

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



Université A. Mira de Béjaïa
Faculté des Sciences Exactes
Département d'Informatique

MÉMOIRE DE MASTER PROFESSIONNEL

En
Informatique

Option
Génie Logiciel

THÈME

Application des techniques d'intelligence artificielle
pour l'analyse prédictive des ventes.
(Cas : Sonelgaz Distribution Direction de BEJAIA)

Présenté par :

M^{elle} LOUALIA ZAHIA

M^{elle} NAIT SLIMANE LYDIA

Devant le jury composé de :

| | | | |
|-----------|-----------------------|-----|--------------------------|
| Président | M. BEDJOU Khaled | MAA | Université A MIRA-BÉJAIA |
| Examineur | M. OUZEGGANE Redouane | MAA | Université A MIRA-BÉJAIA |
| Encadrant | M. MIR Foudil | MAA | Université A MIRA-BÉJAIA |

Béjaïa, Juin 2023.

※ *Remerciements* ※

Au terme de cette aventure académique, nous souhaitons exprimer notre profonde reconnaissance envers toutes les personnes qui ont généreusement apporté leur soutien et leur contribution à la réalisation de ce travail. En premier lieu, nous exprimons notre gratitude la plus sincère à l'égard de la Divinité suprême pour avoir béni notre parcours avec santé, détermination et persévérance, qui ont été les fondements de notre réussite.

Nous tenons tout d'abord à adresser nos remerciements les plus chaleureux à notre encadrant, **Monsieur MIR Foudil**, dont la sollicitude et les conseils avisés ont grandement enrichi notre travail. Sa disponibilité infaillible et son écoute attentive ont été d'une inestimable valeur, éclairant nos réflexions et nous guidant dans la bonne direction.

Nous souhaitons également exprimer notre reconnaissance profonde envers **Monsieur HADJOUT Dalil**, responsable du service statistiques, pour sa disponibilité inébranlable et ses précieux conseils. Nous lui sommes infiniment reconnaissants pour le temps précieux qu'il nous a accordé et pour sa contribution inestimable à notre travail.

Nos remerciements s'étendent également à l'ensemble du personnel de la Sonelgaz Distribution Direction de BEJAIA, dont l'appui et la collaboration ont joué un rôle essentiel dans la réalisation de notre projet.

Nous exprimons également notre gratitude envers le professeur Monsieur Abbache grâce à vos encouragements, j'ai pu développer une confiance en moi et dépasser les obstacles qui ont jalonné mon chemin. Votre conviction en mes capacités a été une source d'inspiration majeure. Vos paroles d'encouragement ont été comme un phare dans les moments de doute, m'incitant à persévérer et à donner le meilleur de moi-même. Encore une fois, merci infiniment pour votre soutien indéfectible. Votre impact durable restera gravé dans nos esprits.

Nous tenons à exprimer notre profonde gratitude envers les membres du jury, qui ont accepté avec honneur d'évaluer notre travail. Leur expertise et leurs évaluations objectives ont contribué à la rigueur et à la qualité de notre projet.

Enfin, nous souhaitons adresser nos remerciements les plus sincères à nos familles, dont le soutien inconditionnel et les encouragements constants ont été une source inépuisable d'inspiration. Leur amour et leur confiance ont été les piliers qui nous ont permis de surmonter les obstacles et d'atteindre nos aspirations.

Nous tenons également à remercier toutes les personnes qui, de près ou de loin, ont contribué à la réalisation de ce mémoire. Leurs conseils éclairés, leurs encouragements sincères et leur soutien inestimable ont enrichi notre parcours académique et ont renforcé notre détermination à réussir.

✧ Dédicaces ✧

*À tous ceux qui ont traversé les méandres de ma vie,
Je désire consacrer ce projet de clôture d'études à toutes les âmes ayant joué un rôle remarquable
dans ma trajectoire académique et personnelle. Votre appui indéfectible, vos encouragements
enflammés et votre amour inconditionnel se sont révélés être des leviers fondamentaux de ma
réussite.*

*À mes géniteurs, vous êtes les piliers inébranlables, les éclaireurs avisés et les plus fervents
partisans.*

*Je vous dédie ces mots sincères. À toi, cher père MOHAMED, incarnation de la force
incomparable, et à toi, tendre mère SALIHA, fontaine de douceur inépuisable.*

À mes soeurs, Kenza, Wissam, Sonia, Dina, je vous adresse mes pensées.

À mes nièces, Malak et Darine, je vous entoure de ma tendresse.

À ma grand-mère Ounissa, que Dieu la préserve, je dédie mes pensées les plus chaleureuses.

*À ma binôme Zahou, chère compagne de route, ta présence à mes côtés a été un soutien
indéfectible tout au long de cette aventure académique. Ensemble, nous avons surmonté les
obstacles, partagé les succès et créé des souvenirs inoubliables. Cette dédicace t'est également
dédiée, en témoignage de notre collaboration exceptionnelle.*

*Et à mes amis, kousseila, Idir, Manel, Asma, Naima, Lyna, Rima, je vous honore tout
particulièrement. Votre présence dans ma vie a été une véritable bénédiction. Vos rires, vos
encouragements et vos épaules réconfortantes ont illuminé mon parcours.*

*Que cette dédicace leur parvienne, emplie de ma gratitude la plus profonde et de mon respect
éternel. Que mes mots soient l'écho de l'amour et de la reconnaissance qui habitent mon coeur
envers chacun d'entre eux.*

M. NAIT SLIMENE LYDIA

※ Dédicaces ※

À tous ceux qui ont traversé les méandres de mon parcours,
Je souhaite dédier ce projet de clôture d'études à toutes les âmes qui ont joué un rôle exceptionnel dans ma trajectoire académique et personnelle. Votre soutien indéfectible, vos encouragements passionnés et votre amour inébranlable ont été des piliers essentiels de ma réussite.
A mes très chers parents, Aujourd'hui, alors que j'achève mon mémoire de fin d'études, je tiens à vous dédier ces quelques lignes.

Maman, tu es la source de ma force et de ma détermination. Tu m'as appris à persévérer dans les moments difficiles, à croire en moi-même et à ne jamais abandonner mes rêves. Ta présence bienveillante et tes conseils avisés ont été essentiels pour surmonter les obstacles qui se sont dressés sur mon chemin. Ta patience infinie m'a toujours donné la confiance nécessaire pour poursuivre mes études et atteindre mes objectifs. **Papa**, tu es le modèle de persévérance et de réussite. Ton travail acharné et ton dévouement m'ont toujours inspirée à donner le meilleur de moi-même. Ta présence bienveillante et ta capacité à me faire sentir capable de tout ont été mes plus grandes motivations. Je suis tellement fière de pouvoir partager avec vous ce mémoire qui représente le fruit de notre travail commun.

A ma binome **Lydia**, ma partenaire extraordinaire, ta précieuse présence et notre étroite coopération sont essentielles à notre succès mutuel. Grâce à notre solide travail d'équipe, nous avons surmonté des obstacles et relevé tous les défis. Je te remercie amplement pour tes conseils et ton enthousiasme contagieux. Ce mémoire est dédié à notre complicité, à notre persévérance et à notre réussite conjointe. Merci d'avoir été là chaque étape, je suis honorée d'avoir partagé cette expérience unique avec toi.

A ma grand-mère **Zahra** et à mon grand-père **Salah**, que Dieu les préserve, je souhaite leur dédier mes pensées les plus chaleureuses.

À ma chère soeur **Sarah** et son conjoint **Said** je vous envoie mes pensées affectueuses. À mon frère, **Salim**, tu occupes une place particulière dans mes réflexions et dans mon coeur. Vos encouragements, votre soutien et vos précieux conseils ont été des piliers essentiels de ma réussite. Je souhaite également ajouter quelques mots pour ma bienveillante voisine, **Ninane**, et son fils adorable, **Ilyane**. Votre amitié et votre soutien ont été d'une valeur inestimable pour moi. Et à mes cher(e)s ami(e)s, **Faiza, Rahma, Asma, Bahia, Nina, Sarah, Koukou, Yanis**, je tiens à vous rendre hommage tout particulièrement. Vos encouragements étaient importants dans mon parcours.

Que cette dédicace leur parvienne, empreinte de ma profonde gratitude et de mon respect éternel.
Que mes paroles reflètent l'amour et la reconnaissance qui habitent mon coeur envers chacun d'entre eux.

M. LOUALIA Zahia

Table des matières

| | |
|--|----------|
| Table des matières | i |
| Table des figures | iv |
| Liste des tableaux | vi |
| Liste des abréviations | vii |
| Introduction générale | 1 |
| 1 Concepts et fondamentaux abordés | 4 |
| 1.1 Introduction | 4 |
| 1.2 Statistique descriptive | 4 |
| 1.2.1 Séries chronologiques | 4 |
| 1.2.2 Description d'une série chronologique | 4 |
| 1.2.3 Décomposition d'une série chronologique | 5 |
| 1.2.4 Série désaisonnalisée | 6 |
| 1.3 Méthodes de prévision | 7 |
| 1.3.1 Moyenne Mobile | 7 |
| 1.3.2 Lissage exponentiel | 10 |
| 1.3.3 Lissage exponentiel simple | 10 |
| 1.3.4 Lissage exponentiel double | 10 |
| 1.3.5 Une série stationnaire | 12 |
| 1.3.6 Vérifier la stationnarité d'une série chronologique | 12 |
| 1.3.7 Transformer une série non-stationnaire en une série stationnaire | 13 |
| 1.3.8 La méthode ARIMA | 13 |
| 1.3.9 La méthode SARIMA | 15 |
| 1.3.10 Équation SARIMA | 16 |
| 1.3.11 Types du Machine Learning | 17 |
| 1.3.12 Quel est le rapport entre le Machine Learning et l'IA ? | 18 |
| 1.3.13 Comment fonctionne le Machine Learning ? | 18 |
| 1.3.14 Modèles du Machine Learning | 19 |

| | | |
|----------|---|-----------|
| 1.4 | Deep Learning | 20 |
| 1.4.1 | Différentes applications du deep learning | 21 |
| 1.4.2 | Différence entre deep learning et machine learning | 21 |
| 1.4.3 | Les réseaux de neurones artificiels | 22 |
| 1.4.4 | Comment se compose un réseau de neurones ? | 22 |
| 1.4.5 | Les différents réseaux de neurones | 22 |
| 1.4.6 | L'architecture de LSTM | 24 |
| 1.4.7 | Gestion de projet | 26 |
| 1.5 | Conclusion | 29 |
| 2 | ANALYSE DES BESOINS ET CONCEPTION | 30 |
| 2.1 | Introduction | 30 |
| 2.2 | Présentation du groupe SONELGAZ | 30 |
| 2.3 | La transformation de SONELGAZ en groupe : | 31 |
| 2.4 | Les différents types de client Sonelgaz | 32 |
| 2.5 | Présentation de l'organisme d'accueil (Sonelgaz Distribution Direction de BEJAIA) : | 33 |
| 2.5.1 | Organisation de la Sonelgaz Distribution Direction de BÉJAIA | 33 |
| 2.6 | Problématique | 35 |
| 2.7 | Solutions proposées | 35 |
| 2.8 | Délimitation du domaine d'étude | 37 |
| 2.8.1 | Répartition des rôles | 37 |
| 2.8.2 | Identification des acteurs | 37 |
| 2.8.3 | Diagramme de contexte | 37 |
| 2.9 | La méthodologie SCRUM | 38 |
| 2.9.1 | Le Product backlog | 38 |
| 2.9.2 | Planification des sprints | 40 |
| 2.10 | Diagramme de cas d'utilisation | 41 |
| 2.11 | Description des cas d'utilisations | 43 |
| 2.11.1 | Cas d'utilisation « S'authentifier » | 43 |
| 2.11.2 | Cas d'utilisation « Visualiser les consommations » | 45 |
| 2.11.3 | Cas d'utilisation « Prédire des ventes » | 47 |
| 2.11.4 | Cas d'utilisation « Comparaison des ventes » | 49 |
| 2.12 | Diagramme de séquence détaillé | 51 |
| 2.12.1 | Diagramme de séquence détaillé du cas d'utilisation "S'authentifier" : | 52 |
| 2.12.2 | Diagramme de séquence détaillé du cas d'utilisation «Visualiser les consommations» | 53 |
| 2.12.3 | Diagramme séquence détaillé du cas d'utilisation «Prédire des ventes» | 54 |
| 2.12.4 | Diagramme séquence détaillé du cas d'utilisation «Comparaison des ventes» | 55 |
| 2.13 | Diagramme d'activité | 56 |

| | | |
|----------|---|-----------|
| 2.14 | Diagrammes de Classes | 57 |
| 2.15 | Conclusion | 59 |
| 3 | RÉALISATION | 60 |
| 3.1 | Introduction | 60 |
| 3.2 | Environnement de développement de l'application | 60 |
| 3.2.1 | L'architecture 3-tier | 60 |
| 3.2.2 | Les langages utilisés | 60 |
| 3.2.3 | Les frameworks | 61 |
| 3.2.4 | Les logiciels | 61 |
| 3.2.5 | Les bibliothèques Python | 63 |
| 3.3 | Modélisation de la prédiction de vente | 64 |
| 3.3.1 | La modélisation avec SARIMA | 64 |
| 3.3.2 | La modélisation avec LSTM | 67 |
| 3.4 | Présentation de l'application | 69 |
| 3.4.1 | Application web | 69 |
| 3.4.2 | Quelques interfaces graphiques | 70 |
| 3.5 | La comparaison entre les méthodes ARIMA, SARIMA et LSTM | 80 |
| 3.5.1 | Discussion des résultats | 81 |
| 3.6 | Conclusion | 81 |
| | Conclusion et perspectives | 82 |
| | Bibliographie | 84 |

Table des figures

| | | |
|------|--|----|
| 1.1 | Trafic ferroviaire sur les Trains à Grande Vitesse (TGV) sur la période janvier 1995-décembre 2005, données mensuelles [24]. | 5 |
| 1.2 | Décomposition d'une série chronologique [24]. | 6 |
| 1.3 | Graphe des moyennes mobiles d'ordre 4 des ventes d'huitres [38]. | 9 |
| 1.4 | Illustrations avec des moyennes mobiles de différents ordres [38]. | 9 |
| 1.5 | Représentation graphique du LED [38]. | 11 |
| 1.6 | Diagramme de la relation entre IA et Machine Learning [21]. | 18 |
| 1.7 | Comment fonctionne le Machine Learning? [2]. | 19 |
| 1.8 | Machine à vecteurs de support [22]. | 19 |
| 1.9 | Schéma de l'agorithme K-means [21]. | 20 |
| 1.10 | Les types de réseaux de neurones récurrents [26]. | 24 |
| 1.11 | Schéma d'un neurone LSTM [36]. | 25 |
| 1.12 | Schéma de la méthode SCRUM. | 29 |
| | | |
| 2.1 | Logo de la Sonelgaz Direction Distribution de BÉJAIA [35]. | 33 |
| 2.2 | Organigramme de la Sonelgaz Direction Distribution de BÉJAIA [35]. | 34 |
| 2.3 | Diagramme de contexte | 37 |
| 2.4 | Diagramme de cas d'utilisation | 42 |
| 2.5 | Diagramme de séquence système du cas d'utilisation «S'authentifier» | 45 |
| 2.6 | Diagramme séquence système du cas d'utilisation «Visualiser les consommations» | 47 |
| 2.7 | Diagramme séquence système du cas d'utilisation «Prédire les ventes» | 49 |
| 2.8 | Diagramme séquence système du cas d'utilisation «Comparaison des ventes» | 51 |
| 2.9 | Diagramme de séquence détaillé du cas d'utilisation «S'authentifier» | 53 |
| 2.10 | Diagramme séquence détaillé du cas d'utilisation «Visualiser les consommations» | 54 |
| 2.11 | Diagramme de séquence détaillé du cas d'utilisation «Prédire des ventes» | 55 |
| 2.12 | Diagramme de séquence détaillé du cas d'utilisation «Comparaison des ventes» | 56 |
| 2.13 | Diagramme d'activité | 57 |
| 2.14 | Diagramme de Classe initial | 58 |
| 2.15 | Diagramme de Classe de l'application | 58 |
| | | |
| 3.1 | Composantes de la série temporelle du client 66 | 65 |
| 3.2 | Les diagnostics du modèle SARIMA | 66 |

| | | |
|------|--|----|
| 3.3 | Interface d'authentification | 70 |
| 3.4 | Interface de tableau de bord | 71 |
| 3.5 | Interface Inscription | 71 |
| 3.6 | Interface de visualisation générale des ventes-Tableau | 72 |
| 3.7 | Interface de visualisation générale des ventes-Graphe | 72 |
| 3.8 | Interface visualisation des ventes par client | 73 |
| 3.9 | Interface localisation des clients | 74 |
| 3.10 | Interface prédiction avec ARIMA-Graphe | 75 |
| 3.11 | Prédiction avec ARIMA-Graphe | 75 |
| 3.12 | Interface prédiction avec ARIMA-Tableau | 76 |
| 3.13 | Interface prédiction avec SARIMA-Graphe | 76 |
| 3.14 | Prédiction avec SARIMA-Graphe | 77 |
| 3.15 | Interface prédiction avec SARIMA-Tableau | 77 |
| 3.16 | Interface prédiction avec LSTM-Graphe | 78 |
| 3.17 | Prédiction avec LSTM-Graphe | 78 |
| 3.18 | Interface prédiction avec LSTM-Tableau | 79 |
| 3.19 | Interface comparaison des ventes | 79 |
| 3.20 | Interface comparaison des ventes (Consommation-Prédiction(SARIMA)) | 80 |

Liste des tableaux

| | | |
|-----|---|----|
| 2.1 | Messages échangés entre le système et les acteurs | 38 |
| 2.2 | Product Backlog | 40 |
| 2.3 | Planification des Sprints | 41 |
| 2.4 | Formalisme de description des cas d'utilisation | 43 |
| 2.5 | Description du cas d'utilisation «S'authentifier» | 44 |
| 2.6 | Description du cas d'utilisation «Visualiser les consommations» | 46 |
| 2.7 | Description du cas d'utilisation «Prédire les ventes» | 48 |
| 2.8 | Description du cas d'utilisation «comparaison des ventes» | 50 |
| 2.9 | Description UML du diagramme de séquence détaillé | 52 |
| 3.1 | Comparaison des méthodes ARIMA, SARIMA et LSTM selon la métrique MAPE. . | 81 |

Liste des abréviations

- AIC** Akaike Information Criterion.
- AR** AutoRégressive.
- ARIMA** AutoRegressive Integrated Moving Average.
- ARMA** AutoRegressive Moving Average.
- ADF** Dickey-Fuller augmenté.
- ACF** Fonction Auto-Corrélation.
- BDD** Base de données.
- CSS** Cascading Style Sheets.
- CVS** Série Corrigée de Variations Saisonnières.
- DL** Deep Learning.
- GRU** Gated Recurrent Unit .
- IA** Intelligence Artificielle.
- LSTM** Long Short Term Memory.
- LES** Lissage Exponentiel Simple.
- LED** Lissage Exponentiel Double.
- ML** Machine Learning.
- PACF** Fonction Auto-Corrélation Partielle.
- RAD** Rapide Application Development.
- RNN** Recurent Neural Network.
- SARIMA** Seasonal AutoRégressionIntegrated Moving Average.
- SC** Série Chronologique.-
- SVM** Support Vector Machine.
- UML** Unified Modeling Language.
- XP** Externe Programming.

Introduction générale

L'anticipation de la production électrique revêt une importance cruciale de nos jours, tant d'un point de vue économique qu'environnemental. Étant donné que l'électricité ne peut être stockée, il est essentiel de répondre à la demande actuelle en ajustant la production de manière proportionnelle. Cela permet d'atteindre un équilibre entre la production et la consommation, réduisant ainsi les pertes et assurant un approvisionnement continu pour les consommateurs. Grâce à l'utilisation d'outils de prédiction, il est possible de connaître à l'avance les besoins énergétiques pour la production. Ces outils reposent sur des approches statistiques et d'intelligence artificielle largement reconnues dans le domaine de l'analyse des séries temporelles, permettant ainsi de prévoir les valeurs futures avec précision. L'évolution de l'informatique et l'introduction de diverses techniques avancées, telles que l'apprentissage automatique (machine learning) et le (deep learning), ont grandement renforcé leur puissance dans le domaine de la prévision.

Notre projet consiste à développer un système novateur pour la **Sonelgaz Distribution Direction de BEJAIA**, qui sera dédié à l'analyse et à la prévision des ventes. L'objectif principal est de simplifier et d'améliorer le processus d'analyse des statistiques de ventes d'électricité pour les clients en Haute Tension A (HTA), tout en permettant également de prédire leur consommation électrique. Actuellement, l'organisme utilise des méthodes traditionnelles telles que les fichiers Excel pour analyser ses ventes, mais il ne dispose pas d'un outil de prévision. Par conséquent, notre projet vise à fournir une solution automatisée et plus efficace pour répondre à leurs besoins.

L'objectif de notre projet est de résoudre la problématique suivante : Comment intégrer efficacement l'intelligence artificielle pour résoudre les problèmes rencontrés par la direction commerciale Sonelgaz de BEJAIA notamment le déséquilibre entre production et besoins entraînant des surplus ou des déficits d'électricité, la détection des fraudes liées aux compteurs d'énergie chez les clients qui est difficile, entraînant des évaluations de consommation erronées et des facturations inexactes, ainsi que l'utilisation de méthodes d'analyse de ventes traditionnelles et l'absence de systèmes d'aide à la décision avancés ?

Dans ce mémoire, nous envisageons d'exploiter deux approches distinctes pour la prédiction de la consommation électrique :

la méthode statistique **ARIMA** (Auto Regressive Integrated Moving Average) et **SARIMA** (Seasonal Auto Regressive Integrated Moving Average), utilisées dans le domaine de l'apprentissage automatique (machine learning), ainsi que la méthode **LSTM** (Long Short Term Memory), qui est une technique du domaine de l'apprentissage profond (deep learning) spécifiquement conçue pour la prédiction des séries temporelles. En ce qui concerne la gestion de notre projet, nous avons choisi d'adopter la méthodologie agile Scrum pour assurer une gestion efficace et flexible.

Organisation du mémoire

Le mémoire est structuré en trois chapitres distincts.

Le premier chapitre donne un aperçu des différentes technologies utilisées dans notre projet. Nous abordons l'apprentissage automatique (Machine Learning), l'apprentissage profond (Deep Learning), ainsi que l'analyse et la prévision des séries temporelles.

Nous examinons également le processus d'analyse prédictive et définissons la gestion de projet. De plus, nous mettons en évidence les différentes méthodes agiles, en mettant particulièrement l'accent sur la méthode Scrum qui sera mise en oeuvre dans notre projet.

Le deuxième chapitre présente en détail l'organisme d'accueil, la problématique traitée, la solution proposée et les objectifs de développement de l'application. Nous identifions les acteurs et décrivons les différents cas d'utilisation. Nous utilisons des diagrammes de séquences pour modéliser ces cas d'utilisation, ainsi que des diagrammes de séquences détaillés pour décrire les interactions entre les objets du système.

Enfin, nous présentons un diagramme de navigation qui permet de visualiser l'architecture de l'application, ainsi qu'un diagramme de classes.

Le troisième et dernier chapitre se concentre sur les étapes de construction des deux modèles de prédiction. Nous détaillons également l'architecture de l'application et expliquons son fonctionnement en utilisant les interfaces graphiques.

En conclusion, nous récapitulons les principaux points abordés dans ce rapport et suggérons des perspectives pour de futures recherches et développements.

Concepts et fondamentaux abordés

1.1 Introduction

Ce chapitre met en évidence l'importance de l'IA dans l'analyse et la prédiction des séries temporelles, ainsi que dans le domaine de l'électricité pour la prédiction de la demande et l'optimisation de la production d'énergie. Il souligne la nécessité de maîtriser les fondamentaux de l'IA et de comprendre les différents types pour minimiser les risques. Le chapitre explore également l'utilisation de méthodologies agiles telles que Scrum pour favoriser la communication transparente et la satisfaction du client. L'objectif global est de fournir une compréhension approfondie des technologies utilisées dans le projet d'IA, y compris le machine learning, le deep learning, l'analyse des séries temporelles et la gestion de projet avec Scrum.

