

République Algérienne Démocratique et Populaire
Ministre de L'Enseignement Supérieur et de la Recherche Scientifique



Université Abderrahmane MIRA – BEJAIA

Faculté des sciences Exactes

Département d'Informatique

Spécialité : GENIE LOGICIEL

THEME

Conception et réalisation d'un système d'information d'aide à la décision pour les prédictions de ventes (Cas : SARL IFRI)

Devant un jury composé de :

Président : M. MIR Foudel

Examineur : M^{elle} HOUHA Amel

Encadrant : M. ACHROUFENE Achour

Réalisé par :

M^{elle} AIROUCHE Cylia

M^{elle} LOUBAR Nacera

Année universitaire 2019/2020

Dédicaces

*Je dédie ce modeste travail aux personnes chères à mon cœur,
À mes parents, qui ont su m'inculquer une éducation de sagesse et de claire voyance.
À mon frère Arezki.
À mes sœurs Rekia, Sonia et kafia.
À ma tante Khadidja.
À mes amis : Nacera, Djoudjou, Hanane, Fatiha, Nawel, Hamid et Lyes.*

Cylia

*Je dédie ce modeste travail aux personnes chères à mon cœur,
À mes parents, qui ont su m'inculquer une éducation de sagesse et de claire voyance.
À mes trois frères Mohamed, Hamid, Kamel et leurs femmes.
À mes sœurs Djedjiga, Samia et leurs maris.
À mes cousines Karima et Nawel.
À mes amis : Cylia, Djoudjou, Hanane, Fatiha, Nawel, Lyes et Hamid.*

Nacera

Remerciements

Au terme de notre travail, on remercie Dieu tout puissant de nous avoir donné le courage et la patience pour réaliser ce modeste travail.

La réalisation de ce mémoire a été un parcours jalonné de nombreuses rencontres, sans lesquelles ce travail n'aurait pas pu aboutir. On n'aurait pas éprouvé autant de plaisir à réaliser ce travail sans ces personnes, qui par leur générosité, leur disponibilité, leur bonne humeur et l'intérêt manifesté à l'égard de notre recherche, ont grandement contribué à l'amélioration de notre travail.

On tient particulièrement à adresser nos remerciements d'abord à notre encadrant monsieur ACHROUFENE Achour, pour nous avoir orienté durant l'élaboration de ce travail, Il a toujours été disponible à l'écoute de nos nombreuses questions.

Notre reconnaissance et notre estime sont également portés à Mr. GUIMAOUI Sofiane et tout le personnel du service informatique qui nous ont bien assistés et nous ont donné leurs conseils précieux.

On tient à remercier les membres du jury qui nous font le grand honneur d'évaluer ce travail.

Enfin, nos remerciements s'adressent à toutes nos familles qui sont toujours là pour nous soutenir. Et les personnes qui ont contribué de près ou de loin à la réalisation de ce modeste travail.

AIROUCHE Cylia

LOUBAR Nacera

Table des matières

Dédicaces	i
Remerciements	ii
Liste des abréviations	ix
Introduction	1
1 Organisme d'accueil & prédiction de ventes	3
1.1 Introduction	3
1.2 Présentation de l'organisme d'accueil	3
1.2.1 Création et évolution de la SARL « IBRAHIM&FILS »	3
1.2.2 Missions d'IFRI	4
1.2.3 Répartition géographique	4
1.2.4 Activités et filières d'IFRI	4
1.2.5 Cadre juridique de la SARL « IFRI »	5
1.2.6 Organisation de la SARL « IBRAHIM&FILS »	5
1.2.7 Présence d'IFRI	7
1.3 Problématique	7
1.4 Marché	7
1.4.1 Étude du marché	8
1.4.2 Approche d'analyse des ventes	8
1.4.3 Prévision des ventes	9
1.5 Solution proposée et objectifs	9
1.6 Prédiction	10
1.6.1 Analyse prédictive	10
1.6.2 Utilisations métier de l'analyse prédictive	11
1.7 Conclusion	12
2 Systèmes d'aide à la décision	13
2.1 Introduction	13
2.2 Décision	13
2.3 Processus de décision	13
2.4 Systèmes Interactifs d'Aide à la Décision	15
2.4.1 Définition d'un SIAD	15
2.4.2 Composants d'un SIAD	16
2.5 Machine Learning	17
2.5.1 Types d'apprentissage automatique	17
2.6 Business Intelligence	19
2.6.1 Collecter, nettoyer et consolider les données	19
2.6.2 Stocker	20
2.6.3 Distribuer	20

2.6.4	Exploiter	21
2.7	Modélisation d'un entrepôt de données	22
2.7.1	Fait	22
2.7.2	Dimension	23
2.7.3	Types des schémas	23
2.8	Démarche de développement de SIAD	25
2.9	Conclusion	27
3	Analyse des besoins et des exigences	28
3.1	Introduction	28
3.2	Délimitation du domaine d'étude	28
3.2.1	Identification des acteurs	28
3.2.2	Diagramme de contexte	29
3.3	Analyse des situations décisionnelles	29
3.4	Identification des cas d'utilisation	30
3.5	Besoins fonctionnels	31
3.6	Besoins non fonctionnels	32
3.7	Élaboration des premières maquettes d'IHM	32
3.8	Diagramme de cas d'utilisation	33
3.9	Description des cas d'utilisation	35
3.9.1	Cas d'utilisation « S'authentifier »	35
3.9.2	Cas d'utilisation « Gérer les ventes »	37
3.9.3	Cas d'utilisation « Prédire les ventes »	38
3.9.4	Cas d'utilisation « Rechercher »	40
3.10	Conclusion	42
4	Conception	43
4.1	Introduction	43
4.2	Diagramme d'interaction	43
4.2.1	S'authentifier	43
4.2.2	Gérer les ventes	44
4.2.3	Prédire les ventes	45
4.2.4	Rechercher	46
4.3	Diagramme de navigation	47
4.4	Diagramme de classes	49
4.5	Conception de l'entrepôt de données	52
4.5.1	Dimensions participantes du modèle	52
4.5.2	Faits ventes	54
4.5.3	Modèle en flocon	54
4.6	Conclusion	56
5	Réalisation	57
5.1	Introduction	57
5.2	Environnement de développement de l'application	57
5.2.1	Environnement matériel	57
5.2.2	Environnement logiciel	57
5.3	ETL	61
5.4	Modélisation de la prédiction de vente	62

5.5	Présentation de l'application	64
5.5.1	Application web	64
5.5.2	Quelques interfaces graphiques	64
5.6	Tests	67
5.7	Conclusion	67
	Conclusion et perspectives	68
	Bibliographie	70

Table des figures

1.1	Caption for LOF	5
1.2	Organigramme de la SARL IFRI.	6
1.3	Processus de l'analyse prédictive [5].	11
2.1	Modèle de processus de décision de Simon [7].	14
2.2	Types d'apprentissage machine.	18
2.3	Extraction, Transformation et Chargement de données au niveau des outils ETL [23].	20
2.4	Les quatre composants de l'informatique décisionnelle [25].	22
2.5	Exemple de table faits.	22
2.6	Exemple de table dimension.	23
2.7	Exemple de schéma en étoile.	23
2.8	Exemple de schéma en flocon.	24
2.9	Exemple de schéma en constellation.	24
3.1	Diagramme de contexte.	29
3.2	Maquette de l'interface d'analyse.	33
3.3	Maquette de l'interface de prédiction.	33
3.4	Diagramme de cas d'utilisation.	34
3.5	Diagramme de séquence système du cas d'utilisation « S'authentifier ».	37
3.6	Diagramme de séquence système du cas d'utilisation « Gérer les ventes ».	38
3.7	Diagramme de séquence système du cas d'utilisation « Prédire les ventes ».	40
3.8	Diagramme de séquence système du cas d'utilisation « Rechercher ».	42
4.1	Diagramme d'interaction du cas d'utilisation « S'authentifier ».	44
4.2	Diagramme d'interaction du cas d'utilisation « Gérer les ventes ».	45
4.3	Diagramme d'interaction du cas d'utilisation « Prédire les ventes ».	46
4.4	Diagramme d'interaction du cas d'utilisation « Rechercher ».	47
4.5	Diagramme de navigation.	48
4.6	Diagramme de classes.	50
4.7	Dimension date.	52
4.8	Dimension produit.	52
4.9	Dimension catégorie.	53
4.10	Dimension client.	53
4.11	Dimension région.	53
4.12	Dimension température.	54
4.13	Faits ventes.	54
4.14	Modèle en flocon de notre entrepôt de données.	55
5.1	Job de collection de données sur TOS.	61

5.2	Application des contraintes d'intégrité pour la tables Faits_ventes. . .	62
5.3	Nuage de points de la quantité mensuel vendue en fonction de la température.	63
5.4	Droite de régression correspondant à la modélisation du nuage de points.	63
5.5	Interface d'authentification.	65
5.6	Interface d'analyses.	65
5.7	Interface de gestion de ventes.	66
5.8	Interface de prédiction des ventes.	66

Liste des tableaux

2.1	Comparaison entre les méthodes de développement dans différents domaines.	26
3.1	Messages échangés entre le système et le décisionnaire.	29
3.2	Analyse des situations décisionnelles.	30
3.3	Les cas d'utilisation du système.	31
3.4	Formalisme de description des cas d'utilisation.	35
3.5	Description du cas d'utilisation « S'authentifier ».	36
3.6	Description du cas d'utilisation « Gérer les ventes ».	37
3.7	Description du cas d'utilisation « Prédire les ventes ».	39
3.8	Description du cas d'utilisation « Rechercher ».	41
4.1	Dictionnaire de données.	51

Liste des abréviations

ADESIAD	Approche de DEveloppement de Systèmes Interactifs d’Aide à la Décision
API	Application Programming Interface
BDD	Base De Données
BI	Business Intelligence ou l’informatique décisionnelle
BL	Béjaïa Logistique
CSS	Cascading Style Sheets
CSV	Comma-separated values
DBA	DataBase Administrator
DSS	diagrammes de séquence système
EIP	Entreprise Information Portal
ETL	Extract, Transform and Load
GP	Général PLAST
GUI	Graphical User Interface
HTML	Hyper Text Markup Language
HTTP	Hypertext Transfer Protocol
IA	Intelligence artificielle
IDE	Integrated Development Environment
IHM	Interfaces Homme/ Machine
ISC	Ingénierie des systèmes de connaissances
ISO	International Organization for Standardization
JDBC	Java DataBase Connectivity
JDK	Java Développment Kit
MVC	Modèle-Vue-Contrôleur
MYSQL	My Structured Query Language
OLAP	Online Analytical Processing
PET	Polyéthylène téréphtalate
PEHD	Polyéthylène Haute Densité
PHP	Hypertext Preprocessor
REST	Representational State Transfer
RH	Ressource Humaines
SAD	Systèmes d’aide à la décision
SADG	Systèmes de décision de groupe
SARL	Société à Responsabilité Limitée

SGBC	Systeme de Gestion de Bases de connaissance
SGBD	Systeme de Gestion de Bases de Données
SGBDR	Systeme de Gestion de Bases de Données Relationnelle
SGBM	Systeme de Gestion de Bases de Modèles
SIAD	Systemes interactifs d'aide à la décision
SID	Systemes d'information pour dirigeants
SIG	Systemes d'information de gestion
SNC	Société Non Collectif
SQL	Structured Query Language
UML	Unified Modeling Language
URI	Uniform Resource Identifier
WSGI	Web Server Gateway Interface
XAMPP	Cross-Platform, Apache, MySQL, PHP et Perl
XML	Extensible Markup Language

Introduction

Cadre général

De nos jours, il existe une forte informatisation couplée à l'intelligence artificielle dans de nombreux secteurs, c'est pour cela que l'informatique ne cesse d'évoluer et de gagner du terrain. Ainsi, elle a pris une importance considérable dans l'entreprise, elle n'est plus seulement l'un des instruments de productivité, mais elle est devenue un outil de gestion et de pilotage de l'entreprise, car cette dernière a trouvé une solution à ces multiples problèmes liés notamment à la communication, à l'enregistrement des données, aux différents calculs comptables et surtout à la prise de décision.

Chaque jour, les managers font en permanence des choix, puisque sans prise de décision l'entreprise ne peut pas fonctionner. L'un des problèmes majeurs de décision que les entreprises rencontrent est l'analyse du marché qui passe d'abord par l'étude de ce dernier afin de collecter des données et d'avoir une idée sur l'évolution du marché. Cela permet certes d'améliorer la technique de vente des produits existants en élaborant de nouvelles stratégies de vente pour l'entreprise. Cependant, ce processus nécessite une collecte de données avec les changements fréquents du marché ce qui demande plus d'effort et beaucoup de temps. Dans le but d'exploiter les différentes sources de données qui influencent l'évolution du marché et de réduire les efforts consentis pour l'étude du marché, le développement des systèmes décisionnels s'avère plus rentable, ces systèmes sont devenus plus performants grâce à l'évolution de l'intelligence artificielle (IA) notamment avec la branche apprentissage automatique (Machine learning). En effet, les dirigeants d'entreprise s'orientent de plus en plus vers les techniques de business intelligence (BI) ou l'informatique décisionnelle qui aide à la prise de décision communément appelé SIAD (Système Interactif d'Aide à la Décision).

Objectifs

C'est dans ce contexte que la SARL IBRAHIM et fils (IFRI) nous a proposé de mettre en œuvre un système d'aide à la décision pour la prédiction des ventes selon le comportement du marché. Cette SARL ne dispose pas d'application de prédiction et se base sur des données stockées sous format excel pour avoir certaines statistiques, d'où des difficultés de prendre des décisions. Nous proposons de développer une application informatique d'aide à la décision (SIAD) pour la prédiction des ventes de la SARL IFRI. Nous allons premièrement mettre en œuvre un entrepôt de données qui va alimenter le modèle de prédiction proposé, ensuite construire ce modèle en utilisant l'apprentissage automatique, plus précisément l'apprentissage supervisé. L'intérêt de l'apprentissage automatique est d'améliorer les prédictions automatiquement avec l'expérience. L'application doit permettre de faire des prédictions sur les quantités des produits à vendre selon le mois, la température et la région. Nous allons également concevoir des tableaux de bord pour visualiser les données et pouvoir faire des analyses sur les ventes et les clients.

Organisation du mémoire

Ce mémoire est structuré comme suit :

Dans le **premier chapitre** nous présenterons l'organisme d'accueil, nous établirons la problématique suivie par la solution ainsi que les objectifs de l'application proposée. Nous étudierons également le processus de l'analyse prédictive.

Le **second chapitre** sera consacré pour l'étude du domaine de la décision, de la technique de Machine Learning, ainsi que pour la présentation de la démarche de développement adoptée.

Dans le **troisième chapitre**, nous passerons à la phase d'analyse des besoins et des exigences, où nous allons délimiter le champ d'étude, identifier les acteurs, les cas d'utilisation et analyser les situations décisionnelles. Par la suite, le diagramme des cas d'utilisation sera établi pour montrer les fonctionnalités du système. Ce chapitre se terminera par la description des cas d'utilisation les plus pertinents par les diagrammes de séquences.

Le **quatrième chapitre** portera l'aspect conceptuel de l'application. Les interactions entre les objets du systèmes seront modélisées par les diagrammes d'interactions. Un diagramme de navigation sera établi pour comprendre l'architecture de notre application. A la fin de ce chapitre nous allons présenter un diagramme de classes à partir duquel nous allons modéliser un entrepôt de données.

Le **cinquième** et dernier chapitre illustrera la phase de réalisation de l'application. Nous présenterons les environnements matériel et logiciel de développement utilisés. Ensuite, nous allons énumérer les étapes de la construction de l'entrepôt de données et celles du modèle Machine Learning. Enfin, nous allons expliquer l'architecture de l'application et son fonctionnement à l'aide des interfaces graphiques.

Le mémoire sera clos par une conclusion générale et la présentation de quelques perspectives.

Chapitre 1

Organisme d'accueil & prédiction de ventes

1.1 Introduction

Pour répondre aux différentes problématiques de l'entreprise, comme la prédiction de ventes, il faut étudier sa structure et suivre la circulation des informations entre ses divers services.

Dans ce chapitre nous présenterons tout d'abord l'entreprise d'accueil la SARL IBRAHIM&FILS « IFRI », qui est spécialisée dans la production des eaux minérales et de boissons diverses. Ensuite, nous donnerons une définition du marché. Puis, nous exposerons brièvement la problématique et nous proposerons une solution informatique d'aide à la décision pour la prédiction de ventes. Enfin, nous décrirons le processus d'analyse prédictive.

1.2 Présentation de l'organisme d'accueil

La SARL IBRAHIM&FILS « IFRI » est une société à caractère industriel, elle est spécialisée dans la production des eaux minérales et des boissons diverses, elle contribue au développement du secteur agroalimentaire à l'échelle national.

1.2.1 Création et évolution de la SARL « IBRAHIM&FILS »

Cette société est créée par les fonds propres de M. IBRAHIM Laid en 1986, elle était « LIMONADERIE IBRAHIM » spécialisée dans la production de boissons gazeuses en emballage verre¹.

Depuis la date de création, l'organisation a capitalisé une expérience dans le domaine des boissons ; ce n'est que dix ans plus tard, en 1996, que l'entreprise hérite un statut juridique de SNC (Société Non Collectif) puis le statut de la SARL (Société à Responsabilité Limitée) composée de plusieurs associées².

À cette dernière date, la marque « IFRI » est connue et exploitée dans le monde industriel, ça a été le point de départ de la première unité de fabrication d'eau minérale naturelle en Algérie sous un emballage en bouteilles en polyéthylène téréphtalate

1. Document de l'entreprise Ifri

2. Document de l'entreprise Ifri : "Forme juridique"

(PET). Plus de vingt (20) millions de bouteilles ont été commercialisées sur l'ensemble du territoire national dans la même année. Ce chiffre atteint 48 millions d'unités en 1999, puis 252 millions de litres en 2004. La production franchira le cap des 541378351 millions de litres dans toutes les gammes des produits IFRI en 2012³.

1.2.2 Missions d'IFRI

L'entreprise IFRI a pour mission essentielle la production et la commercialisation des produits agroalimentaires. IFRI est spécialisée dans la production d'eau minérale et de boissons diverses en emballage verre et PET.

La finalité de l'entreprise est d'être leader dans le domaine des eaux minérales tout en renforçant progressivement ses positions dans le segment des boissons diverses et de développer ses capacités à l'international.

1.2.3 Répartition géographique

La SARL IFRI est répartie sur deux sites qui sont⁴ :

1.2.3.1 Site IGHZER AMOKRANE

L'activité principal et la direction de la SARL IBRAHIM & Fils « IFRI » est située dans la commune d'IGHZER – AMOKRANE, Daira IFRI OUZELLAGUEN dans la wilaya de Béjaïa dans le nord de l'Algérie. Elle est implantée à l'entrée-Est de la vallée de la Soummam dans la zone « AHRİK IGHZER AMOKRANE », en contre bas du massif montagneux de Djurdjura qui constitue son réservoir naturel d'eau.

1.2.3.2 Site Zone activité TAHARACHT AKBOU

L'activité secondaire de production de JUS IFRUIT est implantée à la Zone TAHARACHT AKBOU sur un site de 20 HA destiné à recevoir les projets d'extension dans la gamme soda, jus etc.

1.2.4 Activités et filières d'IFRI

La société travaille 24/24 heures avec des lignes de production automatisées et équipées des systèmes de contrôle de qualité de dernière génération dans toutes les unités et étapes de la production. Grâce aux options technologiques qui ont prévalu lors du choix des équipements de production et de contrôle, IFRI accroît sans cesse ses capacités. Elle veille au respect des normes d'hygiène, de sécurité et environnementales et de qualité les plus strictes afin de diversifier sa gamme de production.

Le groupe IFRI a diversifié ses filières, il est composé de quatre (4) sociétés⁴ :

- IFRI : qui a pour mission de produire une gamme diversifiée de boissons (eau minérale naturelle, eau minérale gazéifiée, les sodas, les boissons fruitées, les boissons fruitées au lait) ;
- GENERAL PLAST : créée en 1999, l'entreprise GP (Général PLAST) s'est spécialisée dans la fabrication de la préforme en PET (Polyéthylène Téréphtalate) et bouchon en PEHD (Polyéthylène Haute Densité) ;

3. Document de l'entreprise Ifri : "Service commercial"

4. Document de l'entreprise Ifri

- BEJAIA LOGISQTIQUE : fondée en 2008, la SARL BEJAIA LOGISTIQUE (BL) est la référence Algérienne dans le domaine du transport routier, son activité est étendue dans le transport public de marchandises, location d'engins et matériels pour bâtiments, travaux publics et manutention, location de véhicules avec ou sans chauffeur et dans le transport des produits pétroliers ;
- HUILLERIES OUZELLAGUEN : lancée en 2008 avec la création de la filiale oléicole dénommée SARL Huilleries Ouzellaguen. Spécialisée dans la transformation (trituration) d'olives et mise en bouteille d'huile d'olive extra vierge, le produit est commercialisé sous le nom de Numidia.

1.2.5 Cadre juridique de la SARL « IFRI »

Les informations relatives à la SARL « IFRI » sont présentées dans la fiche signalétique⁵ de la figure 1.1.

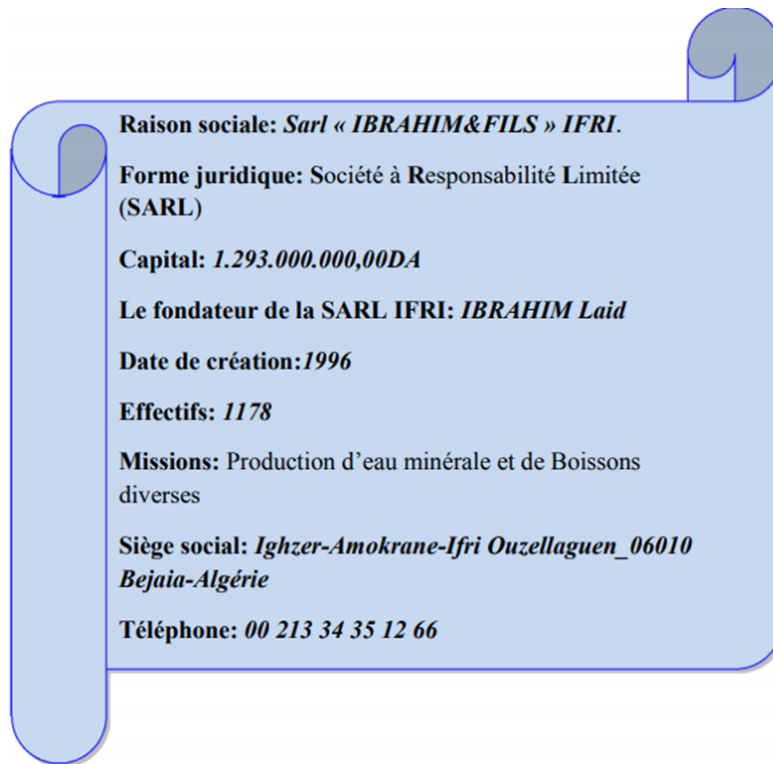


FIGURE 1.1 : Fiche signalétique de la SARL « IBRAHIM&FILS » IFRI.

D'autres informations sont :

- Numéro de registre de commerce est : 98B0182615 ;
- Numéro d'article d'imposition : 06360646615 ;
- Numéro d'identifiant fiscale : 099806018261598.

