre, nous avons présenté les différentes phases et tâches de notre projet, c'est ce découpage en phases et en tâches qui nous a permis de mener à bien ce travail que nous avons par la suite résumé dans le temps à travers des diagrammes de Gantt prévisionnel et réel, avant de présenter par la suite les apports de ce projet et les perspectives.

CONCLUSION GÉNÉRALE

Arrivé au terme de notre travail, où nous avions pour mission de concevoir et développer une plateforme de Machine Learning de mise en relation des entrepreneurs locaux désireux de nouer des partenariats et des investisseurs du monde à la recherche d'opportunités nouvelles.

Après une présentation du contexte dans lequel s'inscrit notre sujet, nous avons présenté les concepts clés nécessaires tout aussi bien à la compréhension du sujet, mais aussi à sa résolution.

Par la suite nous nous sommes appesantis sur la présentation du Machine Learning, et ses algorithmes, où nous avons pu identifier la classe d'algorithmes pouvant être utilisée pour résoudre notre problématique. Ayant fait face à la difficulté de collecte des informations sur les Entreprises, nous avons pour les besoins du mémoire généré et traité des données tout en se rassurant qu'elles reflètent la réalité enfin d'entrainer notre modèle. Après cette phase, nous nous sommes attelés à concevoir la plateforme proprement dite en se basant sur la méthodologie 2TUP. Le Framework utilisé est Django et le langage de programmation principal est python.

En perspective, nous prévoyons accentuer la collecte des informations auprès des Entreprises, ceci dans l'optique d'enrichir notre dataset et mieux entraîner notre modèle. Aussi, nous prévoyons réaliser des prédictions en présentant un taux (valeur numérique), de viabilité de l'Entreprise évaluée.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [O1] Pascal Roques, « UML 2 par la pratique : Études de cas et exercices corrigés », 6e édition. Avril 2008.
- [O2] Pascal Roques, « UML 2 : Modéliser une application web, 2e édition ». Juin 2008.
- [O3] Pascal Roques, Franck Vallée, « UML 2 en action : De l'analyse des besoins à la conception », 4e édition. Février 2007.
- [O4] Guillaume Saint-Cirgue, « Apprendre le Machine Learning en une semaine ». 2019
- [O5] Vincent Bouchet, « Machine learning en finance : vers de nouvelles stratégies ? ». 2017

WEBOGRAPHIE

- [W1] <https://openclassrooms.com/fr/courses/2304731-learn-python-basics-for-data-analysis>, le 09-08-2020 : Apprentissage de langage Python orienté analyse des données
- [W2] <https://openclassrooms.com/fr/courses/4011851-initiez-vous-au-machine-learning>, le 05-08-20 : Initiation au Machine Learning
- [W3] <https://openclassrooms.com/fr/courses/4470406-utilisez-des-modeles-supervises-non-lineaires>, le 07 -08-2020 : Utilisation des modèles supervisés
- [W4] <https://www.ionos.fr/digitalguide/web-marketing/search-engine-marketing/deep-learning-vs-machinelearning/#:~:text=Tandis%20que%20le%20Machine%20learning,pour%20donner%20des%20r%C3%A9sultats%20fiables>, le 12-09-2020 : Deep learning Vs Machine Learning
- [W5] <https://mrmint.fr/aborder-probleme-de-machine-learning-parte-1>, le 05-09-2020 : Aborder un problème de Machine Learning
- [W6] [https://www.lebigdata.fr/confusion-matrix-definition#:~:text=Une%20Confusion%20Matrix%20\(matrice%20de,dans%20des%20probl%C3%A8mes%20de%20classification](https://www.lebigdata.fr/confusion-matrix-definition#:~:text=Une%20Confusion%20Matrix%20(matrice%20de,dans%20des%20probl%C3%A8mes%20de%20classification), le 07-12-2020 : Matrix de confusion
- [W7] <https://www.appvizer.fr/magazine/relation-client/customer-relationship-management-crm/indicateurs-performance-commerciale>, le 30-08-2020 : les PKI

- [W8] https://fr.wikipedia.org/wiki/Business_case, le 32-08-2020 : Business case
- [W9] <https://www.manager-go.com/gestion-de-projet/business-case.htm>, le 32-08-2020 : Business case
- [W10] <https://www.ginkonote.com/fr/utilisateurs/flo/articles/machine-learning-choix-des-modeles-et-donnees@intelligence-artificielle>, le 06-09-2020 : ML-Choix de modèles
- [W11] <https://fr.slideshare.net/preemptif/mthodologie-2-track-unified-process>, le 02-10-2020 : 2-TUP
- [W12] <https://mrmint.fr/introduction-k-nearest-neighbors>, le 05-10-2020 : Introduction au KNN
- [W13] https://www.tutorialspoint.com/scikit_learn/scikit_learn_knn_learning.htm, le 06-10-2020 : KNN
- [W14] <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>, le 06-10-2020 : le KNN
- [W15] <https://simpleisbetterthancomplex.com/tutorial/2016/11/15/how-to-implement-a-crud-using-ajax-and-json.html>, le 15-05-2020 : CRUD Django
- [W16] <https://docs.djangoproject.com/en/1.8/intro/tutorial01/>, le 06-05-2020 : Installation django
- [W17] <https://django-book.readthedocs.io/en/latest/chapter01.html>, le 06-05-2020 : Django
- [W18] <https://www.planetoscope.com/Internet-/1547-nombre-de-tweets-expedies-sur-twitter.html#:~:text=Chaque%20seconde%20environ%205%20900,humanit%C3%A9%20chaque%20jour%20sur%20internet>, le 15-12-2020 : Big Data

ANNEXES

Annexe 1: Questionnaire sur les informations des Entreprises

Rubrique 1 sur 6

Questionnaire sur le profil des Entreprises.

Afin de permettre aux entreprises locales de lever des fonds et aux investisseurs de réaliser des placements bénéfiques sur des entreprises à fort potentiel, ce questionnaire vous est soumis pour collecte des informations nécessaires à la mise en place de la plate forme qui mettra en relation les entrepreneurs locaux et les investisseurs du monde entier.

Après la section 1 Passer à la section suivante ▾

Rubrique 2 sur 6

Identification de la structure

Description (facultative)

N° Régistre commerce *

Lien : https://docs.google.com/forms/d/e/1FAIpQLSf9EFXjZddH-I39Q4kEeWUpk7YZC-LaSxLgvv8In80jxAEDCg/viewform?usp=sf_link