1.2 Statistique descriptive

1.2.1 Séries chronologiques

1.2.1.1 Définition

Série chronologique ou (une Série temporelle ou bien chronique),noté X_t ensemble des observations d'une variable statistique économique faites à intervalles réguliers (années, trimestre, mois, jour,etc) [24].

telle que :

$$X_t = x_t, t = 1, \dots, T, T \in R, Z, N \quad (1.1)$$

1.2.2 Description d'une série chronologique

On considère qu'une série chronologique (X_t) est la résultante de différentes composantes fondamentales :

La tendance (ou trend) (T) représente l'évolution à long terme de la série étudiée. Elle traduit le comportement "moyen" de la série [24].

La composante saisonnière (ou saisonnalité) (St) correspond à un phénomène qui se répète à intervalles de temps réguliers (périodiques). En général, c'est un phénomène saisonnier d'où le terme de variations saisonnières [24].

La composante résiduelle (ou bruit ou résidu) notée (ε_t) correspond à des fluctuations irrégulières, en général de faible intensité mais de nature aléatoire. On parle aussi d'aléas. (Tout ce qui n'est pas pris en compte par la tendance et le mouvement saisonnier) [24].

Des phénomènes accidentels (grèves, conditions météorologiques exceptionnelles, crash financier) peuvent notamment intervenir.

Une autre composante parfois étudiée de manière spécifique à trait au **phénomène cyclique** : c'est souvent le cas en climatologie et en économie (exemple : récession et expansion...). Il s'agit d'un phénomène se répétant mais contrairement à la saisonnalité sur des durées qui ne sont pas fixes et généralement plus longues. Sans informations spécifiques, il est généralement très difficile de dissocier tendance et cycle. Dans le cadre de ce cours, la composante correspondant aux phénomènes accidentels sera intégrée aux fluctuations irrégulières de la série et la composante tendance regroupera à la fois la tendance et le cycle [24].

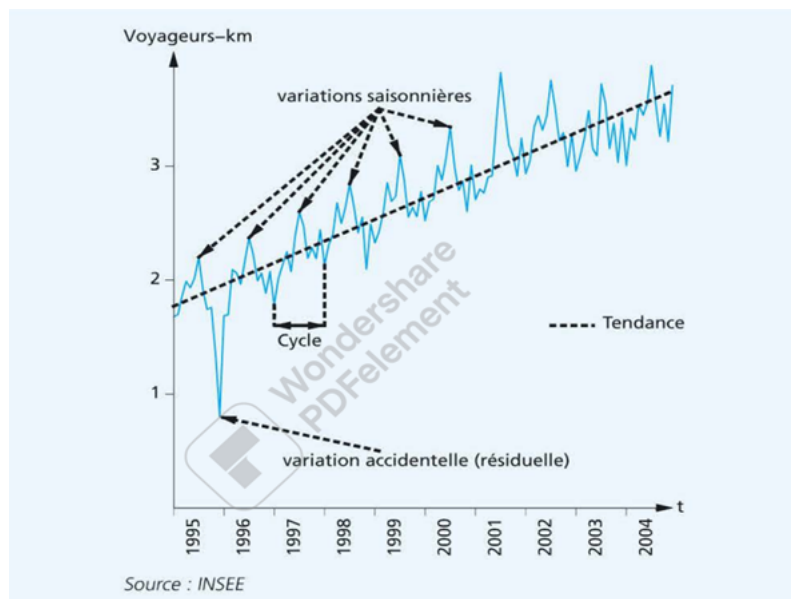


FIGURE 1.1 – Trafic ferroviaire sur les Trains à Grande Vitesse (TGV) sur la période janvier 1995-décembre 2005, données mensuelles [24].

1.2.3 Décomposition d'une série chronologique

Les deux schémas classiques de décomposition

On distingue traditionnellement le schéma de décomposition additif et le schéma de décomposition multiplicatif :

1. Le modèle additif : C'est le modèle classique de décomposition dans le traitement des modèles d'ajustement [24].

La variable X_t s'écrit comme le somme de trois termes :

$$X_t = T + S_t + (\varepsilon_t) \quad (1.2)$$

2. Le modèle multiplicatif : La variable X_t s'écrit au terme d'erreur près comme le produit de la tendance et d'une composante de saisonnalité [24] :

$$X_t = T(1 + S_t)(1 + \varepsilon_t). \quad (1.3)$$

Exemples de schémas additif et multiplicatif

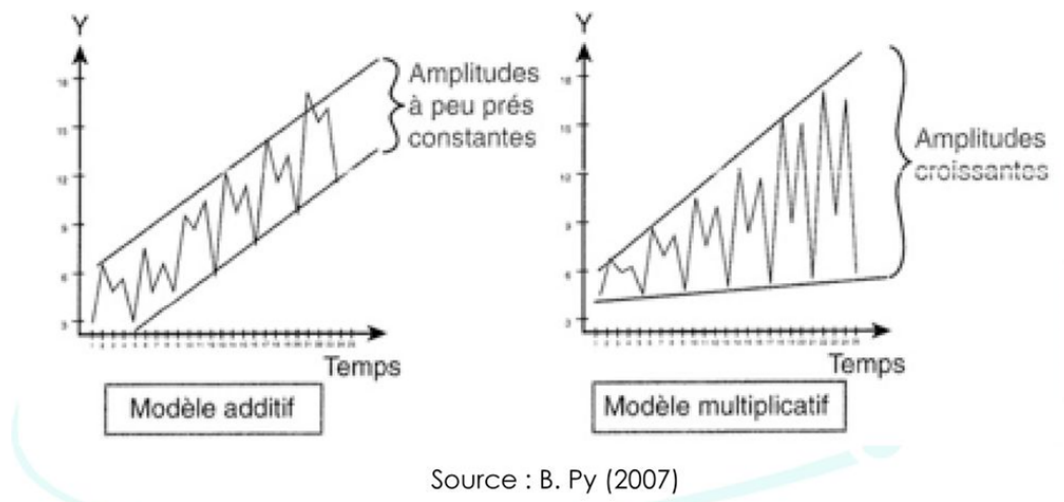


FIGURE 1.2 – Décomposition d'une série chronologique [24].

1.2.4 Série désaisonnalisée

On appelle série désaisonnalisée ou série corrigée des variations saisonnières notée série CVS, la série chronologique Y_t à laquelle on a enlevé les variations saisonnières.

1. Nous traçons le graphe de (Y_t) .
2. Nous choisissons le modèle de composition : additif ou multiplicatif.
3. Nous estimons la tendance C_t , et on la trace.
4. Nous déterminons les valeurs de trend (YI)

$$YI = ax + b \quad (1.4)$$

5. Nous déterminons les rapports (D_t) :

Modèle additif :

$$Dt = yi - YI \quad (1.5)$$

Modèle multiplicatif :

$$Dt = yi/YI \quad (1.6)$$

6. Nous calculons les coefficients saisonniers (st) .

7. Nous calculons la série CVS (Xt) :

Modèle additif :

$$Xt = yi - st \quad (1.7)$$

Modèle multiplicatif :

$$Xt = yi/st \quad (1.8)$$

avec y_i = la valeur observée.

Le but de la décomposition d'une série chronologique est de distinguer dans l'évolution de la série, une tendance générale , des variations saisonnières qui se répètent chaque année, et des variations accidentelles imprévisibles.

1.3 Méthodes de prévision

1.3.1 Moyenne Mobile

1.3.1.1 Définition

Les moyennes sont des méthodes efficaces pour lisser les variations aléatoires, détecter les tendances et prévoir les valeurs futures dans une séries chronologique, elles peuvent être d'ordre pair ou d'ordre impair[33].

1.3.1.2 Avantages

Outil simple à mettre en oeuvre qui met en évidence l'allure de la tendance en supprimant la composante saisonnière et en atténuant le bruit[24].

1.3.1.3 Propriétés d'un lissage par moyennes mobiles

1. Suppression de la composante saisonnière

Si la série y_t possède une composante saisonnière de période p alors l'application d'une moyenne mobile d'ordre p supprime cette saisonnalité. La série $m_{p,t}$ ne possède plus de composante saisonnière de période p [24].

2. Atténuation de la composante résiduelle

Par construction, une moyenne mobile consiste à faire des moyennes partielles de proche en proche. On obtient donc un "lissage" de la série. L'effet de la composante irrégulière est d'autant plus atténué que l'ordre de la moyenne mobile est grand [33].

3. Conservation de la tendance

Pour des moyennes mobiles simples ou centrées, l'application d'une moyenne mobile (d'ordre quelconque) ne modifie pas une tendance constante. En particulier l'application d'une moyenne mobile conserve une tendance linéaire [33].

1.3.1.4 Moyennes mobiles simples

La série des moyennes mobiles d'ordre p , notée $m_{p,t}$, est la série des moyennes de p observations consécutives et elle prend ses valeurs aux dates moyennes correspondantes.

Plus précisément, on calcule les moyennes de p termes consécutifs

Si p est impair : $p = 2k + 1$, la série moyenne mobile est calculée aux mêmes instants que les observations initiales [33]. La formule de la moyenne mobile d'ordre impair est représentée comme suit par l'équation (1.9) :

$$[m_{p,t} = \frac{1}{p} \sum_{i=-k}^k x_{t+i}] \quad (1.9)$$

Pour comparer la série lissée avec la série initiale, on cherchera à obtenir les valeurs pour les mêmes dates d'observation.

On définit les moyennes mobiles centrées pour pallier l'inconvénient des moyennes mobiles d'ordre pair.

Dans tous les cas on fait la moyenne de p observations ($1/p$). Pour obtenir une moyenne mobile pour le temps t , on retiendra $2k$ observations, en pondérant les deux extrêmes par 0.5[33].

La formule de la moyenne mobile d'ordre pair pour les moyennes mobiles centrées est donnée par l'équation (1.10) :

$$[m_{p,t} = \frac{1}{p} \left(\frac{1}{2} \sum_{i=-k}^{k-1} x_{t+i} + \frac{1}{2} \sum_{i=-k+1}^k x_{t+i} \right)] \quad (1.10)$$

où t est l'indice du terme de la moyenne mobile, k est le nombre d'observations de chaque côté de l'indice t , p est le nombre total d'observations utilisées pour la moyenne mobile, et x_i représente les valeurs des termes à prendre en compte.

1.3.1.5 Choix pratique de l'ordre d'une moyenne mobile

Le but d'un lissage par moyenne mobile est de faire apparaître l'allure de la tendance, en gommant les variations saisonnières, faire disparaître la composante saisonnière de période p avec une moyenne mobile d'ordre p .

En pratique on doit trouver le meilleur compromis pour le choix de l'ordre p de lissage optimal [33].

1.3.1.6 Représentations graphiques des moyennes mobiles

Lorsque l'on représente la série initiale et la moyenne mobile d'ordre 4 sur le même graphique on constate que la courbe des moyennes mobiles représente la tendance.

On peut interpréter cette courbe comme la moyenne trimestrielle des ventes de l'année qui entoure chaque valeur [38].

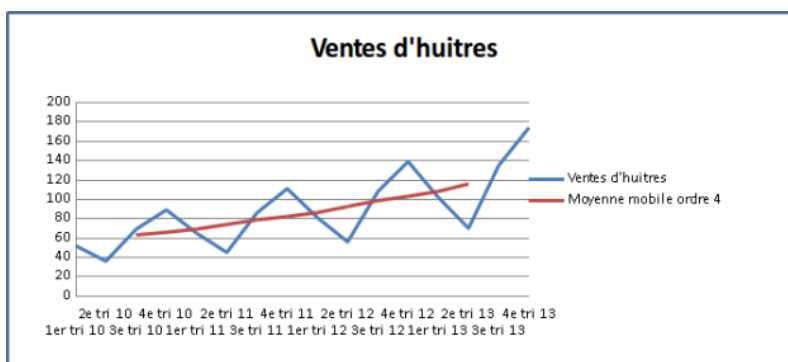


FIGURE 1.3 – Graphe des moyennes mobiles d'ordre 4 des ventes d'huîtres [38].

A titre d'illustration, il est possible de calculer les moyennes mobiles d'ordre 2 et 3 de la même façon.

On constate sur le graphique suivant que seul l'ordre 4 donne une série lissée [38].

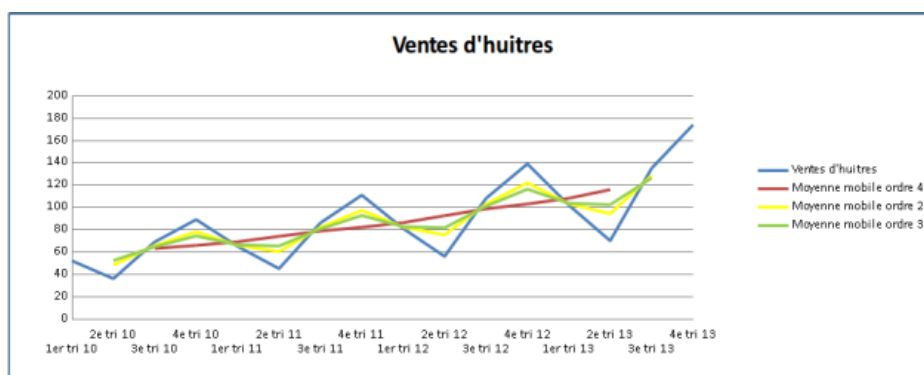


FIGURE 1.4 – Illustrations avec des moyennes mobiles de différents ordres [38].

1.3.2 Lissage exponentiel

1.3.3 Lissage exponentiel simple

1.3.3.1 Définition

C'est une technique très simple de prévision à $t+1$. Elle s'applique à des SC sans tendance. Le principe consiste à donner plus d'importance aux dernières observations. On ne prolonge pas une série comme on le ferait avec une régression simple mais on cherche à obtenir une valeur lissée en t pour la reporter tout simplement en $t + 1$. Elle est plus réactive que les Moyennes Mobiles ou les modèles utilisant la régression car elle prend rapidement en compte une modification de tendance [31].

La formule nous notons la formule de la façon suivante :

$$[\text{LES}(t) = \alpha x_t + (1 - \alpha)\text{LES}(t - 1)] \quad (1.11)$$

où $\text{LES}(t)$ est la valeur lissée à l'instant t , x_t est la valeur observée à l'instant t , et α est le coefficient de lissage exponentiel

La présentation de cette formule est déstabilisante puisque t est le moment où la prévision a été faite et non celui où elle doit se réaliser. Le coefficient α , compris entre 0 et 1, s'applique à la dernière réalisation. α s'appelle la constante de lissage (ou coefficient α de lissage). Évidemment, si elle est égale à 1, on ne fait que reporter en $t + 1$ l'observation de la période t . Le coefficient $(1 - \alpha)$ s'applique quant à lui à la prévision précédente [31].

La prévision initiale : En raison de la formule récurrente du LES, on est obligé de choisir une valeur à partir de laquelle les prévisions seront effectuées. Cette valeur n'a que peu d'importance si la série est longue. On prend souvent la moyenne des deux ou trois premières observations mais ce choix est arbitraire. On peut également prendre la première valeur [31].

Le choix de la constante de lissage

On peut confronter les observations avec ce qu'aurait donné un LES utilisant une constante de lissage de 0,1 puis 0,2 puis 0,3 et ainsi de suite. On se donne un indicateur d'écart pour comparer les séries (On va prendre la somme des carrés des erreurs) [31].

1.3.4 Lissage exponentiel double

1.3.4.1 Définition

Le LES permet d'établir une prévision à $t + 1$ lorsqu'il n'existe pas de tendance. Lorsqu'il y en a une, on peut effectuer un LED. Lorsqu'on souhaite établir une prévision par lissage sur une série avec tendance, on doit alors établir les paramètres d'une tendance linéaire $y = at + b$. Celle-ci ne résume pas les valeurs d'une SC de façon indifférenciée. Selon le principe du lissage exponentiel,

les dernières valeurs ont un poids plus important que les plus anciennes. Du coup, la prévision est recalculée pour chaque observation supplémentaire [31].

La prévision en h

Le LED nécessite une constante de lissage alpha, comprise entre 0 et 1, qui permet de plus ou moins pondérer la dernière observation par rapport aux précédentes.

Comment déterminer les paramètres a et b ?

1. L'estimation de la " constante " b est égale à deux fois la première valeur lissée moins une fois la seconde valeur lissée.

2. Le coefficient a est égal à la différence entre les deux valeurs lissées (la première moins la seconde), multipliée par un coefficient alpha/ (1 - alpha). Ces paramètres sont déterminés par minimisation des carrés des erreurs Pour initialiser le lissage double, on utilise les deux premières valeurs[38].

Un graphique permet de visualiser tout ceci :

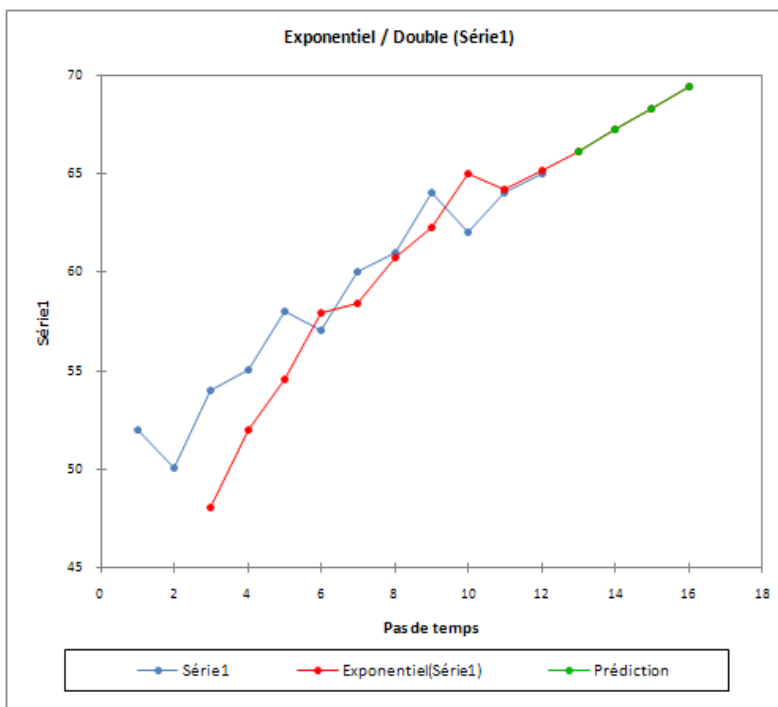
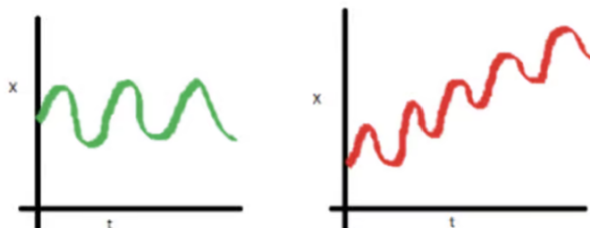


FIGURE 1.5 – Représentation graphique du LED [38].

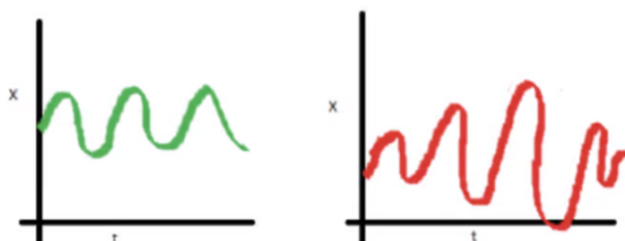
On perçoit bien les limites du LED, surtout pour les prévisions situées au-delà de $t + 1$. Non seulement elles dépendent fortement de la constante de lissage, mais elles peuvent beaucoup varier d'un mois sur l'autre (contrairement à une régression simple où le coefficient directeur reste constant) [38].

1.3.5 Une série stationnaire

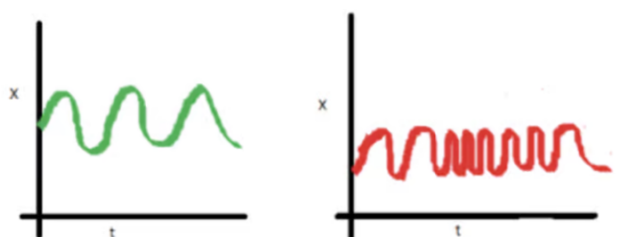
1. La moyenne de la série ne doit pas être fonction du temps. Le graphique rouge ci-dessous n'est pas stationnaire car la moyenne augmente avec le temps [33].



2. La variance de la série ne doit pas être une fonction du temps. Notez dans le graphique rouge ci-dessous la variance des données qui varient dans le temps [33].



3. Enfin, la covariance du i -ème terme et du $(i + m)$ -ième terme ne doit pas être fonction du temps. Dans le graphique suivant, vous remarquerez que l'écart se rapproche à mesure que le temps augmente. Par conséquent, la covariance n'est pas en rapport avec le temps pour la " série rouge " [33].



1.3.6 Vérifier la stationnarité d'une série chronologique

Avant de pouvoir construire un modèle, nous devons nous assurer que la série temporelle est stationnaire. Il y a deux façons principales de déterminer si une série temporelle donnée est stationnaire :

Statistiques roulantes tracer la moyenne mobile et l'écart-type mobile. La série temporelle est stationnaire si elle reste constante dans le temps (à l'oeil nu, regardez si les lignes sont droites et parallèles à l'axe des x) [33].

Test de Dickey-Fuller augmenté (ADF) : La série temporelle est considérée comme stationnaire si la valeur p est faible (selon l'hypothèse nulle) et si les valeurs critiques à des intervalles de confiance de 1 % 5% 10% sont aussi proches que possible des statistiques de l'ADF (Augmented Dickey-Fuller)[33].

1.3.7 Transformer une série non-stationnaire en une série stationnaire

Il existe de multiples transformations que nous pouvons appliquer à une série temporelle pour la rendre stationnaire [33].

Soustraction de la moyenne mobile

Cette méthode consiste à calculer la moyenne mobile sur une fenêtre donnée et à soustraire cette moyenne mobile de la série temporelle. Cela permet de réduire les tendances et les variations lentes présentes dans la série.

La décroissance exponentielle

Cette transformation implique l'application d'une décroissance exponentielle à la série temporelle. Elle est utile pour réduire les variations à long terme et les tendances non linéaires.

le décalage temporel

Cette technique consiste à décaler la série temporelle d'un certain nombre d'intervalles. Cela peut aider à réduire les variations saisonnières ou à rendre la série plus stationnaire en ajustant les retards dans les données.

1.3.8 La méthode ARIMA

Dans le domaine du Machine Learning, il existe un ensemble spécifique de méthodes et de techniques particulièrement bien adaptées pour prédire la valeur d'une variable dépendante en fonction du temps. Dans cette partie, nous aborderons la moyenne mobile intégrée autorégressive (ARIMA).

Il permet en effet de déterminer les valeurs intégrées aux séries temporelles en s'appuyant sur les précédentes valeurs observées. Avec ce système autorégressif, les fonctions prédictives peuvent être appliquées à un cas isolé. Ce modèle statistique permet surtout de suivre et d'anticiper l'évolution d'un phénomène. C'est une notion mathématique que l'on retrouve souvent dans le domaine de l'économétrie et dans l'analyse statistique.

L'algorithme de moyenne mobile intégrée autorégressive est également très présent dans les processus d'action liés aux intelligences artificielles (IA). Il permet de déterminer une perspective de transformation ou de changement. Le modèle de machine learning qui utilise Arima prend en charge des algorithmes et des calculs spécifiques. Ces algorithmes sont nécessaires pour exprimer une prédiction en se basant sur les précédentes valeurs, sur l'échelle de temps utilisée et sur leurs fluctuations.

Avant d'appliquer un modèle statistique sur une série temporelle, nous devons nous assurer qu'elle est stationnaire [33].