1.2.6 Organisation de la SARL « IBRAHIM&FILS »

La structure organisationnelle des différentes fonctions de l'entreprise est présentée par la figure 1.2⁶ :

5. Document de l'entreprise Ifri : "Forme juridique"

6. Document de l'entreprise Ifri

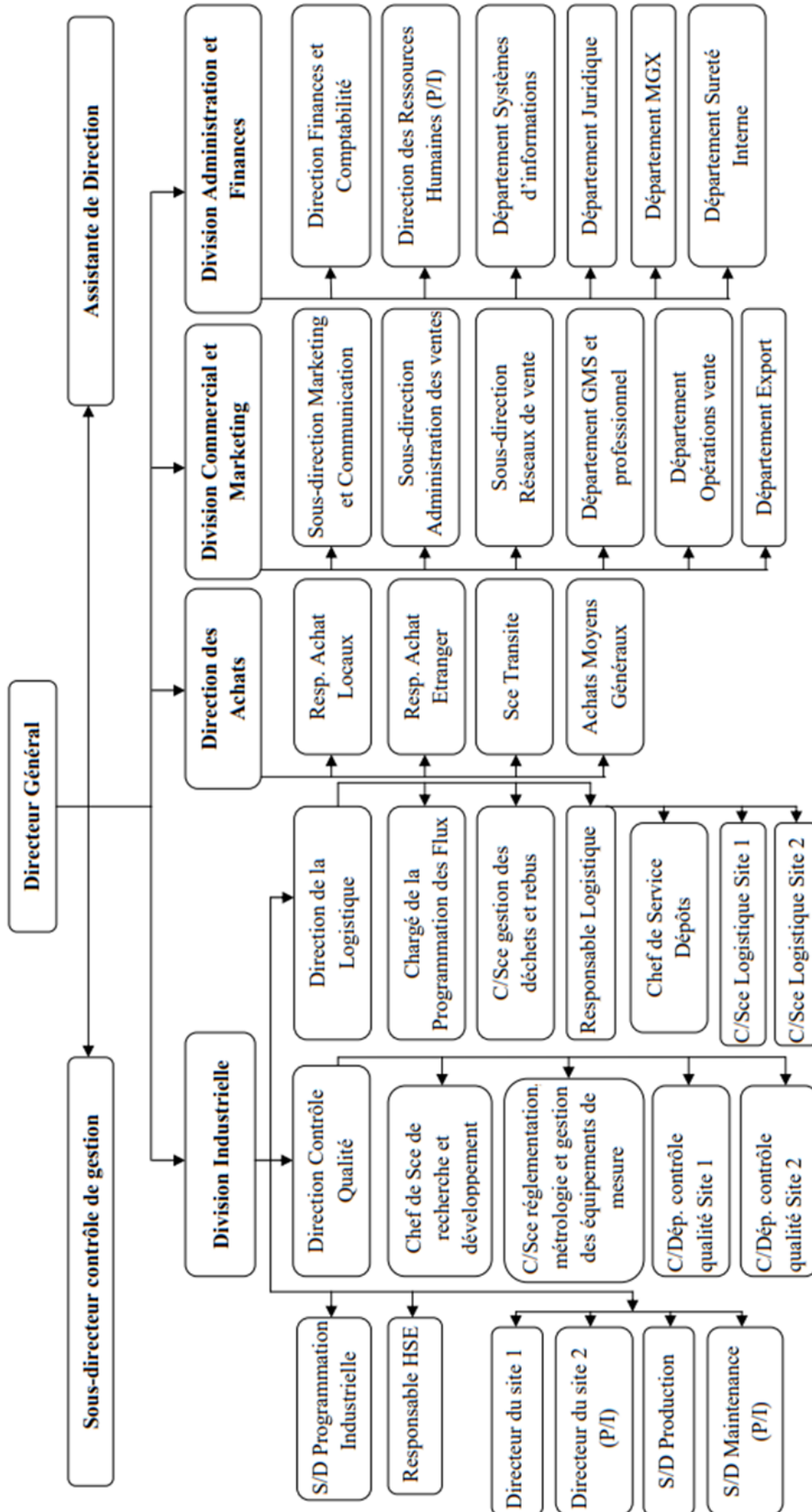


FIGURE 1.2 : Organigramme de la SARL IFRI.

1.2.7 Présence d'IFRI

La SARL IFRI est présente à l'échelle nationale et internationale.

1.2.7.1 IFRI à l'échelle nationale

Dans un esprit de proximité du consommateur, le produit IFRI figure sur tout le territoire national. La société touche les 48 wilayas, ayant couvert les besoins du marché local. Elle commence à satisfaire le marché algérien avec 500 millions de bouteilles par an (emballage PET et Verre)

1.2.7.2 IFRI à l'échelle internationale

L'établissement IFRI se lance dans la conquête du marché mondial, grâce à la stratégie du groupe en matière de développement des exportations par sa gamme élargie de boissons. Aujourd'hui, la SARL IFRI exporte ses produits vers des pays européens : la *France* (qui est son premier client à travers notamment la grande distribution), l'*Espagne*, l'*Italie*, l'*Allemagne*, et la *Belgique*, ... Ainsi que le *Soudan*, le *Mali* et les *Émirats arabes unis* qui sont ses principaux partenaires.

1.3 Problématique

La vie de l'entreprise est marquée par de nombreuses décisions. Elles sont prises chaque jour, depuis sa création jusqu'à sa mort. Elles ne sont pas toutes de même importance ni de même nature.

Malheureusement l'entreprise a du mal à prendre des décisions quelles que soient leurs niveaux d'importances (stratégiques, tactiques ou opérationnelles), dont on trouve celles de ventes et marketing, ressources humaines, techniques etc. et parmi les facteurs qui influencent la prise de décision : les caractéristiques de l'entreprise, l'**évolution du marché**, les logiques financières etc.

D'après les informations acquises durant notre stage au sein du service informatique et les interviews que nous avons eu avec le personnel du service commercial, nous avons recensé quelques problèmes :

- Il paraît que l'entreprise ne satisfait pas le vrai besoin du client, c'est-à-dire l'entreprise rencontre le problème de ruptures de stocks ;
- L'entreprise utilise des méthodes classiques pour les analyses des ventes et les prédictions (manuellement sous Excel) ;
- L'entreprise ne possède pas des systèmes d'aide à la décision.

L'entreprise a besoin d'écouler sa production. Pour ce faire, elle doit prendre en compte l'évolution du marché.

1.4 Marché

Un marché est un lieu physique ou virtuel sur lequel sont échangés des biens et services de nature diverse. Par extension, le marché est qualifié comme l'ensemble des consommateurs réels et/ou potentiels, souhaitant procéder à un échange, leur permettant de satisfaire un besoin [1].

L'entreprise IFRI a une forte caractérisation ; elle est leader dans le marché national ; elle répond aux besoins du marché dont elle offre une gamme complète des produits, elle crée des nouveaux besoins grâce à l'innovation des produits (un nouveau produit chaque année). IFRI est certifiée sous les normes ISO 9001⁷ et ISO 22000⁸ (concernant l'eau minérale et les jus). Ce qui apporte la confiance à leurs clients et permet sa présence en force sur le marché national et international.

1.4.1 Étude du marché

L'étude de marché définit l'ensemble des procédures techniques mises en œuvre pour produire et fournir de l'information utile et fiable en vue de réduire l'incertitude et d'aider à la prise de décision dans tous les champs. Cette information peut être utilisée pour analyser un problème et suggérer un certain nombre de solutions, ou pour vérifier l'efficacité de décisions prises [2].

L'étude de marché est l'une des clés essentielles du marketing, de ce fait c'est une stratégie que l'entreprise suit pour son lancement ou bien pour son développement. En d'autres termes, c'est un travail d'exploration marketing destiné à analyser, mesurer et comprendre le fonctionnement réel des forces à l'œuvre dans le cadre d'un marché.

Donc, l'étude de marché a pour principal objectif de réduire les risques d'échec, permettant à l'entreprise de prendre les mesures adéquates pour s'insérer durablement sur son marché, garantir la satisfaction du client et à plus long terme de mieux cerner les forces en présence.

1.4.2 Approche d'analyse des ventes

L'analyse des ventes est le processus de traitement des données relatives aux ventes dans une période donnée, qui permet d'orienter les actions commerciales et marketing de l'entreprise, elle repose sur l'utilisation de tableau de bord commercial et des techniques statistiques.

L'analyse des ventes peut se faire d'une manière permanente, ou du moins avec une périodicité diverse : chaque mois, chaque trimestre ou à la rigueur chaque année, en fonction de différents critères : produits ou modèles, clients, régions et circuits de distribution.

L'analyse des ventes passée et actuelle d'une entreprise demeure une fonction primordiale à savoir [3] :

- L'analyse des ventes permet d'évaluer la situation commerciale, de mesurer la performance, suivre la réalisation des objectifs et la rentabilité de l'entreprise ;
- La connaissance et l'analyse des ventes représentent une source d'information interne très importante et utile à la prévision des ventes futures ;
- Elle permet de dégager les forces et les faiblesses dans l'implantation des ventes de l'entreprise ;
- Elle permet aussi d'identifier les produits, les clients, les régions et les vendeurs les plus rentables pour l'entreprise.

7. ISO 9001 : définit des exigences pour la mise en place d'un système de management de la qualité pour les organismes souhaitant

8. ISO 22000 : est une norme internationale, relative à la sécurité des denrées alimentaires. Elle est applicable pour tous les organismes de la filière agroalimentaire

1.4.3 Préviation des ventes

La préviation des ventes est l'activité où l'entreprise cherche à calculer ou prédire un évènement futur, sur la base d'une analyse rationnelle des données disponibles, de l'expérience passée et de tout autre évènement pertinent. La préviation recouvre également un ensemble de méthodes statistiques très diverses qui ont pour objectif de chercher à réduire l'incertitude liée à la non connaissance du futur.

L'activité de préviation est essentielle pour les entreprises, car elle est utile pour la planification des approvisionnements et de la production. Les prévisions des ventes ont donc pour principal objectif d'aider les entreprises à anticiper la demande des produits ou des services qu'elles offrent et optimiser leurs ressources, dans le but d'avoir estimé les quantités qui sont à prévoir dans le futur.

Une bonne préviation des ventes permet de :

- Établir quelle capacité de production est requise afin d'ajuster l'offre à la demande ;
- Conditionner l'optimisation des ressources ;
- Minimiser les délais de réaction ;
- Orienter la politique et les stratégies de gestion des stocks ;
- Déterminer les meilleures stratégies de production.

1.5 Solution proposée et objectifs

Comme l'entreprise rencontre des problèmes de décision concernant les ventes, donc elle doit mettre en œuvre des méthodes d'analyse et de préviation de la structure de ses ventes et s'adapter aux exigences du marché tout en tenant compte de l'aspect comportemental du marché et de l'environnement concurrentiel.

Bien évidemment les entreprises utilisent plusieurs méthodes d'analyses et de prévisions des ventes. Parmi les méthodes d'analyse nous citons ABC (Activity Based Costing)⁹ basée sur les coûts par activité, matrice BCG (Boston Consulting Group) d'allocation de ressources par activité¹⁰ et les tableaux de bord. Les prévisions de ventes reposent souvent sur les opinions d'experts et des vendeurs et la méthode des moindres carrés ou encore la méthode de la moyenne mobile. En conséquence, l'exactitude des prévisions dépend de la compétence des experts choisis, de plus cela demande un grand nombre de valeurs historiques, donc elles sont assez couteuses.

Ainsi, les entreprises ont exploité d'autres procédures et solutions qui répondent à leurs besoins dont le développement des systèmes d'aide à la décision qui deviennent de plus en plus évolutifs grâce à l'intelligence artificielle. En effet, ce genre de système peut être utilisé pour mieux gérer le problème de décision de ventes en l'appliquant sur l'étude du marché qui est une étape très importante et rentable ; car elle permet de tester une idée et de réduire les incertitudes et les risques. De ce fait l'utilisation des solutions informatiques intégrant le principe de l'aide à la décision se révèle précieux.

Notre objectif est de développer une application informatique d'un système d'aide à la décision pour les prédictions de ventes. Principalement, nous allons :

9. La méthode ABC permet d'identifier les facteurs de coûts réels et les économies potentielles et d'améliorer la rentabilité des produits et des clients.

10. La matrice BCG permet de justifier des choix d'allocation de ressources entre les différentes activités d'une entreprise diversifiée.

- Mettre en place un entrepôt de données tout en rassemblant toutes les sources de données disponibles ;
- Développer un modèle de Machine Learning pour les prédictions des ventes, le déployer puis l'intégrer dans l'application.
- Concevoir un tableau de bord pour mieux visualiser les données et augmenter la précision des prévisions ;

Afin de comprendre la prédiction de ventes, la section suivante aborde le processus de la prédiction d'une manière générale.

1.6 Prédiction

La prédiction est une action d'annoncer par avance, de prédire l'avenir [4].

1.6.1 Analyse prédictive

L'analyse prédictive, parfois appelée analyse avancée, est un terme utilisé pour décrire une série de techniques analytiques et statistiques permettant de prédire des actions ou des comportements futurs. Dans les entreprises, l'analyse prédictive est utilisée pour prendre des décisions proactives et déterminer des actions, au moyen de modèles statistiques permettant de découvrir des schémas dans des données historiques et transactionnelles, dans le but d'identifier des risques potentiels et des opportunités.

L'analyse prédictive intègre plusieurs activités comme : l'accès aux données, l'analyse exploratoire des données et la visualisation, l'élaboration d'hypothèses et de modèles de données, l'application de modèles prédictifs, ainsi que l'estimation et/ou la prédiction de résultats futurs. Un exemple de processus d'analyse prédictive est donné dans la référence [5]. Ce processus est illustré par la figure 1.3 qui montre et explique ses différentes étapes.

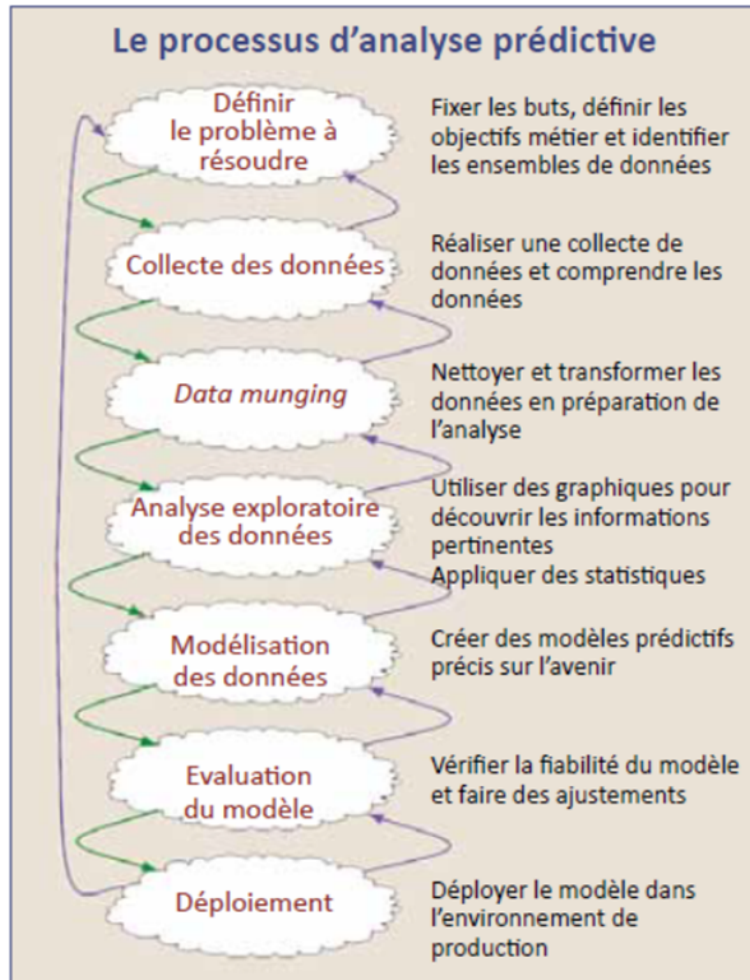


FIGURE 1.3 : Processus de l'analyse prédictive [5].

1.6.2 Utilisations métier de l'analyse prédictive

Il est évidemment nécessaire d'utiliser l'analyse prédictive dans l'entreprise, car elle offre une analyse plus intelligente, qui permet d'optimiser les prises de décision, d'accroître la compétitivité sur le marché, d'envisager plus directement les opportunités du marché et les menaces, de réduire l'incertitude et de gérer les risques, d'amener une approche de planification et d'action proactives, de découvrir des schémas judicieux et des moyens d'anticiper et de réagir aux tendances émergentes.

De nombreux problèmes de l'entreprise peuvent être résolus, dans diverses catégories, grâce à l'analyse prédictive. Voici quelques exemples [5] :

- Prévisions des ventes ;
- Détection des fraudes ;
- Optimisation des campagnes de vente ;
- Analyse marketing et clientèle ;
- Analyse RH ;
- Gestion des risques.

1.7 Conclusion

Dans ce chapitre nous avons premièrement présenté l'organisme d'accueil, puis soulevé les problèmes auxquels il fait face en particulier l'évolution du marché. Deuxièmement, nous avons défini certaines notions telles que l'étude du marché, les approches d'analyse et de prévision des ventes pour comprendre comment appréhender l'évolution du marché. Nous avons alors proposé de développer une application informatique d'aide à la décision pour la prédiction des ventes basée sur la technique de Machine Learning.

Le chapitre suivant sera consacré à l'étude du domaine de l'aide à la décision et la présentation des démarches de développement des applications informatiques des systèmes d'aide à la décision.

Chapitre 2

Systemes d'aide à la décision

2.1 Introduction

La prise de décision est un élément d'une importance capitale dans toute entreprise. Avec les nouvelles technologies, de plus en plus de décisions sont prises par des algorithmes tels que ceux proposés par Machine Learning. Certaines entreprises pour confronter la problématique de qualité de la décision et de l'analyse prédictive adoptent les systèmes interactifs d'aide à la décision intelligents.

Dans ce chapitre nous allons définir le domaine de la décision : le processus de décision et les systèmes d'information d'aide à la décision. Ensuite, nous donnons des aperçus de Machine Learning et de l'informatique décisionnelle. Enfin, nous allons présenter la démarche de développement à suivre durant notre projet.

2.2 Décision

La décision est définie comme étant un [6] « acte par lequel un ou des décideurs opèrent un choix entre plusieurs options permettant d'apporter une solution satisfaisante à un problème donné ».

2.3 Processus de décision

La décision ne peut être étudiée que via la notion de processus de décision, qui aide à identifier les problèmes qui se posent aux décideurs et à comprendre leur comportement individuel, mais aussi au niveau organisationnel afin de les assister dans leurs décisions. En somme, les décisions ne sont pas prises après avoir posé le problème et collecté toutes les informations, mais progressivement durant un long processus d'action et de planification.

Simon¹ a été l'un des premiers à aborder les problématiques liées aux processus de décision dans les organisations. Il a proposé un schéma général de prise de décision illustré par la figure 2.1.

1. Herbert Alexander Simon est un économiste et sociologue américain ayant reçu le prix dit Nobel d'économie en 1978. Il s'est d'abord intéressé à la psychologie cognitive et la rationalité limitée qui constitue le cœur de sa pensée.

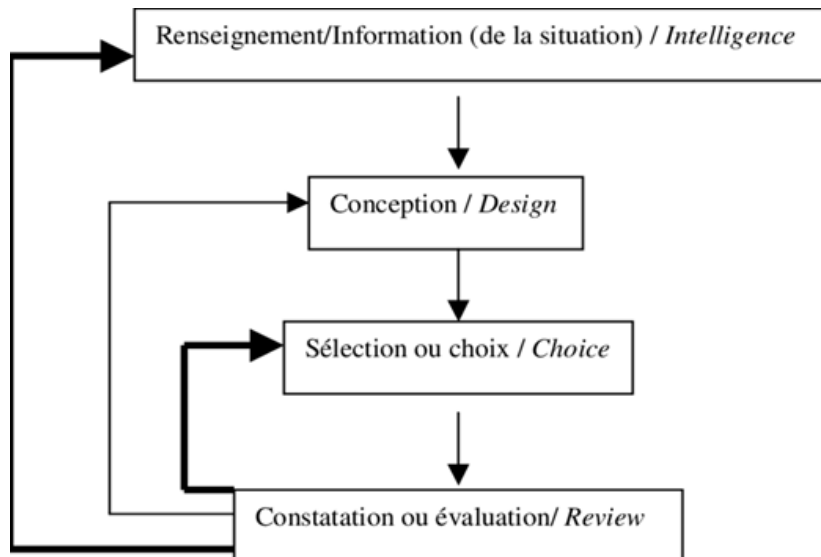


FIGURE 2.1 : Modèle de processus de décision de Simon [7].

Ce processus est appliqué en quatre phases [8] qui sont :

- Recherche d'information (Intelligence) : elle vise à recenser l'ensemble des informations utiles et prioritaires dont le décideur aura besoin lors de la prise de décision. Il s'agit d'identifier les objectifs ou priorités du décideur et de définir le problème à résoudre. Pour cela, il est nécessaire de rechercher les informations pertinentes en fonction des préoccupations du décideur. Néanmoins, l'acquisition d'informations adaptées pendant cette phase peut se poser elle-même en termes de décision. En effet, sur ces informations pertinentes s'appuient toutes les autres phases du processus de décision, leur choix est donc essentiel.
- Conception (Design) : cette phase comprend la génération, le développement et l'analyse des différentes suites possibles d'actions. Le décideur construit des solutions, imagine des scénarios, ce qui peut l'amener à rechercher des informations complémentaires. Pour cela, il va être nécessaire de choisir un ou plusieurs modèles de décision en fonction de la complexité du problème à traiter.
- Sélection (Choice) : ici le décideur choisit entre les différentes suites d'actions (solutions) qu'il a construit et identifié pendant la phase précédente. Il s'agit dans cette phase de déterminer les critères d'évaluation des différentes solutions envisageables et d'étudier ou mesurer les conséquences de chaque alternative. L'évaluation des alternatives et le choix final dépendent du type des critères utilisés.
- Évaluation (Review) : cette phase conduit à la recommandation d'une solution appropriée au modèle. Elle peut amener à la réactivation de l'une des trois phases précédentes ou, au contraire, à la validation de la solution. De nouvelles informations pertinentes peuvent influencer tel ou tel choix, voir le modifier complètement. Une rétroaction (feedback) intelligente permet de corriger bien des erreurs sur le déroulement d'un processus décisionnel, elle conduit à des performances aussi bonnes que des stratégies compliquées sans rétroaction. La recommandation finale doit traduire le résultat fourni par le modèle d'évaluation dans le langage courant du client et du processus de décision dans lequel il est impliqué.

Sur un autre plans, la décision apparait comme le résultat d'un processus global de résolution de problèmes. Il existe cinq types de problèmes distincts au cours du processus

de décision [9] :

- Description : problèmes associés à la caractérisation réelle de l'état courant de l'organisation ;
- Explication : problèmes associés aux relations entre deux ou plusieurs éléments de données ou phénomènes ;
- Diagnostic : problèmes associés à l'établissement d'une relation de cause à effet ;
- Prédiction : problèmes associés à une projection basée sur des données historiques ;
- Prescription : problèmes associés à la projection normative basée sur des données historiques.

Dans le cadre de notre travail, le problème de décision à résoudre est celui de prédiction, en implémentant un système d'aide à la décision.

2.4 Systèmes Interactifs d'Aide à la Décision

Le concept d'aide à la décision peut se définir comme l'ensemble des activités destinées à aider le décideur dans sa prise de décision en lui proposant d'effectuer un choix, un tri ou un rangement [10].

Un système d'aide à la décision (SAD) est conçu pour aider et remédier aux déficiences cognitives humaines en mettant en œuvre une ou plusieurs des fonctions d'aide à la décision et faciliter le processus de structuration et l'optimisation des décisions. En pratique, le concept de SAD recouvre des réalités diverses et relève de champs conceptuels différents : la théorie de la décision, l'informatique, l'intelligence artificielle, la psychologie cognitive et l'ergonomie [11]. Les SAD ont été classés en quatre types selon le niveau et le type de la décision à prendre dans l'entreprise qui sont :

- Systèmes d'Information de Gestion (SIG) ;
- Systèmes d'Information pour Dirigeants (SID) ;
- Systèmes de Décision de Groupe (SADG) ;
- Systèmes Interactifs d'Aide à la Décision (SIAD).

Les SIADs permettent aux managers d'entreprise d'avoir le contrôle de toutes les étapes du processus de décision, ce qui leur permet de prendre des décisions facilement. Notre principal objectif est de mettre en œuvre un tel système. Ce qui nous amène à faire une présentation des SIADs et leurs composants.

2.4.1 Définition d'un SIAD

Les SIADs peuvent être caractérisés comme des systèmes informatisés, interactifs, qui aident les décideurs en utilisant des données et des modèles pour résoudre des problèmes « mal ou non structurés ». Leur définition repose sur les mots « données » et « modèles » qui fondent leur architecture.

Une caractéristique commune à tous les SIADs est l'interactivité. La notion d'interactivité dans un SIAD renvoie au rôle indispensable de l'utilisateur dans son fonctionnement, rôle non passif qui sous-entend le terme aide à la décision, mais aussi à la qualité de l'intégration des différents composants du système et à la nature de l'interface homme/machine.

Il faut souligner, d'un côté, que les SIADs n'ont pas pour objectif la recherche d'une solution optimale, mais permettent plutôt d'orienter le décideur vers des points qu'il ne pourrait pas toujours observer seul. D'un autre côté, l'utilisation d'outils en provenance

de l'intelligence artificielle [12] a permis de renforcer le traitement de l'information en éliminant les informations non pertinentes et en présentant des faits déduits à partir des données présentes dans les systèmes d'information.

2.4.2 Composants d'un SIAD

Quatre composants fondamentaux d'un SIAD sont évoqués [13] :

Système de gestion et de génération de dialogue (l'interface homme-machine)

Est un élément clé des fonctionnalités du système global. Il représente la caractéristique principale du système « l'interactivité homme-machine ». Le décideur accède aux données et aux fonctions de calcul au travers cette interface et le système utilise le même vecteur pour lui communiquer le résultat des manipulations qu'il a effectué. C'est une session interactive dynamique et un échange d'informations en temps réel entre l'utilisateur et le système. Cette interface doit posséder un certain nombre de caractéristiques spécifiques qui sont importantes à considérer lors de sa conception : l'accessibilité, la flexibilité, l'interactivité, l'ergonomie, etc.

Système de Gestion de Bases de Données (SGBD)

Permet le stockage, l'organisation et le tri de données pour un contexte particulier de décision. Dans un SIAD, un système de base de données, doit mettre en corrélation des données de différentes sources, permettre une recherche rapide par des requêtes et des rapports, réaliser des tâches de recherche et de manipulation complexe de données.

Système de Gestion de Bases de Modèles (SGBM)

A une fonction similaire à celle du SGBD si ce n'est qu'il gère des modèles quantitatifs de l'organisation. On entend par « modèles » l'ensemble des procédures de calcul utilisées dans les différents traitements standards des données mis à disposition de l'utilisateur. Il peut s'agir d'outils de programmation mathématiques, de modèles de prédiction et de simulation, de modèles financiers et de calcul standards, etc.

Système de Gestion de Bases de Connaissances (SGBC)

Permet d'effectuer les tâches relatives à la reconnaissance de problèmes et à la génération de solutions finales ou intermédiaires aussi bien que des fonctions relatives à la gestion du processus de résolution de problèmes. L'architecture même de ces systèmes fait apparaître une partie technologique issue de l'intelligence artificielle intégrant une représentation des connaissances dans le problème à résoudre. L'intérêt de cette architecture réside dans l'accent mis sur le raisonnement dans la prise de décision.

A noter que le couple **utilisateur/décideur** est une partie intégrante du processus de résolution de problèmes à l'aide d'un SIAD.

Le lien entre l'intelligence artificielle et les SIADs n'a fait que se renforcer au cours des années et donne naissance à une autre génération de SIAD dits intelligents [14]. De ce fait, la section suivante présente un aperçu de la technique de Machine Learning, branche de l'intelligence artificielle, qui nous permettra d'aborder facilement le développement d'un modèle de prédictions des ventes par la suite.

2.5 Machine Learning

Le « Machine Learning » ou « apprentissage automatique » en français est un concept qui fait de plus en plus parler de lui dans le monde de l'informatique et qui se rapporte au domaine de l'intelligence artificielle.

Selon Arthur Samuel² « L'apprentissage automatique est la science qui permet aux ordinateurs d'apprendre sans être explicitement programmés ». En d'autres mots, il s'agit d'une sorte de programmes permettant à un ordinateur ou à une machine un apprentissage automatisé, de façon à pouvoir réaliser un certain nombre d'opérations très complexes.

Une autre définition qui se rapporte aux systèmes d'informations stipule que [15] : l'objectif visé par l'apprentissage automatique est de rendre la machine ou l'ordinateur capable d'apporter des solutions à des problèmes compliqués par le traitement d'une quantité astronomique d'informations. Comme conséquence, il peut être utilisé comme un moyen pour analyser et mettre en évidence les corrélations qui existent entre deux ou plusieurs situations données et de prédire leurs différentes implications.

2.5.1 Types d'apprentissage automatique

Il existe quatre types d'apprentissage automatique qui sont les suivants (voir la figure 2.2) :

2.5.1.1 Apprentissage supervisé

Un processus au sein du Machine Learning pour enseigner à un programme une certaine tâche à travers des exemples préparés et organisés par des scientifiques, ingénieurs etc. L'un des objectifs principaux de ce type d'apprentissage est de faire des prédictions [16]. Il existe deux types de sous-problèmes d'apprentissage supervisé [17] :

Régression : les tâches de régression se concentre sur la recherche à prédire la valeur d'une variable continue, c'est-à-dire une variable qui peut prendre une infinité de valeurs.

Classification : dans un problème de classification, on cherche à classer un objet dans différentes classes, c'est-à-dire que l'on cherche à prédire la valeur d'une variable discrète (qui ne prend qu'un nombre fini de valeurs).

2. Arthur Samuel : c'est un mathématicien américain qui a développé un programme pouvant apprendre tout seul comment jouer aux Dames en 1959.

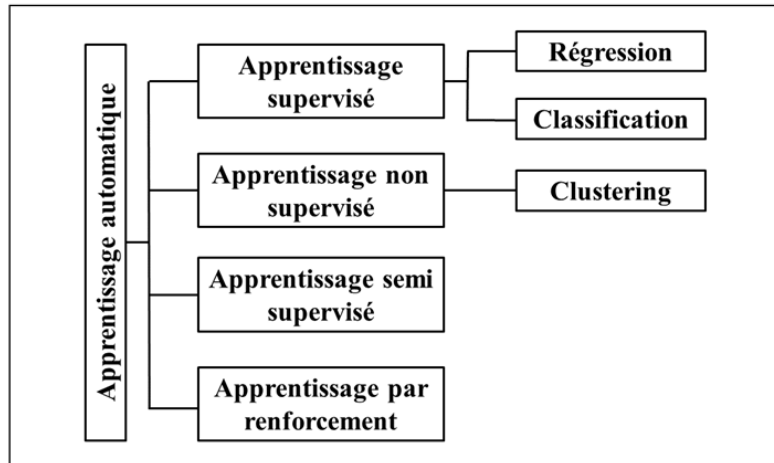


FIGURE 2.2 : Types d'apprentissage machine.

2.5.1.2 Apprentissage non supervisé

Dans le cadre de l'apprentissage non supervisé, les données ne sont pas étiquetées. Il s'agit alors de modéliser les observations pour mieux les comprendre.

Regroupement : un sous-problème d'apprentissage non supervisé qui fait un regroupement des données en classes homogènes. Généralement, il consiste à représenter un nuage de points d'un espace quelconque en un ensemble de groupes appelé Clusters d'où le nom de Clustering en anglais. D'après [18], c'est un traitement sur un ensemble d'objets qui n'ont pas été étiquetés par un superviseur.

2.5.1.3 Apprentissage semi-supervisé

L'apprentissage semi-supervisé est l'hybridation des approches d'apprentissage supervisé et non supervisé. Ils classent l'ensemble de données sans étiquette à l'aide de l'ensemble de données étiquetées [19].

2.5.1.4 Apprentissage par renforcement

Processus ou implémentation de Machine Learning où un logiciel apprend à effectuer une tâche à partir de ses échecs et de ses succès [16].

D'après cet aperçu, nous constatons que l'apprentissage supervisé et plus précisément les algorithmes de régression sont plus adéquats pour le cas de prédiction. Ainsi, nous allons utiliser ces algorithmes pour la prédiction des ventes dans l'application d'aide à la décision développée.

De plus, pour offrir une bonne aide à la décision, certaines entreprises s'orientent vers la business intelligence, car elle couvre un ensemble de technologies et d'outils pour la mise en place des systèmes décisionnels. De ce fait, il est important de définir ce domaine.

2.6 Business Intelligence

Le système d'information décisionnel est un ensemble de données organisées de façon spécifiques, facilement accessibles et appropriées à la prise de décision ou encore une représentation intelligente de ces données au travers d'outils spécialisés. Les systèmes décisionnels sont dédiés au management de l'entreprise pour l'aider au pilotage de l'activité et sont indirectement opérationnels car n'offrant que rarement le moyen d'appliquer les décisions [8]. Ce type d'application repose sur une architecture commune [20] :

- Les données opérationnelles sont extraites périodiquement de sources hétérogènes : fichiers plats, fichiers Excel, base de données (Oracle, SQL Server, etc.), service web, données massives et stockées dans un entrepôt de données (à définir dans 2.6.2).
- Les données sont restructurées, enrichies, agrégées, reformatées pour être présentées à l'utilisateur sous une forme sémantique ; rapports pré-préparés paramétrables, de tableaux de bords plus synthétiques et interactifs.
- Ces données sont livrées aux divers domaines fonctionnels (direction stratégique, finance, production, comptabilité, ressources humaines, etc.).

Il est coutumier de présenter les éléments composants l'informatique décisionnelle en quatre catégories correspondantes chacune à une phase de processus décisionnel. C'est ceux que nous allons présenter dans ce qui suit.

2.6.1 Collecter, nettoyer et consolider les données

C'est un processus automatisé qui consiste à extraire les données de différentes sources et les faire adapter à un usage décisionnel. L'ETL (Extract, Transform and Load) prend les données brutes, extrait l'information nécessaire à l'analyse, la transforme en un format qui peut répondre aux besoins opérationnels et la charge dans un entrepôt de données (voir la figure 2.3). L'ETL résume généralement les données afin de réduire leur taille et d'améliorer leur performance pour des types d'analyse spécifiques [21]. Les étapes de l'ETL sont [22] :

- Extraction de données : l'extraction est la première étape du processus d'apport de données à l'entrepôt de données. Extraire, cela veut dire lire et interpréter les données sources et les copier dans la zone de préparation en vue de manipulations ultérieures.
- Transformation de données : une fois les données extraites, elles peuvent être transformées. Les opérations les plus courantes à ce niveau incluent des opérations de filtrage de données, de dérivation de valeurs, de transformation de formats de données, de génération automatique de numéros de séquence, etc.
- Chargement de données : il s'agit de prendre la sortie de l'étape de transformation et de les placer dans l'emplacement approprié dans l'entrepôt de données. C'est une étape très délicate et exige une certaine connaissance des structures de système de gestion de la base de données (tables et index) afin d'optimiser au mieux le processus.

L'ETL peut prendre en charge différentes natures de sources de données, tant en entrée qu'en sortie, les principales étant bien sûr les SGBDR et les flux XML, mais il peut s'agir également de fichiers à formats fixes ou avec séparateurs (CSV) [23].

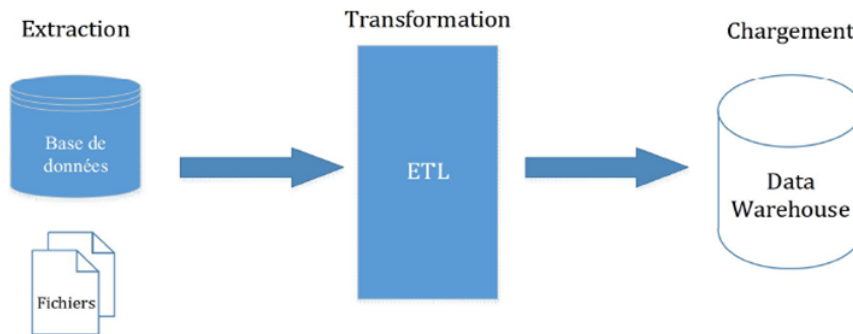


FIGURE 2.3 : Extraction, Transformation et Chargement de données au niveau des outils ETL [23].

2.6.2 Stocker

Centraliser les données structurées et les traitées afin qu'elles soient disponibles pour un usage décisionnel.

2.6.2.1 Datawarehouse

Un entrepôt de données (Datawarehouse) est un vaste dépôt centralisé de données qui contient de l'information provenant de nombreuses sources au sein d'une organisation [21]. Ces données sont orientées sujet, intégrées et contenant des informations historiques, non volatiles permettant l'analyse des indicateurs pertinents pour faciliter la prise de décisions [24]. Les données recueillies sont utilisées pour guider les décisions d'affaire au moyen d'outils d'analyse, de production de rapports et d'exploration de données (Datamining) [21].

2.6.2.2 Datamart

Un Datamart est un magasin de données. Il s'agit d'un sous-ensemble du Datawarehouse qui ne contient que les données d'un métier de l'entreprise, alors qu'un Datawarehouse contient toutes les données de l'entreprise pour tous les métiers. L'avantage de la mise en œuvre d'un tel outil est plus aisé vu sa portée et son étendue réduites dans l'organisation [8].

2.6.3 Distribuer

Faciliter l'accessibilité des informations selon les fonctions et les types d'utilisation. Avant d'exploiter les données via divers outils, l'information est distribuée auprès de l'ensembles des partenaires à partir du portail d'information d'entreprise (Entreprise Information Portal, EIP). Si le Datawarehouse est la solution pour stocker et organiser l'information, le portail est incontournable à son tour pour en banaliser l'accès. Les portails d'entreprise (EIP) ont initié une nouvelle approche de la conception des systèmes d'information en plaçant au premier plan les besoins des utilisateurs en matière d'accès aux informations essentielles et de partage avec leurs pairs [25].

2.6.4 Exploiter

Assister de mieux possible l'utilisateur pour qu'il puisse extraire la substance de l'information des données stockées à cet usage [26].

Parmi les différents types d'outils d'exploitation de données existantes nous citons [27] :

2.6.4.1 Reporting (Générateur de rapport)

Permet la production, la livraison et la gestion des rapports. Les rapports utilisent des requêtes prédéfinies, comme les requêtes demandant des informations dans un format spécifique qui sont effectuées sur une base régulière.

2.6.4.2 Datamining

Ensemble de techniques permettant d'extraire des modèles d'une base de données historiques, Datawarehouse, etc. afin de décrire le comportement actuel et/ou de prédire le comportement futur d'un procédé.

2.6.4.3 OLAP

OLAP (On-Line Analytical Processing) est défini comme « une catégorie de logiciels axés sur l'exploration et l'analyse rapides des données selon une approche multidimensionnelle à plusieurs niveaux d'agrégation » [28].

2.6.4.4 Tableau de bord

Le tableau de bord est un outil de pilotage et ensemble des indicateurs qui permettent de mesurer la réalisation des objectifs. Il sélectionne, hiérarchise et présente l'information de façon synthétique et ciblée. Il met en évidence les résultats significatifs, les écarts et les tendances. Le tableau de bord est l'instrument qui, se suffisant à lui-même, permet au manager de disposer d'une vision claire de l'état de la situation et de prendre des décisions [29].

La figure 2.4 résume les concepts et les composants de business intelligence.

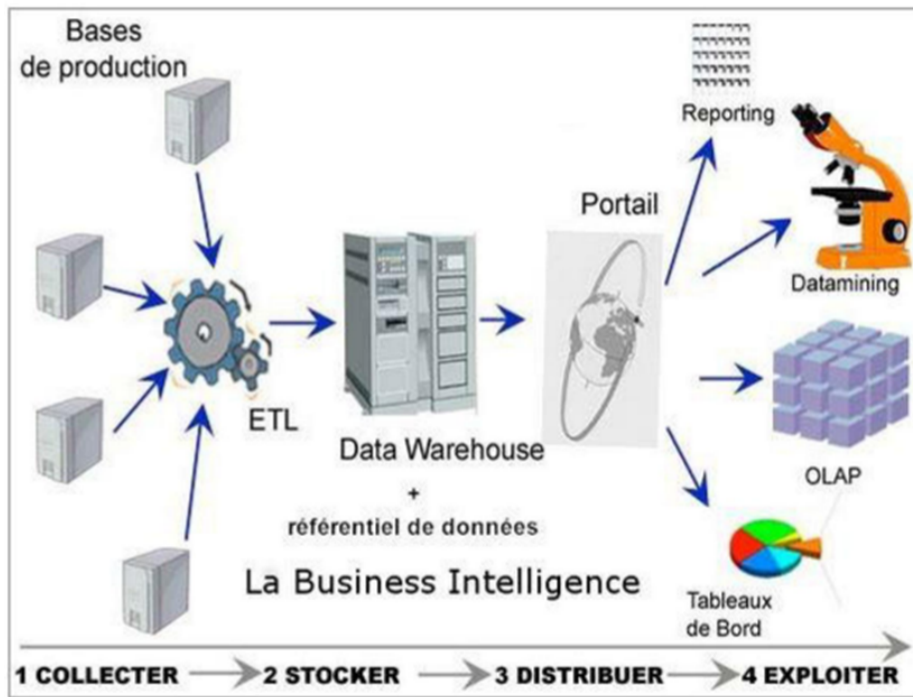


FIGURE 2.4 : Les quatre composants de l'informatique décisionnelle [25].

La prochaine partie de ce chapitre sera consacrée pour la modélisation d'un entrepôt de données.

2.7 Modélisation d'un entrepôt de données

La modélisation multidimensionnelle consiste à représenter les sujets d'analyse appelés faits autour d'axes d'analyse appelés dimensions.

2.7.1 Fait

Un fait est tout ce qu'on voudra analyser. C'est le sujet ou le thème analysé. Il présente un centre d'intérêts de l'entreprise et est considéré comme un concept clé sur lequel repose le processus de prise de décision. Un fait est formé de mesures et d'attributs du fait qui correspondent aux informations liées au thème analysé [30]. Un exemple de fait est donné par la figure 2.5.

Table de faits des ventes	
Clés étrangères vers les dimensions	Clé date (CE)
	Clé produit (CE)
	Clé magasin (CE)
Faits	Quantité vendue
	Coût
	Montant des ventes

FIGURE 2.5 : Exemple de table faits.

2.7.2 Dimension

Une dimension est tout ce qu'on utilisera pour faire nos analyses. Cela représente un contexte d'analyse d'un fait [31]. Les dimensions se présentent sous forme d'une liste d'éléments organisés de façon hiérarchique [30]. La figure 2.6 représente un exemple de table de dimension.

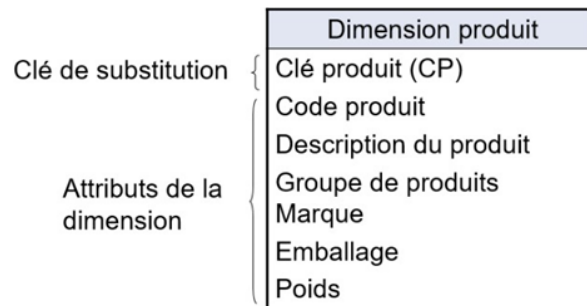


FIGURE 2.6 : Exemple de table dimension.

2.7.3 Types des schémas

Il existe trois types de schémas relationnels pour la modélisation des entrepôts de données.

2.7.3.1 Schéma en étoile

Un schéma relationnel dans lequel une table centrale contenant les faits analysés référence des tables de dimensions, chaque dimension étant décrite par une seule table dont les attributs représentent les diverses granularités (voir la figure 2.7).

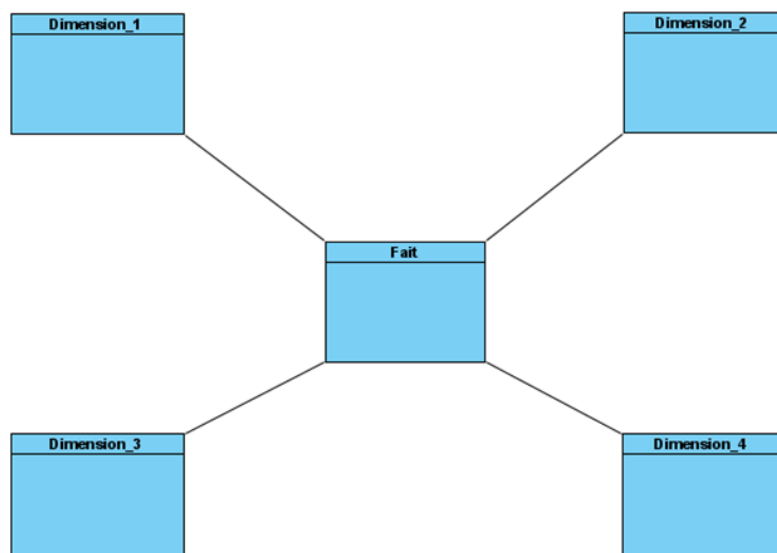


FIGURE 2.7 : Exemple de schéma en étoile.

Dans ce schéma, les informations associées à une hiérarchie de dimension sont représentées dans une seule table [27] :

- La table de faits contient une clé composée des clés des tables de dimensions.
- Les tables de dimensions contiennent quant à elle une clé primaire unique correspondant exactement à l'un des composants de la clé de la table des faits.
- Les faits sont généralement numériques alors que les dimensions sont qualitatives, elles contiennent des informations textuelles pour décrire les faits.

2.7.3.2 Schéma en flocon

Le principe est le même que pour le modèle en étoile, mais en plus, les dimensions sont décomposées comme le montre la figure 2.8. Le but est d'économiser de la place. Cela permet également d'instaurer une hiérarchie au sein des dimensions, mais engendre par contre une complexification du modèle [32].

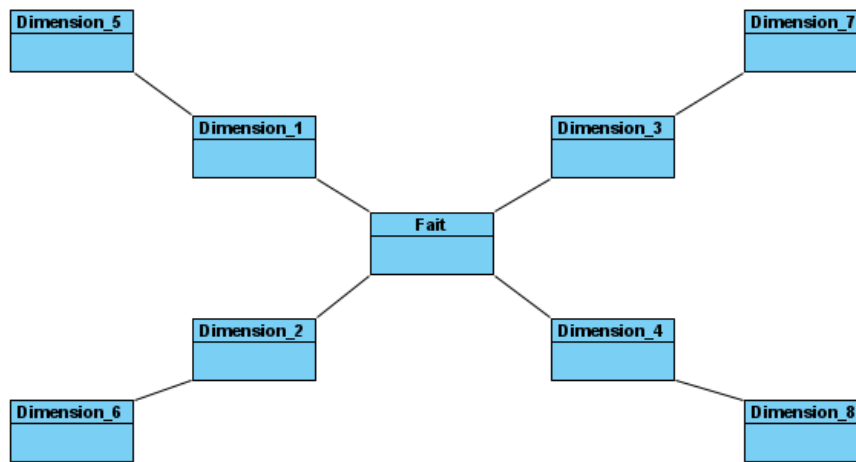


FIGURE 2.8 : Exemple de schéma en flocon.

2.7.3.3 Schéma en constellation

Un schéma de constellation comporte plusieurs tables de faits qui partagent des tables de dimensions [33]. Un exemple de schéma de constellation est présenté par la figure 2.9.

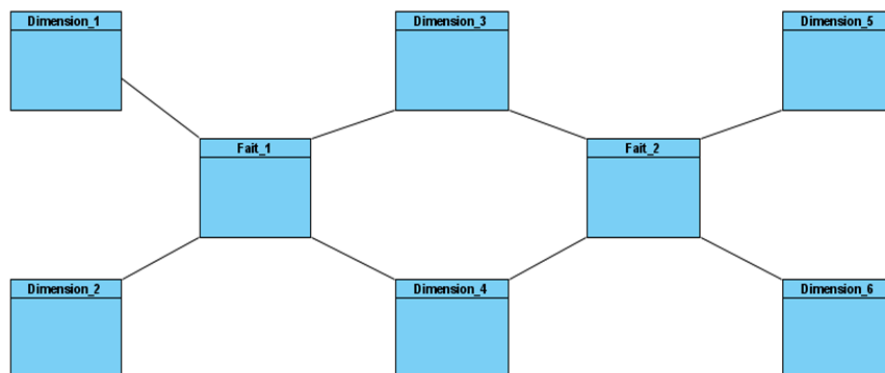


FIGURE 2.9 : Exemple de schéma en constellation.

Comme pour toute application informatique, le développement d'un SIAD se fait suivant une succession d'étapes. De ce fait la dernière partie de ce chapitre sera consacrée pour la présentation de la démarche de développement que nous allons adopter par la suite.

2.8 Démarche de développement de SIAD

Un cycle ou processus de développement logiciel a pour objectif d'énoncer les étapes dans un ordre logique et temporel pour la production de logiciels de qualité.

Il n'existe pas de modèles propres au développement des SIADs autre que l'Approche de Développement de SIAD (ADESIAD) pour les SIADs orientés connaissances. Cette approche a la particularité d'être centrée sur les décideurs, mais sa finalité est de développer un système à base de connaissances qui vise l'automatisation de décision (pas d'interaction homme-machine). A mentionner également que certaines méthodes de conception spécifiques à chaque type de SIAD ont été utilisées.

Par contre, plusieurs démarches de développement de systèmes informatiques ont été proposées : les processus classiques de génie logiciel (en cascade, en V et en spirale) sont bien trop souvent orientés vers la partie technique (le code) et non vers l'utilisateur. Ces aspects sont laissés à l'appréciation du développeur en particulier. Avec le temps, il est remarquable que la tendance allait vers des processus plus itératifs (les méthodes Agiles). Cependant, même si les utilisateurs sont mentionnés pour les étapes d'analyse et de validation de prototype, les processus s'accompagnent le plus souvent de peu d'explications relatives à la prise en compte des utilisateurs. D'un autre côté, les processus enrichis sous l'angle de l'IHM sont très orientés utilisateurs mais ne présentent pas clairement l'aspect itératif et incrémental indispensable au développement du système interactif. De plus, ils ne prennent pas en compte les connaissances des utilisateurs et leurs problèmes de décision.

Nous constatons que les démarches existantes ont montré des avantages mais aussi souvent des insuffisances pour développer un système « interactif » et « d'aide à la décision ». Pour combler ces insuffisances une nouvelle approche proposée dans [9] permet de mieux prendre en compte les utilisateurs/décideurs ainsi que leurs problèmes de décision durant tout le cycle de développement. Concrètement, ce processus doit être à la fois : itératif, incrémental et participatif (des caractéristiques empruntées des méthodes agiles) ; centré utilisateurs (emprunté du domaine de l'IHM) ; dirigé par les problèmes de décision des utilisateurs décideurs (emprunté du domaine de l'ingénierie des connaissances) et enfin collaboratif.