1.3.8.1 Modèle Auto-Régressif (AR)

Les modèles auto-régressifs fonctionnent en partant du principe que les valeurs passées ont un effet sur les valeurs actuelles. Les modèles AR sont couramment utilisés pour analyser la nature, l'économie et d'autres processus variables dans le temps. Tant que l'hypothèse tient, nous pouvons construire un modèle de régression linéaire qui tente de prédire aujourd'hui la valeur d'une variable dépendante, compte tenu des valeurs qu'elle avait les jours précédents.

L'ordre du modèle AR correspond au nombre de jours incorporés dans la formule. On peut le représenter de la manière suivante :

$$[\text{Modèle AR}(t) = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p}] \quad (1.12)$$

où t est l'instant de temps actuel, x_t est la valeur observée à l'instant t , $\phi_1, \phi_2, \dots, \phi_p$ sont les coefficients du modèle AR correspondant aux jours précédents jusqu'à p jours, et p est l'ordre du modèle AR.

1.3.8.2 Modèle de la Moyenne Mobile (MA)

Supposez que la valeur de la variable dépendante du jour en cours dépend des termes d'erreur des jours précédents. La formule peut être exprimée sous cette forme :

$$[\text{Modèle MA}(t) = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}] \quad (1.13)$$

où t est l'instant de temps actuel, $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ sont les termes d'erreur des jours précédents jusqu'à q jours, et $\theta_1, \theta_2, \dots, \theta_q$ sont les coefficients du modèle MA correspondant aux termes d'erreur.

1.3.8.3 Modèle de la Moyenne Mobile Auto-Régressive (ARMA)

Le modèle de la Moyenne Mobile Auto-Régressive (ARMA) combine à la fois des termes auto-régressifs (AR) et des termes de la moyenne mobile (MA). La formule générale pour un modèle ARMA est donnée par :

$$[\text{Modèle ARMA}(t) = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}] \quad (1.14)$$

où t est l'instant de temps actuel, x_t est la valeur observée à l'instant t , $\phi_1, \phi_2, \dots, \phi_p$ sont les coefficients auto-régressifs, $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ sont les termes d'erreur des jours précédents, et $\theta_1, \theta_2, \dots, \theta_q$ sont les coefficients de la moyenne mobile.

1.3.8.4 Modèle de la Moyenne Mobile Auto-Régressive Intégrée (ARIMA)

Le modèle ARIMA ajoute une différence à un modèle ARMA. La différenciation soustrait la valeur actuelle de la précédente et peut être utilisée pour transformer une série temporelle en une série stationnaire. Par exemple, la différenciation du premier ordre traite des tendances linéaires et utilise la transformation $z_i = y_i - (y_{i-1})$ [29].

La différenciation du second ordre traite des tendances quadratiques et utilise une différence du premier ordre sur une différence du premier ordre, à savoir $z_i = (y_i - (y_{i-1})) - ((y_{i-1}) - (y_{i-2}))$, et ainsi de suite [29].

Trois entiers (**p**, **d**, **q**) sont généralement utilisés pour paramétrer les modèles ARIMA :

p : nombre de termes autorégressifs (ordre AR)

d : nombre de différences non saisonnières (ordre de différenciation)

q : nombre de termes moyens mobiles (ordre MA)

1.3.8.5 Fonctions d'Auto-Corrélation et d'Auto-Corrélation partielle

Fonction d'Auto-Corrélation ACF

La fonction d'autocorrélation est une mesure de l'association entre des valeurs des séries passées et actuelles. Il indique quelles valeurs de séries passées sont les plus importantes pour prédire les valeurs futures. L'ordre des processus d'un modèle ARIMA peut être déterminé grâce à ces données [27].

Fonction d'Auto-Corrélation partielle PACF

Elle permet de calculer et de faire un graphique de la corrélation entre les observations des séries chronologiques. C'est une mesure de la corrélation entre les observations d'une série chronologique séparée par k unité de temps. Les données des séries devront être collectées et consignées par ordre chronologique. Ces données doivent être enregistrées dans une feuille de travail dans l'ordre de la collecte [22].

1.3.9 La méthode SARIMA

SARIMA est une technique largement utilisée dans l'analyse des séries chronologiques pour prédire les valeurs futures sur la base de données historiques ayant une composante saisonnière. C'est une ARIMA saisonnière, ou simplement ARIMA avec une composante saisonnière [19].

Il permet de modéliser les séries temporelles comportant une composante saisonnière et désigné par 7 paramètres :

p, d, q : les même que ceux de ARIMA.

P : ordre de la partie autorégressive saisonnière.

D : ordre de la différence saisonnière.

Q : ordre de la moyenne mobile saisonnière.

m : la période de la composante saisonnière.

En dehors de ces 7 paramètres, l'implémentation de SARIMA permet d'intégrer une variable exogène qui permet à la fois d'améliorer les performances du modèle et d'expliquer le comportement de la série étudiée [19].

1.3.10 Équation SARIMA

$$[\text{SARIMA} = (p, d, q) \times (P, D, Q)] \quad (1.15)$$

où (p, d, q) représente l'ordre ARIMA (autoregressive, intégré, moyenne mobile) et (P, D, Q) représente l'ordre des termes saisonniers.

Les paramètres de ces types de modèles sont les suivants :

p et P saisonnier : indiquent le nombre de termes AR (décalages de la série stationnaire)
d et saisonnier D : indiquent les différenciations qu'il faut faire aux séries stationnaires q et Q
saisonnier : indiquent le nombre de termes MA (décalages des erreurs de prévision) lag : indique la durée saisonnière dans les données [19].

1.3.10.1 Modélisation Sarima

La modélisation d'une série chronologique de données est un processus hautement subjectif et individuel. On peut avoir des paramètres différents pour la même série temporelle. Il n'y a donc pas de solution fixe. La meilleure solution est celle qui répond avec succès aux exigences de l'entreprise. En raison de ce niveau de subjectivité impliqué, il devient parfois difficile de comprendre le processus de construction du modèle [19].

Les étapes du processus de construction du modèle de manière structurée sont les suivantes :

- 1- Tracer la série (pour vérifier les valeurs aberrantes)
- 2- Transformer les données (pour rendre la moyenne et la variance constantes)
- 3- Appliquer des tests statistiques pour vérifier si la série est stationnaire (à la fois tendance et saisonnalité)
- 4- Si non stationnaire (avec une tendance ou une saisonnalité), rendez-le stationnaire en différenciant
- 5- Tracer l'ACF des séries stationnaires pour l'ordre MA, l'ordre MA saisonnier aux pics saisonniers
- 6- Tracer PACF des séries stationnaires pour l'ordre AR, ordre AR saisonnier aux pics saisonniers
- 7- Exécuter SARIMA avec ces paramètres

8- Vérifier la validité du modèle à l'aide de parcelles résiduelles [19].

1.3.11 Types du Machine Learning

Le Machine Learning est généralement divisé en trois types :

1.3.11.1 L'apprentissage supervisé

L'apprentissage supervisé est un type d'apprentissage automatique dans lequel un algorithme apprend à partir d'un ensemble de données étiquetées (c'est-à-dire des données avec une réponse connue). L'algorithme utilise ces données pour prédire la réponse pour de nouvelles données.

Par exemple, un algorithme d'apprentissage supervisé peut apprendre à reconnaître les chiffres manuscrits en utilisant des images de chiffres manuscrits étiquetées.

1.3.11.2 L'apprentissage non supervisé

L'apprentissage non supervisé est un type d'apprentissage automatique dans lequel un algorithme apprend à partir d'un ensemble de données non étiquetées (c'est-à-dire des données sans réponse connue). L'algorithme utilise ces données pour découvrir des motifs et des tendances dans les données.

Par exemple, un algorithme d'apprentissage non supervisé peut regrouper des clients similaires en fonction de leurs habitudes d'achat.

1.3.11.3 L'apprentissage par renforcement

L'apprentissage par renforcement est un type d'apprentissage automatique dans lequel un algorithme apprend à partir de ses interactions avec un environnement.

L'algorithme prend des décisions et reçoit des récompenses ou des pénalités en fonction des résultats de ces décisions.

L'algorithme utilise ces récompenses et pénalités pour apprendre à prendre de meilleures décisions à l'avenir.

En résumé, le Machine Learning est une technique d'analyse de données qui utilise des algorithmes pour apprendre à partir de données et pour effectuer des prévisions ou des classifications sur de nouvelles données.

Les différents types d'apprentissage automatique sont l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement [21].

1.3.12 Quel est le rapport entre le Machine Learning et l'IA ?

Le Machine Learning, ainsi que l'apprentissage profond et les réseaux de neurones artificiels qui le composent, constituent des sous-ensembles concentriques de l'IA.

L'IA traite les données pour prendre des décisions et émettre des prévisions. Les algorithmes de Machine Learning permettent non seulement à l'IA de traiter ces données, mais aussi de les exploiter pour apprendre et devenir plus intelligente, sans avoir à recourir à une programmation supplémentaire. L'intelligence artificielle est la mère du Machine Learning et de tous les sous-ensembles qui le composent.

Le premier sous-ensemble englobe le domaine du Machine Learning, qui, à son tour, comprend l'apprentissage profond. Au sein de l'apprentissage profond se trouvent les réseaux de neurones artificiels [21].

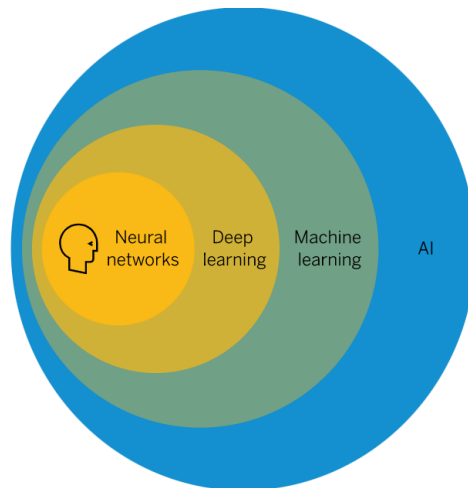


FIGURE 1.6 – Diagramme de la relation entre IA et Machine Learning [21].

1.3.13 Comment fonctionne le Machine Learning ?

Le Machine Learning se décline sous différents types de modèles, qui emploient chacun des techniques algorithmiques différentes. Selon la nature des données et le résultat souhaité, l'un de ces quatre modèles d'apprentissage peut être utilisé : supervisé, non supervisé, semi-supervisé ou par renforcement.

Dans chacun de ces modèles, une ou plusieurs techniques algorithmiques peuvent être appliquées. Tout dépend des ensembles des données qui seront utilisés et de l'objectif visé au niveau des résultats. Par nature, les algorithmes de Machine Learning sont conçus pour classifier des éléments, repérer des patterns, prévoir des résultats et prendre des décisions éclairées. Les algorithmes peuvent être mis en œuvre individuellement ou en groupe dans le but d'atteindre la plus grande précision possible lorsque les données utilisées sont complexes et imprévisibles [21].

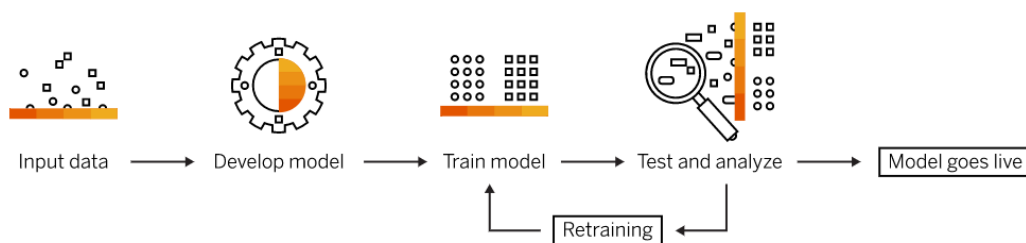


FIGURE 1.7 – Comment fonctionne le Machine Learning? [2].

1.3.14 Modèles du Machine Learning

Il existe de nombreux modèles machine learning, mais certains sont plus populaires que d'autres.

Voici quelques-uns des modèles les plus populaires :

1.3.14.1 Machine à vecteurs de support (SVM)

Le principe des SVM consiste à ramener un problème de classification ou de discrimination à un hyperplan dans lequel les données sont séparées en plusieurs classes dont la frontière est la plus éloignée possible des points de données [22].

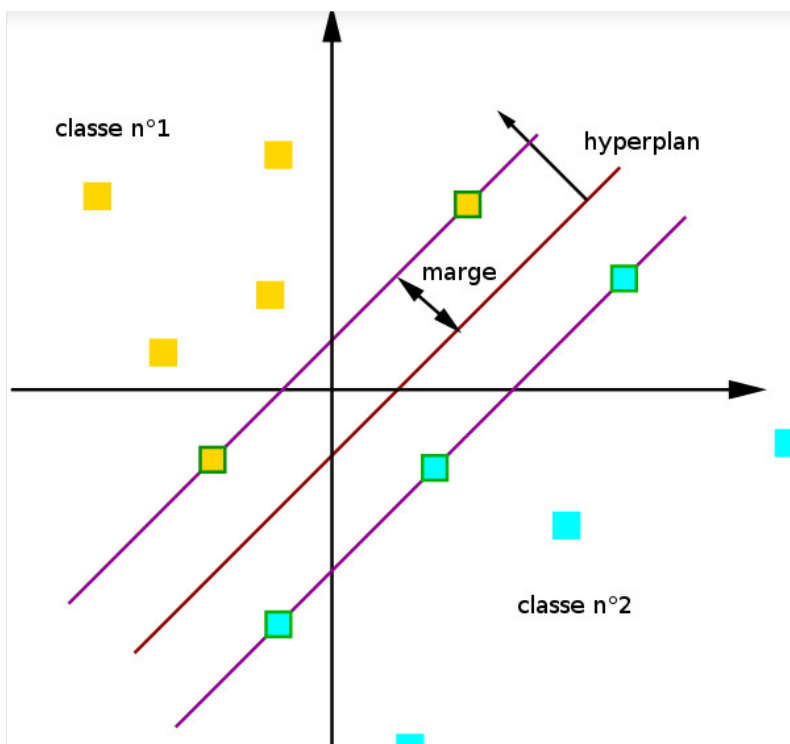


FIGURE 1.8 – Machine à vecteurs de support [22].

1.3.14.2 La répartition en K-moyennes (K-means)

L'algorithme K-means s'appuie sur une méthode itérative pour trier des points de données en groupes basés sur des caractéristiques similaires. La répartition en K-moyennes a la réputation d'être précise, tout en étant capable de gérer des groupes de données en un laps de temps relativement court. Ce type d'algorithme est aussi bien utilisé par les éditeurs de moteur de recherche pour proposer des résultats pertinents ou par des entreprises qui veulent classifier les comportements des utilisateurs. Cette technique s'avère également efficace dans le contexte d'analyse de performances informatiques [21].

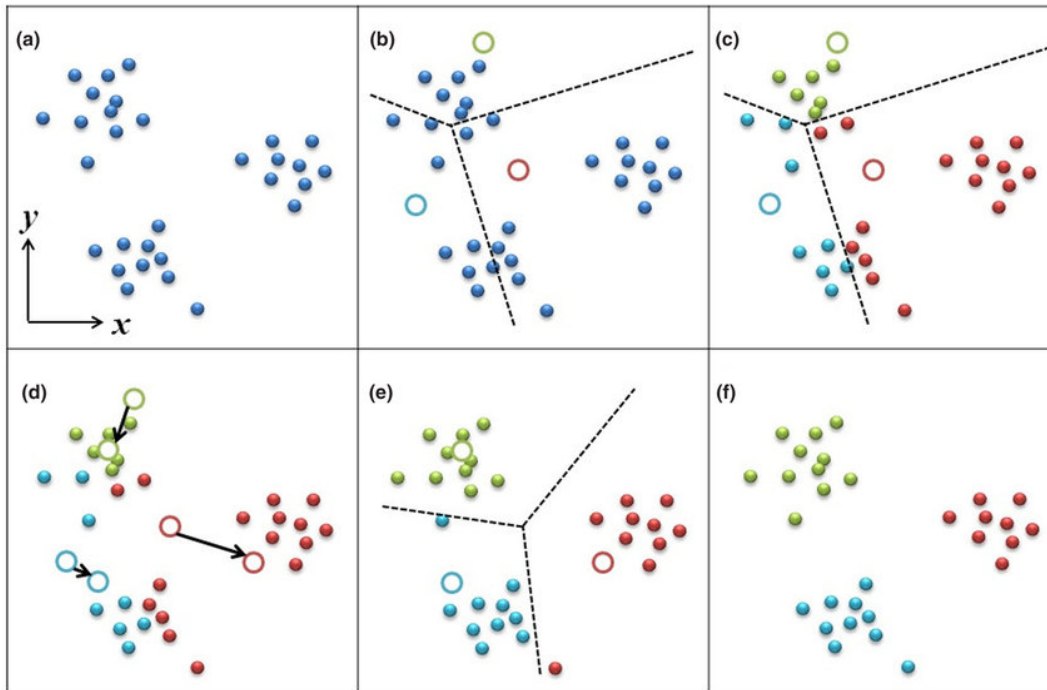


FIGURE 1.9 – Schéma de l'algorithme K-means [21].

1.3.14.3 Régression linéaire

Les algorithmes de régression linéaire sont les plus utilisés par les équipes de data science. Il s'agit d'effectuer des corrélations simples entre deux variables dans un jeu de données. Un ensemble d'entrées et les sorties correspondantes sont examinés et quantifiés pour montrer une relation, par exemple comment le changement d'une variable affecte une autre. Les régressions linéaires sont représentées sous forme de lignes sur un graphique [21].

1.4 Deep Learning

Le Deep Learning, ou apprentissage profond en français, est une technique d'apprentissage automatique qui permet aux ordinateurs d'apprendre à partir de données de manière autonome.

Cette technique est basée sur des réseaux de neurones artificiels qui sont capables de traiter des données complexes, tels que des images, des vidéos ou du langage naturel.

Le Deep Learning est différent des méthodes d'apprentissage automatique traditionnelles en ce qu'il utilise des réseaux de neurones artificiels qui sont constitués de plusieurs couches. Chaque couche du réseau de neurones traite les données de manière progressive, en extrayant des caractéristiques de plus en plus complexes de la donnée d'entrée. Cela permet au réseau de neurones d'apprendre des représentations de plus en plus abstraites des données.

1.4.1 Différentes applications du deep learning

Les applications du Deep Learning sont nombreuses et diverses, notamment :

- La reconnaissance de la parole
- La reconnaissance d'images
- La prédiction de la demande
- La détection d'anomalies
- La prédiction de la santé
- La reconnaissance de la fraude
- La reconnaissance faciale
- La traduction automatique

Il est souvent considéré comme la pierre angulaire de l'intelligence artificielle.

1.4.2 Différence entre deep learning et machine learning

Le deep learning constituant un sous-domaine du machine learning, il convient de connaître les différences entre ces deux types d'intelligence artificielle.

Le Machine Learning (apprentissage automatique) est une technologie permettant à un système d'évoluer à partir d'un retour humain. Cela implique donc de fournir des données organisées. À partir de celles-ci, l'ordinateur est capable de comprendre ces données, de les catégoriser et d'exécuter des actions programmées. Les algorithmes sont optimisés grâce aux développeurs et les bases de données sont contrôlables.

Le Deep Learning (apprentissage profond), quant à lui, est la section du machine learning pouvant fonctionner à partir de données non structurées. En effet, ces algorithmes s'appuient sur un réseau neuronal de plusieurs couches, similaire au cerveau humain. Cela permet au système de traiter des données sans même que celles-ci n'aient été préparées par des humains en amont.

Un IA à apprentissage profond possède donc la capacité d'apprendre en toute autonomie, contrairement à un IA à apprentissage automatique, qui nécessite l'intervention humaine [23].

1.4.3 Les réseaux de neurones artificiels

Un réseau de neurones artificiels est représenté sous la même forme que des neurones dans un cerveau biologique. Les neurones artificiels sont appelés " noeuds " et sont regroupés en plusieurs couches, qui fonctionnent en parallèle [26].

1.4.4 Comment se compose un réseau de neurones ?

Un réseau de neurones artificiels est composé de trois éléments principaux : les neurones, les poids et les biais, et les fonctions d'activation.

1.4.4.1 Les neurones

Les neurones reçoivent les données en entrée et effectuent des calculs pour générer des sorties. Les neurones sont organisés en couches, avec des couches d'entrée, des couches cachées et des couches de sortie. Les neurones d'entrée reçoivent les données du monde extérieur, les neurones de traitement effectuent des calculs sur les données et les neurones de sortie génèrent les sorties du réseau [26].

1.4.4.2 Les poids et biais

Les poids et les biais sont des paramètres utilisés pour ajuster les entrées des neurones et influencer leur comportement. Les poids sont utilisés pour pondérer les entrées des neurones et les biais sont utilisés pour ajuster les valeurs d'entrées auxquelles les poids ont été appliqués. Les poids et les biais sont continuellement mis à jour pendant l'apprentissage pour améliorer la précision du réseau [26].

1.4.4.3 La fonction d'activation

La fonction d'activation est utilisée pour normaliser les données d'entrée avant de les transmettre aux neurones suivants. Elle est activée lorsque la valeur calculée atteint le seuil requis fixé. La fonction d'activation peut être **linéaire** ou **non linéaire**, et elle peut être différente pour chaque neurone [26].

Enfin, un réseau de neurones peut avoir plusieurs couches cachées, ce qui donne lieu à ce qu'on appelle le Deep Learning [37].

1.4.5 Les différents réseaux de neurones

On peut distinguer 4 grandes familles dans lesquelles les réseaux de neurones artificiels peuvent être regroupés.

1.4.5.1 Les réseaux de neurones feed-forwarded

Au sein des architectures de type feed-forwarded, les informations sont transmises exclusivement vers l'avant, sans jamais faire demi-tour. Cette famille de réseaux neuronaux comprend deux catégories de neural networks : **Le perceptron simple** et **Le perceptron multicouche**, tous deux caractérisés par leur fonctionnement bien spécifique [26].

Le perceptron simple est un neural network qui se compose de seulement deux couches de neurones : une couche d'entrée et une couche de sortie, toutes deux étroitement reliées. [26].

Le perceptron multicouche quant à lui, se caractérise par la présence d'une ou de plusieurs couches cachées entre la couche d'entrée et celle de sortie. L'algorithme intègre ainsi plusieurs matrices de poids, ce qui le rend particulièrement adapté au traitement des fonctions non linéaires [26].

1.4.5.2 Les réseaux de neurones à résonance

Les réseaux de neurones à résonance se distinguent par leur fonctionnement complexe, caractérisé par des oscillations provoquées par la répercussion de l'activation d'un neurone sur l'ensemble de ses homologues. Cette architecture, aux contours multiples et variés, se décline en de nombreuses formes, dont la complexité peut atteindre des sommets vertigineux [26].

1.4.5.3 Les réseaux de neurones récurrents

Les réseaux de neurones récurrents (RNN) sont utilisés pour traiter des séries temporelles (langue, vidéo, données numériques). ils gardent la mémoire des données passées pour prédire des séries de données futures proches [26].

1.4.5.4 Les types de réseaux de neurones récurrents

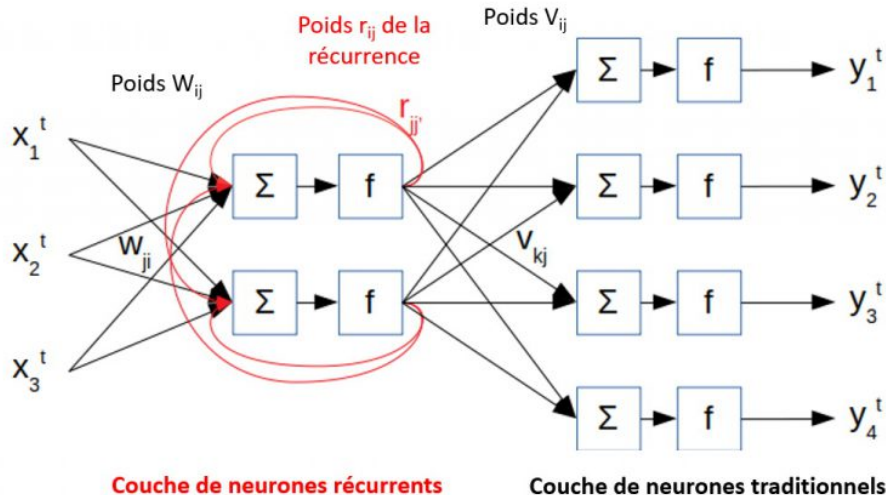


FIGURE 1.10 – Les types de réseaux de neurones récurrents [26].