Le tableau 2.1 présente une comparaison entre les méthodes de développement classiques, Agiles, ingénierie IHM et des systèmes à base de connaissances et ainsi que la démarche de développement adoptée basée de l'approche proposée dans [9]. A noter que les phases « Allocation des besoins » et « Optimisation » ne sont recommandés que pour les moyens et grands projets où l'application développée est répartie et/ou exécutée sur divers types de matériel et utilisée par différents utilisateurs. La démarche de développement proposée sera appliquée en se basant sur le langage UML (Unified Modeling Language) et ces différents diagrammes. UML [34] est un langage graphique conçu pour représenter, spécifier et construire les artefacts d'un système à dominante logicielle.

Phase de développement	Génie Logiciel (Processus Classiques)	Génie Logiciel (Processus Agile)	Ingénierie des IHM	Ingénierie des systèmes de connaissances	Processus de développement SIAD adopté
Analyse des besoins	Analyse des exigences système	Analyse des besoins utilisateurs	Analyse de l'utilisateur et de ses tâches	Caractérisation du problème de l'utilisateur	Analyse des utilisateurs et leurs problèmes de décision communs pour déduire les exigences système
Allocation des besoins	Matériel/ Logiciel	Matériel/ Logiciel	Humaine/ Machine	Expert/ Intelligence artificielle	Humaine/ Machine + Intelligence Artificielle
Conception préliminaire	Conception de l'architecture	Conception de l'architecture	Conception du dialogue	Spécification des connaissances	Conception de l'architecture en tenant compte du dialogue HM
Conception détaillée	Conception logique	Conception logique	Conception Des écrans	Conception de la base de connaissance	Conception logique
Implémentation	Codage	Codage	Codage	Codage	Codage
Test d'implémentation	Tests unitaires et d'intégration	Tests unitaires et d'intégration	Tests unitaires et d'intégration	Tests d'opérationnalité des traitements sur les Connaissances	Tests unitaires d'intégration et d'utilisabilité
Test du système	Test du système	Test du système	Observation contextuelle	Comparaison résultats produits/ attendus par l'expert	Test du système en comparant les résultats produits/ attendus
Optimisation	Performance de la machine	Performance de la machine	Performance de la machine	Performance de la machine	Performance machine homme

TABLE 2.1 : Comparaison entre les méthodes de développement dans différents domaines.

2.9 Conclusion

Dans ce chapitre nous avons introduit le domaine de la décision dans le contexte des SIADs, dans lequel nous avons présenté les éléments composants de l'informatique décisionnelle. Puis, nous avons donné les différents types de Machine Learning avant d'opter pour l'apprentissage supervisé. Ensuite, nous avons défini la modélisation dimensionnelle et ses différents schémas. Enfin nous avons présenté la démarche de développement que nous allons suivre pour concevoir le SIAD après avoir fait une comparaison des méthodes de développement logicielles existantes.

Dans le chapitre suivant, nous ferons une analyse approfondie du fonctionnement de l'organisme d'accueil afin d'identifier les fonctionnalités de notre future application.

Chapitre 3

Analyse des besoins et des exigences

3.1 Introduction

Ce chapitre sera consacré pour l'analyse des besoins et des exigences. Il se focalisera sur la délimitation du domaine d'étude, l'identification des acteurs et l'analyse des situations décisionnelles. Puis, il traduira les besoins informationnels en spécifications fonctionnelles et non fonctionnelles ainsi que en interfaces homme/machine. Ces besoins seront également modéliser à l'aide d'un diagramme de cas d'utilisation et quelques cas d'utilisation seront expliqués par une description textuelle et diagramme de séquence.

3.2 Délimitation du domaine d'étude

La délimitation du périmètre de l'intervention comprend la compréhension du contexte du projet et l'identification des acteurs impliqués.

Notre domaine d'étude implique uniquement le service commercial, car tout ce qui concerne les analyses et les prévision de ventes se fait dans ce service. Il reste maintenant à recenser les acteurs intervenant dans ce même service. En effet, le succès de la mise en œuvre d'une démarche de conception centrée utilisateur repose sur l'implication de l'équipe de projet et l'identification des différents acteurs qui interagissent avec le système.

3.2.1 Identification des acteurs

Un acteur est une entité externe qui joue un rôle dans le système. Il peut être un utilisateur humain, un dispositif matériel ou un autre système qui interagit avec l'application [35].

Le responsable des statistiques est l'unique acteur qui est sensé interagir avec notre système, c'est lui qui joue le rôle de décisionnaire. Il aura accès aux différentes fonctionnalités du système après avoir être authentifié, donc il pourra par exemple effectuer l'analyse et la consultation des statistiques ou de faire la prédiction.

Il s'agit également de recueillir les données concernant l'acteur du SIAD qui seront utiles vis-à-vis des futures interfaces homme-machine relatives au système à développer.

3.2.2 Diagramme de contexte

Le diagramme de contexte présente les acteurs qui interagissent avec le système avec des différents messages.

La figure 3.1 représente le diagramme de contexte de notre système.

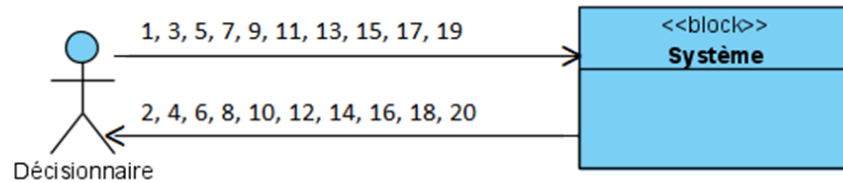


FIGURE 3.1 : Diagramme de contexte.

La description des messages est donnée dans le tableau 3.1.

Numéro	Message	Numéro	Message
1	Demande d'authentification	2	Affichage de l'interface d'accueil
3	Demande d'analyse statistiques	4	Affichage de tableau de bord
5	Demande de gestion des ventes	6	Affichage de l'interface de gestion des ventes
7	Demande de consultation d'une vente	8	Affichage de détails de la vente
9	Effectuer un tri des ventes	10	Affichage du résultat de tri
11	Demande de gestion des clients	12	Affichage de l'interface de gestion des clients
13	Demande de consultation des informations d'un client	14	Affichage de détails du client
15	Effectuer une recherche	16	Affichage du résultat de la recherche
17	Demande de prédiction	18	Affichage du formulaire
19	Remplir et valider les champs du formulaire	20	Affichage du résultat de prédiction

TABLE 3.1 : Messages échangés entre le système et le décisionnaire.

3.3 Analyse des situations décisionnelles

Elle consiste en une étude préalable du problème décisionnel concerné par le développement. Notre étude vise l'identification des situations décisionnelles, ce qui revient à extraire les problèmes à partir des connaissances expertes du responsable des statistiques de la SARL. L'acquisition de ces problèmes s'est faite principalement lors des réunions en s'appuyant sur les techniques d'entretien.

Dans le cadre de notre projet, l'analyse est portée sur les situations décisionnelles relatives à l'application des différentes techniques de l'étude de marché, l'analyse des ventes et des clients comme le présente le tableau 3.2.

Phase	Spécificités
Traitement et stockage des données	Les données utiles collectées pour les situations décisionnelles sont de différentes sources et leur degré de complexité dépend principalement de la quantité et de la nature des données en question. De ce fait, la sélection, le nettoyage et la transformation des données seront faits facilement avec un outil ETL qui les rassemble sous forme d'un entrepôt de données.
Analyse des résultats	Il s'agit des situations décisionnelles relatives aux analyse des ventes et des clients. La mise en place d'un tableau de bord est très utile pour ce genre d'analyse. Cela aidera les décideurs d'entreprise à prendre des meilleures décisions.
Estimation	Il s'agit de l'incapacité à mettre en œuvre des prévisions précises et efficaces en termes de ventes. L'utilisation d'un model Machine Learning en appliquant l'algorithme de régression linéaire est une solution pour avoir des résultats de prévisions plus précises. Généralement les prévisions des ventes se font une fois par année, mais dans notre projet elles seront faites une fois par mois.

TABLE 3.2 : Analyse des situations décisionnelles.

La définition des situations décisionnelles du SIAD permet de cerner les objectifs, définir les cibles et vérifier les besoins.

3.4 Identification des cas d'utilisation

Un cas d'utilisation est une unité cohérente représentant une fonctionnalité visible de l'extérieur du système. Il réalise un service de bout en bout, avec un déclenchement, un déroulement et une fin, pour l'acteur qui l'initie. Un cas d'utilisation modélise donc un service rendu par le système, sans imposer le mode de réalisation de ce service [36].

Le tableau 3.3 représente les différents cas d'utilisation réalisé par le décisionnaire.

N° Cas	Cas d'utilisation
1	S'authentifier
2	Analyser les statistiques
3	Analyser les statistiques nationales
4	Analyser les statistiques internationales
5	Gérer les ventes
6	Gérer les clients
7	Trier les ventes
8	Trier par date
9	Trier par région
10	Trier par client
11	Consulter une vente
12	Consulter un client
13	Prédire les ventes
14	Rechercher

TABLE 3.3 : Les cas d'utilisation du système.

3.5 Besoins fonctionnels

Il faut souligner que durant cette activité, les deux étapes présentées précédemment (analyse des situations décisionnelles et l'identification des cas d'utilisation) aident à faire ressortir les besoins fonctionnels. Le recueil et la collecte des besoins se font d'une façon itérative jusqu'à leur validation. Ces besoins expriment des scénarios pouvant prendre la forme de cas d'utilisation.

Les besoins fonctionnels expriment les actions que doit effectuer le système en réponse à une demande [37].

Le système que nous allons développer doit couvrir les points suivants :

Authentification : pour que le décisionnaire accède aux différentes fonctionnalités du système, il doit d'abord s'authentifier avec son nom d'utilisateur et son mot de passe (cas 1) ;

Analyse : le décisionnaire pourra analyser globalement les informations stockées dans la base de données et représentées sous forme de graphiques (cas de 2 à 6) ;

Tri : le décisionnaire pourra trier les ventes selon trois facteurs : par client, par date ou par région (cas de 7 à 10) ;

Consultation : le décisionnaire doit pouvoir consulter un élément précis et avoir les détails concernant cet élément (cas 11 et 12) ;

Prédiction : le décisionnaire pourra afficher le résultat des prédictions effectuées après avoir rempli et validé les informations nécessaires pour que l'API lui renvoie les résultats de prédiction (cas 13) ;

Recherche : le décisionnaire doit pouvoir effectuer une recherche d'un élément (cas 14).

3.6 Besoins non fonctionnels

Ce sont les besoins en matière de performance ou de type de matérielle [38]. Dans le cadre de ce travail le système devra répondre aux caractéristiques suivantes :

Performance : l'application doit répondre aux exigences de l'utilisateur dans un temps de réponse minimale (un entrepôt de données possède beaucoup d'informations, il faut donc faire en sorte que les résultats s'affichent le plus rapidement possible) ;

Ergonomie : l'interface doit être conviviale et simple à utiliser, les informations doivent être lisibles pour que l'utilisateur pourra mieux gérer son espace de travail ;

Modularité du code : le code doit être lisible pour faciliter la maintenance ;

Sécurité : le système doit permettre de gérer l'accès de l'utilisateur selon un privilège.

La démarche centrée utilisateur est basée sur l'expression des besoins. Elle consiste en l'analyser des objectifs du SIAD en termes des besoins fonctionnels et non fonctionnels comme nous venons de le voir et en terme des besoins relatifs à l'IHM

3.7 Élaboration des premières maquettes d'IHM

Nous avons élaboré nos premières maquettes en vue d'aider à identifier les fonctions principales du SIAD et à définir la structure globale des données ainsi que les actions que chaque utilisateur est censé exécuter en utilisant l'interface Homme-Machine appropriée.

L'élaboration de premières maquettes d'interfaces du SIAD permettent d'impliquer le plus rapidement possible les futurs utilisateurs en leur donnant un aperçu de solutions possibles. Ce type de maquettage permet de définir toutes les contraintes ergonomiques liées à la mission de l'utilisateur et à son contexte de travail en termes de besoins, car la maquette dérive souvent d'une analyse de l'activité ou d'un recueil des besoins utilisateurs.

Deux exemple de maquettes élaborées sont donnés par les figures 3.2 et 3.3. La figure 3.2 représente la maquette de l'interface d'analyse et la figure 3.3 représente la maquette de l'interface de prédiction.

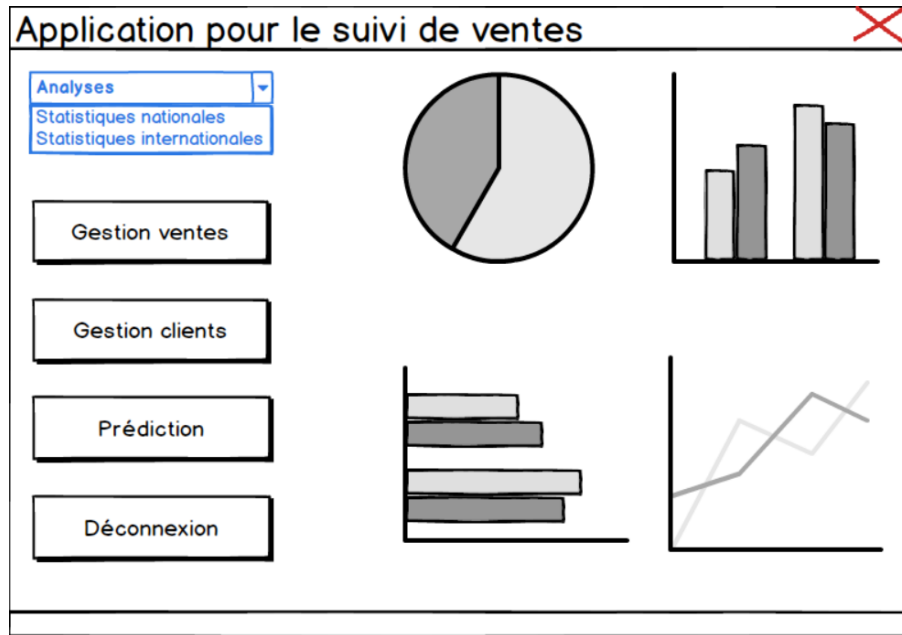


FIGURE 3.2 : Maquette de l'interface d'analyse.

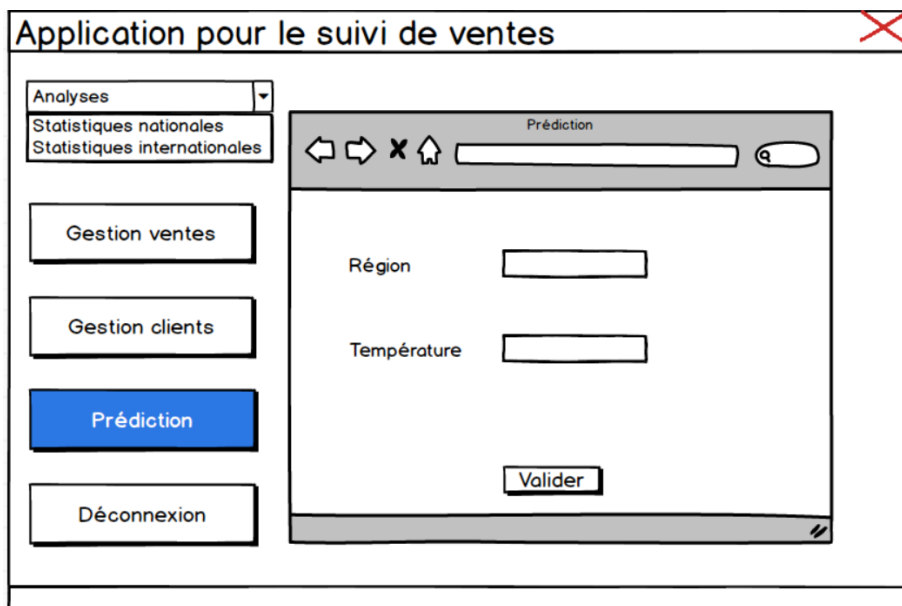


FIGURE 3.3 : Maquette de l'interface de prédiction.

3.8 Diagramme de cas d'utilisation

C'est un formalisme qui permet de modéliser le fonctionnement d'un système par un découpage de celui-ci en fonctionnalités. Il illustre de plus la nature des interactions avec ces fonctionnalités offertes à titre de services à des acteurs externes au système. Le découpage peut être global ou très éclaté. La nature des interactions peut être décrite de manière sommaire ou détaillée selon le niveau de détail recherché par le modélisateur [39]. Le diagramme des cas d'utilisation est représenté dans la figure 3.4.

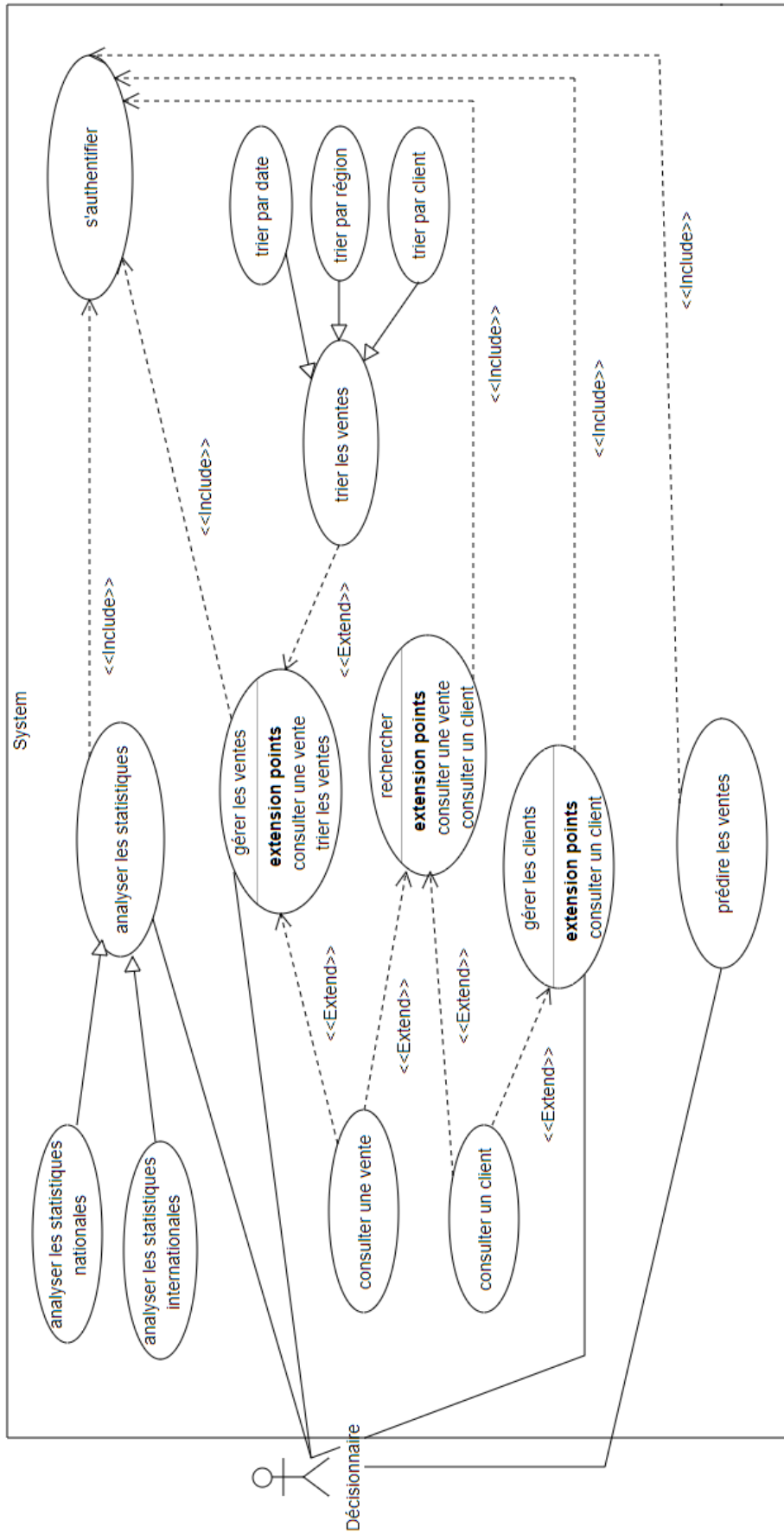


FIGURE 3.4 : Diagramme de cas d'utilisation.

3.9 Description des cas d'utilisation

Nous réalisons une fiche descriptive et un diagramme de séquences pour les cas d'utilisation pertinents de façon à présenter le déroulement des différentes actions.

Description textuelle

Le formalisme suivi pour la description textuelle des cas d'utilisation de notre système est décrit dans le tableau 3.4.

Nom du cas d'utilisation	
Acteurs	Les acteurs participants au cas d'utilisation.
Description	Une description succincte du cas d'utilisation.
Auteurs	Ceux qui sont chargés de rédiger la fiche.
Date	La date de rédaction ou de mise à jours de la fiche.
Préconditions	Les conditions obligatoires au bon déroulement du cas d'utilisation.
Description des scénarios	
Scénario nominal	La description de déroulement idéal des actions, où tout va pour le mieux.
Scénario alternatif/exception	La description d'une suite d'étapes qui se termine : soit par la satisfaction, soit par l'abandon de l'objectif.
Post-conditions	Le résultat tangible qui est vérifiable après l'arrêt du cas d'utilisation et qui pourra témoigner du bon fonctionnement.

TABLE 3.4 : Formalisme de description des cas d'utilisation.

Diagramme de séquence

Chaque cas d'utilisation peut donné lieu à un (ou plusieurs) diagramme de séquences simple représentant graphiquement la chronologie des interactions entre les acteurs et le système vu comme une boîte noire. Il montre également les événements systèmes déclenchés par les acteurs. L'ordre chronologique des messages doit suivre la séquence décrite dans le cas d'utilisation [39].

3.9.1 Cas d'utilisation « S'authentifier »

La description textuelle de ce cas est donnée dans le tableau 3.5.

S'authentifier	
Acteurs	Le décisionnaire.
Description	Permet d'accéder aux différentes fonctionnalités de l'application.
Auteurs	AIROUCHE Cylia LOUBAR Nacera
Date	01/07/2020
Préconditions	L'utilisateur existe dans la base de données.
Description des scénarios	
Scénario nominal	<ol style="list-style-type: none"> 1. Le système affiche l'interface d'authentification après son lancement par l'utilisateur ; 2. L'utilisateur fournit un nom d'utilisateur et un mot de passe ; 3. Le système vérifie les informations envoyées, puis il donne l'accès à l'utilisateur afin d'accéder à son espace de travail.
Scénario alternatif/exception	<ol style="list-style-type: none"> 3.1. L'utilisateur n'a pas saisi les bons identifiants ; 3.2. Le système renvoie un message d'erreur et signale à l'utilisateur de recommencer (revenir à 1).
Post-conditions	<ul style="list-style-type: none"> — Utilisateur authentifié ; — La page d'accueil s'affiche.

TABLE 3.5 : Description du cas d'utilisation « S'authentifier ».

Le diagramme de la figure 3.5 représente le scénario lorsque l'utilisateur essaye de se connecter au système.

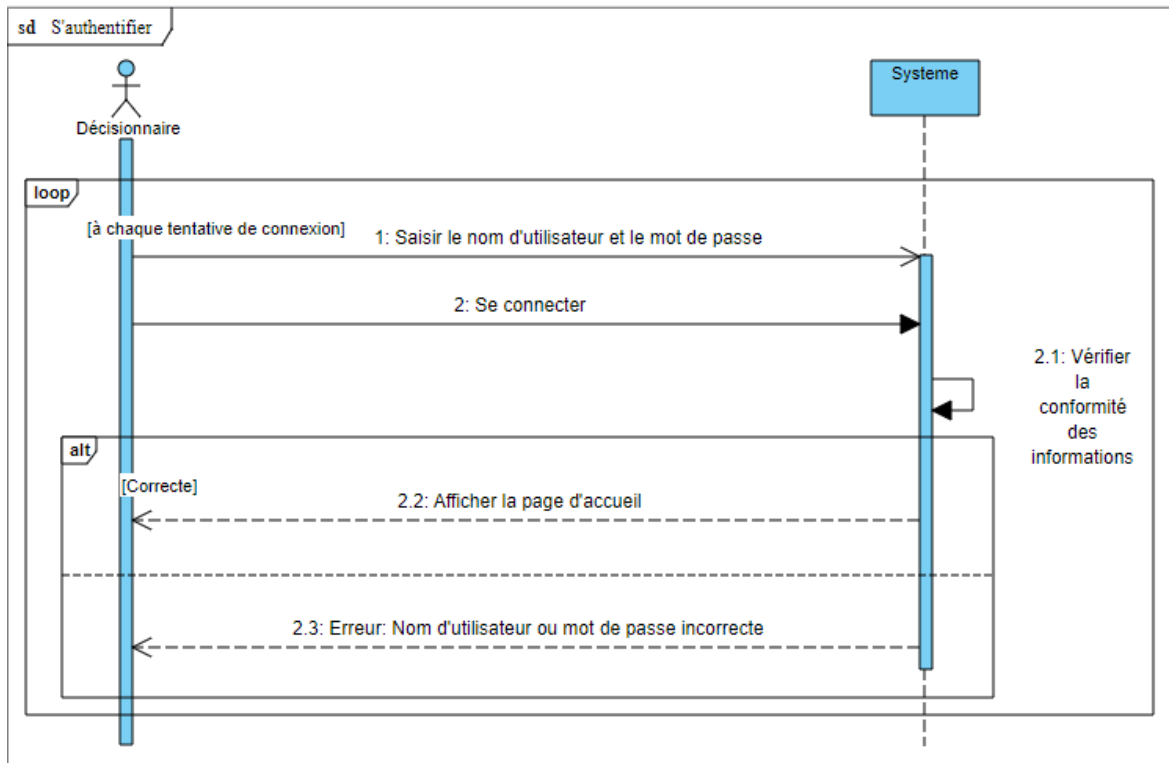


FIGURE 3.5 : Diagramme de séquence système du cas d'utilisation « S'authentifier ».