Les réseaux de neurones traditionnels ne sont pas adaptés à la prédiction de séries chronologiques. Pour cela, nous utilisons des réseaux de neurones récurrents, tels que les RNN simples, LSTM et GRU. Les RNN simples sont peu utilisés en pratique. En revanche, les LSTM et GRU, avec leurs portes qui permettent de mémoriser sélectivement les informations des séries temporelles précédentes, sont plus couramment employés. Les réseaux de neurones récurrents peuvent contenir plusieurs couches pour capturer davantage de non-linéarités dans les données, mais cela augmente le temps de calcul pendant l'apprentissage. Nous pouvons également combiner des couches récurrentes avec des couches classiques telles que les couches denses ou convolutives [26].

1.4.5.5 Les réseaux de neurones autoorganisés

Les réseaux de neurones auto-organisés sont utilisés pour le traitement de l'information spatiale. Ils sont capables d'étudier la distribution de données à grande échelle afin de fournir des solutions aux problèmes de classification. La carte auto-organisée de Kohonen est le modèle de réseau de neurones artificiels auto-organisé le plus célèbre [26].

1.4.6 L'architecture de LSTM

Les LSTM traitent à la fois de la mémoire à long terme (LTM) et de la mémoire à court terme (STM).

les cellules LSTM (Long Short Term Memory), qui possèdent une mémoire interne appelée cellule (ou cell). La cellule permet de maintenir un état aussi longtemps que nécessaire. Cette cellule consiste en une valeur numérique que le réseau peut piloter en fonction des situations. Pour rendre les calculs simples et efficaces, ils utilisent le concept de portes [36].

La figure suivante présente le schéma d'un seul neurone LSTM :

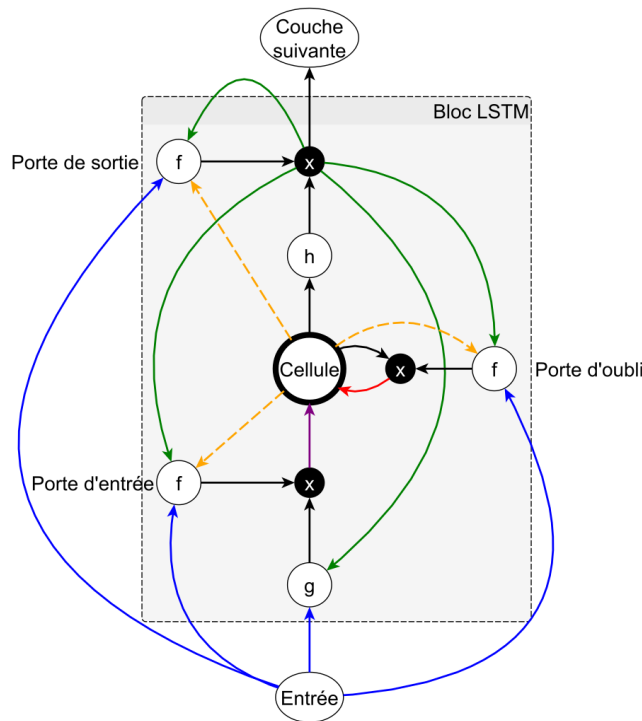


FIGURE 1.11 – Schéma d'un neurone LSTM [36].

Comme nous pouvons le constater sur le schéma, la cellule mémoire peut être pilotée par trois portes de contrôle qu'on peut voir comme des vannes :

La porte d'entrée : décide si l'entrée doit modifier le contenu de la cellule ;

La porte d'oubli : décide s'il faut remettre à 0 le contenu de la cellule ;

La porte de sortie : décide si le contenu de la cellule doit influencer sur la sortie du neurone ;[36].

Le mécanisme des trois portes est strictement similaire.

L'ouverture/la fermeture de la vanne est modélisée par une fonction f qui est généralement une sigmoïde. Cette sigmoïde est appliquée à la somme pondérée des entrées (en bleu), des sorties (en vert) et de la cellule (en orange), par des poids w_{ab} spécifiques à chaque connexion entre les signaux a et b [36].

Soient x_i et z_h les entrées et les sorties de la cellule, s_c la valeur de la cellule et l, w, c les indices décrivant respectivement les signaux issus de la porte d'entrée, d'oubli et de la cellule. w_{il} sont donc les poids reliant les entrées à la cellule d'entrée, w_{wh} les poids reliant les sorties à la porte d'oubli [36]. Les équations régissant les trois portes de contrôle sont donc les suivantes ; elles sont l'application de la somme pondérée suivie de l'application d'une activation f , typiquement la sigmoïde :

Porte d'entrée

$$[i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)] \quad (1.16)$$

Le vecteur de la porte d'entrée à l'instant t est représenté par \mathbf{i}_t .

La fonction d'activation sigmoïde est σ .

La matrice de poids est \mathbf{W}_i et elle est appliquée aux vecteurs d'entrée \mathbf{h}_{t-1} et \mathbf{x}_t .

Le biais associé à la porte d'entrée est \mathbf{b}_i .

Porte d'oubli

$$[f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)] \quad (1.17)$$

Le vecteur de la porte d'oubli à l'instant t est représenté par \mathbf{f}_t .

La fonction d'activation sigmoïde est σ .

La matrice de poids est \mathbf{W}_f et elle est appliquée aux vecteurs d'entrée \mathbf{h}_{t-1} et \mathbf{x}_t .

Le biais associé à la porte d'oubli est \mathbf{b}_f .

Porte de sortie

$$[o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)] \quad (1.18)$$

Le vecteur de la porte de sortie à l'instant t est représenté par \mathbf{o}_t .

La fonction d'activation sigmoïde est σ .

La matrice de poids est \mathbf{W}_o et elle est appliquée aux vecteurs d'entrée \mathbf{h}_{t-1} et \mathbf{x}_t .

Le biais associé à la porte de sortie est \mathbf{b}_o .

1.4.6.1 Utilisation des LSTM

L'entraînement des LSTM supprime le problème du gradient de fuite (les poids deviennent trop petits qui ne correspondent pas au modèle), mais il est toujours confronté au problème du gradient explosif (les poids deviennent trop grands qui sur-correspondent au modèle). La formation des LSTM peut être facilement effectuée en utilisant des frameworks Python comme Tensorflow, Pytorch, Theano, etc. et le piège est le même que RNN, nous aurions besoin d'un GPU pour former des réseaux LSTM plus profonds [37].

1.4.7 Gestion de projet

La gestion de projet, ou le management de projet, consiste à organiser le déroulement d'un projet de A à Z, de sa phase de conception à sa phase finale. Pour ce faire, il faut définir les objectifs, les ressources humaines et matérielles nécessaires, le budget, les délais et les contraintes éventuelles [32].

1.4.7.1 parmi les méthodes adoptées pour la gestion de projet

La méthode Agile La méthode Agile les besoins du customer sont placés au coeur des ententes. Cette méthode flexible consiste à découper un projet en plusieurs petits ententes, de sorte que le customer puisse valider chaque étape au fur et à mesure. La méthode Agile est importante pour assurer la satisfaction de ses clients à la fin du projet [32].

Parmi les méthodes Agiles les plus utilisées :

XP (eXtreme Programming)

RAD (Rapid Application Development)

Scrum

Gérer des projets ne s'improvise pas. Chaque projet étant totalement différent d'un autre, il n'existe pas de méthode miracle pour réussir un projet. Il convient de sélectionner la méthode la plus appropriée à chaque projet. Certaines grandes étapes se retrouvent toutefois généralement d'une technique de gestion de projet à une autre, dans notre cas, nous avons opté pour la méthode SCRUM, en voici quelques détails [32][6] :

1.4.7.2 Scrum

Le Scrum ou « mêlée » est un terme emprunté au rugby. En rugby, la mêlée a pour objectif « de reprendre le jeu rapidement, en toute sécurité et équitablement, après une faute mineure ou un arrêt de jeu » [32].

Dans la méthode Scrum, le Scrum désigne la réunion quotidienne (Daily Scrum Meeting) qui réunit l'ensemble des acteurs du projet pour dresser, en un temps limité à quinze minutes, le bilan de la journée passée, planifier celle qui commence et repérer les éventuels obstacles rencontrés par chacun. Le Scrum symbolise la solidarité et la force qui lient les membres de l'équipe pour le succès de l'itération [32][6].

1.4.7.3 Les trois piliers de Scrum

Scrum est un processus empirique il se base sur l'expérience du terrain. Il s'appuie sur trois piliers. Ils suivent également les principes des méthodes agiles [32][6].

La transparence

L'inspection

L'adaptation

1.4.7.4 Les différents rôles dans la méthode Scrum

1. **Le rôle du Scrum Master (Facilitateur agile)** Le Scrum Master est le guide de l'avancement du projet, celui qui s'assure que les principes et les valeurs du Scrum sont

respectés. C'est le coordinateur des équipes qui vérifie que la communication est au top. Il améliore aussi la productivité et il lève les obstacles.

2. **Le rôle du Product Owner** c'est l'expert qui collabore avec le client. Il définit à la suite des feedbacks clients les spécificités fonctionnelles du produit, Il les priorise ensuite avec l'équipe et Il valide les fonctionnalités développées [32][6].
3. **Les membres de l'équipe** elle est constituée des développeurs. Dans la méthode SCRUM, il n'est pas censé y avoir de hiérarchie entre eux, quand bien même leur savoir-faire et compétence seraient différents. Idéalement une équipe contient 6 à 10 personnes pour être la plus efficace possible.[32][6].

1.4.7.5 Les grandes étapes de la méthode Scrum

Étape 1 le Product Backlog Dans cette phase, le Product Owner rencontre le client et analyse son besoin. Il identifie toutes les fonctionnalités dont le produit devra être composé (les user stories) dans ce qui s'appelle le Product Backlog. Ce « cahier des charges » n'est pas fixé pour toujours, et pourra évoluer en fonction des besoins du client et l'avancement du projet. L'équipe décide de ce qu'elle peut faire et dans quel ordre le faire [32][6].

Étape 2 le sprint La méthode SCRUM se caractérise par une répartition de chacune des tâches à faire. L'équipe trie les fonctionnalités et tâches qu'elle répartit dans des Sprint (durée de cycle de deux semaines). Et pendant ce cycle, l'équipe s'occupera par exemple uniquement de coder une fonctionnalité du produit qu'elle devra livrer à la fin de cette phase [32][6].

Le Daily SCRUM la mêlée quotidienne Dans l'optique de respecter la méthode SCRUM, chaque matin, quand l'équipe est au complet on réalise le daily scrum : une réunion de 5-10 minutes où l'on parle de trois choses : Ce qu'on a fait hier ? Quels problèmes on a rencontrés ? Que va-t-on faire aujourd'hui ?

Cette réunion permet de motiver et stimuler les équipes et sans cesse se poser la question du sens de ses actions [32][6].

Étape 3 sprint Review et Sprint rétrospective Chaque période de sprint s'achève avec deux bilans qui permettent de partager les retours d'expérience de chacun et de préparer le prochain sprint.

Le premier est la Sprint review dédiée au " quoi " animée par le Product Owner (le fameux Monsieur quoi), il s'agit d'une revue de qui a été fait. Celle-ci s'accompagne souvent d'une démonstration du produit, ou d'une présentation des améliorations [32][6].

Le second est la Sprint rétrospective consacré au " comment " L'équipe doit réfléchir aux obstacles de la dernière période, des points faibles et des points forts. Le Scrum Master joue le modérateur. Si besoin, il veille à recentrer la réunion sur les points d'amélioration à mettre en oeuvre pour la suite [32][6].

Enfin, l'équipe met à jour le Backlog et planifie le prochain sprint

Schéma récapitulatif de la méthode SCRUM :

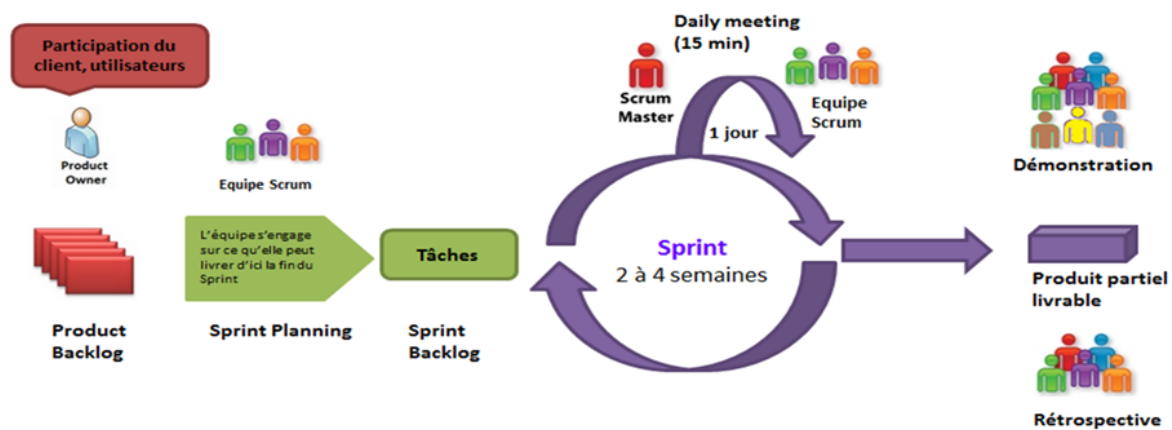


FIGURE 1.12 – Schéma de la méthode SCRUM.

1.5 Conclusion

En somme, ce chapitre nous a permis de dresser un panorama des technologies incontournables dans la réalisation de notre projet. Nous avons abordé la notion d'application web et les séries temporelles, en présentant les différents modèles et types existants. Nous avons également évoqué l'importance de l'IA, du ML et du DL dans l'analyse et la prédiction de la demande d'électricité, et comment ces technologies peuvent être mises à profit pour optimiser l'efficacité énergétique. Enfin, nous avons souligné l'importance de la gestion de projet, un aspect crucial dans tout projet informatique, pour assurer la bonne coordination de l'équipe de développement et la réussite du projet.

Dans le chapitre suivant, nous allons aborder l'entreprise d'accueil de notre projet afin d'identifier les fonctionnalités spécifiques qu'elle requiert.

ANALYSE DES BESOINS ET CONCEPTION

2.1 Introduction

L'analyse des besoins et la conception constituent une phase cruciale dans le processus d'analyse d'un projet, car elles permettent de créer des applications de qualité.

Ce chapitre sera consacré à l'étude des besoins et à la conception de l'application développée. Nous commencerons par identifier l'équipe Scrum du projet, les acteurs principaux de l'application, puis nous aborderons le backlog de produit et la planification des sprints. Enfin, nous procéderons à la modélisation des différents diagrammes qui représentent notre système.

2.2 Présentation du groupe SONELGAZ

Sonelgaz est l'opérateur historique dans le domaine de la fourniture des énergies électrique et gazière en Algérie. Créé en 1969, Sonelgaz, oeuvre depuis un demi-siècle au service du citoyen algérien en lui apportant cette source énergétique essentielle à la vie quotidienne.

A la faveur de la promulgation de la loi sur l'électricité et la distribution du gaz par canalisations, Sonelgaz est passée d'une entreprise verticalement intégrée à une holding pilotant un Groupe industriel multi-sociétés et multi-métiers.

Sonelgaz a toujours joué un rôle majeur dans le développement économique et social du pays. Sa contribution dans la concrétisation de la politique énergétique nationale est à la mesure des importants programmes réalisés, en matière d'électrification rurale et de distribution publique gaz ; ce qui a permis de hisser le taux de couverture en électricité à 98% pour 10 983 538 clients et un taux de pénétration du gaz à 65% pour 6 886 407 clients. Aujourd'hui, le groupe Sonelgaz est composé de 14 sociétés filiales, gérées directement par la holding et de 12 sociétés en participations avec des tiers [35].

2.3 La transformation de SONELGAZ en groupe :

A partir de 2004, le processus de transformation de SONELGAZ relative à la loi N° 02-01 du 05 février 2002 permet non seulement l'ouverture de la production de l'électricité à la concurrence mais aussi la séparation, sous forme de filiales érigées en SPA, des fonctions de production, de transport de l'électricité et du gaz ainsi que la distribution, en effet, ce processus a donné lieu à partir du 2004 à la création de nouvelles filiales [35].

En 2004

Naissance de :

- Une société de transport du gaz : gestionnaire du réseau de transport du gaz (GRTG) ;
- Une société de production constituée des centrales existantes : Sonelgaz production de l'électricité(SPE) ;
- Une société de transport de l'électricité transitoirement gestionnaire du système production transport : gestionnaire du réseau de transport de l'électricité (GRTE) ;
- Première modification des statuts du fonds des oeuvres sociales et culturelles (FOSC)

En 2005

- Création de la commission de régulation d'électricité et du gaz (CREG) qui est un organisme indépendant et autonome doté de la personnalité juridique et investi de trois missions principales :

- 1) Réalisation et contrôle du service public ;
- 2) Conseil auprès des pouvoirs publics en ce qui concerne l'organisation et le fonctionnement du marché de l'électricité et du marché intérieur du gaz ;
- 3) Surveillance et contrôle du respect des lois et règlements relatifs au marché de l'électricité et au marché intérieur du gaz.

- Création de la société civile de médecine du travail (SMT) ;
 - Création d'une société de recherche et développement de l'électricité et du gaz (CREDEG) ;
 - Fusion en une société des quatre (04) sociétés de maintenance et de prestations véhicules (MPV) ;
 - Fusion des trois (03) sociétés de maintenance des transformateurs et une seule société (SMT) ;
- Préparation de la filialisation de la distribution : création de quatre (04) directions générales régionales.

En 2006

- Création de quatre (04) sociétés de Distribution :
 1. Sonelgaz distribution Alger (SDA) ;
 2. Sonelgaz distribution Centre(SDC) ;
 3. Sonelgaz distribution Est (SDE) ;

4. Sonelgaz distribution Ouest (SDO).

- Mise en place de la société opérateur système production-transport : OS, SPT ;
- Nouvelle modification des statuts du fonds des oeuvres sociales et culturelles ;
- Réintégration des cinq (05) sociétés travaux : KAHRAKIB, KANAGAZ, KAHRIF, ETTER-KIB et INERGA [35].

En 2007

- Création de l'industrie de formation en électricité et gaz (IFEG) ;
- Réorganisation de la SPE en quatre (04) pôles nationaux de production d'électricité dont un pôle sud.

En 2009

- Création de la société de gestion du patrimoine immobilier (SOPIG) ;
- Création de la compagnie d'engineering de l'électricité et du gaz (CEEG) ;
- Création de la société des systèmes (ELIT).

2.4 Les différents types de client Sonelgaz

Les types de clients sont définis en fonction du type de tension (basse tension ou haute tension) qui lui sont attribués. Parmi ces derniers on distingue :

- **Clients HTB (Haute Tension B)** : Il s'agit des clients qui utilisent des lignes HTB qui constituent le réseau de répartition ou d'administration régionale et permettent le transport à l'échelle régionale ou locale. Ces lignes acheminent l'électricité aux industries lourdes, aux grands consommateurs électriques comme les transports ferroviaires et font le lien avec le second réseau. Leur tension est de 63 kV ou 90 kV.

- **Client HTA (Haute Tension A)** : C'est le type client bénéficiant des lignes HTA qui permettent le transport de l'électricité à l'échelle locale vers les petites industries, les PME et les commerces. Ces lignes ont une tension comprise entre 15kV et 30kV [35].

- **Client BT(Basse Tension)** : Ce genre de client utilise les plus petites lignes du réseau qui sont les lignes BT. Leur tension est de 230V ou 400V. Ce sont celles qui nous servent tous les jours pour alimenter nos appareils ménagers. Elles permettent donc la distribution d'énergie électrique vers les ménages et les artisans [35].

2.5 Présentation de l'organisme d'accueil (Sonelgaz Distribution Direction de BEJAIA) :



FIGURE 2.1 – Logo de la Sonelgaz Direction Distribution de BÉJAIA [35].

La direction de distribution de Bejaia est rattachée à la Société Algérienne de Distribution de l'Electricité et du GAZ de l'Est (SDE), dont le siège se trouve à Constantine, cette dernière est composée d'une Direction à laquelle sont reliés directement [35] :

- Le secrétariat.
- Les assistants du Directeur de Distribution.
- Le chargé des affaires juridiques.
- Le chargé de la communication.
- Le chargé de sécurité.
- 9 Divisions et 10 Agences commerciales (BEJAIA CITE TOBAL, BEJAIA 4 CHEMINS, EL KSEUR, AMIZOUR, SIDI AICH, SEDDOUK, AKBOU, TAZMALT ; AOUKAS, KHERRATA) [35].
- 5 Districts (BEJAIA, AKBOU, SIDI AICH, AMIZOUR, KHERRATA)

2.5.1 Organisation de la Sonelgaz Distribution Direction de BÉJAIA

La structure organisationnel des différentes divisions de la Sonelgaz Direction de BÉJAIA est présentée dans la figure

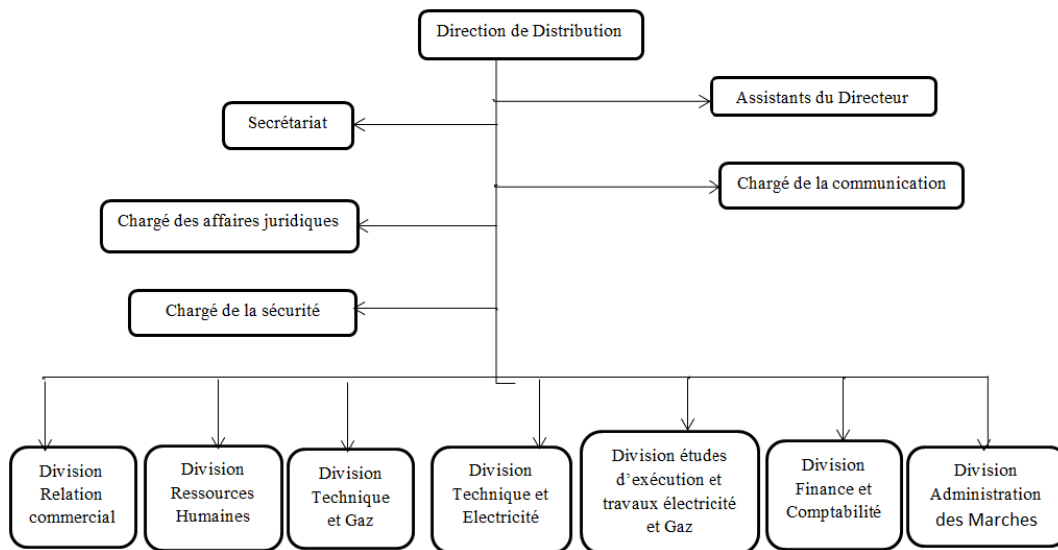


FIGURE 2.2 – Organigramme de la Sonelgaz Direction Distribution de BÉJAIA [35].

2.5.1.1 Divisions relations commerciales :

Cette division joue un rôle primordial au sein de la Direction de Distribution, ses missions principales sont :

- La vente de l'électricité et du gaz.
- La gestion des relations clients.
- Facturation des FSM (facturation sur mémoire).
- Analyse statistique (taux de pertes).
- Assurer la qualité de service et la satisfaction client.

Elle est scindée en 3 services [35] :

1. Service développement des ventes (RCN) :

- Prise en charge de toutes les demandes de raccordements des clients en électricité .
- Orienter les conseiller les clients MT/MP en matière de mode de raccordement, la détermination du niveau de tension / pression.
- Facturation des devis RCN et de toutes les prestations.
- Etablir les ordres d'exécution travaux après paiement des devis par les clients.