3.9.2 Cas d'utilisation « Gérer les ventes »

La description textuelle de la gestion des ventes est donnée dans le tableau 3.6.

Gérer les ventes	
Acteurs	Le décisionnaire.
Description	Permet à l'utilisateur d'analyser les données à travers des graphiques.
Auteurs	AIROUCHE Cylia LOUBAR Nacera
Date	01/07/2020
Préconditions	L'utilisateur doit être authentifié.
Description des scénarios	
Scénario nominal	<ol style="list-style-type: none"> 1. Le décisionnaire demande l'accès à l'espace d'analyse des ventes ; 2. Le système lui affiche la page demandée.
Scénario alternatif/exception	Aucun
Post-conditions	Les ventes s'affichent.

TABLE 3.6 : Description du cas d'utilisation « Gérer les ventes ».

Le diagramme de la figure 3.6 représente le scénario qui s'exécute lorsque le décisionnaire demande de gérer les ventes.

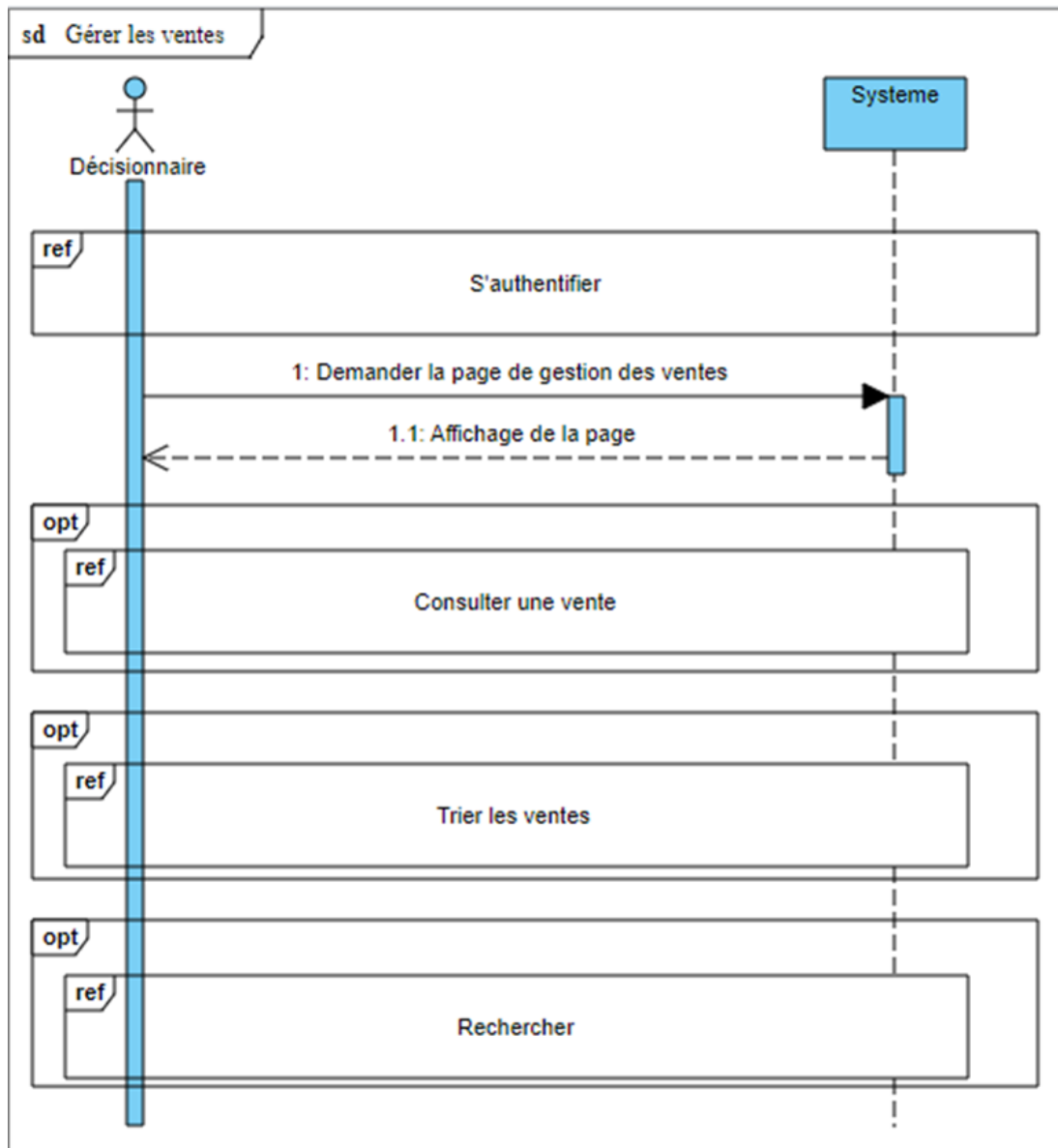


FIGURE 3.6 : Diagramme de séquence système du cas d'utilisation « Gérer les ventes ».

3.9.3 Cas d'utilisation « Prédire les ventes »

La description textuelle de la prédiction des ventes est donnée dans le tableau 3.7.

Prédire les ventes	
Acteurs	Le décisionnaire.
Description	Permet de prédire la quantité du produit vendu en fonction des différents facteurs.
Auteurs	AIROUCHE Cylia LOUBAR Nacera
Date	01/07/2020
Préconditions	Les informations doivent être remplies et validées par le décisionnaire.
Description des scénarios	
Scénario nominal	<ol style="list-style-type: none"> 1. Le décisionnaire demande d'accéder à l'espace de prédiction ; 2. Le système affiche le formulaire ; 3. Le décisionnaire remplit et valide les champs du formulaire ; 4. L'API renvoie le résultat de la prédiction ; 5. Le système reçoit et affiche le résultat.
Scénario alternatif/exception	<ol style="list-style-type: none"> 3.1. L'utilisateur a mal rempli les champs ; 3.2. Le système renvoi un message d'erreur et demande de ressaisir (revenir à 2).
Post-conditions	Résultat envoyé et affiché.

TABLE 3.7 : Description du cas d'utilisation « Prédire les ventes ».

Le diagramme de la figure 3.7 représente le scénario qui s'exécute entre le système et le décisionnaire lorsqu'il veut afficher une prédiction.

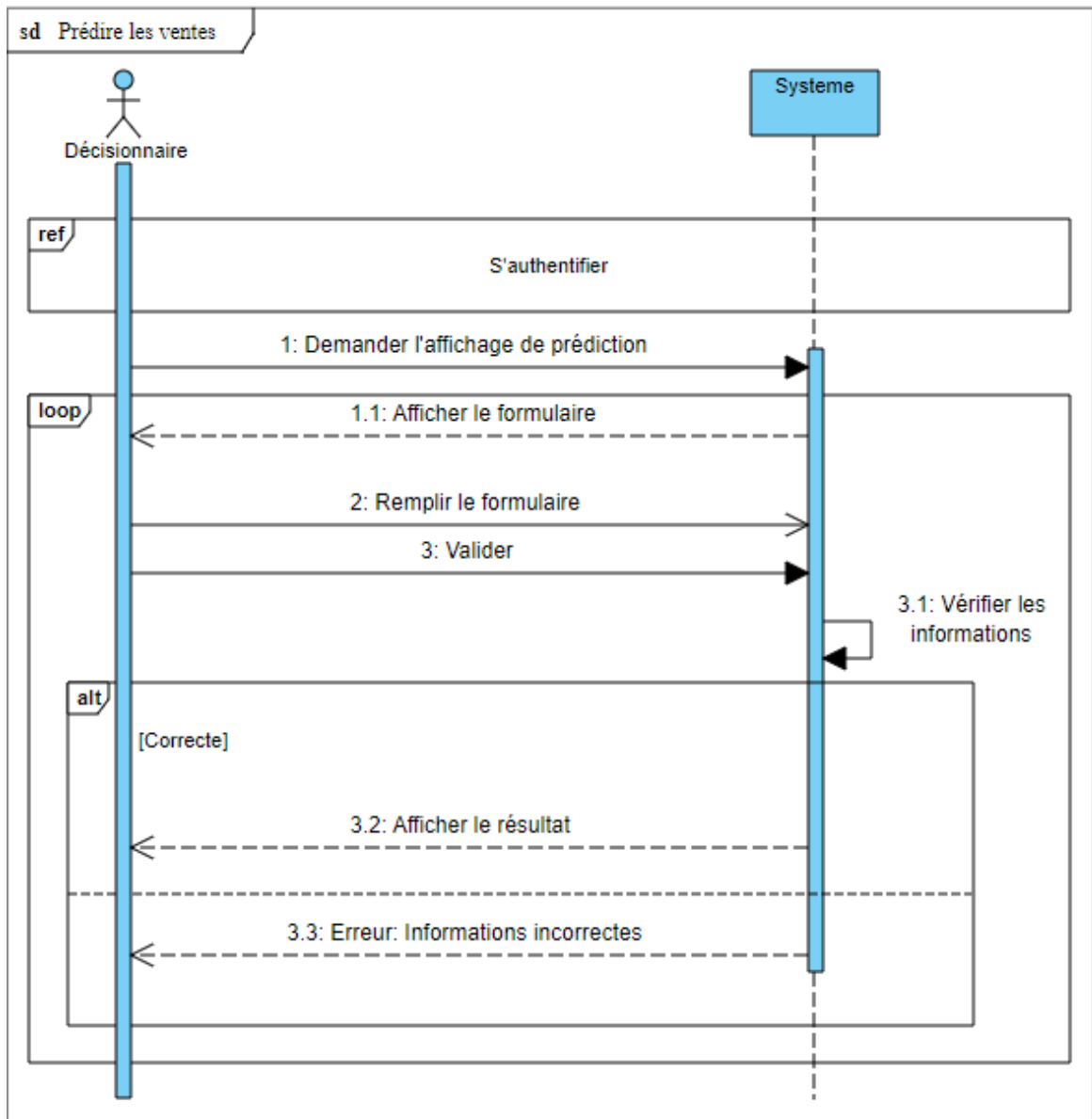


FIGURE 3.7 : Diagramme de séquence système du cas d'utilisation « Prédire les ventes ».

3.9.4 Cas d'utilisation « Rechercher »

La description textuelle de l'opération de recherche est donnée dans le tableau 3.8.

Rechercher	
Acteurs	Le décisionnaire.
Description	Permet à l'utilisateur de faire une recherche concernant la vente ou le client.
Auteurs	AIROUCHE Cylia LOUBAR Nacera
Date	01/07/2020
Préconditions	L'utilisateur doit être authentifié et accédé à l'espace de gestion des ventes ou des clients.
Description des scénarios	
Scénario nominal	<ol style="list-style-type: none"> 1. L'utilisateur saisie un mot clé dans la barre de recherche puis valide la recherche ; 2. Le système affiche le résultat de la recherche.
Scénario alternatif/exception	<ol style="list-style-type: none"> 2.1. L'utilisateur n'a effectué aucune recherche ; 2.2. Le système demande à l'utilisateur de saisir un mot clé ; 2.3. L'utilisateur a saisi un mot inexistant dans la BDD ; 2.4. Le système demande à l'utilisateur de saisir un mot clé correcte (revenir à 1).
Post-conditions	Résultat affiché.

TABLE 3.8 : Description du cas d'utilisation « Rechercher ».

Le diagramme de la figure 3.8 représente le scénario qui s'exécute lorsque le décisionnaire fait une recherche depuis les interfaces de gestion des ventes ou des clients.

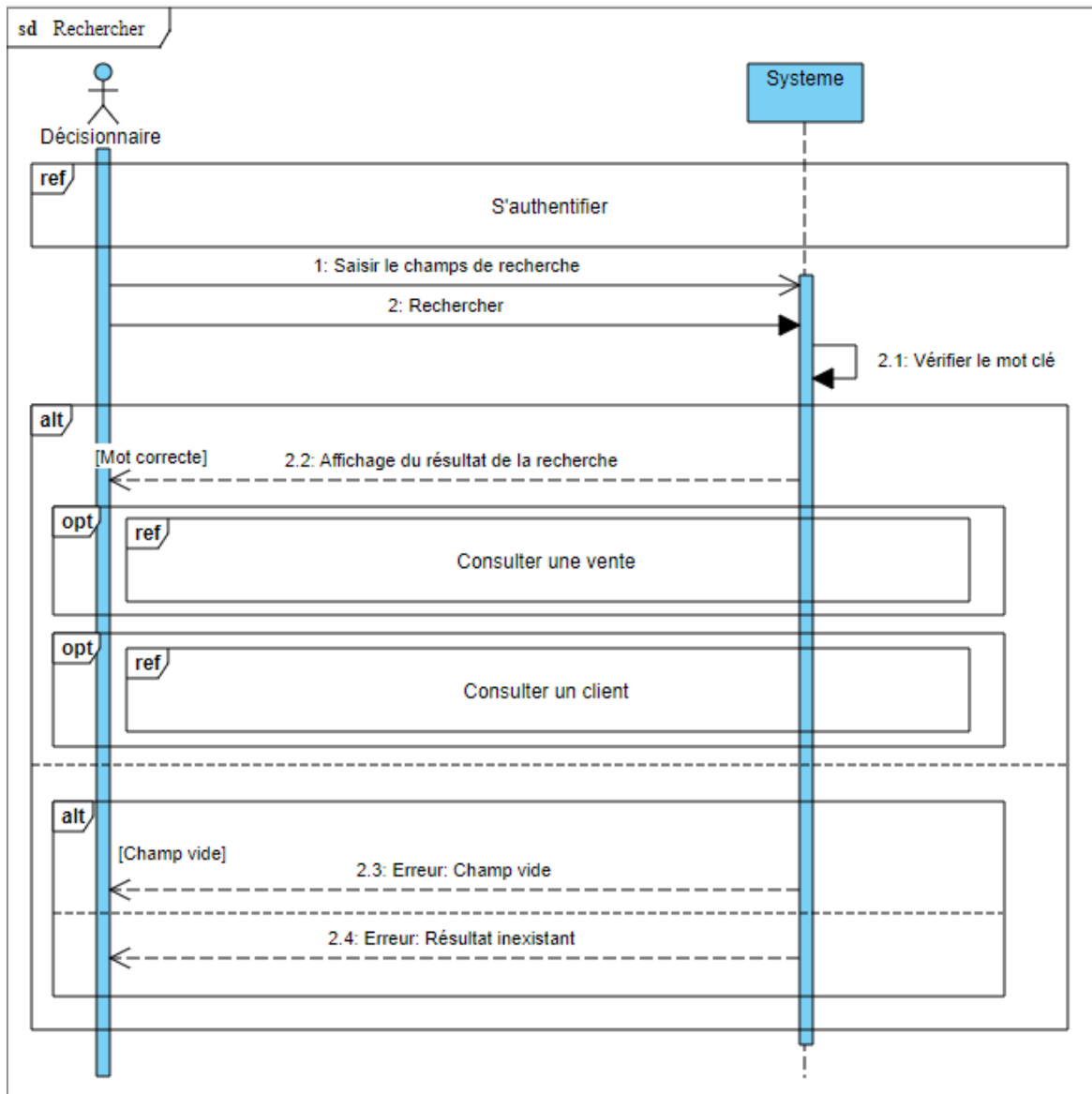


FIGURE 3.8 : Diagramme de séquence système du cas d'utilisation « Rechercher ».

3.10 Conclusion

Dans ce chapitre nous nous sommes focalisés sur l'étude des besoins de notre système, ce qui nous a aidé à identifier ses principales fonctionnalités en tenant compte de la situation décisionnelle de l'entreprise et les besoins en terme de l'IHM du système.

Ce chapitre nous servira de point de base à l'élaboration de la prochaine phase, qui est conception (préliminaire et détaillée).

Chapitre 4

Conception

4.1 Introduction

Ce chapitre présentera la phase de conception qui représente une phase primordiale et déterminante pour produire une application de haute qualité. Après avoir vu les diagrammes de séquences, nous allons présenter les diagrammes d'interaction qui leur sont associés, ainsi que le diagramme de navigation. Ensuite, nous allons concevoir un diagramme de classes à partir duquel nous établirons un Datawarehouse suivant le schéma en flocon.

4.2 Diagramme d'interaction

Chaque diagramme d'interaction va ainsi représenter un ensemble d'objets de classes différentes collaborant dans le cadre d'un scénario d'exécution du système. Dans ce genre de diagramme, les objets communiquent en s'envoyant des messages qui invoquent des opérations sur les objets récepteurs. Ces objets sont des instances des trois types de classes d'analyse (interfaces, contrôles et entités), ils remplacent le système qui est présenté comme une boîte noire [40].

4.2.1 S'authentifier

L'utilisateur doit saisir ses identifiants dans le formulaire d'authentification. Suivant le résultat du contrôle effectué par le système, il se retrouve soit sur la page d'accueil, soit de nouveau sur le formulaire avec un message d'erreur. Le scénario qui s'exécute est représenté par la figure 4.1.

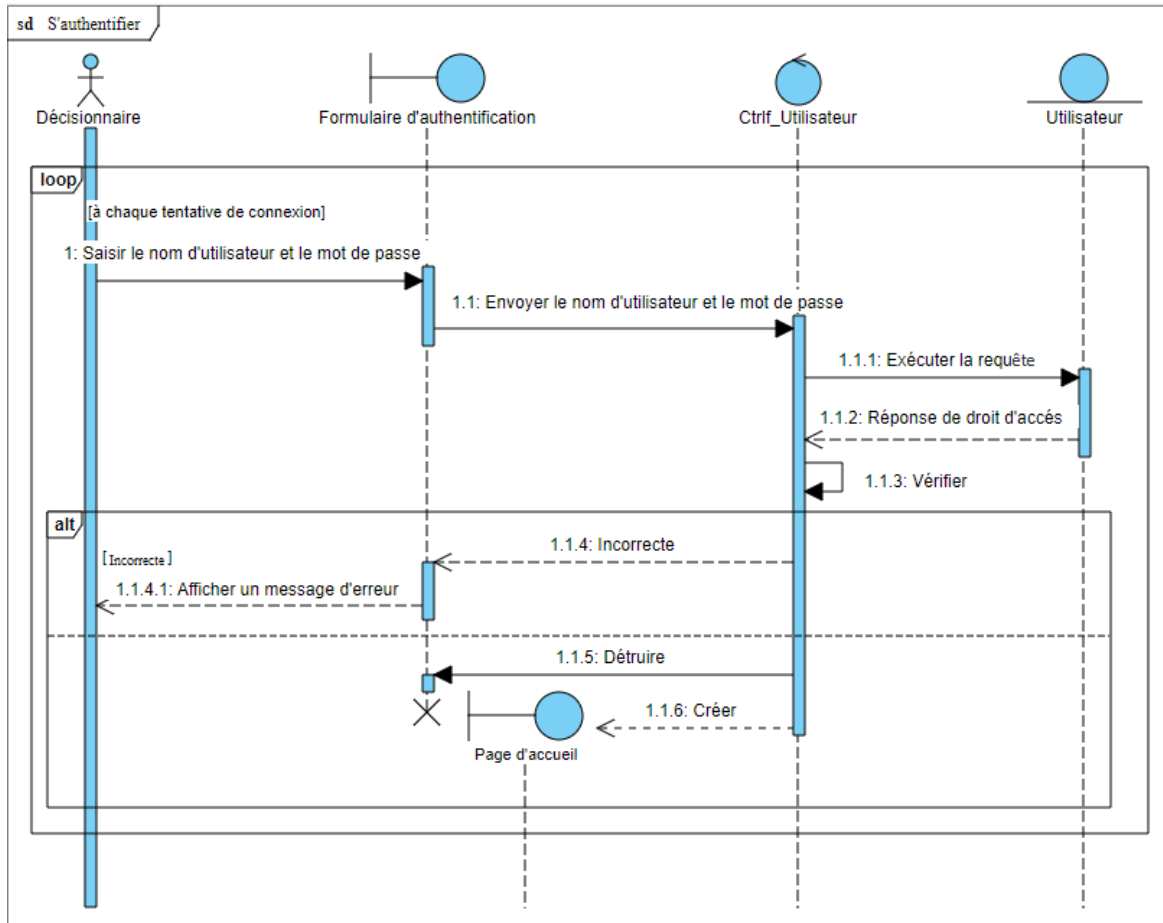


FIGURE 4.1 : Diagramme d'interaction du cas d'utilisation « S'authentifier ».

4.2.2 Gérer les ventes

Après l'authentification, l'utilisateur se retrouve dans la page d'accueil. Il accédera à la page de gestion des ventes pour faire les analyses, il pourra effectuer une recherche, une consultation ou un tri. Le scénario qui s'exécute est représenté dans la figure 4.2.

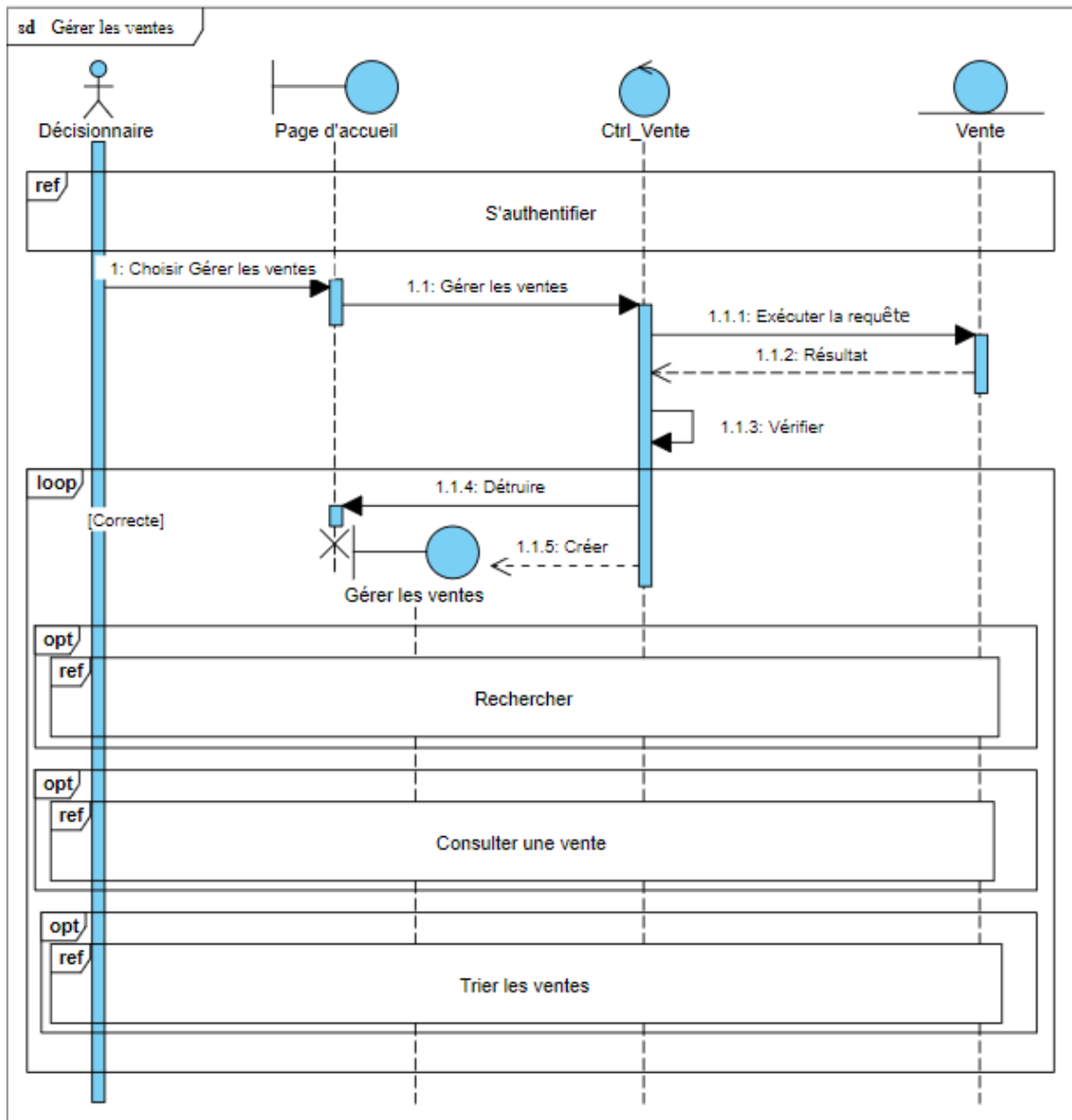


FIGURE 4.2 : Diagramme d'interaction du cas d'utilisation « Gérer les ventes ».