2. Service recouvrement :

- Assurer la gestion des comptes clients IIT/IIP, MT/MP , BT/BP.
- Respecterles délais de facturation des groupes en HT/IIP, MT/MP, BT/BP.
- Avertissement des clients lors des retards de paiements de factures (ordres d coupures, mise en demeure...)

- Contrôler et archiver toutes les pièces comptables.
- Assurer le suivi des clients grands comptes.
- Suivi de la facturation et du revouvrement FSM (facturation sur mémoire).

3. Service grands comptes :

- Assurer la gestion, la télé relève et la facturation mensuelle des clients MT/MP, HT/HP.
- Etablir le contrat de 5 ans à signer avant la mise en service d'un poste client (livraison) MT/MP.
- Veiller à la qualité de service.

Nous avons effectué notre stage au sein du département de la Gestion des Systèmes Informatiques (DGSI) qui est chargée de la gestion du centre informatique au niveau de Direction de la Distribution de Béjaia (DD).

2.6 Problématique

Pendant notre stage, nous avons identifié plusieurs problèmes majeurs rencontrés par l'entreprise. Tout d'abord, lorsqu'elle ne produit pas suffisamment d'énergie, elle ne répond pas aux besoins réels des clients. En revanche, si elle produit trop d'électricité, elle se retrouve confrontée à un nouveau problème : la perte d'énergie, car l'électricité ne peut pas être stockée. Ensuite, l'entreprise doit faire face à des difficultés de détection de fraudes. Lorsqu'un système de comptage d'énergie électrique est défaillant chez un client, il devient difficile d'évaluer la quantité d'énergie consommée et donc de facturer correctement le client. Enfin, l'entreprise utilise des méthodes d'analyse de ventes classiques et ne dispose pas de systèmes d'aide à la décision sophistiqués.

En intégrant l'intelligence artificielle, ces entreprises pourraient améliorer leur performance dans ces domaines clés. Les systèmes d'analyse prédictive pourraient aider à prédire la demande en électricité et ainsi optimiser la production pour répondre aux besoins des clients. La détection de fraudes pourrait également être améliorée grâce à des systèmes de surveillance sophistiqués basés sur l'analyse de données. Enfin, l'utilisation de l'intelligence artificielle pour l'analyse des ventes permettrait d'obtenir des informations précises et en temps réel sur les habitudes de consommation des clients, ce qui pourrait aider à prendre des décisions plus éclairées.

En résumé, l'objectif serait d'étudier les avantages potentiels de l'intelligence artificielle pour améliorer la performance des entreprises du secteur électrique dans ces domaines clés.

2.7 Solutions proposées

Lorsqu'une entreprise est confrontée à des difficultés de prise de décision en matière de ventes, il est important pour elle de mettre en place des méthodes d'analyse et de prévision de ses ventes. Ces dernières années, l'émergence de l'intelligence artificielle, notamment le machine learning et

le deep learning, a incité de nombreuses entreprises à intégrer ces nouvelles technologies pour améliorer leur gestion et relever les défis pratiques.

Il est évident que l'utilisation de l'intelligence artificielle, notamment le machine learning et le deep learning, est essentielle pour développer des méthodes d'analyse et de prévision des ventes et ainsi résoudre les problèmes de prise de décision liés à ces dernières. Le développement d'une application informatique de système intelligent pour l'analyse et la prédiction des ventes énergétiques des clients IITA répondra aux besoins de la Direction de BEJAIA de Sonelgaz Distribution en matière de gestion des ventes et permettra une meilleure compréhension des habitudes de consommation des clients.

Dans cette optique, la Direction de BEJAIA de Sonelgaz Distribution souhaite mettre en place un système intelligent qui permettra d'analyser les ventes des clients HTA afin de mieux gérer le problème de décision. L'objectif de notre travail est de répondre à cette demande en développant une application informatique qui permettra l'analyse et la prédiction des ventes énergétiques des clients IITA.

Le but principal de ce mémoire est de fournir à l'entreprise un système qui lui permettra de prévoir la consommation future des clients HTA afin de minimiser la perte d'énergie et de détecter la fraude. Ce système permettra également aux analystes de Sonelgaz de mieux visualiser les ventes des clients afin de mieux connaître leur clientèle, de comprendre l'évolution de la consommation électrique et de déduire la période d'accélération des ventes. Les principales fonctionnalités de notre application seront les suivantes :

1. Authentification et inscription des utilisateurs.
2. Localisation des clients sur une carte géographique.
3. Visualisation des ventes électriques des différents clients.
4. Prédiction des consommations électriques à l'aide des modèles statistiques utilisés dans le machine learning.
5. Prédiction des consommations électriques à l'aide d'un modèle statistique utilisé dans le deep learning.
6. Analyse comparative des clients en fonction de leur consommation et de la prédiction de leur consommation.

En résumé, notre application fournira à Sonelgaz une solution efficace pour résoudre les problèmes de prise de décision liés aux ventes. En utilisant les dernières technologies en matière d'intelligence artificielle, elle permettra une meilleure gestion des ventes et une compréhension plus fine des habitudes de consommation des clients.

2.8 Délimitation du domaine d'étude

2.8.1 Répartition des rôles

- **Scrum master : Mr. Foudil MIR** il agit en tant que facilitateur entre le responsable du produit et l'équipe principalement il assure la bonne communication entre les membres de l'équipe, lève tous les obstacles qui peuvent empêcher l'équipe d'atteindre les objectifs fixés pour chaque sprint de travail, et cherche à améliorer la productivité et le savoir-faire de son équipe.

- **Product owner : Mr. Dalil HADJOUT** il est le chef du produit, il définit les spécifications fonctionnelles et établit la liste des priorités de ce qu'il faut développer. C'est celui qui valide les fonctionnalités.

- **L'équipe** : Est chargée de transformer les besoins du client en produit ou service. Leurs rôles consistent à s'occuper du code : l'écrire, bien le comprendre, le modifier, gérer son existence, sa sauvegarde et ses versions et bien sur sa transformation en code exécutable. Elle est Composée de : Mlle Lydia NAIT SLIMANE et Mlle Zahia LOUALIA .

2.8.2 Identification des acteurs

Nous allons maintenant citer les différents acteurs susceptibles d'interagir avec le système, mais d'abord nous donnons une définition de ce que c'est un acteur :

Un acteur est une entité qui définit le rôle joué par un utilisateur ou par un système qui interagit avec le système modélisé.

Nous allons ensuite citer l'acteur principal de notre système :

-**Utilisateur** : ce sont des personnes déjà enregistrées dans la base de données qui se connectent au site.

2.8.3 Diagramme de contexte

Le diagramme de contexte est une modélisation permettant de mettre en avant les acteurs à haut niveau du système.

Les acteurs qui interagissent avec le système avec des différents messages dans notre application sont :

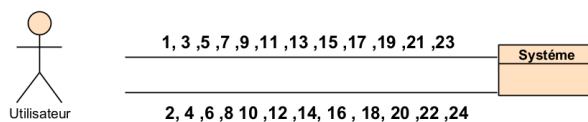


FIGURE 2.3 – Diagramme de contexte

La description des messages est donnée dans le tableau :

| Acteur | Utilisateur Numéro | Message | Numéro | Message |
|--------|--------------------|--|--------|--|
| | 1 | Demande d'authentification | 2 | afficher l'interface d'accueil |
| | 3 | Demande d'analyse statistiques | 4 | Afficher de tableau de bord |
| | 5 | Demande d'afficher la localisation des clients | 6 | Affichage de la carte avec la localisation des clients |
| | 7 | Demande de consultation de ventes | 8 | Affichage d'un choix de format |
| | 9 | Effectuer un choix | 10 | Affichage de la page de consommation choisie |
| | 11 | Effectuer une recherche | 12 | Affichage du résultat de la recherche |
| | 13 | Demande de prédiction | 14 | Afficher un choix |
| | 15 | Effectuer un choix de prédiction | 16 | Affichage du formulaire |
| | 13 | Demande de comparaison | 14 | Afficher un choix |
| | 15 | Effectuer un choix de comparaison | 16 | Affichage du formulaire |
| | 17 | Remplir et valider les champs du formulaire | 18 | Affichage du résultat de prédiction |
| | 19 | Demande de gestion des utilisateurs | 20 | Affichage d'un choix |
| | 21 | Demande d'ajout de profils utilisateurs | 22 | Affichage de l'interface d'ajout |
| | 23 | Demande de consulter la liste des utilisateurs | 24 | Affichage de la liste des utilisateurs |

TABLEAU 2.1 – Messages échangés entre le système et les acteurs

2.9 La méthodologie SCRUM

2.9.1 Le Product backlog

Il s'agit d'une liste hiérarchisée des exigences initiales du client concernant le produit à réaliser. Il est exprimé sous forme d'un ensemble d'user story. Ce document évolue sans cesse durant le projet, en fonction des besoins du client. En d'autre terme c'est une sorte de carnet de commande

pour le produit sous la responsabilité du Product owner. Dans notre cas le client a exprimé ses besoins de la manière suivante :

| ID du User Story | Nom du User Story | Description du User Story | Priorité | Statut |
|------------------|-----------------------------|--|----------|--------|
| US01 | Authentification | <ul style="list-style-type: none"> - En tant qu'administrateur, je dois m'authentifier avec mon nom d'utilisateur et mon mot de passe pour accéder à l'application. - En tant que client, je souhaite m'inscrire, me connecter et me déconnecter en toute sécurité . | Haute | Fait |
| US02 | Gestion de la localisation | - En tant qu'administrateur, je veux pouvoir visualiser la localisation de mes clients du secteur économique sur une carte pour mieux comprendre leur emplacement géographique et leur répartition. | Faible | Fait |
| US03 | Gestion des comptes clients | <ul style="list-style-type: none"> - En tant qu'administrateur, je veux pouvoir gérer les comptes client. - En tant qu'administrateur je dois garantir la sécurité des données des clients HTA. - En tant qu'administrateur, je veux pouvoir supprimer des clients du secteur économique de l'application, afin de ne plus suivre leur consommation électrique. | Moyenne | Fait |
| US04 | Visualisation des données | - En tant qu'administrateur, je veux pouvoir visualiser les données de consommation électrique des clients du secteur économique pour une période donnée sous format graphique et tableau, afin d'analyser les tendances de consommation. | Haute | Fait |

| ID du User Story | Nom du User Story | Description du User Story | Priorité | Statut |
|------------------|--|--|----------|--------|
| US05 | Comparaison des consommations | - En tant qu'administrateur, je veux pouvoir comparer les consommation électrique des clients du secteur économique entre deux périodes de temps données . | Haute | Fait |
| US06 | Prédiction des ventes avec modèle Machine Learning | - En tant qu'administrateur, je veux pouvoir sauvegarder les données de consommation d'électricité des clients et les résultats de prédiction dans une base de données pour une utilisation ultérieure. - En tant qu'administrateur, je veux pouvoir utiliser un modèle de Machine Learning pour prédire la consommation future d'électricité pour chaque client en utilisant les données historiques de consommation. - En tant qu'administrateur, je souhaite pouvoir visualiser les prévisions de consommation électrique de mes clients du secteur économique sous forme de graphiques et de tableaux pour mieux les comprendre. | Haute | Fait |
| US07 | Gestion du compte client | En tant que client, je veux pouvoir gérer mon compte pour accéder à l'interface client et aux données de prédiction. | Faible | Fait |

TABLEAU 2.2 – Product Backlog

2.9.2 Planification des sprints

Les «User stories» précédents définis dans le «Product Backlog» sont triés par ordre de priorité. Le travail sera planifié selon des sprints que nous avons définis ,

Un Sprint est une itération. Il s'agit d'une période de 2 à 4 semaines maximum pendant laquelle une version terminée et utilisable du produit est réalisée. Un nouveau sprint commence dès la fin du précédent. Chaque sprint a un objectif et une liste de fonctionnalités à réaliser.

| Sprint | Nom du sprint | Période |
|---------------|---|----------------|
| Sprint 1 | Authentification | 2 semaines |
| Sprint 2 | Gestion de la localisations | 1 semaines |
| Sprint 3 | Visualisation des Consommations | 1 semaine |
| Sprint 4 | Prédiction des Ventes avec un modèle machine learning | 2 semaines |
| Sprint 5 | Comparaison des Consommations | 1 semaines |
| Sprint 6 | Prédiction des Ventes avec un modèle deep learning | 2 semaines |

TABLEAU 2.3 – Planification des Sprints

2.10 Diagramme de cas d'utilisation

Un cas d'utilisation est une unité cohérente représentant une fonctionnalité visible de l'extérieur. 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 [32][6].

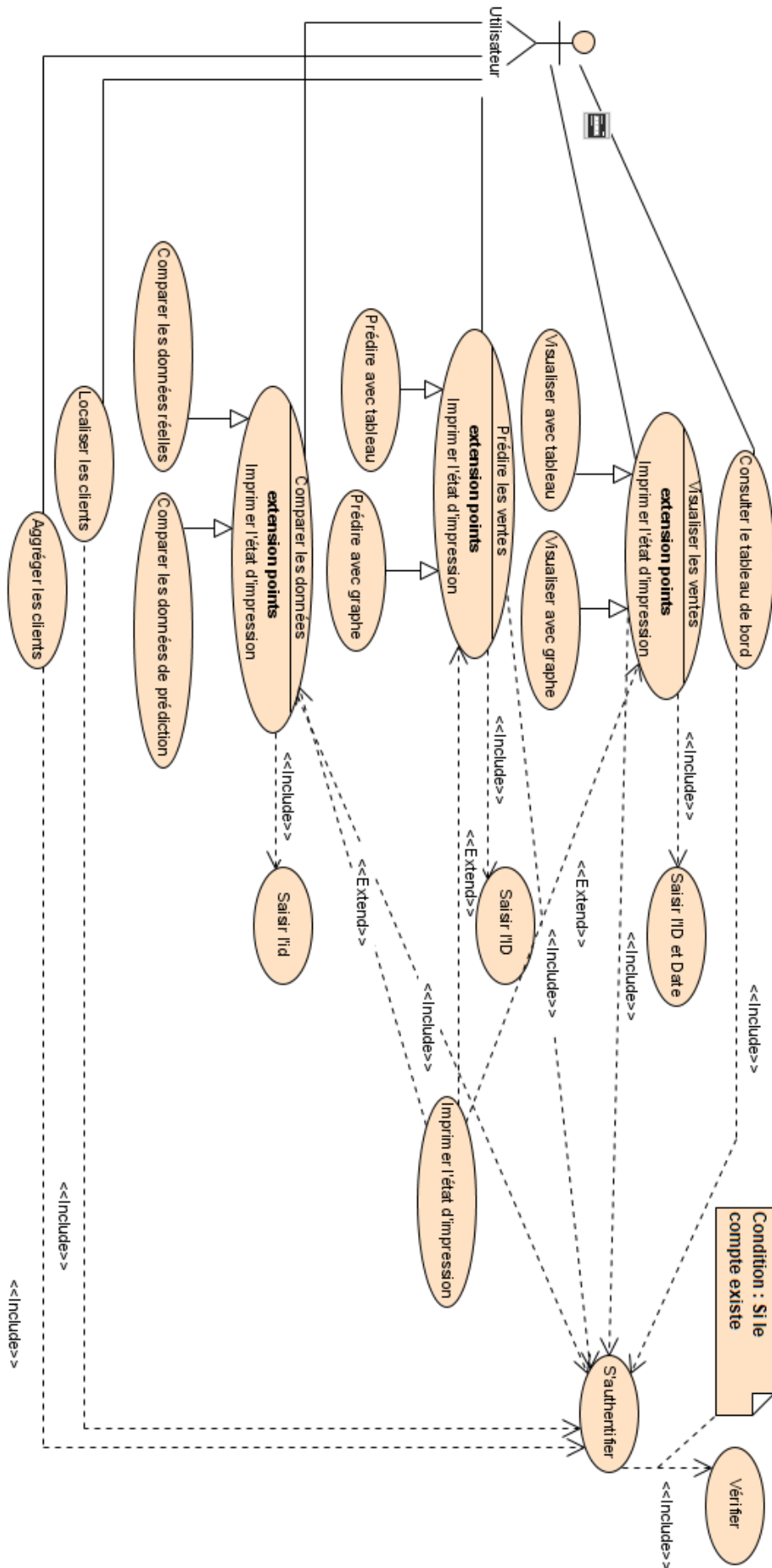


FIGURE 2.4 – Diagramme de cas d'utilisation

2.11 Description des cas d'utilisations

Une description textuelle des cas d'utilisation est un tableau qui décrit les différentes étapes ou interactions entre les différents acteurs impliqués dans un scénario d'utilisation d'un système, d'une application ou d'un service.

Ces étapes sont détaillées et organisées dans un ordre logique pour décrire comment un utilisateur peut utiliser le système pour accomplir une tâche [20].

Chaque cas d'utilisation doit inclure les éléments suivants :

| Nom du cas d'utilisation | |
|---------------------------|---|
| Acteurs | Les acteurs qui vont réaliser le cas d'utilisation |
| Description | Une description résumé de cas d'utilisation |
| Auteurs | Les intervenants de l'élaboration de la fiche |
| Date | La date de rédaction ou mise à jour de la fiche |
| Pré-conditions | Les conditions nécessaires pour déclencher le cas d'utilisation |
| Description des scénarios | |
| Scénario nominal | La description des interactions entre acteur et système, dans le meilleur des cas ou plus précisément le meilleur des scénarios |
| Scénario alternatif | La description des interactions entre acteur et système, dans le meilleur des cas ou plus précisément le meilleur des scénarios |
| Post-conditions | État du système après réalisation du cas d'utilisation et qui pourra témoigner du bon fonctionnement |

TABLEAU 2.4 – Formalisme de description des cas d'utilisation

Diagrammes de séquences

Les diagrammes de séquences montrent les différentes interactions échangées entre l'utilisateur et le système. Dans cette étude, les cas d'utilisation qui nous paraissent utiles de les représenter par leurs diagrammes de séquences sont : Authentification, Visualisation des consommations, Prédiction des ventes, Comparaison des consommations.

2.11.1 Cas d'utilisation « S'authentifier »

La description textuelle de ce cas est :

| S'authentifier | |
|---------------------------|--|
| Acteurs | Utilisateur |
| Description | Permet à l'Utilisateur de s'authentifier en utilisant son nom d'utilisateur et son mot de passe. |
| Auteurs | NAIT SLIMANE Lydia, LOUALIA Zahia |
| Date | 24 Avril 2023 |
| Pré-conditions | L'Utilisateur doit avoir un compte d'utilisateur avec un nom d'utilisateur et un mot de passe valides. |
| Description des scénarios | |
| Scénario nominal | <ol style="list-style-type: none"> 1. L'Utilisateur ouvre l'application et est redirigé vers l'écran d'authentification. 2. L'Utilisateur entre son nom d'utilisateur et son mot de passe. 3. Le système vérifie que le nom d'utilisateur et le mot de passe correspondent à un compte utilisateur valide. 4. Le système redirige l'utilisateur vers l'écran principal de l'application. |
| Scénario alternatif | <ol style="list-style-type: none"> 1. L'Utilisateur ouvre l'application et est redirigé vers l'écran d'authentification. 2. L'Utilisateur entre un nom d'utilisateur ou un mot de passe invalide. 3. Le système affiche un message d'erreur correspondant et invite l'agent de vente à saisir de nouveau ses informations d'identification. 4. L'Utilisateur corrige les erreurs de saisie et valide à nouveau ses informations d'identification. 5. Le système vérifie à nouveau les informations d'identification et redirige l'agent de vente vers l'écran principal de l'application si elles sont valides. |
| Post-conditions | L'Utilisateur est authentifié et peut accéder aux fonctionnalités de l'application. |

TABLEAU 2.5 – Description du cas d'utilisation «S'authentifier»

Le diagramme de la figure 2.4 représente le scénario lorsque l'agent de vente essaye de se connecter :

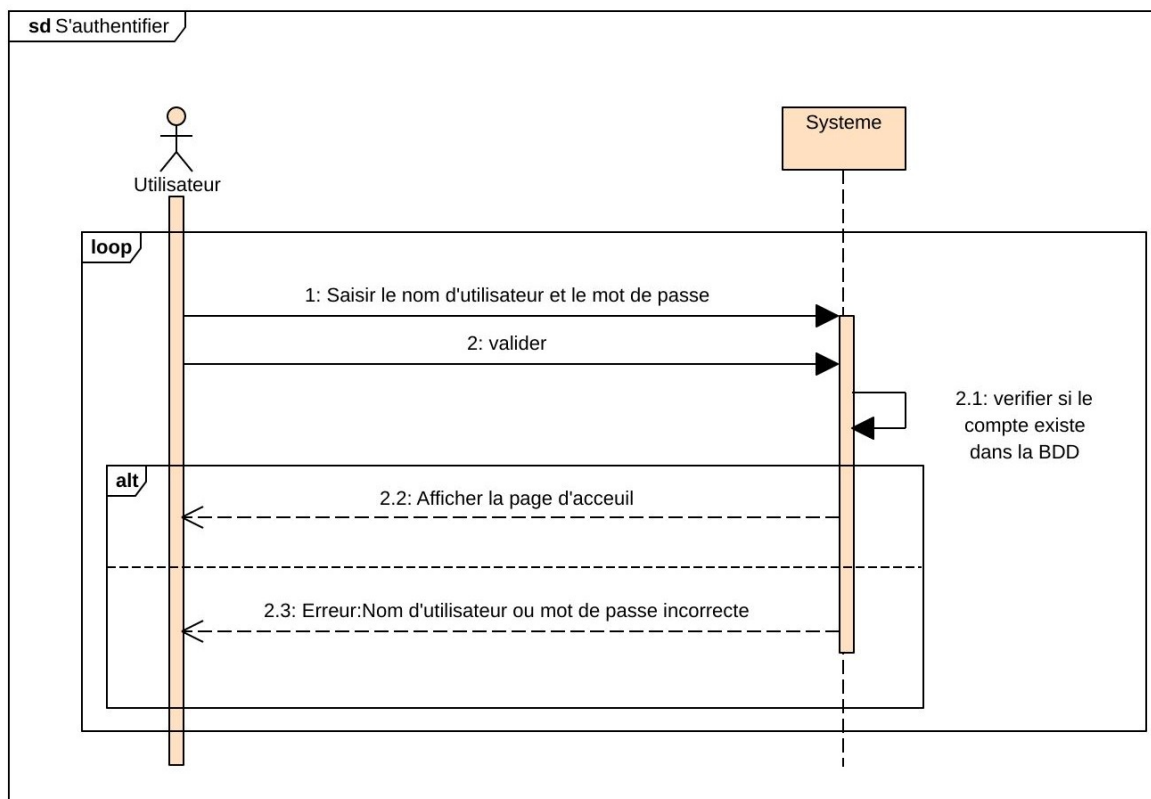


FIGURE 2.5 – Diagramme de séquence système du cas d'utilisation «S'authentifier»

2.11.2 Cas d'utilisation « Visualiser les consommations »

La description textuelle de ce cas est :

| Visualiser les consommations | |
|------------------------------|--|
| Acteurs | Utilisateur |
| Description | Permet à l'Utilisateur de visualiser les consommations des clients avec le format graphique et tableau. Le format graphique affiche les données sous forme de graphique, tandis que le format tableau affiche les données sous forme de tableau. L'Utilisateur peut choisir le format qui convient le mieux à ses besoins. |
| Auteurs | NAIT SLIMANE Lydia, LOUALIA Zahia |
| Date | 2 Mai 2023 |
| Pré-conditions | L'Utilisateur doit être authentifié. |
| Description des scénarios | |
| Scénario nominal | <ol style="list-style-type: none"> 1. L'Utilisateur sélectionne l'option "Visualiser les consommations" dans le menu. 2. Le système affiche un choix de type de visualisation. 3. L'Utilisateur effectue un choix " . 4. L'Utilisateur remplit et valide un formulaire . 5. Le système affiche les Résultats. |
| Scénario alternatif | <ol style="list-style-type: none"> 1. L'Utilisateur sélectionne l'option "Visualiser les consommations" dans le menu. 2. Le système affiche un choix de type de visualisation. 3. L'Utilisateur effectue un choix de type de visualisation. 4. L'Utilisateur remplit et valide un formulaire. 5. Le système essaie d'accéder aux données de consommation correspondantes. 6. Le système détecte une erreur d'accès aux données, soit en raison d'une panne de serveur ou d'une erreur de connexion de base de données. 7. Le système affiche un message d'erreur informant l'Utilisateur qu'il y a un problème technique et qu'il doit réessayer plus tard ou contacter le support technique. |
| Post-conditions | Le résultat s'affiche |