4.2.3 Prédire les ventes

Après l'authentification, l'utilisateur se retrouve dans la page d'accueil. Il accédera à la page de prédiction des ventes, il aura le résultat de prédiction après avoir remplis le formulaire correctement. Le scénario qui s'exécute est représenté dans la figure 4.3.

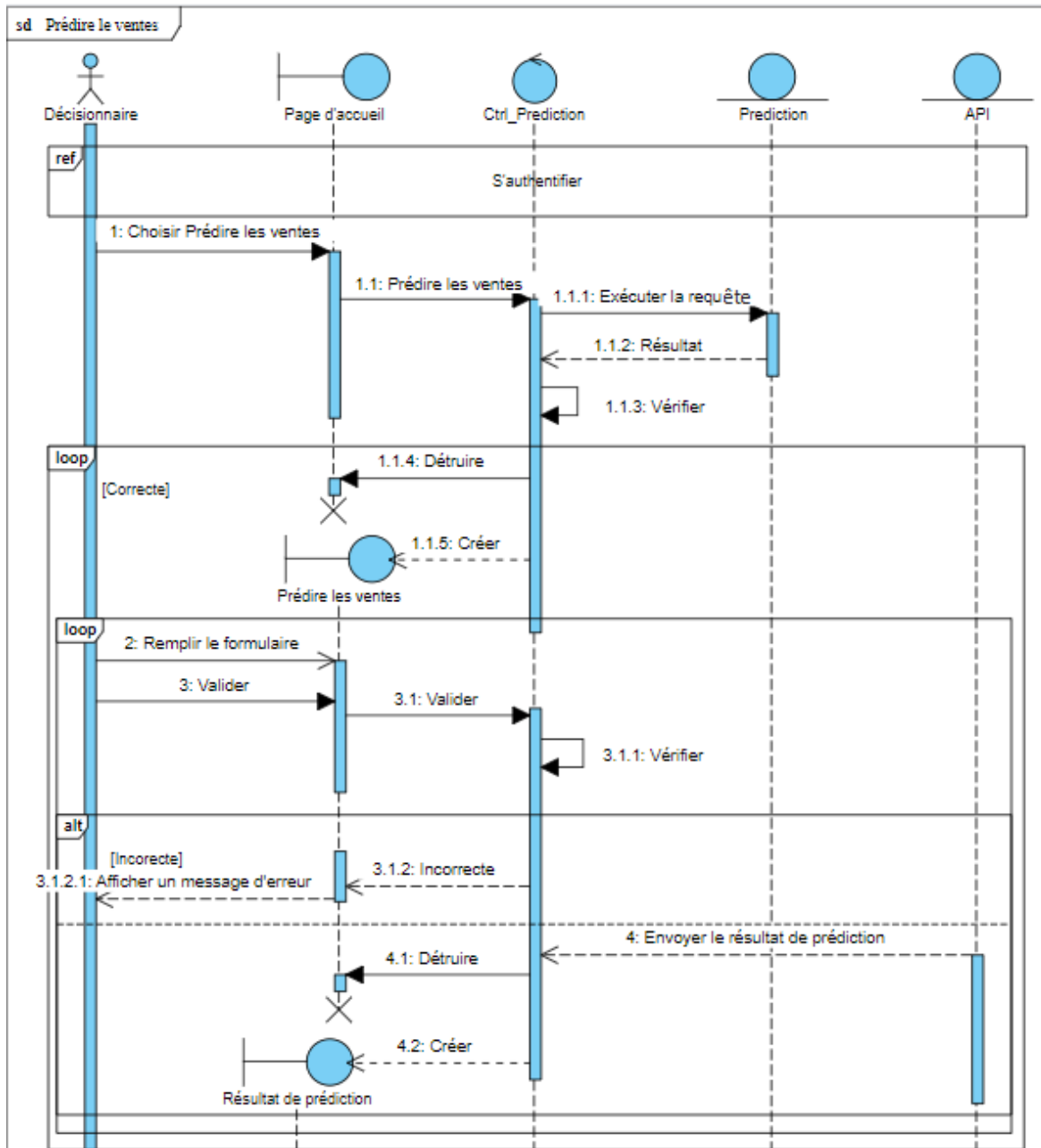


FIGURE 4.3 : Diagramme d'interaction du cas d'utilisation « Prédire les ventes ».

4.2.4 Rechercher

L'utilisateur doit s'authentifier et il doit accéder à la page de gestion de ventes ou celle de gestion des clients pour effectuer une recherche. Suivant le mot clé saisi, il se retrouve soit sur la page de résultat de recherche, soit de nouveau sur l'une des dites interfaces avec un message d'erreur. Le scénario qui s'exécute est représenté dans la figure 4.4.

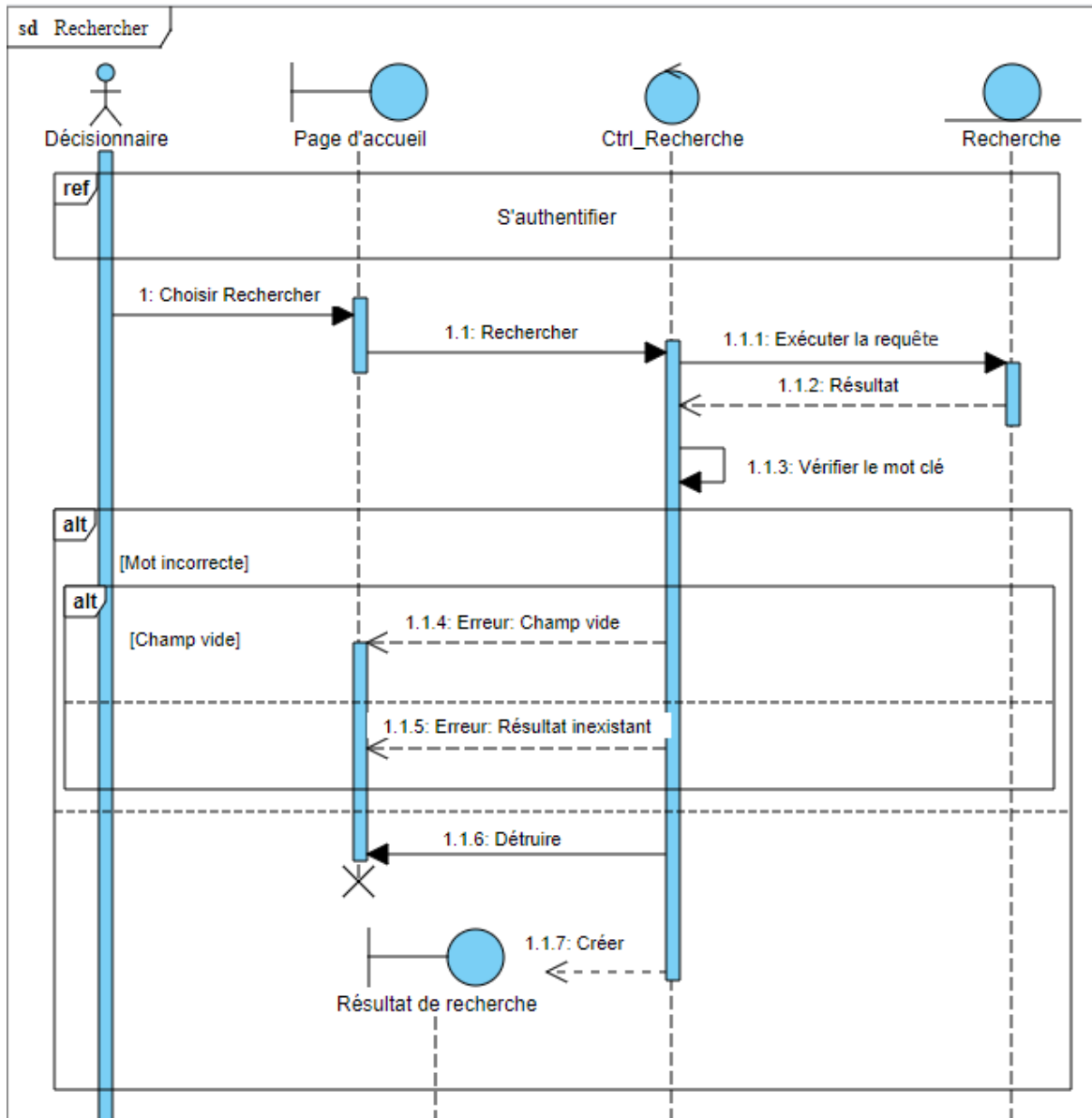


FIGURE 4.4 : Diagramme d'interaction du cas d'utilisation « Rechercher ».

4.3 Diagramme de navigation

Le diagramme de navigation est un diagramme dynamique représentant de manière formelle l'ensemble des chemins possibles entre les principales interfaces proposées à l'utilisateur. Il est considéré comme une vue concrète mais non définitive de l'IHM [41].

La figure 4.5 représente le diagramme de navigation de notre système.

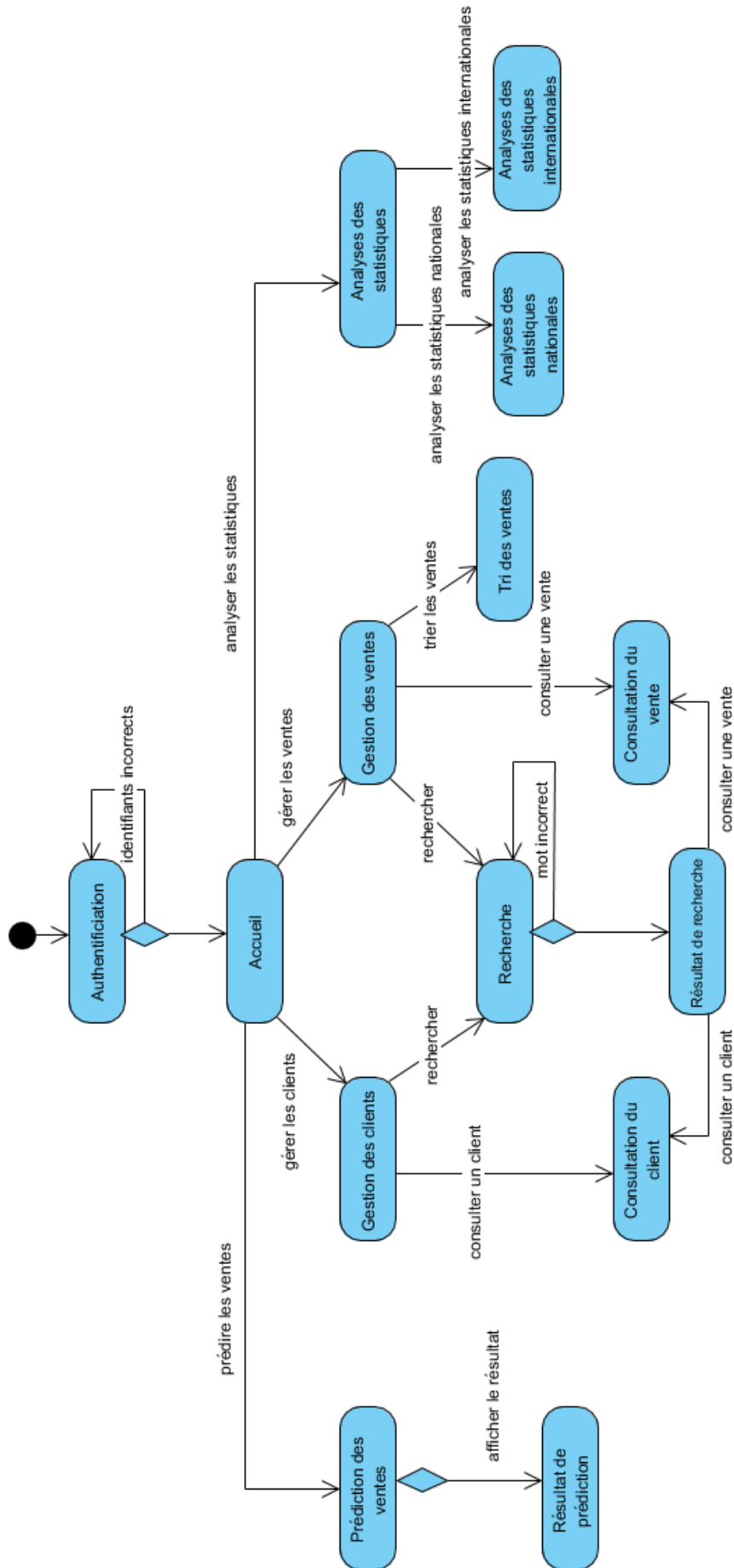


FIGURE 4.5 : Diagramme de navigation.

4.4 Diagramme de classes

Le diagramme de classes est un diagramme structurel qui décrit un ensemble de classes, d'interfaces et leurs relations [42].

Le diagramme de classes est obtenu à partir des diagrammes précédents et représenté par la figure 4.6. Le dictionnaire des données qui vient pour compléter le diagramme de classes est donné dans le tableau 4.1. A noter que les ventes peuvent changer d'une saison à une autre dans une région en fonction de la température. En effet, généralement, les ventes des quantités d'eau (ou des boissons) croissent avec l'augmentation des températures. Ces relations peuvent être facilement tirées du diagramme de classes. Cependant, elles seront modélisées d'une manière explicite dans l'entrepôt de données.

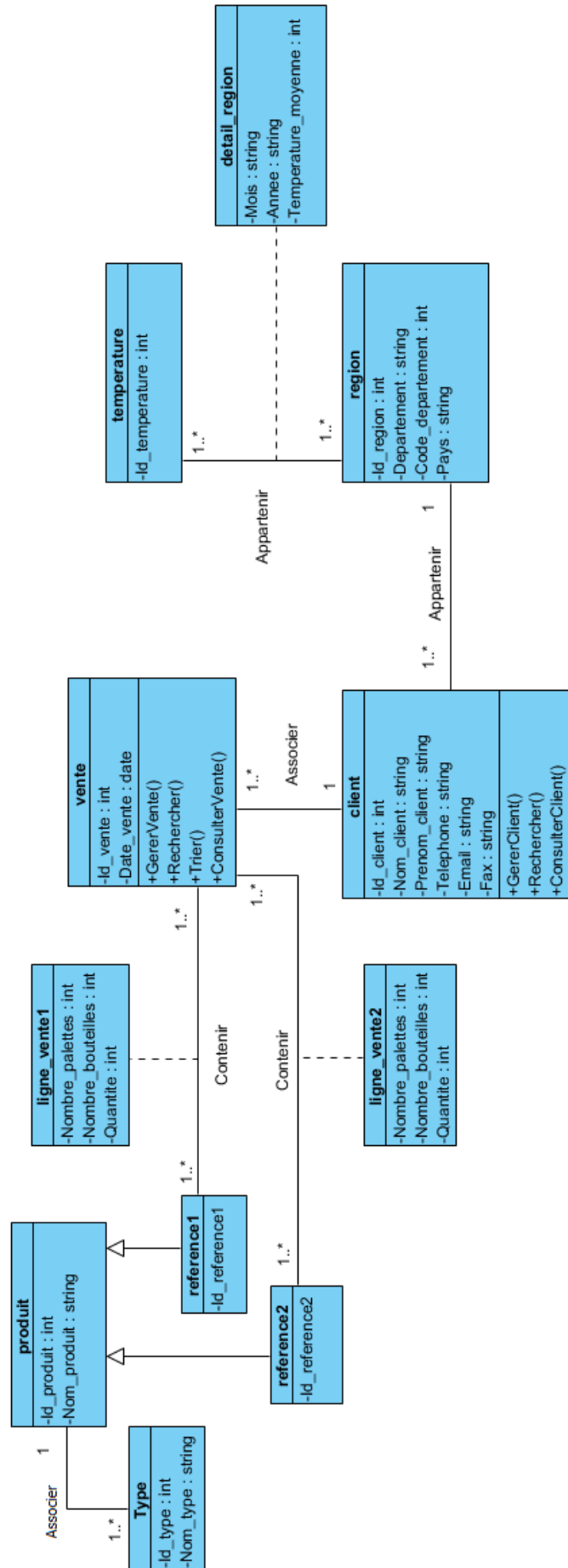


FIGURE 4.6 : Diagramme de classes.

Classe	Attribut	Description	Type
Produit	Id_Produit	L'identifiant du produit	Entier
	Nom_produit	Le nom du produit	Texte
Type	Id_type	L'identifiant du type	Entier
	Nom_type	Type de bouteille du produit (verre/PET ...)	Texte
Reference1	Id_reference1	L'identifiant de la référence 1,5 litre	Texte
Reference2	Id_reference2	L'identifiant de la référence 0,5 litre	Texte
Client	Id_client	L'identifiant du client	Entier
	Nom_client	Le nom du client	Texte
	Prenom_client	Le prénom du client	Texte
	Telephone	Le numéro de téléphone du client	Texte
	Email	L'adresse email du client	Texte
	Fax	Le numéro de fax du client	Texte
Région	Id_region	L'identifiant de la région	Entier
	Departement	Le nom du département/ wilaya de la région	Texte
	Code_departement	Le numéro du département/ wilaya	Entier
	Pays	Le nom du pays	Texte
Vente	Id_vente	L'identifiant de vente	Entier
	Date_vente	La date de la vente	Date(JJ/MM/AAAA)
Temperature	Id_temperature	L'identifiant de la température	Entier
	Nombre_palettes	Le nombre de palettes	Entier
ligne_vente1	Nombre_bouteilles	Le nombre de bouteilles	Entier
	Quantité	La quantité vendue	Entier
ligne_vente2	Nombre_palettes	Le nombre de palettes	Entier
	Nombre_bouteilles	Le nombre de bouteilles	Entier
Detail_region	Quantité	La quantité vendue	Entier
	Mois	Le mois correspondant à la température et la région	Texte
	Annee	L'année correspondante	Texte
	Temperature_moyenne	La température moyenne de la région	Entier

TABLE 4.1 : Dictionnaire de données.

La section suivante est consacré pour la conception de l'entrepôt de données avec la modélisation des tables dimensions et la table faits.

4.5 Conception de l'entrepôt de données

L'entrepôt de données est considéré comme étant un support décisionnel robuste, il permet des pratiques commerciales supérieures et un traitement analytique efficace. Nous allons faire une modélisation en flocon qui semble plus adéquate dans notre cas, puisque la dimension température dépend de deux autres dimensions qui sont la date et la région.

4.5.1 Dimensions participantes du modèle

Les dimensions sont un concept essentiel dans les bases de données multidimensionnelles. Elles sont les critères suivant lesquels on souhaite évaluer, quantifier et qualifier le fait. Elles représentent les centres d'intérêts décisionnels analysés, c'est pour cela on va accumuler toutes les informations nécessaires pour les dimensions participantes du modèle.

4.5.1.1 Dimension date

La dimension date de la figure 4.7 est la seule dimension qui figure systématiquement dans tout entrepôt de données puisque le temps est toujours l'une des perspectives d'analyse.



FIGURE 4.7 : Dimension date.

4.5.1.2 Dimension produit

La dimension produit de la figure 4.8 contient des informations désirables pour une analyse. Elle décrit les différentes caractéristiques d'un produit.

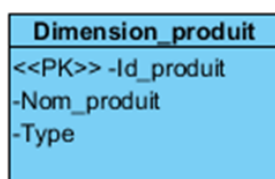


FIGURE 4.8 : Dimension produit.

4.5.1.3 Dimension catégorie

Cette dimension (voir la figure 4.9) représente un axe qui permet de faire l'analyse de l'ensemble des indicateurs selon la catégorie du produit.

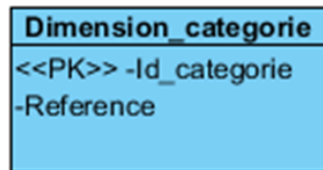


FIGURE 4.9 : Dimension catégorie.

4.5.1.4 Dimension client

La dimension client de la figure 4.10 représente l'un des axes d'analyse dans l'opération de vente. L'analyse du comportement du client peut aider l'entreprise à mieux le satisfaire.

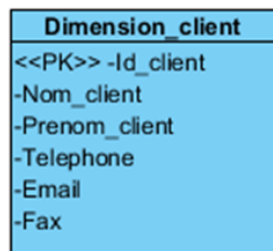


FIGURE 4.10 : Dimension client.

4.5.1.5 Dimension région

Cette dimension représente un axe qui permet de faire l'analyse de l'ensemble des indicateurs selon la région (voir la figure 4.11).

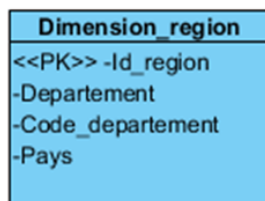


FIGURE 4.11 : Dimension région.

4.5.1.6 Dimension température

Cette dimension de la figure 4.12 représente un axe qui permet de faire l'analyse de l'ensemble des indicateurs selon la température.

Dimension_temperature
<<FK>> <<PK>> -ld_date
<<FK>> <<PK>> -ld_region
-Temperature_moyenne

FIGURE 4.12 : Dimension température.

4.5.2 Faits ventes

Le choix des indicateurs et la déduction des dimensions d'analyse nous ont mené à élaborer une table de faits qui est représentée dans la figure 4.13.

Faits_ventes
<<FK>> <<PK>> -ld_client
<<FK>> <<PK>> -ld_produit
<<FK>> <<PK>> -ld_date
<<FK>> <<PK>> -ld_region
<<FK>> <<PK>> -ld_categorie
-Nombre_palettes
-Nombre_bouteilles
-Quantite

FIGURE 4.13 : Faits ventes.

4.5.3 Modèle en flocon

La figure 4.14 montre notre modèle en flocon complet de notre entrepôt de données y compris toutes les dimensions et indicateurs :

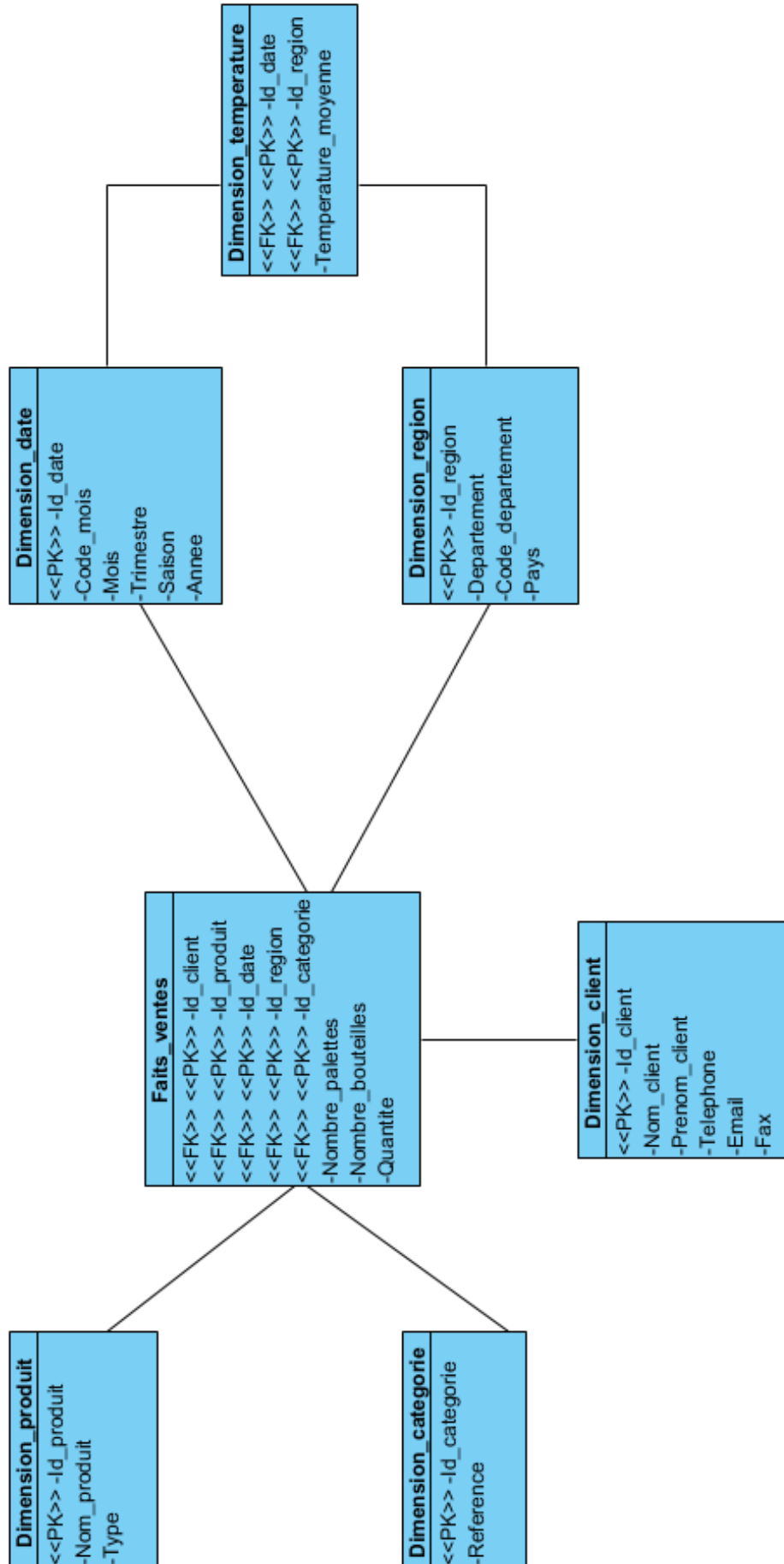


FIGURE 4.14 : Modèle en flocon de notre entrepôt de données.