TABLEAU 2.6 – Description du cas d'utilisation «Visualiser les consommations»

Le diagramme de la figure 2.5 représente le scénario lorsque l'utilisateur essaye de visualiser les consommations :

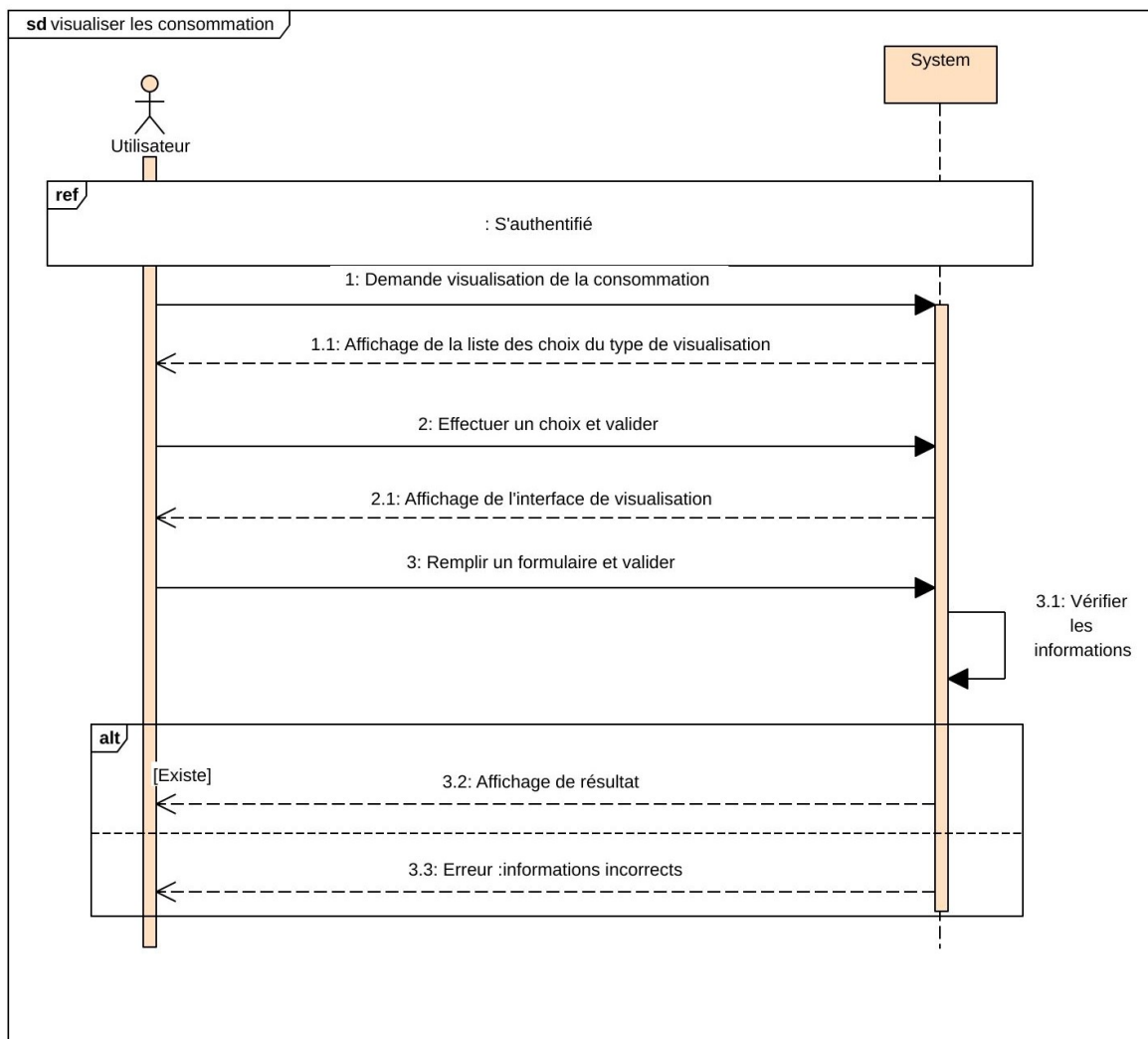


FIGURE 2.6 – Diagramme séquence système du cas d'utilisation «Visualiser les consommations »

2.11.3 Cas d'utilisation « Prédire des ventes»

La description textuelle de se cas est :

| Prédiction des ventes | |
|---------------------------|---|
| Acteurs | Agent de vente |
| Description | Permet à l'Utilisateur de visualiser les prévisions de vente pour leur consommations . Le système affiche les prévisions de vente sous forme graphique et tableau |
| Auteurs | NAIT SLIMANE Lydia, LOUALIA Zahia |
| Date | 15 Mail 2023 |
| Pré-conditions | L'Utilisateur doit être authentifié et avoir accès aux données de vente. |
| Description des scénarios | |
| Scénario nominal | <ol style="list-style-type: none"> 1. L'Utilisateur sélectionne l'option "Prédiction des ventes" dans le menu. 2. Le système affiche un choix de prediction. 3. L'Utilisateur effectue un choix . 4. Le système lui affiche l'intefrace de prédiction . 5. L'Utilisateur remplir un formulaire et valider 6. Le système affiche le résultat |
| Scénario alternatif | <ol style="list-style-type: none"> 1. L'Utilisateur sélectionne l'option "Prédiction des ventes" dans le menu. 2. Le système affiche un choix de prévisions. 3. L'agent de vente effectue un choix. 4. Le système affiche l'interface de prédiction. 5. L'Utilisateur remplit le formulaire avec des valeurs erronées ou manquantes. 6. Le système détecte l'erreur de saisie et affiche un message d'erreur correspondant. 7. L'Utilisateur corrige les erreurs de saisie et valide à nouveau le formulaire. 8. Le système génère les prévisions de vente en utilisant les données saisies et les affiche à l'Utilisateur. |
| Post-conditions | Les prévisions de vente s'affichent. |

TABLEAU 2.7 – Description du cas d'utilisation «Prédire les ventes»

Le diagramme de la figure 2.6 représente le scénario lorsque l'agent de vente essaye de prédire les ventes :

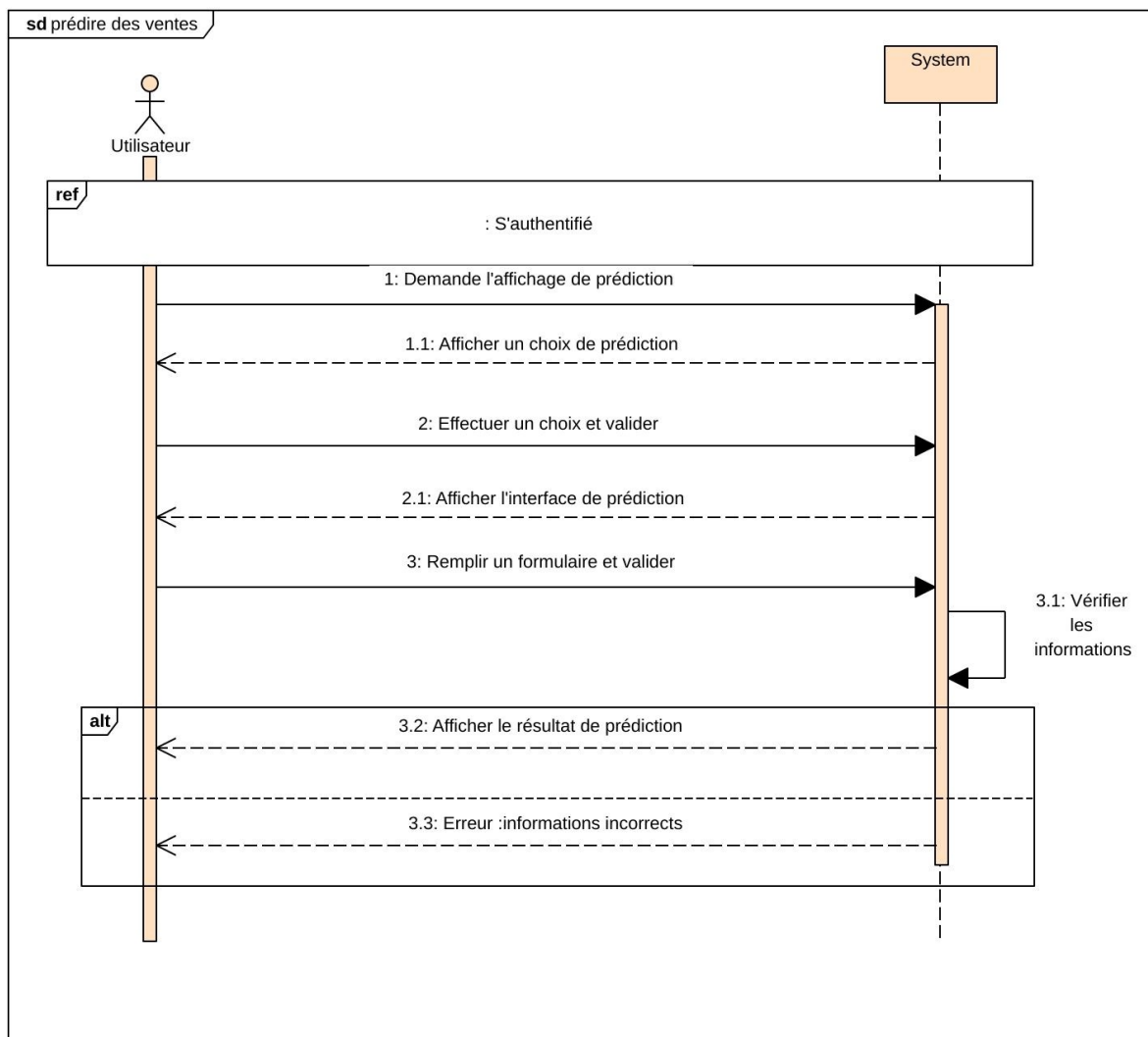


FIGURE 2.7 – Diagramme séquence système du cas d’utilisation «Prédire les ventes»

2.11.4 Cas d’utilisation « Comparraison des ventes »

La description textuelle de se cas est :

| Comparaison des ventes | |
|---------------------------|---|
| Acteurs | Utilisateur |
| Description | Permet à l'utilisateur de comparer les ventes entre deux périodes de temps données. Le système affiche les données sous forme de graphique et de tableau. |
| Auteurs | NAIT SLIMANE Lydia LOUALIA Zahia |
| Date | 25 Mai 2023 |
| Pré-conditions | L'utilisateur doit être authentifié. |
| Description des scénarios | |
| Scénario nominal | <ol style="list-style-type: none"> 1. L'Utilisateur s'électionne l'option "Comparaison des ventes" dans le menu. 2. Le système affiche un formulaire de sélection de périodes de temps. 3. L'Utilisateur sélectionne deux périodes de temps qu'il souhaite comparer. 4. L'utilisateur valide ses s'élections en cliquant sur le bouton "Valider". 5. Le système affiche les données de vente pour les deux périodes de temps sélectionnées. |
| Scénario alternatif | <ol style="list-style-type: none"> 1. L'Utilisateur sélectionne l'option "Comparaison des ventes" dans le menu. 2. Le système affiche un formulaire de sélection de périodes de temps. 3. L'Utilisateur sélectionne deux périodes de temps qu'il souhaite comparer. 4. L'Utilisateur valide ses sélections en cliquant sur le bouton "Valider". 5. Le système essaie d'accéder aux données de vente correspondantes. 6. Le système détecte une erreur d'accès aux données, soit en raison d'une panne de serveur ou d'une erreur de connexion de base de données. 7. Le système affiche un message d'erreur informant l'Utilisateur qu'il y a un problème technique et qu'il doit réessayer plus tard ou contacter le support technique. |
| Post-conditions | Les données de vente pour les deux périodes de temps sélectionnées s'affichent ' |

TABLEAU 2.8 – Description du cas d'utilisation «comparaison des ventes»

Le diagramme de la figure 2.7 représente le scénario lorsque l'utilisateur essaye de comparer les ventes :

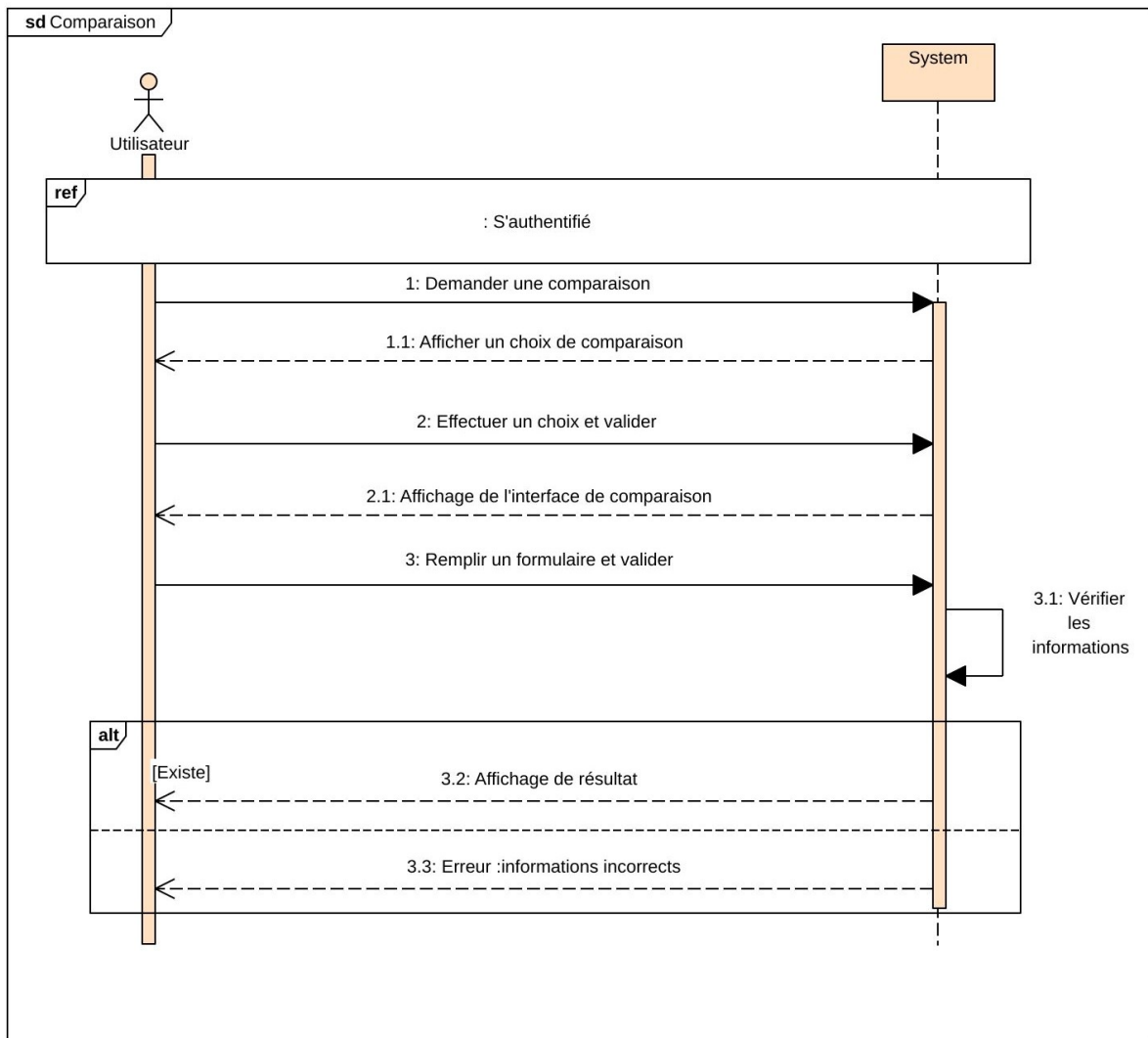


FIGURE 2.8 – Diagramme séquence système du cas d'utilisation «Comparaison des ventes»

2.12 Diagramme de séquence détaillé

Les diagrammes de séquence détaillés permettent de visualiser clairement l'ordre d'exécution des actions, les conditions et les boucles, les appels de méthode et les retours, ainsi que les interactions entre les différents objets du système. Ils sont utiles pour comprendre en détail le fonctionnement d'un scénario ou d'un processus, et peuvent être utilisés pour la conception, le débogage et la documentation des systèmes logiciels. [20]

Le tableau 2.7 représente les trois types de classe utilisés dans le diagramme de séquence détaillé.

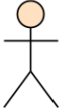
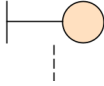


| Rôle | Représentation | Signification |
|-----------|---|---|
| Acteur |  | une entité qui définit le rôle joué par un utilisateur ou par un système qui interagit avec le système modélisé |
| Interface |  | servent à modéliser les interactions entre le système et ses acteurs |
| Contrôle |  | utilisées pour représenter la coordination, l'enchaînement |
| Entité |  | servent à modéliser des informations durables persistantes |

TABLEAU 2.9 – Description UML du diagramme de séquence détaillé

2.12.1 Diagramme de séquence détaillé du cas d'utilisation "S'authentifier" :

L'acteur "Utilisateur" démarre le processus d'authentification en entrant ses identifiants dans l'interface utilisateur. Le composant "Système d'authentification" reçoit le message et vérifie les informations d'identification fournies. Il effectue une vérification en interrogeant une base de données. Le composant "Système d'authentification" envoie un message de retour à l'interface utilisateur indiquant si l'authentification a réussi ou échoué. L'interface utilisateur reçoit le message de retour et affiche le résultat à l'Utilisateur. Si l'authentification a échoué, l'interface utilisateur peut afficher un message d'erreur et permettre à l'Utilisateur de réessayer. Si l'authentification a réussi, l'interface utilisateur peut rediriger l'Utilisateur vers la page d'accueil. Le processus d'authentification se termine.

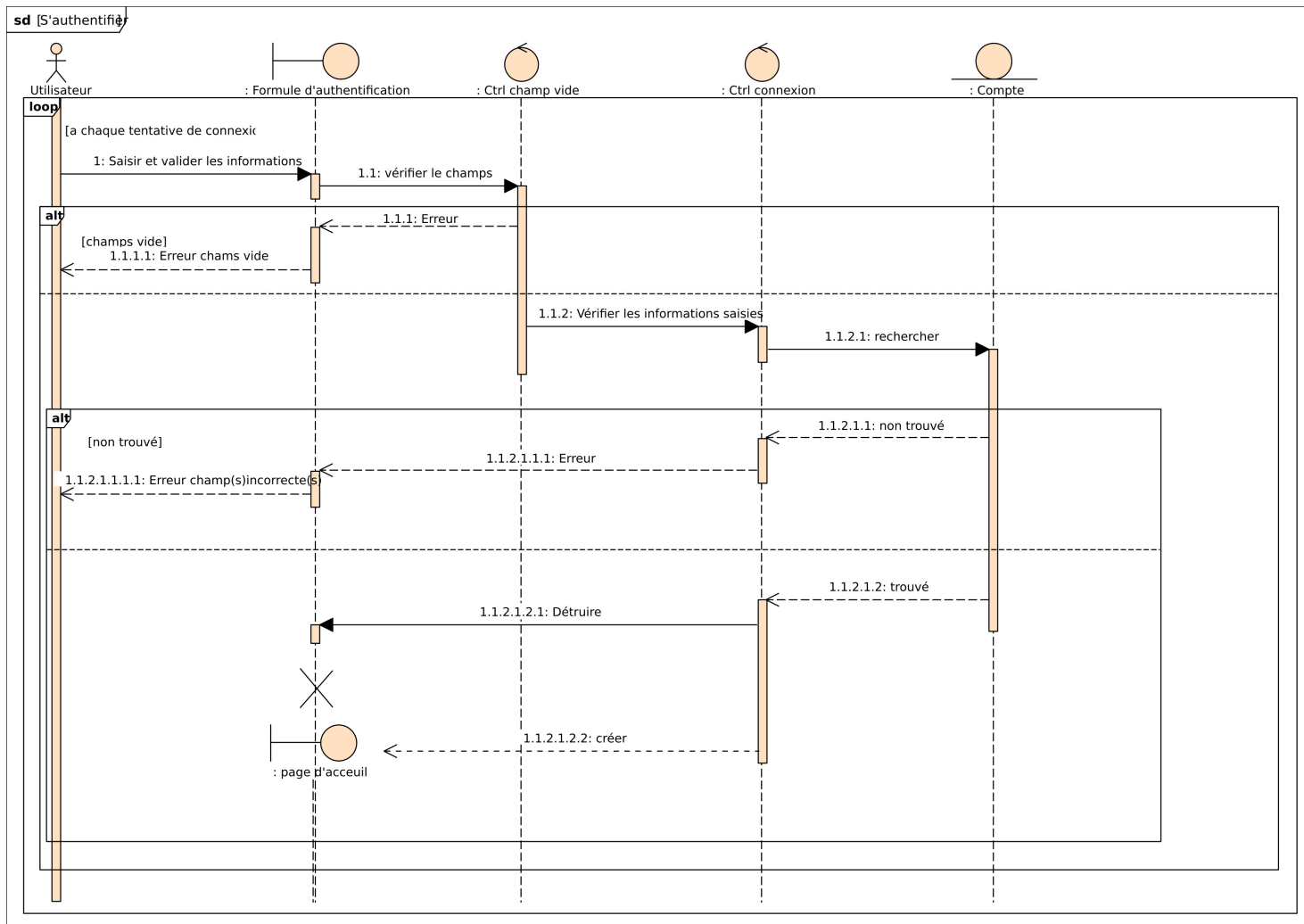


FIGURE 2.9 – Diagramme de séquence détaillé du cas d'utilisation «S'authentifier»

2.12.2 Diagramme de séquence détaillé du cas d'utilisation «Visualiser les consommations»

Le diagramme de séquence détaillé du cas d'utilisation «Visualiser les consommations» représente la séquence d'actions et d'échanges de messages entre les acteurs et les composants du système lorsqu'un utilisateur demande à visualiser les informations de consommation. Le diagramme met en évidence les étapes clés, telles que la demande de visualisation, la récupération des informations de consommation, l'affichage des données à l'utilisateur et la possibilité d'accéder aux détails spécifiques d'une consommation [20].

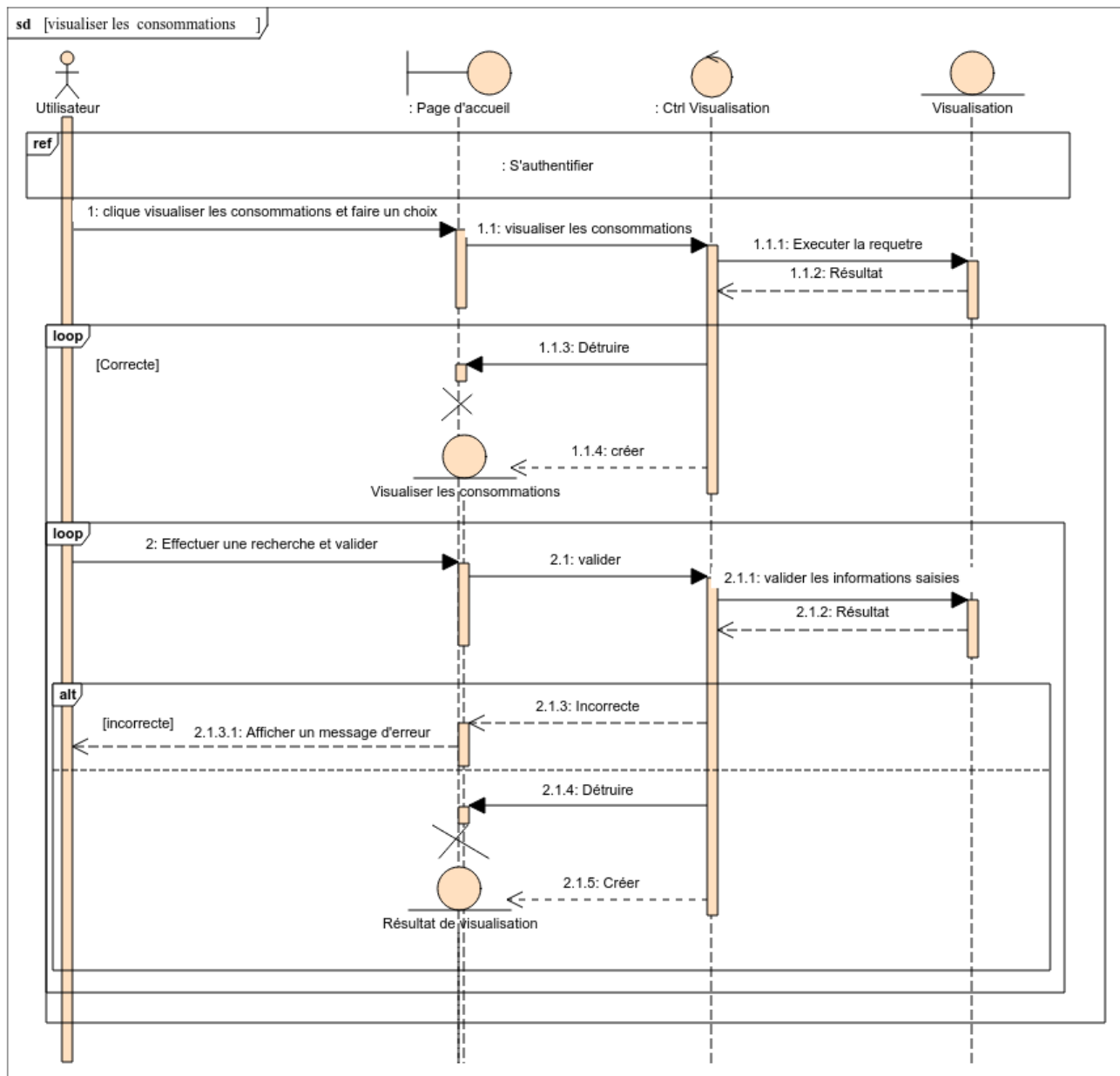


FIGURE 2.10 – Diagramme séquence détaillé du cas d'utilisation «Visualiser les consommations»

2.12.3 Diagramme séquence détaillé du cas d'utilisation «Prédire des ventes»

Le diagramme de séquence détaillé du cas d'utilisation «Prédire les ventes» représente la séquence d'actions et d'échanges de messages entre les acteurs et les composants du système lorsqu'un utilisateur demande une prédiction des ventes. Le diagramme met en évidence les étapes clés, telles que la saisie des données de vente, la transmission des données au composant de prédiction, le traitement des données, la génération de la prédiction et la présentation des résultats à l'Utilisateur. [20] Il offre une représentation visuelle précise de la séquence chronologique des

interactions, permettant de comprendre comment le système utilise les données disponibles pour effectuer une prédiction précise des ventes.

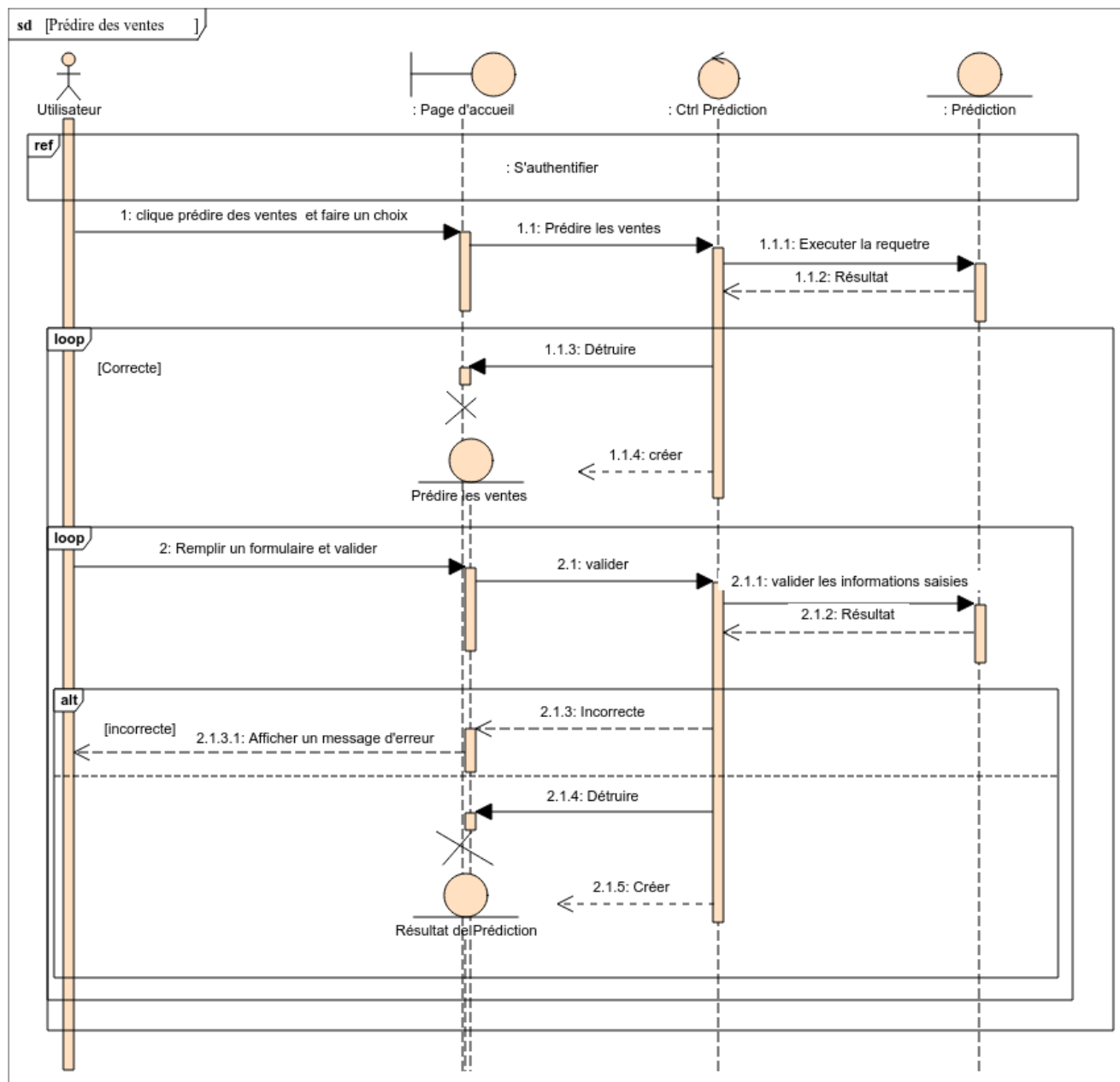


FIGURE 2.11 – Diagramme de séquence détaillé du cas d'utilisation «Prédire des ventes»

2.12.4 Diagramme séquence détaillé du cas d'utilisation «Comparaison des ventes»

Le diagramme de séquence détaillé du cas d'utilisation «Comparaison des ventes» représente la séquence d'actions et d'échanges de messages entre les acteurs et les composants du système lorsqu'un utilisateur souhaite comparer les données de vente. Le diagramme met en évidence les

étapes clés, telles que la sélection des critères de comparaison, la récupération des données de vente correspondantes, le traitement des données, la génération des résultats de comparaison et la présentation des résultats à l'Utilisateur. Il offre une représentation visuelle précise de la séquence chronologique des interactions, permettant de comprendre comment le système répond aux requêtes de comparaison de l'agent de vente en fournissant des informations pertinentes et significatives sur les ventes, facilitant ainsi l'analyse comparative des données de vente. [20]

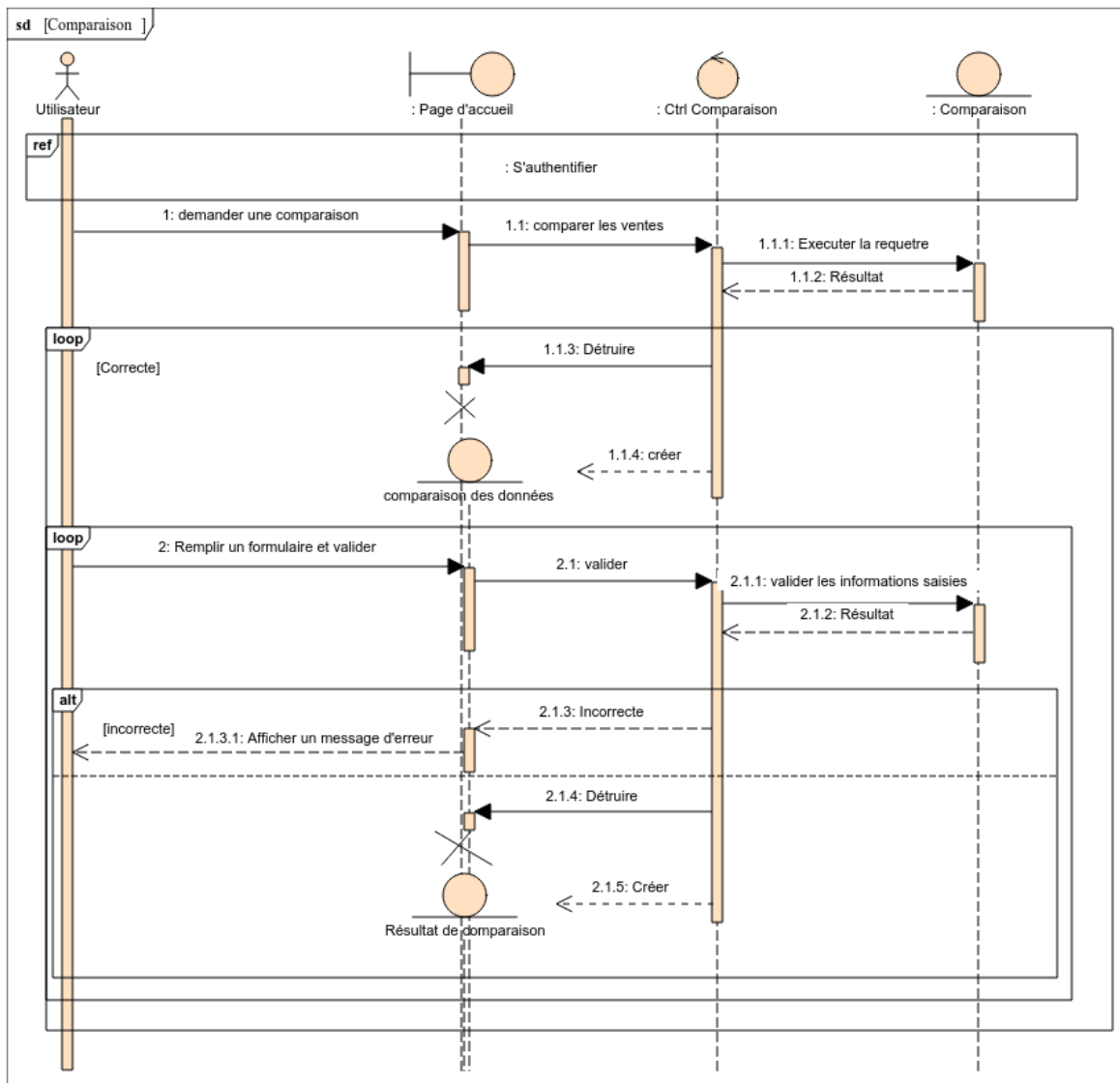


FIGURE 2.12 – Diagramme de séquence détaillé du cas d'utilisation «Comparaison des ventes»

2.13 Diagramme d'activité

Le diagramme d'activité est un type de diagramme utilisé dans la conception d'interfaces utilisateur pour représenter la structure et le flux de navigation entre les différentes pages d'une

application. Les diagrammes d'activités sont utilisés pour comprendre, planifier, améliorer et documenter la structure de navigation d'une application, contribuant ainsi à une expérience utilisateur cohérente [20].

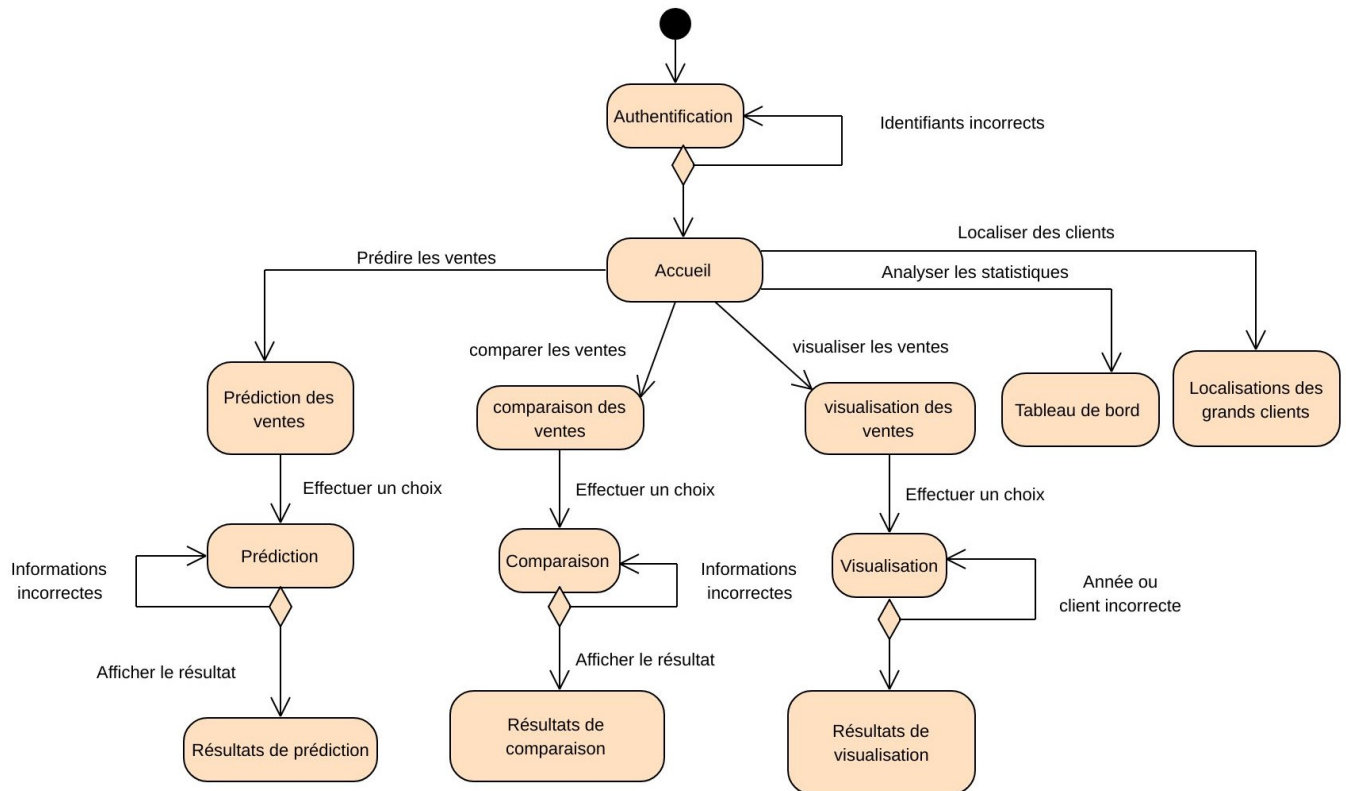


FIGURE 2.13 – Diagramme d'activité

2.14 Diagrammes de Classes

Le diagramme de classe est un type de diagramme utilisé en génie logiciel pour représenter les classes, les objets, les attributs et les relations entre eux dans un système logiciel. Le diagramme de classe permet de visualiser l'organisation et les interactions des objets dans un système, ainsi que les attributs et les méthodes qui leur sont associés. De plus, il sert de base pour la conception et la mise en oeuvre du code source, en aidant à la modélisation des classes, des relations et de la logique de programmation [20].

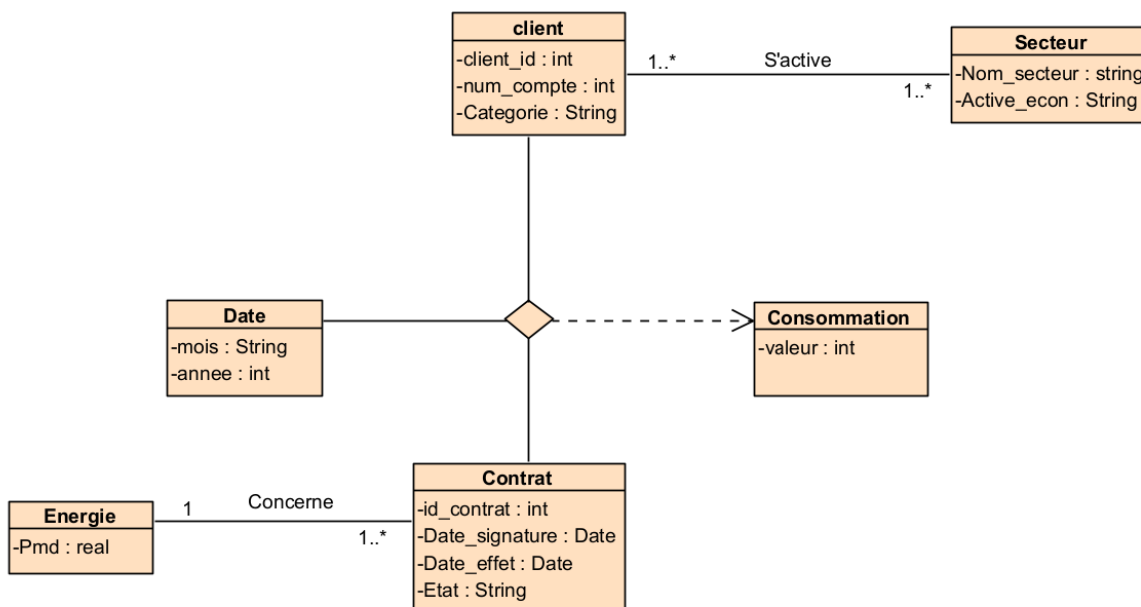


FIGURE 2.14 – Diagramme de Classe initial

Pour des raisons de confidentialités, ils nous ont communiqué uniquement les données et informations nécessaires dont nous aurons besoin durant notre stage. Donc nous nous sommes basées uniquement des informations qu'ils nous ont donné qui sont représentées dans le diagramme dans la figure suivante :

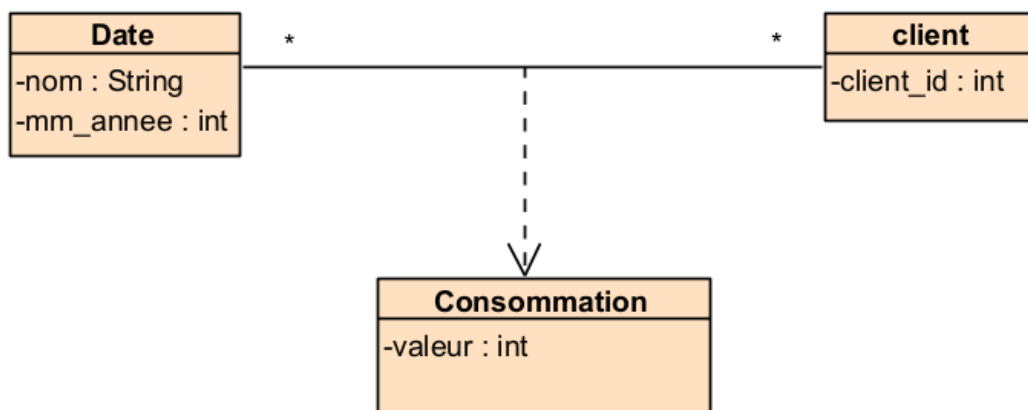


FIGURE 2.15 – Diagramme de Classe de l'application

2.15 Conclusion

Ce chapitre a joué un rôle essentiel dans l'identification des principales fonctionnalités de notre système et dans la mise en oeuvre de la méthode Scrum pour notre projet.

Il a également permis une présentation détaillée des différentes fonctionnalités de notre système à travers des diagrammes de séquence détaillés, ainsi que la conception de diagrammes de navigation et de classes.

Ce chapitre peut être considéré comme une préparation de l'étape de réalisation de notre système, qui sera présentée dans le chapitre suivant.

RÉALISATION

3.1 Introduction

Dans le dernier chapitre, nous présenterons tout le travail qui a été fait. Dans un premier temps, nous décrivons l'environnement matériel et logiciel dans lequel l'application a été développée, en nous concentrons sur les technologies et les langages de programmation utilisés. Ensuite, nous détaillerons les différentes étapes suivies pour modéliser une prévision à l'aide de techniques d'apprentissage automatique et d'apprentissage profond.

Enfin, nous terminerons le chapitre en donnant un aperçu du projet et en partageant quelques captures d'écran illustrant le fonctionnement de l'application. Ces captures d'écran vous permettront de visualiser concrètement l'interface et les différentes fonctions proposées par nos applications.

3.2 Environnement de développement de l'application

3.2.1 L'architecture 3-tier

Dans l'architecture à 3 niveaux (appelée architecture 3-tier), il existe un niveau intermédiaire, c'est-à-dire que l'on a généralement une architecture partagée entre :

Le client : le demandeur de ressources

Le serveur d'application (appelé aussi middleware) : le serveur chargé de fournir la ressource mais faisant appel à un autre serveur.

Le serveur secondaire : (généralement un serveur de base de données), fournissant un service au premier serveur

3.2.2 Les langages utilisés

Python : Python est le langage de programmation open source le plus utilisé par les informaticiens. En fait, parmi les nombreuses fonctionnalités de Python, les développeurs peuvent se concentrer sur ce qu'ils font, plutôt que sur la façon dont ils le font. Il libère les développeurs des

contraintes formelles qui tourmentaient leur temps avec les langages plus anciens. Par conséquent, développer du code en Python est plus rapide que dans d'autres langages [1].

CSS : CSS hacks, également connus sous le nom de "CSS tricks" ou "CSS workarounds", font référence à des techniques non standard ou des contournements utilisés pour résoudre des problèmes spécifiques liés à la mise en page et à l'apparence des éléments sur une page web en utilisant CSS (Cascading Style Sheets) [8].

3.2.3 Les frameworks

Streamlit : Streamlit est une bibliothèque Python open source créée en 2018. Cela signifie que tous les utilisateurs peuvent le réutiliser pour créer leur propre logiciel, ce qui permet une grande flexibilité pour répondre aux besoins de chacun. L'open source est devenu un véritable mouvement dans le développement logiciel, encourageant la production collaborative et améliorant ainsi la qualité des logiciels.

Streamlit vous permet de créer des applications en écrivant simplement du code Python. Par conséquent, cela devient une méthode pratique et très accessible pour quiconque connaît Python et souhaite créer des applications Web. C'est un outil simple et rapide pour intégrer la visualisation de données dans les applications.

L'intégration de modèles d'apprentissage automatique dans l'interface est également facile à réaliser. Grâce à l'application, n'importe qui peut prédire et tester ses modèles par rapport aux données avec des clients ou des collaborateurs. Le point fort de Streamlit est qu'il n'y a pas besoin de se soucier de l'interface, car son interface de programmation d'application (API) nécessite très peu de ligne de commande, donc son utilisation est simplifiée [10].

3.2.4 Les logiciels

- **Anaconda**

Type : Distribution de logiciels pour l'analyse de données.

Date de création : La première version d'Anaconda est sortie en 2012.

Développeur : Anaconda, Inc.