4.6 Conclusion

Ce chapitre a permis de présenter en détail les différentes fonctionnalités de notre système à travers les diagrammes d'interaction, ainsi que de concevoir les diagrammes de navigation et classes. Enfin, nous avons modélisé l'entrepôt de données en précisant le rôle de la table des faits et des tables des dimensions.

Ce chapitre est une ébauche de l'étape de la réalisation de notre système que nous allons présenter dans le chapitre suivant.

Chapitre 5

Réalisation

5.1 Introduction

Dans ce dernier chapitre, nous allons décrire la phase d'implémentation et réalisation ; qui est une étape très importante dans le cycle de vie d'une application. Nous allons d'abord définir les outils de développement et les langages de programmation utilisés, puis décrire la mise en œuvre de l'entrepôt de données. Nous allons également présenter les étapes suivies pour la modélisation des prédictions à l'aide de l'apprentissage automatique. Nous terminons le chapitre par la présentation de quelques interfaces graphiques et jeux de tests.

5.2 Environnement de développement de l'application

5.2.1 Environnement matériel

Les principales caractéristiques des machines utilisées pour implémenter et tester notre application sont :

- PC 1
 - Type de la machine : ASUS
 - Processeur : Intel(R) Preinium (R) CPU N3540 @2.16GHz 2.16GHz
 - RAM : 4 Go
 - Système d'exploitation : Windows 10 (64 bits)
- PC 2 :
 - Type de la machine : ACER
 - Processeur : Intel(R) core(TM) i7-3632QM CPU @2.20GHz 2.20 GHz
 - RAM : 8 Go
 - Système d'exploitation : Windows 10 (64 bits)

5.2.2 Environnement logiciel

Cette sous-section présente les différent outils et langages utilisés pour la réalisation de l'application.

5.2.2.1 Outils de développement

Les différents outils de développement utilisés pour la réalisation du système sont définis comme suit :

Balsamic mockups

Balsamiq est un excellent outil, répondant à toutes les exigences du Wireframing, de la collaboration et de la créativité. Son ensemble unique de fonctionnalités permet au membre de l'équipe de faire du Wireframing rapide, d'obtenir un consensus sur la fonctionnalité à développer. Cela finira par aligner l'équipe sur les fonctionnalités en termes de mises en page [43].

Visual paradigm

Permet la création des diagrammes UML et des modèles qui en sont à l'origine. Ceux-ci peuvent alors générer du code dans un langage de programmation déterminé. Il propose également la création d'autres types de diagrammes, comme celui qui permet la modélisation des bases de données pouvant, lui aussi, générer des canevas d'applications basé sur des Framework et Pattern mais en plus, générer du code SQL qu'il peut ensuite déployer automatiquement dans différents environnements [44].

Talend Open Studio

Talend Open Studio (TOS), est la solution d'intégration de données la plus ouverte, la plus innovante et la plus puissante qui soit disponible sur le marché. Téléchargeable sous forme d'exécutable prêt à installer, Talend Open Studio répond à la plupart des besoins en intégration de données des entreprises, quelle que soit leur taille ou leur niveau d'expertise en intégration de données [45].

NetBeans

NetBeans IDE est un environnement de développement intégré (IDE) modulaire basé sur des normes, écrit dans le langage de programmation Java. En plus de Java, il permet également de supporter différents autres langages, comme C, C++, XML et HTML [46].

Le projet de NetBeans IDE consiste en un IDE Open Source complet et en une plateforme d'application cliente riche, qui peut être utilisée comme structure générique pour créer n'importe quel type d'application [47].

Scene builder

Scene Builder est un outil de présentation visuelle qui permet aux utilisateurs de concevoir rapidement des interfaces utilisateur d'application Javafx, sans codage. Les utilisateurs peuvent glisser et déposer des composants de l'interface utilisateur dans une zone de travail, modifier leurs propriétés, appliquer des feuilles de style et le code FXML de la mise en page qu'ils créent est automatiquement généré en arrière-plan. Le résultat est un fichier FXML qui peut ensuite être combiné avec un projet Java en liant l'interface utilisateur à la logique de l'application [48].

Sublime text

Sublime text est un éditeur de texte générique codé en C++ et Python, disponible sur Windows, Mac et Linux. Le logiciel prend en charge des langages de programmation

majeurs, tandis que des plugins sont souvent disponibles pour les langages plus rares [49].

JDK

Le JDK (Java Développment Kit) représente la collection de ces bibliothèques, y compris le " javac " compilateur Java. Le JDK contient le compilateur, les bibliothèques de programmation et le JRE afin de présenter aux programmeurs une plateforme qui permet de compiler et exécuter des programmes Java [47].

JDBC

La technologie JDBC (Java DataBase Connectivity) est une API fournie avec Java permettant de se connecter à des bases de données, c'est-à-dire que JDBC constitue un ensemble de classes permettant de développer des applications capables de se connecter à des serveurs de bases de données (SGBD) [50].

XAMPP

XAMPP signifie Cross-Platform (X), Apache (A), MySQL (M), PHP (P) et Perl (P). C'est un ensemble de logiciels permettant de mettre en place facilement un serveur Web et un serveur FTP. Il s'agit d'une distribution de logiciels libres (X Apache MySQL Perl PHP) facile à installer offrant une bonne souplesse d'utilisation permettant l'exploitation d'un serveur Apache, de l'SGBD MySQL et l'interpréteur PHP. XAMPP est également multiplateforme [51].

MySQL

C'est un Système de Gestion de Bases de Données (SGBD) client/serveur open source qui gère stocke et manipule une grande quantité de données en les organisant sous forme de tables, il prend aussi en charge les fonctionnalités de protection et de sécurité et fournit un ensemble d'interfaces de programmation (dont une avec PHP) facilitant l'accès aux données [52].

Serveur Flask

Flask est un Framework d'application Web Server Gateway Interface (WSGI) léger. Il est conçu pour permettre une mise en route rapide et facile du développement Web, avec la possibilité d'évoluer vers des applications complexes. Il est devenu l'un des cadres d'application Web Python les plus populaires. Flask propose des suggestions, mais n'impose aucune dépendance ni structure de projet. Il appartient au développeur de choisir les outils et les bibliothèques qu'il souhaite utiliser, grâce aux nombreuses extensions fournies par la communauté qui facilitent l'ajout de nouvelles fonctionnalités [53].

REST

REST (REpresentational State Transfer) ou RESTful est un style d'architecture pour les systèmes hypermédia distribués, utilisant les spécifications originelles du protocole

http, et permettant de construire des applications (Web, Intranet, Web Service). Il s'agit d'un ensemble de conventions et de bonnes pratiques à respecter [54].

Anaconda

C'est une distribution Python particulièrement populaire pour l'analyse de données et le calcul scientifique.

- Projet open source développé par Continuum Analytics, Inc ;
- Disponible pour Windows, Mac OS et Linux ;
- Comprend de nombreux packages populaires : NumPy, SciPy, Matplotlib, Pandas, IPython, Cython ;
- Inclut Spyder, un environnement de développement Python ;
- Inclut conda, un gestionnaire de packages indépendant de la plateforme ;
- Inclut Jupyter Notebook [55].

Jupyter Notebook

C'est un environnement de programmation interactif permettant l'exécution du code et l'insertion d'image et de texte. Il supporte de nombreux langages de programmation. Il est abondamment utilisé dans la data science [17].

5.2.2.2 Langages de programmation

Les différents langages de programmation utilisés dans l'implémentation de l'application système sont :

JavaFX

Selon Oracle, JavaFX est « un ensemble de packages graphiques et multimédias qui permet aux développeurs de concevoir, créer, tester, déboguer et déployer des applications clientes riches qui fonctionnent de manière cohérente sur diverses plates-formes. » En d'autres termes, JavaFX est le dernier moyen de créer des applications GUI (Graphical User Interface) avec Java [56].

FXML

C'est un format de fichier propre à Javafx et servant, entre autres, à définir des interfaces graphiques. Ce format utilise tout simplement la syntaxe XML [48].

SQL

SQL (Structured Query Language) permet de déclarer, manipuler et d'interroger les données, mais aussi de contrôler leur accès et d'être intégré dans des programmes écrits dans un langage quelconque [57].

Python

Python est un langage de programmation polyvalent à usage général. C'est un langage interprété et peut s'exécuter de manière interactive à partir d'une console. Il ne nécessite pas de compilateur, donc le temps de développement a tendance à être plus

court. Il est disponible en téléchargement gratuit et peut être installé sur de différents systèmes d'exploitation. Il est particulièrement populaire pour les applications scientifiques et mathématiques et bénéficie donc d'un grand nombre de bibliothèques complémentaires utiles développées par sa grande communauté. Python est relativement facile à apprendre par rapport aux autres langages [58].

5.3 ETL

La mise en œuvre de l'entrepôt de données est faite avec l'outil ETL Talend Open Studio (TOS) qui nous a permis de faire un modèle en flocon.

La première étape est de collecter toutes les données qui se trouvent dans de différents fichiers Excel et les mettre dans une seule base de données. La figure 5.1 représente le job de collection de données exécuté sur TOS.

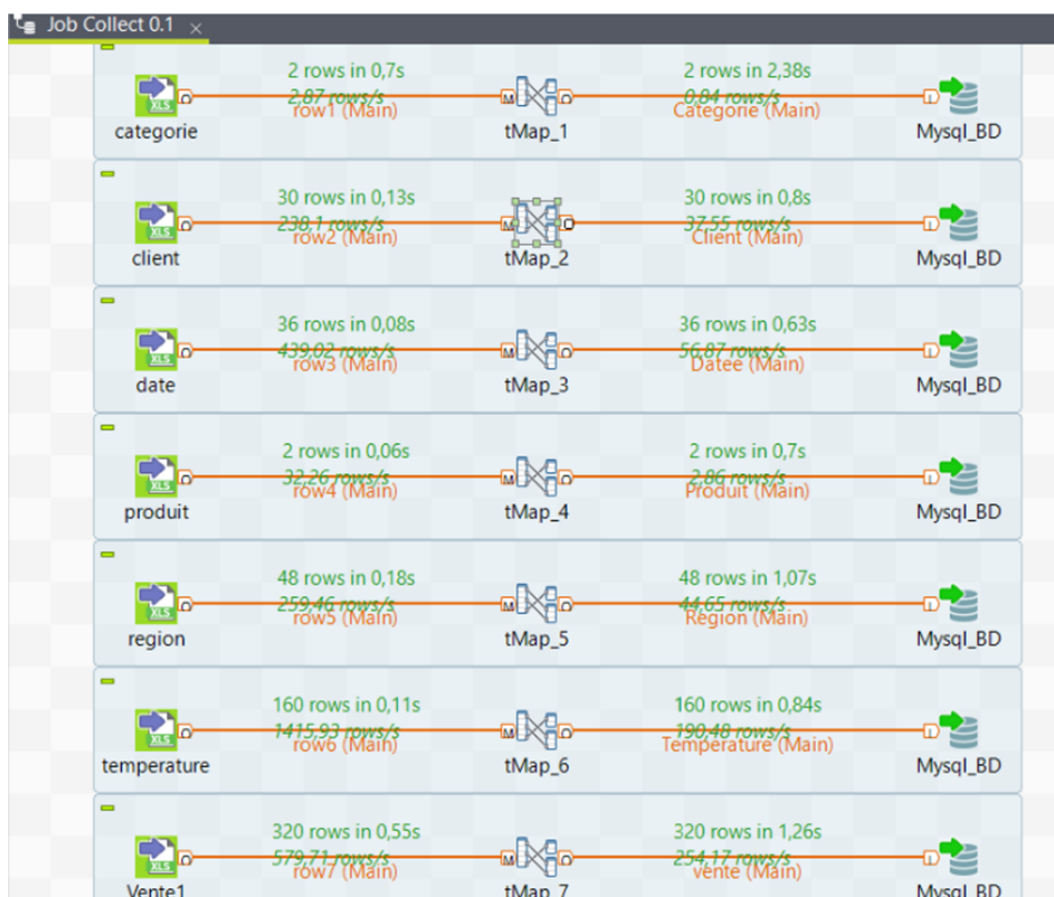


FIGURE 5.1 : Job de collection de données sur TOS.

Par la suite, nous sommes passées à l'étape de transformation des données dans laquelle nous avons changé les types de quelques attributs, et aussi appliqué des contraintes d'intégrités pour certaines d'autres.

L'application des contraintes d'intégrité (clé primaire, clés étrangères) pour la table faits_ventes est présentée dans la figure 5.2.

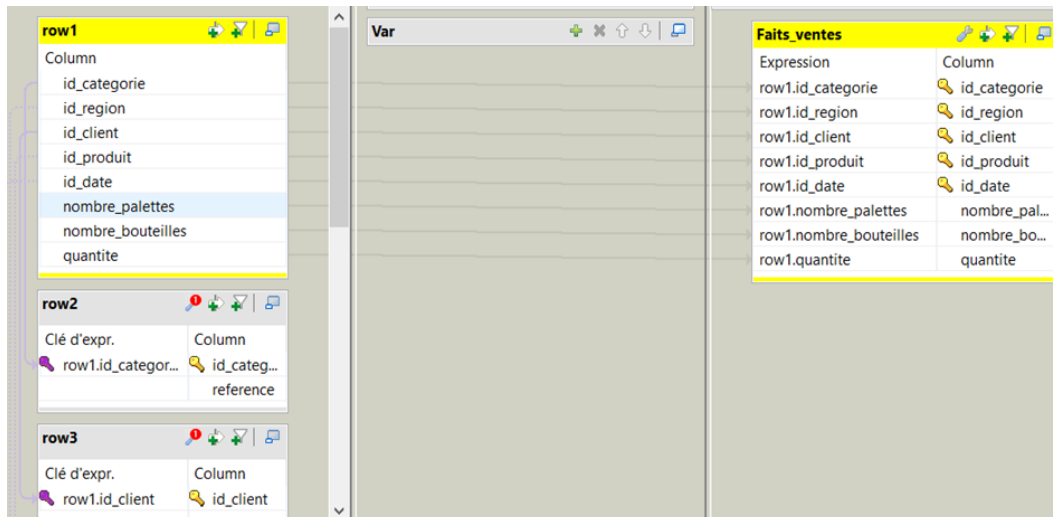


FIGURE 5.2 : Application des contraintes d'intégrité pour la tables Faits_ventes.

Enfin, nous avons créé les tables dimensions et la table faits et nous les avons chargés dans un entrepôt de données.

5.4 Modélisation de la prédiction de vente

Nous rappelons que l'objectif principal de notre application est l'aide à la décision pour la prédiction des ventes (développement d'un SIAD). Pour ce faire, nous nous intéressons aux modèles de régression dans l'apprentissage supervisé pour les raisons évoquées dans la section 6 du chapitre I. La modélisation d'un modèle d'apprentissage se fait suivant les différentes étapes expliquées dans ce qui suit.

1. **Collecte et récupération des données** : nous avons exploré toutes les pistes possibles pour récupérer les données nécessaires constituent l'expérience et les exemples que nous allons fournir à l'algorithme Machine Learning afin qu'il puisse apprendre.
2. **Nettoyage de données** : dans cette étape nous avons d'abord importé le fichier Excel qui contient notre Dataset grâce à la librairie *Pandas*, pour s'assurer que les données sont consistantes, sans valeurs manquantes, la même librairie est utilisée pour visualiser les données.
3. **Exploration des données** : nous avons sélectionné les variables (Features) et cible (Target). Puisque nous avons des variables catégorielles, nous les avons codées par l'encodage *OneHotEncoder* de la librairie *sklearn* ; il permet de créer des colonnes binaires qui réfèrent la donnée par un 0 ou 1 afin qu'elles soient utilisables par l'algorithme.
4. **Découpage des données** : nous avons découpé le jeu de données avec la librairie *Sklearn* via la fonction *train_test_split* en deux parties :
 - a) Données d'entraînement (Train set) : représentent 80% de données destinées à l'apprentissage du modèle.
 - b) Données de test (Test set) : représentent les 20% des données destinées à l'évaluation du modèle. Ces données ne doivent en aucun cas être utilisées lors de la conception du modèle.

5. **Modélisation** : c'est la phase de création du modèle statistique, où l'algorithme à appliquer est choisi après avoir visualisé les différents points sur un graphe comme celui de figure 5.3 qui représente la quantité d'eau vendue en fonction de la température.

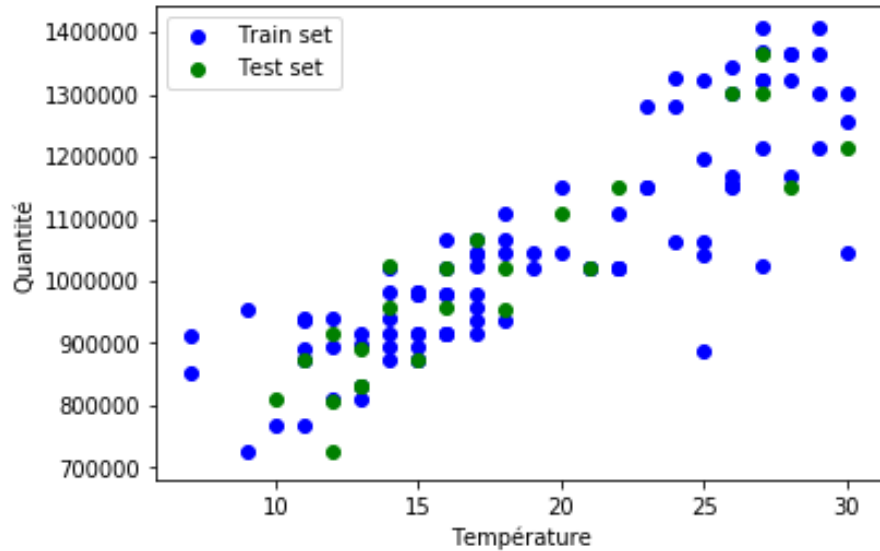


FIGURE 5.3 : Nuage de points de la quantité mensuel vendue en fonction de la température.

D'après ce graphe, nous avons remarqué une augmentation relativement linéaire de la quantité par rapport à la température. De ce fait nous avons appliqué sur les données d'entraînement l'algorithme de régression linéaire (LinearRegression) intégré dans *Sklearn*, ce qui nous a donné un modèle qui représente la droite la plus proche de l'ensemble des points (figure 5.4).

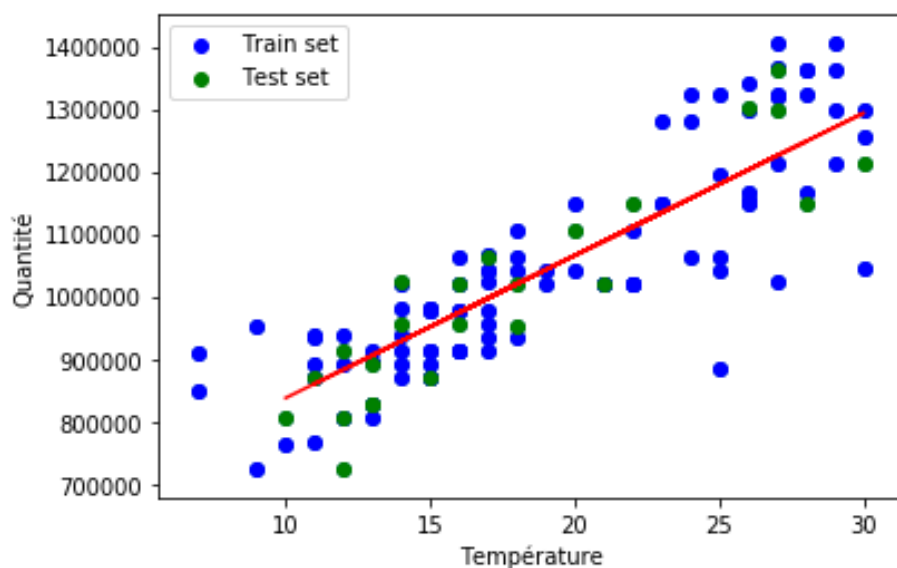


FIGURE 5.4 : Droite de régression correspondant à la modélisation du nuage de points.

6. **Évaluation** : pour évaluer notre modèle de régression nous avons calculé le coefficient de corrélation entre les valeurs prédites et les vraies valeurs en utilisant la fonction `r2_score` (de la librairie `sklearn`).
7. **Exportation du modèle** sous format binaire avec `joblib` (de la librairie `sklearn`).
8. **Déploiement du modèle** sous forme d'une API REST à l'aide du Framework Flask.

5.5 Présentation de l'application

Dans cette partie nous allons présenter tous ceux qui concerne la réalisation de notre application) savoir le module d'apprentissage et les interfaces (IHM).

5.5.1 Application web

Afin que l'API REST du modèle Machine Learning soit intégrée dans notre système et que nous puissions gérer l'état du système de façon optimale, il est nécessaire de développer une interface homme-machine pour pouvoir consommer l'API qui se présentera sous forme d'une application web, cette dernière représente un sous-système de notre application SIAD où il est intégré à l'aide d'un URI (Uniform Resource Identifier) qui va permettre de spécifier la cible grâce à un nom d'hôte, un port, un chemin et un paramètre de requête [59].

L'architecture de l'application web développée est basée sur la communication entre le client (sous navigateur) et le serveur Flask en utilisant des requêtes HTTP (Hypertext Transfer Protocol). Les interfaces graphiques de la partie client de l'application web sont réalisées et développées par le langage de balisage HTML (Hyper Text Markup Language) et les feuilles de style CSS (Cascading Style Sheets)¹, et pour la partie serveur, elle est réalisée à l'aide du Framework web Flask et des modules et classes écrits en langage Python.

5.5.2 Quelques interfaces graphiques

Nous présentons quelques interfaces de l'application développée :

Interface d'authentification : la figure 5.5 représente l'interface d'authentification où l'utilisateur saisi son nom et son mot de passe pour pouvoir accéder aux fonctionnalités du système.

1. <https://www.w3.org/standards/webdesign/htmlcss>

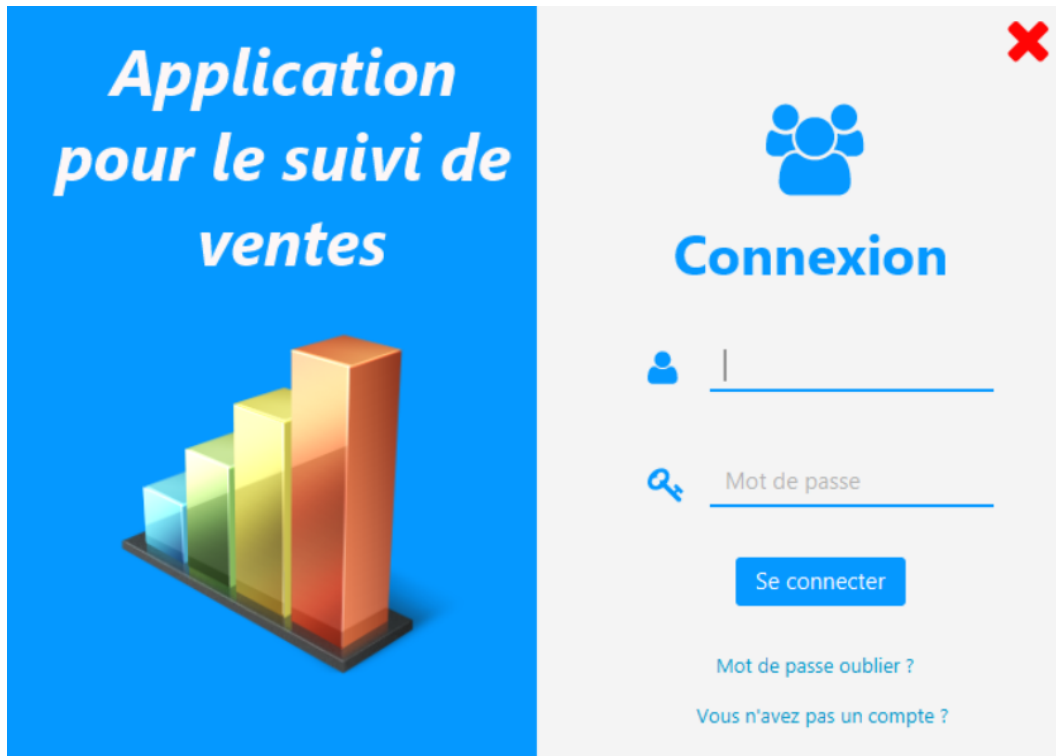


FIGURE 5.5 : Interface d'authentification.

Interface d'analyses : la figure 5.6 correspond à l'interface d'analyse qui est la première interface affichée à l'utilisateur après l'authentification. Cette interface contient un tableau de bord des statistiques nationales et internationales.

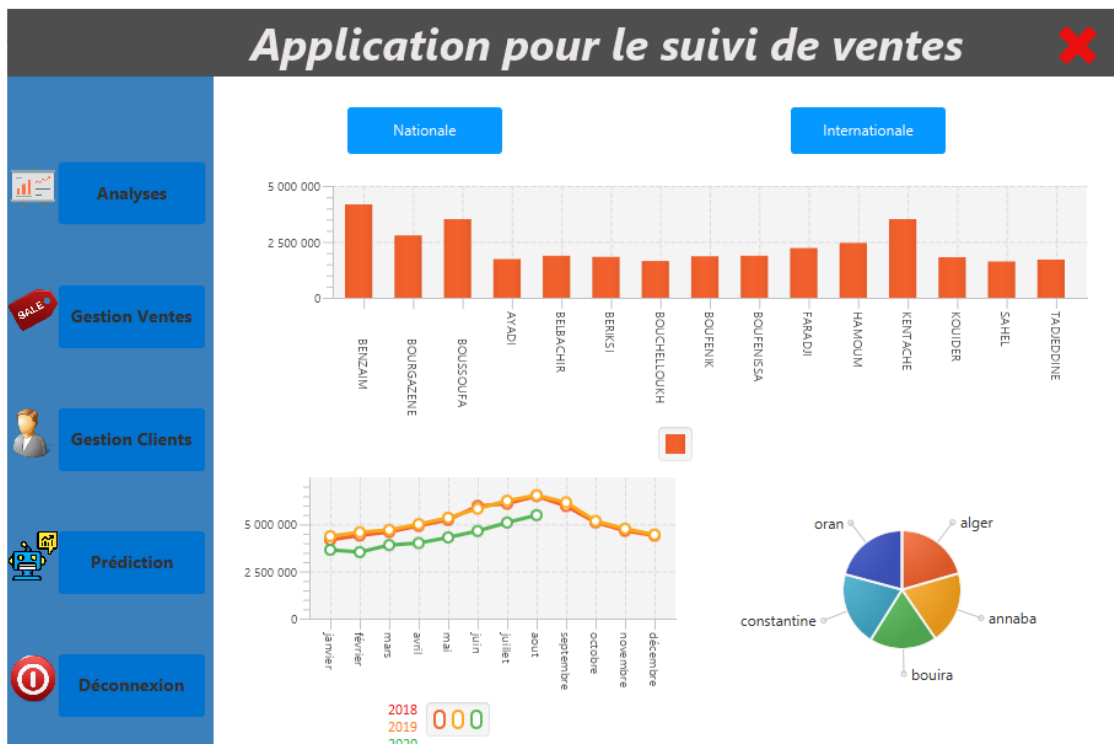


FIGURE 5.6 : Interface d'analyses.

Interface de gestion de ventes

L'interface de gestion de ventes est donnée par la figure 5.7. Elle permet de faire une analyse, un tri par client, date ou région et d'effectuer la recherche.

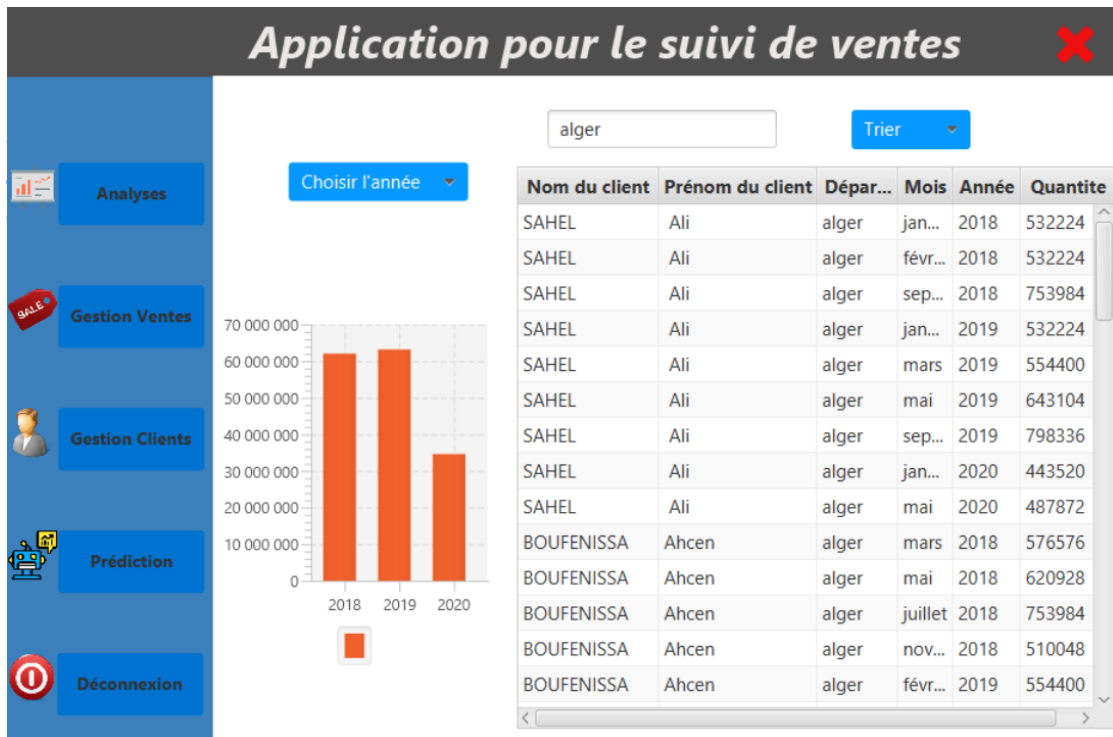


FIGURE 5.7 : Interface de gestion de ventes.

Interface de prédiction de ventes

La figure 5.8 représente l'interface où les prévisions s'effectuent en faisant appel à l'API de prédiction.

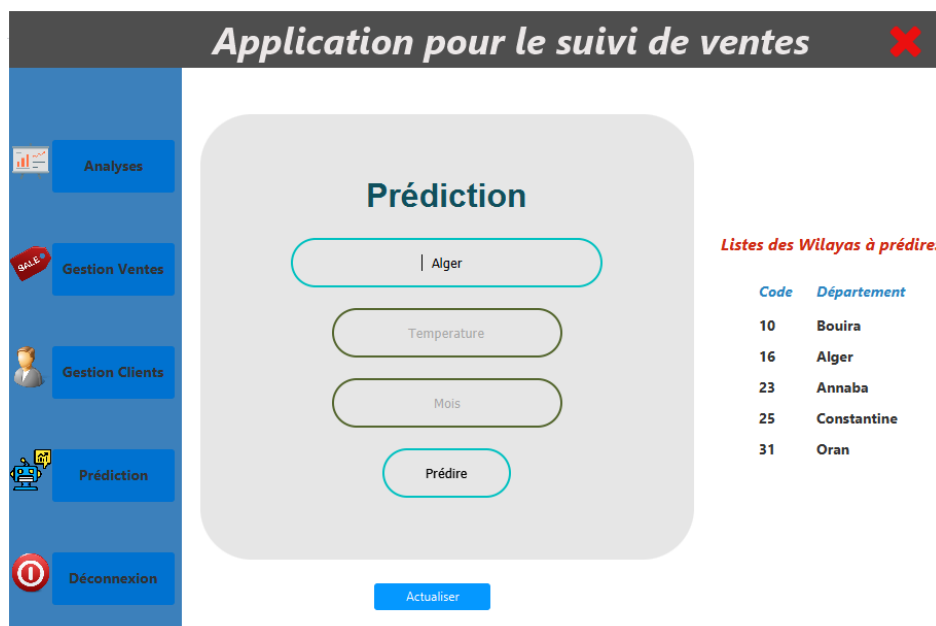


FIGURE 5.8 : Interface de prédiction des ventes.

5.6 Tests

Pour finir et valider notre système il faut passer par la phase de test afin de vérifier la conformité du logiciel et tester ces fonctionnalités. Pour ce faire, nous nous sommes réunis avec le responsable des statistiques pour tester les fonctionnalités du système décrites dans le premier chapitre. La plupart de ces fonctionnalités correspondent à ces attentes.

Néanmoins, nous avons constaté que l'utilisation des toutes les données, y compris celles de l'année 2020, dans la construction du modèle d'apprentissage produit des résultats inattendus dans la partie de prédiction. En conséquence, nous avons considéré deux cas : avec les données de l'année 2020 et sans ces données. Les tests réalisés avec les deux modèle montrent des performances très différentes. Mais le fait d'ignorer complètement les données 2020 ne correspond pas à la réalité. Alors, nous avons pensé à reprogrammer le modèle en donnant un poids faible pour un mois ou une année exceptionnelle comme c'est le cas de la pandémie COVID-19².

5.7 Conclusion

Au cours de ce chapitre, nous avons présenté les détails de la réalisation de l'application d'aide à la décision. Nous avons vus les étapes de la construction de l'entrepôt de données ainsi que les étapes de la modélisation des prédiction de vente en utilisant l'approche d'apprentissage supervisée en occurrence le modèle de régression. De plus, nous avons présenté l'architecture de l'application et quelques interfaces.

2. <https://www.who.int/fr/emergencies/diseases/novel-coronavirus-2019>

Conclusion et perspectives

Rappel des objectifs

Ce projet de fin de cycle avait pour objectif de développer un système d'aide à la décision pour la prédiction des ventes de la SARL IFRI, plus précisément pour le service commercial. En effet, l'étude de l'existant menée au sein de l'organisme d'accueil, grâce notamment aux interviews avec le responsable du service commercial, a fait ressortir les problèmes liés à l'évolution du marché et le manque d'outil automatique qui permet de faire des prévisions sur les ventes autre que les méthodes manuelles et traditionnelles coûteuses en effort et en temps. C'est dans ce sens que nous avons proposé de réaliser une application informatique décisionnelle, un SIAD qui intègre un module de prédiction basé sur l'apprentissage automatique.

Bilan des travaux effectués

Afin de concrétiser ce projet, nous avons suivi une démarche de développement centrée utilisateur adaptée aux SIADs. Cette démarche est tirée des processus de développement proposés en génie logiciel, IHM, ingénierie de connaissances. Bien évidemment, pour la modélisation des différentes facettes ou fonctionnalités de l'application, nous nous sommes appuyés sur le langage UML.

Notre projet est développé en trois grandes parties : la première a consisté en la construction d'un Datawarehouse à partir de différents fichiers Excel avec l'outil ETL Talend Open Studio. Par la suite, nous avons proposé un modèle d'apprentissage supervisé pour pouvoir faire des prédictions sur la quantité d'eau à vendre en fonction de la température, la région, etc. Ce modèle est déployé comme étant une API web et intégré dans notre système. La troisième partie a porté sur le reste de l'application programmée avec le langage java FX ; dont laquelle nous avons construit les interfaces et des tableaux de bord pour faciliter l'analyse des données.

Enfin, l'application développée a été confrontée à un ensemble de tests pour comparer le produit résultat par rapport aux résultats attendus.

Apports, limites et perspectives

Ce projet a été une expérience très riche pour nous, car il nous a permis de mettre en application nos connaissances informatique et d'en acquérir de nouvelles notamment dans le domaine de l'intelligence artificielle et de l'informatique décisionnelle.

Dans ce présent travail nous avons constaté que l'intelligence artificielle nous a bien servi comme un outil d'aide à la prise de décision et que l'une de ses branches, l'apprentissage automatique, nous a permis de réaliser un SIAD prédictif assurant une meilleure utilisation de l'expérience de l'employé chez IFRI.

La solution que nous avons présentée dans ce mémoire répond au problème de la prise de décision uniquement pour la vente d'eau et dans le service commercial de la SARL IFRI. Cette solution n'est en réalité qu'une ouverture vers d'autres travaux car

notre système peut encore évoluer et s'étendre pour le reste des produits et des services d'IFRI. Pour cela, plusieurs perspectives se présentent dont :

- Étendre la solution pour tous les produits.
- Donner la main à l'utilisateur de choisir les poids des variables du modèle d'apprentissage.
- Amélioration du processus d'apprentissage avec les nouvelles données.

Nous souhaitons également élargir le champ d'application de notre proposition en créant d'autres projets là où le génie logiciel et l'intelligence artificielle pourront contribuer à la réussite des entreprises.

Bibliographie

- [1] N CARRE. *L'étude de votre marché : vos clients et vos concurrents*. 2015.
- [2] Daniel CAUMONT et al. *Les études de marché*. Dunod, 2016.
- [3] Jacques LENDREVIE et Julien LÉVY. *Mercator 2013 : Théories et nouvelles pratiques du marketing*. Dunod, 2012.
- [4] LAROUSSE. *Dictionnaire de français*. Sous la dir. de MALESHERBES. 2011.
- [5] Gutierrez DANIEL. *Guide InsideBIGDATA de l'analyse prédictive*. Inside BigData, 2015. URL : <https://www.celge.fr/wp-content/uploads/2015/12/guide-analyse-pr%5C'edictive.pdf>.
- [6] S EILLON. "What is a decision?" In : *Management Science* 16.4 (1969).
- [7] William TAGGART et Daniel ROBEY. "Minds and managers : On the dual nature of human information processing and management". In : *Academy of management review* 6.2 (1981), p. 187-195.
- [8] Abdelkader ADLA. "Aide à la facilitation pour une prise de décision collective : proposition d'un modèle et d'un outil". Thèse de doct. Toulouse 3, 2010.
- [9] Lynda ATIF. "P@, une approche collaborative d'analyse des besoins et des exigences dirigée par les problèmes : le cas de développement d'une application Analytics RH". Thèse de doct. Paris Sciences et Lettres, 2017.
- [10] Farid KADRI. "Contribution à la conception d'un système d'aide à la décision pour la gestion de situations de tension au sein des systèmes hospitaliers. Application à un service d'urgence". Thèse de doct. Valenciennes, 2014.
- [11] Michel VOLLE. "La place de l'aide à la décision dans le système d'information". In : *Revue MODULAD* 57.32 (2005).
- [12] Pierre LÉ VIGNE et Jean-Charles POMEROL. *Systèmes interactifs d'aide à la décision et systèmes experts*. LIVRE. 1990.
- [13] Thomas W MALONE et Kevin CROWSTON. "The interdisciplinary study of coordination". In : *ACM Computing Surveys (CSUR)* 26.1 (1994), p. 87-119.
- [14] Jean-Charles POMEROL. "Artificial intelligence and human decision making". In : *European Journal of Operational Research* 99.1 (1997), p. 3-25.
- [15] Marina MEILA. "Comparing clusterings : an axiomatic view". In : *Proceedings of the 22nd international conference on Machine learning*. 2005, p. 577-584.
- [16] Paolo MESSINA. *The perfect intelligence guide for managers and innovators innodemia*. 2019.
- [17] Saint-Cirgue GUILLAUME. *Apprendre le Machine Learning en une semaine*. LIVRE. 2019.
- [18] Gilbert SAPORTA. *Probabilités, analyse des données et statistique*. Editions Technip, 2006.

- [19] Khalid ANWAR, Jamshed SIDDIQUI et Shahab SAQUIB SOHAIL. “Machine Learning Techniques for Book Recommendation : An Overview”. In : *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India*. 2019.
- [20] Salim BELMEHDI. “Analyse de l’historique de chiffres d’affaires et génération des rapports prévisionnels”. Mém. de mast. SUPINFO Clermont-Ferrand, 2018.
- [21] Sonia BELABBAS et Djerada LYES. “Conception et réalisation d’un système d’information d’aide à la décision pour l’analyse du processus d’achat”. Mém. de mast. Abderrahmane Mira Béjaia, 2019.
- [22] Faten ATIGUI. “Approche dirigée par les modèles pour l’implantation et la réduction d’entrepôts de données”. Thèse de doct. Toulouse 1, 2013.
- [23] Raouf AISSA et Bouaichi KOUSSEILA. “Mise en place d’une solution Open Source de Business Intelligence (Cas : entreprise portuaire de Béjaia)”. Mém. de mast. Abderrahmane Mira Béjaia, 2015.
- [24] Bruno ESTEVEZ. *Informatique décisionnelle opérationnelle*. 2014.
- [25] Alain FERNANDEZ. *Les nouveaux tableaux de bord des managers : le projet décisionnel dans sa totalité*. Editions Eyrolles, 2010.
- [26] Alain FERNANDEZ. *Les nouveaux tableaux de bord des managers : Le projet Business Intelligence clés en main*. Editions Eyrolles, 2013.
- [27] Georges GARDARIN. “Internet, Intranet et bases de données”. In : (1999).
- [28] S IKKEN et Benali S. “Conception et réalisation d’un système décisionnel pour le suivi des projets d’investissent (cas rtc sonatrach”. Mém. de mast. abderrahmane mira béjaia, 2010.
- [29] Bernard LEBELLE. *Construire un tableau de bord pertinent sous Excel*. Editions Eyrolles, 2013.
- [30] Hicham ASSALI et Ghilas BELKACI. “Conception et réalisation d’un entrepôt de données (cas d’étude : gestion de stock)”. Mém. de mast. abderrahmane mira béjaia, 2018.
- [31] S KHOURI. “Modélisation conceptuelle à base ontologique d’un entrepôt de données”. In : *Mémoire du magistère, Ecole doctorale des Sciences et Technologies de l’Information et de la Communication, Oued-Smar Alger* (2009).
- [32] Susanne BUSSE et al. “Federated information systems : Concepts, terminology and architectures”. In : *Forschungsberichte des Fachbereichs Informatik* 99.9 (1999), p. 1-38.
- [33] Alejandro VAISMAN et Esteban ZIMÁNYI. “Extraction, transformation, and loading”. In : *Data Warehouse Systems*. Springer, 2014, p. 285-327.
- [34] Grady BOOCH, James RUMBAUGH, Ivar JACOBSON et al. *Le guide de l’utilisateur UML*. T. 3. Eyrolles, 2000.
- [35] Pascal ROQUES. *UML 2 : Modéliser une application web*. Editions Eyrolles, 2008.
- [36] Laurent AUDIBERT. *UML 2 : De l’apprentissage à la pratique*. T. 298. Ellipses, 2009.
- [37] Pascal ROQUES et Franck VALLÉE. “UML 2 en action”. In : *De l’analyse des besoins à la conception J2EE, 3ème édition Eyrolles* (2004).
- [38] josep GABAY et David GABAY. *Analyse et conception*. Dunod, 2008.
- [39] Gilles ROY. *Conception de bases de données avec UML*. PUQ, 2007.

- [40] Zahir MEGRI et Mohamed RAHMANI. “Conception et réalisation d’une application d’apprentissage en ligne de la programmation par une évaluation automatisée”. Mém. de mast. abderrahmane mira béjaia, 2017.
- [41] Ouacila BELMEHDI. “Conception et réalisation d’un site webtv pour la fédération des associations des parents d’élèves de la wilaya de béjaia”. Mém. de mast. 2017.
- [42] Syla ANKI et Nassim AZI. “Conception et réalisation d’une application de géolocalisation(Cas d’étude :Entreprise Bejaia Logistique”. Mém. de mast. 2016.
- [43] Peter W SZABO. *User experience mapping*. Packt Publishing Ltd, 2017.
- [44] Mechrafi FARID. “Etude du problème de transport à la demande”. Mém. de mast. abderrahmane mira béjaia, 2013.
- [45] Benôt KALUNZNY. “intégration de données financières avec un outil etl open source”. Thèse de doct. Lille, 2011.
- [46] Khadidja KEBIR. “Conception et réalisation d’une application de gestion de réservation d’hotels (Avec service REST/SOAP)”. Mém. de mast. Tlemcen, 2017.
- [47] R YAGOUB et Y ZAHRAOUI. “Interface graphique Java de la gestion d’une banque”. Mém. de mast. blhadj bouchaib d’ain témouchent, 2015.
- [48] Nasser ABOUBACAR BEIDARI ABDOUL et Mohamed BOUREZG. “Développement d’une application de fiable détection et de reconnaissance.” Mém. de mast. Université de Tlemcen, 2018.
- [49] Hadjer Selsabil BENMANSOUR et Ahhmed Djazila BABA. “Développement d’une solution ERP pour la gestion de maintenance assistée par ordinateur.” Mém. de mast. Tlemcen, 2018.
- [50] L STEPHANE. *Transition vers l’agilité à l’échelle d’une organisation*. 2012.
- [51] Ammar KHALLEF et Abd Ed Daim TENACHI. “Conception et réalisation d’un système d’apprentissage en ligne avec personnalisation dynamique du parcours d’apprentissage”. In : (2017).
- [52] Philippe RIGAUX. *Pratique de MySQL et PHP : Conception et réalisation de sites web dynamiques*. Dunod, 2009.
- [53] Omar Abderrezak AZZOUZI et Ahmed CHAREF AISSA. “interfaçage d’un instrument de mesure de puissance optique et d’un controleur de diode laser pour la commande à distance”. Mém. de mast. abdelhamid ibn badis moustaganem, 2019.
- [54] Sylvain CHERRIER. “Architecture et protocoles applicatifs pour la chorégraphie de services dans l’Internet des objets”. Thèse de doct. 2013.
- [55] Weston STEPHEN et Bjornson ROBERT. *Introduction to Anaconda*. 2016. URL : <https://research.computing.yale.edu/sites/default/files/files/anaconda.pdf>.
- [56] A. A. *Programming with JavaFX, Tecnica di programmazione*. 2016/2017. URL : <https://elite.polito.it/files/courses/03FYZ/2017/slide/02-01-javafx-programming.pdf>.
- [57] Soutou CHRISTIAN. *De UML à SQL conception de bases de données*. Eyrolles, 2002.
- [58] David JULIAN. *Designing machine learning systems with Python*. Packt Publishing Ltd, 2016.
- [59] Carsten BORMANN, Angelo P CASTELLANI et Zach SHELBY. “Coap : An application protocol for billions of tiny internet nodes”. In : *IEEE Internet Computing* 16.2 (2012), p. 62-67.

Résumé

La prédiction des ventes en fonction de l'évolution des marchés est l'un des problèmes majeurs des entreprises. L'utilisation des outils d'aides à la décision, comme les Systèmes Interactifs d'Aide à la Décision (SIADs), est une solution prometteuse pour ce problème. C'est dans ce contexte que s'insère le travail présenté dans ce mémoire. Il consiste en la conception et la réalisation d'une application SIAD pour les prédictions des ventes selon le marché en utilisant la technique de l'apprentissage automatique (Machine Learning) pour le service commercial de l'entreprise IFRI. Pour ce faire, nous avons suivi une démarche de développement hybride centrée utilisateur qui combine entre les processus de développement existants en génie logiciel et en ingénierie de connaissances. Nous nous sommes appuyés sur le langage UML pour modéliser les différentes fonctionnalités et facettes de l'application au cours des étapes du développement. L'application développée est structurée en trois grandes parties: la première est un Datawarehouse construit à partir de différents fichiers Excel avec l'outil ETL Talend Open Studio. La seconde partie est un modèle d'apprentissage pour les prédictions des ventes en fonction de la température, la région, etc. Ce modèle est programmé avec le langage Python et déployé grâce au Framework Flask sous forme d'une API web. La troisième partie de l'application programmée avec le langage java FX, concerne les interfaces graphiques et les tableaux de bord permettant de faciliter l'observation et l'analyse de données des ventes et des clients aidant ainsi à la prise de décision.

Mot clés : Aide à la décision, Prédiction de ventes, Machine Learning, Datawarehouse, Processus de développement hybride, UML, ETL Talend Open Studio, Python, Flask, Java FX.

Abstract

Predicting sales based on the evolution of markets is one of the major problems of companies. The use of decision support tools, such as Interactive Decision Support Systems (IDSS), is a promising solution to this problem. The work presented in this report is part of this context. It consists in the design and implementation of an IDSS application for sales predictions according to the market using the machine learning technique for the commercial department of the company IFRI. To do so, we followed a user-centric hybrid development approach that combines existing development processes in software engineering and knowledge engineering. We used the UML language to model the different functionalities and facets of the application during the development stages. The developed application is structured in three main parts: the first is a data warehouse built from various Excel files using the ETL tool Talend Open Studio. The second part is a learning model for sales predictions based on temperature, region, etc. This model is programmed with the Python language and deployed using the Flask Framework in the form of a web API. The third part of the application, programmed with the java FX language, concerns graphical interfaces and dashboards to facilitate the observation and analysis of sales and customer data, thus helping in decision making.

Key words: Decision Support, Sales Prediction, Machine Learning, Datawarehouse, Hybrid development process, UML, ETL Talend Open Studio, Python, Flask, Java FX.