Environnement : Multiplateforme .

Langage de développement : Python.

Langues : Principalement anglais, mais la communauté propose également des traductions dans d'autres langues.

Licence : Anaconda est basé sur des logiciels open-source, et la plupart des packages inclus sont également open-source.

Cependant, certaines fonctionnalités et outils spécifiques peuvent être soumis à des licences propriétaires. Anaconda est populaire parmi les scientifiques des données et les développeurs Python en raison de sa facilité d'utilisation et de sa vaste collection de packages et d'outils préinstallés.

Il comprend également des fonctionnalités pour gérer des environnements virtuels et isolés, ce qui facilite la création et la gestion de configurations spécifiques pour différents projets ou applications [7].

- **Pycharm**

Pycharm est un logiciel

Type : EDI.

Date de création : 2010.

Développeur : JetBrains.

Environnement : multiplateforme.

Langage de développement : Java et Python.

Langues : Anglais.

Licence : Libre et propriétaire.

PyCharm est un EDI dédié à Python. Il permet l'analyse du code, les tests unitaires, l'utilisation d'un système de versionnage (Git, Subversion, Mercurial, Perforce ou CVS) et permet l'utilisation des frameworks Django, Streamlit, web2py et Flask. Il intègre un débogueur et une fonction de refactoring [9].

- **Visual paradigm**

Visual Paradigm est un logiciel de modélisation visuelle et de conception de logiciels.

Type : Logiciel de modélisation et de conception de logiciels.

Date de création : Visual Paradigm a été fondé en 1994.

Développeur : Visual Paradigm International Ltd.

Environnement : Multiplateforme .

Langage de développement : Java.

Langues : Disponible dans de nombreuses langues, dont l'anglais, le français, l'allemand, le chinois, etc [11].

- **Jupyter Nootbook**

Jupyter Notebook est une application web qui permet de créer et partager des documents interactifs contenant du code, des visualisations, du texte et d'autres médias. Il offre un environnement convivial pour l'exploration de données, la programmation, la communication des résultats. Les notebooks Jupyter sont largement utilisés dans les domaines de la science des données, de la recherche et de l'apprentissage automatique pour faciliter l'analyse interactive et la présentation des informations [14].

- **Base de données : SQLite3**

SQLite est une bibliothèque en langage C qui met en oeuvre un moteur de base de données SQL petit, rapide, autonome, fiable et complet. SQLite est le moteur de base de données le plus utilisé au monde.

Le format de fichier SQLite est stable, multiplateforme et rétrocompatible, et les développeurs s'engagent à le maintenir ainsi jusqu'en 2050.

Le code source de SQLite est dans le domaine public et est libre d'utilisation pour tout le monde et à des fins quelconques.

3.2.5 Les librairies Python

Pandas

Pandas est une librairie python qui permet de manipuler facilement des données à analyser : manipuler des tableaux de données avec des étiquettes de variables (colonnes) et d'individus (lignes). ces tableaux sont appelés DataFrames [13].

NumPy

NumPy est une bibliothèque Python très populaire qui est principalement utilisée pour effectuer des calculs mathématiques et scientifiques. Elle offre de nombreuses fonctionnalités et outils qui peuvent être utiles pour les projets de Data Science [15].

Matplotlib

Matplotlib est une bibliothèque Python open source permettant de créer des visualisations de données[12].

Statsmodels

StatsModels est un package dédié à la modélisation statistique. Il incorpore un grand nombre de techniques économétriques telles que la régression linéaire, le modèle linéaire généralisé, le traitement des séries temporelles (AR, ARIMA, etc.) [16].

Keras

Keras est une API de réseau de neurones écrite en langage Python. Il s'agit d'une bibliothèque Open Source, exécutée par-dessus des frameworks tels que Theano et TensorFlow.

Conçue pour être modulaire, rapide et simple d'utilisation, Elle offre une façon simple et intuitive de créer des modèles de Deep Learning.

Aujourd'hui, Keras est l'une des APIs de réseaux de neurones les plus utilisées pour le développement et le testing de réseaux de neurones. Elle permet de créer très facilement des « layers » pour les Neural Networks ou de mettre en place des architectures complexes [2].

TensorFlow

TensorFlow est une bibliothèque open source compatible Python pour le calcul numérique qui rend l'apprentissage automatique et le développement de réseaux neuronaux plus rapides et plus faciles [5].

Scikit-Learn

encore appelé sklearn, est la bibliothèque la plus puissante et la plus robuste pour le machine learning en Python. Elle fournit une sélection d'outils efficaces pour l'apprentissage automatique

et la modélisation statistique, notamment la classification, la régression et le clustering via une interface cohérente en Python [4].

Folium

Folium est une des nombreuses bibliothèques open source complémentaires de Python qui permet d'étendre ses fonctionnalités. Dans son cas, Folium est un puissant outil de visualisation puisque cette bibliothèque rend possible la conception de cartes interactives [5].

3.3 Modélisation de la prédiction de vente

Notre application vise principalement à élaborer un système sophistiqué permettant de prédire les ventes (c'est-à-dire la consommation électrique des clients HTA). À cet effet, nous nous penchons sur le modèle SARIMA, mentionné dans le chapitre 1, ainsi que sur le modèle LSTM, évoqué dans le même chapitre. Ces deux modèles sont employés pour prévoir les tendances des séries temporelles.

Dans les paragraphes suivants, nous allons détailler minutieusement ces deux approches que nous avons utilisées pour cette étude.

3.3.1 La modélisation avec SARIMA

La méthode SARIMA a été employée pour prédire la consommation des cinq principaux clients HTA. Cette approche requiert un modèle dédié à chaque client. La modvlistation du modèle SARIMA s'est déroulée en suivant les différentes étapes exposées ci-après :

1. Collecte et récupération des données

La première étape consistait à constituer un jeu de données sur lequel nous pourrions tester et vérifier les algorithmes. Nous avons exploré toutes les possibilités pour obtenir les données nécessaires qui composent notre expérience et les exemples sur lesquels les algorithmes pourraient apprendre.

2. Nettoyage des données

Durant cette phase, nous avons utilisé la bibliothèque Pandas. Tout d'abord, nous avons importé le jeu de données contenant notre ensemble de données (dataset). Ensuite, nous avons procédé au nettoyage des données en remplaçant les cases vides par des zéros et en éliminant les valeurs aberrantes. La même bibliothèque a été utilisée pour visualiser les données.

3. Découpage des données

Nous avons divisé le jeu de données en deux parties à l'aide de la bibliothèque Pandas et de la méthode `dataframe.iloc()` :

- Ensemble d'entraînement (Train set) : qui représente 90 % des données destinées à l'apprentissage du modèle.
- Ensemble de test (Test set) : qui représente 10% des données utilisées pour évaluer le modèle.

Ces données ne doivent en aucun cas être utilisées lors de la conception du modèle.

4. **Modélisation** Le processus de construction du modèle SARIMA est divisé en plusieurs étapes. Tout d'abord, nous avons examiné le graphique des composantes de la série temporelle à l'aide de la bibliothèque `statsmodels.tsa.seasonals` et de la fonction `seasonal decompose`, afin d'avoir une vision globale de la chronique étudiée et à réaliser. Ensuite, nous avons effectué le test de stationnarité sur cette série à l'aide du test de Dickey-Fuller, un test statistique permettant d'évaluer si une série donnée est stationnaire ou non.

Nous l'avons utilisé avec la bibliothèque `statsmodels.tsa.stattools` et la fonction `adfuller`.

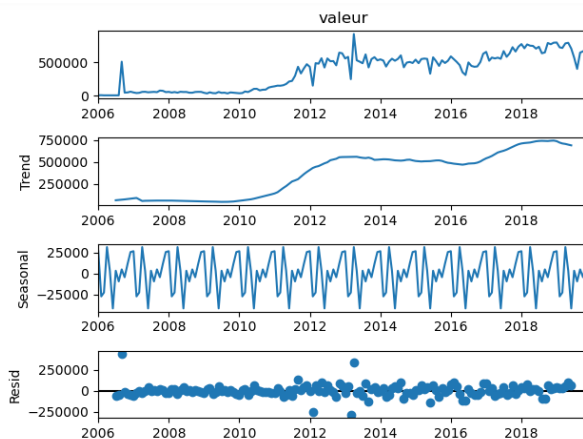


FIGURE 3.1 – Composantes de la série temporelle du client 66

Après avoir analysé le graphique de la figure 3.1, nous concluons que la série chronologique présente une tendance et une saisonnalité, dans notre cas la série n'est pas stationnaire (valeur de p proche de 1).

La non-stationnarité fait référence au fait que la structure d'un processus stochastique change avec le temps. SARIMA est un modèle non stationnaire qui gère la non stationnarité et la saisonnalité.

Compte tenu de toutes les informations ci-dessus, nous sommes arrivés à la conclusion que, dans notre cas, SARIMA est la méthode la plus appropriée pour traiter nos données ou séries.

Par la suite, nous avons appliqué le modèle SARIMA intégré dans `statsmodels.api` aux données d'apprentissage, après avoir sélectionné les ordres p , d , q , P , D et Q du modèle selon le critère d'information minimum de l'AIC.

5. **Evaluation** Afin d'évaluer la performance du modèle résultant, nous devons nous assurer que le résidu Notre modèle est non corrélé et normalement distribué. Pour cela, nous montrons Les diagnostics présentés ci-dessous correspondent à la fonction de graphe de diagnostic de la bibliothèque `matplotlib` nous permet de générer rapidement des diagnostics de modèle et d'enquêter Concernant tout comportement inhabituel.

Comme nous le voyons dans les figure 3.2, la bibliothèque `matplotlib` :

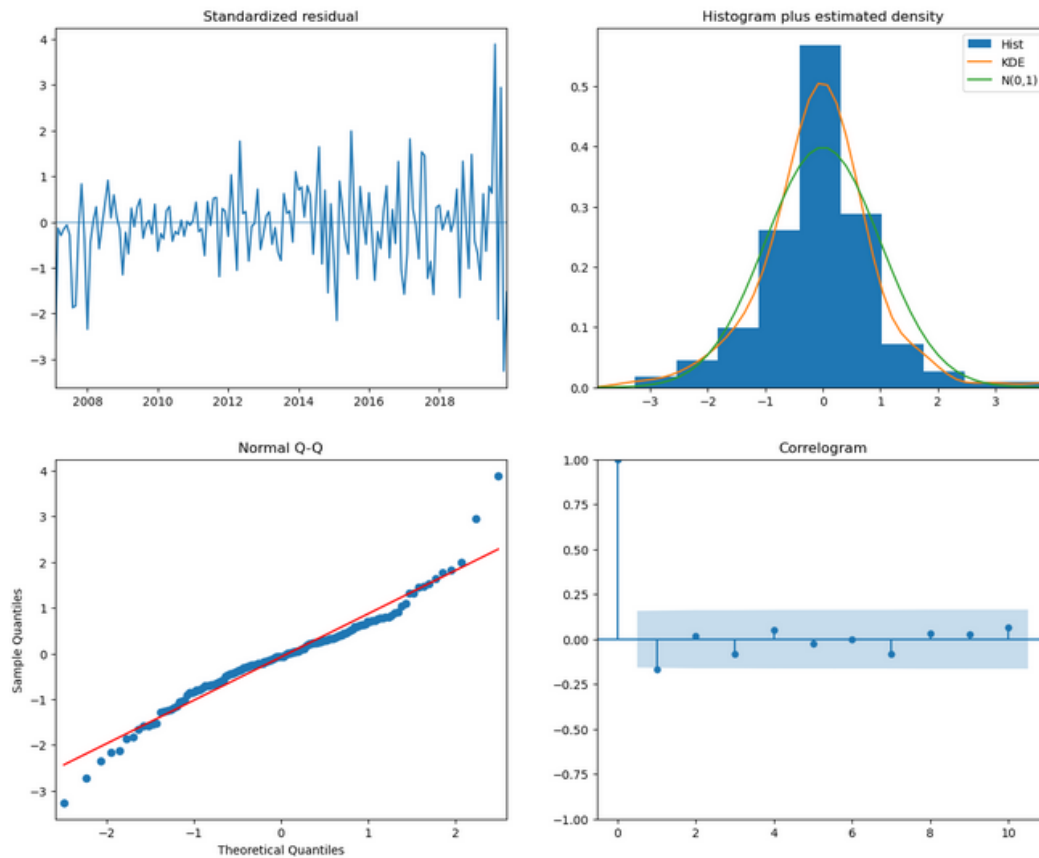


FIGURE 3.2 – Les diagnostics du modèle SARIMA

• **Histogramme plus estimation de la densité de noyau** Le graphique Histogramme Plus Estimated Density représente un histogramme des résidus du modèle SARIMA et une estimation de la densité du noyau (KDE).

Dans notre cas, nous observons que la courbe KDE (ligne rouge) suit de près la ligne de la distribution normale $N(0,1)$ (où $N(0,1)$ représente la moyenne 0 et l'écart type 1), ce qui montre que le modèle résiduel La différence est approximativement distribuée normalement. Une distribution normale des résidus est une caractéristique souhaitable d'un modèle bien ajusté.

Enfin, le diagramme Histogramme plus Densité estimée confirme que les résidus du modèle SARIMA sont normalement distribués, ce qui est un bon indicateur de la qualité de l'ajustement du modèle.

• **Graphique Q-Q** La distribution ordonnée des résidus (représentée par les points bleus) suit la tendance linéaire des échantillons tirés d'une distribution normale standard avec une moyenne de 0 et un écart type de 1 ($N(0,1)$). Cette observation renforce l'idée que les résidus du modèle sont normalement distribués.

• **Standardized residual** Les résidus au fil du temps ne présentent pas de motif saisonnier évident et semblent être un bruit blanc. Cela est confirmé par le graphique d'autocorrélation (également appelé corrélogramme) où aucune corrélation significative n'est observée.

- **Corrélogramme (graphique ACF)** Le graphique indique que les résidus de la série chronologique présentent une corrélation faible.

Comme nous pouvons le voir sur les graphiques ci-dessous, cela semble être beaucoup plus précis. Nous pouvons ensuite utiliser le modèle pour prévoir le nombre des consommations des clients au cours des 5 prochaines années.

6. **Exportation des résultats** : Nous avons enregistré les résultats des prédictions dans un fichier CSV en utilisant la bibliothèque pandas, en utilisant la méthode `dataframe.to_csv()`. Nous avons suivi la même approche décrite précédemment pour construire Onze modèles pour les Onze principaux clients, chaque client ayant son propre modèle. Les résultats de prédiction ont été enregistrés dans un fichier CSV, que nous avons intégré dans la base de données que nous interrogeons à l'aide du framework Streamlit.

3.3.2 La modélisation avec LSTM

Nous avons utilisé cette méthode pour prédire la consommation de tous les clients HTA. Contrairement à SARIMA, qui nécessite un modèle distinct pour chaque client, cette méthode nécessite un seul modèle pour l'ensemble des clients. Nous avons opté pour LSTM car nous souhaitons effectuer des prédictions basées sur le temps, notamment des séquences longues. De plus, notre ensemble de données comprend des échantillons individuels pour chaque client, et la prédiction dépend de plusieurs variables en entrée.

Le processus de construction du modèle LSTM se déroule en suivant les différentes étapes expliquées ci-dessous :

1. Les deux premières étapes, à savoir la préparation des données et le nettoyage des données, sont identiques à celles mentionnées dans la section précédente concernant SARIMA.
2. **Découpage des données**
Nous avons divisé le jeu de données par année à l'aide de la bibliothèque Pandas en utilisant la méthode `dataframe.loc()`. Ce découpage a été effectué en deux parties :
 - a. **Données d'entraînement** : Les données allant de l'année 2006 jusqu'à 2018 sont destinées à l'entraînement du modèle.
 - b. **Données de test** : Les données des années 2018 et 2019. L'année 2019 a été utilisée pour effectuer le test, tandis que l'année 2018 est indispensable pour prédire l'année 2019.
3. **Supervision des données** La supervision des données est effectuée dans la fonction `create-sequences`.

Les séquences d'entrée (X) et de sortie (Y) sont créées de manière à ce que chaque séquence d'entrée corresponde au point de sortie correspondant.

Par exemple, si `sequence-length` est fixé à 14, la première séquence d'entrée contiendra les 14 premiers points de données, et le premier point de sortie correspondant sera également inclus dans y.

4. Normalisation des données

La normalisation des données est effectuée à l'aide du MinMaxScaler fourni par la classe `sklearn.preprocessing`.

Dans le code, les données de consommation (df-cons) sont mises à l'échelle en utilisant cette normalisation avant d'être utilisées pour l'entraînement du modèle LSTM.

La normalisation est importante pour mettre les données à la même échelle, ce qui facilite l'apprentissage pour le modèle..

5. Modélisation (conception du modèle et entraînement du modèle)

La modélisation est réalisée en utilisant la bibliothèque Keras.

Le modèle LSTM est construit en utilisant la classe `Sequential`, qui permet de créer un modèle séquentiel couche par couche :

Une couche LSTM de 200 unités qui renvoie des séquences (`return sequences=True`) et prend des entrées de forme (`X.shape[1]`, `X.shape[2]`). Cette couche est suivie d'une couche de régularisation des pertes avec un taux de 0,2.

La deuxième couche LSTM a 200 unités et renvoie également une séquence, suivie d'une couche de régularisation des gouttes avec un taux de 0,2.

La troisième couche LSTM a 100 unités et renvoie une seule valeur (pas de séquence) suivie d'une couche d'abandon régularisée avec un taux de 0,2.

Couche dense (entièrement connectée) avec 1 unité pour la prédiction finale.

Après la conception du modèle, le modèle est compilé en spécifiant l'optimiseur (Adam), la fonction de perte (mean-squared-error) et les métriques à suivre pendant l'entraînement (accuracy et mse). Ensuite, le modèle est entraîné sur les données d'entrée (X) et de sortie (y) à l'aide de la fonction `fit`. le modèle est entraîné pendant 500 epochs avec une taille de lot de 32.

6. Evaluation

Dans notre modèle nous avons utilisé les métriques, le MSE (Mean Squared Error) et le MAPE (Mean Absolute Percentage Error) pour évaluer les performances du modèle.

MSE (Mean Squared Error) : Le MSE est une mesure couramment utilisée pour évaluer la qualité d'un modèle de régression.

L'équation MSE est représentée comme suit :

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{\text{pred}} - y_{\text{true}})^2 \quad (3.1)$$

Où :

- n est le nombre d'échantillons
- y-pred est la valeur prédite par le modèle
- y-true est la valeur réelle

MAPE (Mean Absolute Percentage Error) : Le MAPE est une mesure de l'erreur relative moyenne exprimée en pourcentage. Il est utilisé pour évaluer la précision des prédictions par rapport aux valeurs réelles.

L'équation MAPE est représentée comme suit :

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_{\text{true}} - y_{\text{pred}}}{y_{\text{true}}} \right| \times 100 \quad (3.2)$$

Où :

- n est le nombre d'échantillons
- y_{pred} est la valeur prédite par le modèle
- y_{true} est la valeur réelle

7. **Sauvegarde du modèle** : Nous avons utilisé la méthode `save` de l'objet modèle pour enregistrer le modèle entraîné .

3.4 Présentation de l'application

3.4.1 Application web

L'architecture de l'application web développée est basée sur le framework Streamlit, qui combine à la fois le côté customer et le côté serveur en une seule entité.

Streamlit permet de créer des opérations web interactives directement en utilisant Python, sans avoir besoin d'un serveur externe.

Dans ce modèle, le navigateur web du customer effectue une demande vers l'application Streamlit, qui exécute le code Python correspondant. Streamlit traite ensuite la demande et génère la réponse, qui est renvoyée au navigateur du customer sous forme d'une interface utilisateur interactive.

L'interface graphique de la partie customer de l'application web est réalisée en utilisant les fonctionnalités intégrées de Streamlit.

Streamlit fournit des contraptions interactifs tels que des curseurs, des listes déroulantes et des boutons, qui permettent aux utilisateurs de manipuler les données et d'interagir avec l'application.

En ce qui concerne le développement de la partie serveur, Streamlit simplifie le processus en utilisant une approche de développement simplifiée.

Il n'est pas nécessaire d'utiliser un framework web externe comme Flask.

Le code Python de l'application Streamlit sert à la fois de serveur et de modèle, avec la logique métier et les opérations de traitement de données intégrées directement dans le code.

En résumé, l'architecture de l'application web développée avec Streamlit est une armature simplifiée où le navigateur du customer envoie des demandes directement à l'application Streamlit, qui traite ces demandes et renvoie les réponses correspondantes.

L'interface utilisateur est créée à l'aide des fonctionnalités intégrées de Streamlit, tandis que la logique métier et le traitement des données sont gérés dans le code Python de l'application Streamlit.

3.4.2 Quelques interfaces graphiques

Nous allons maintenant présenter les différentes interfaces de notre application et expliquer leur utilité respective :

3.4.2.1 Interface d'authentification

L'interface d'authentification, illustrée dans la figure 3.3, permet à l'utilisateur de saisir son nom d'utilisateur et son mot de passe afin de pouvoir accéder aux fonctionnalités du système.

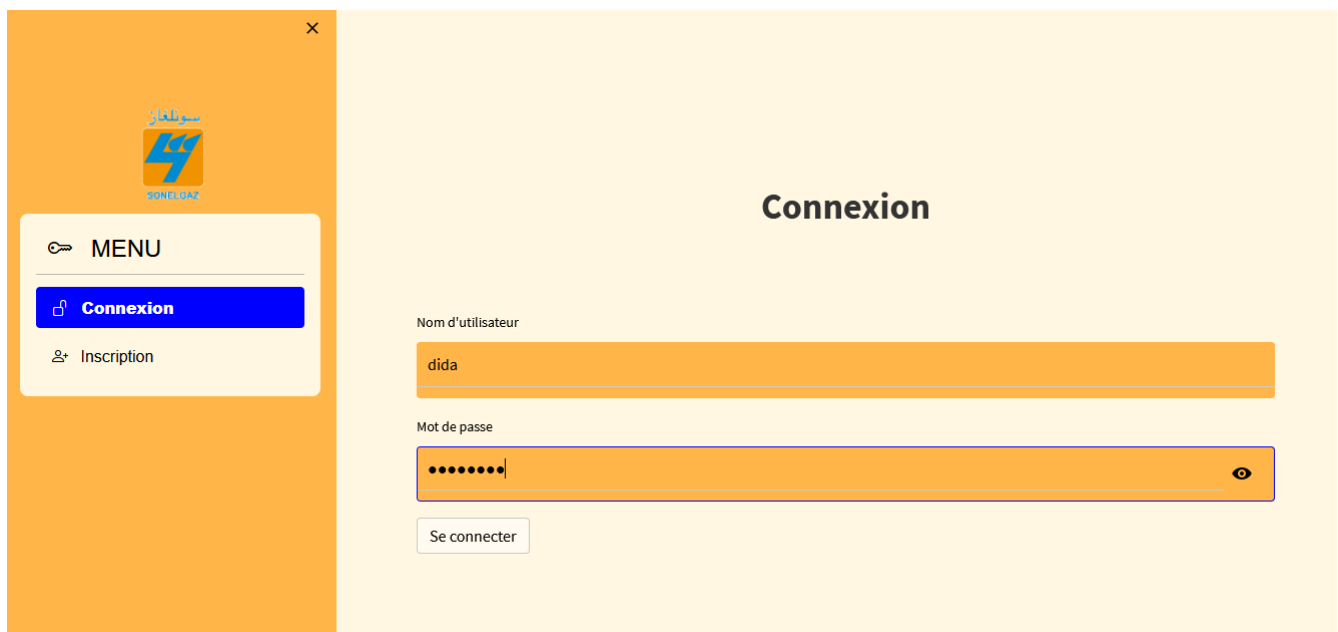


FIGURE 3.3 – Interface d'authentification

3.4.2.2 Interface du tableau de bord

L'interface de tableau de bord, représentée dans la figure 3.4, est la première interface qui apparaît après que l'utilisateur s'est connecté avec succès.

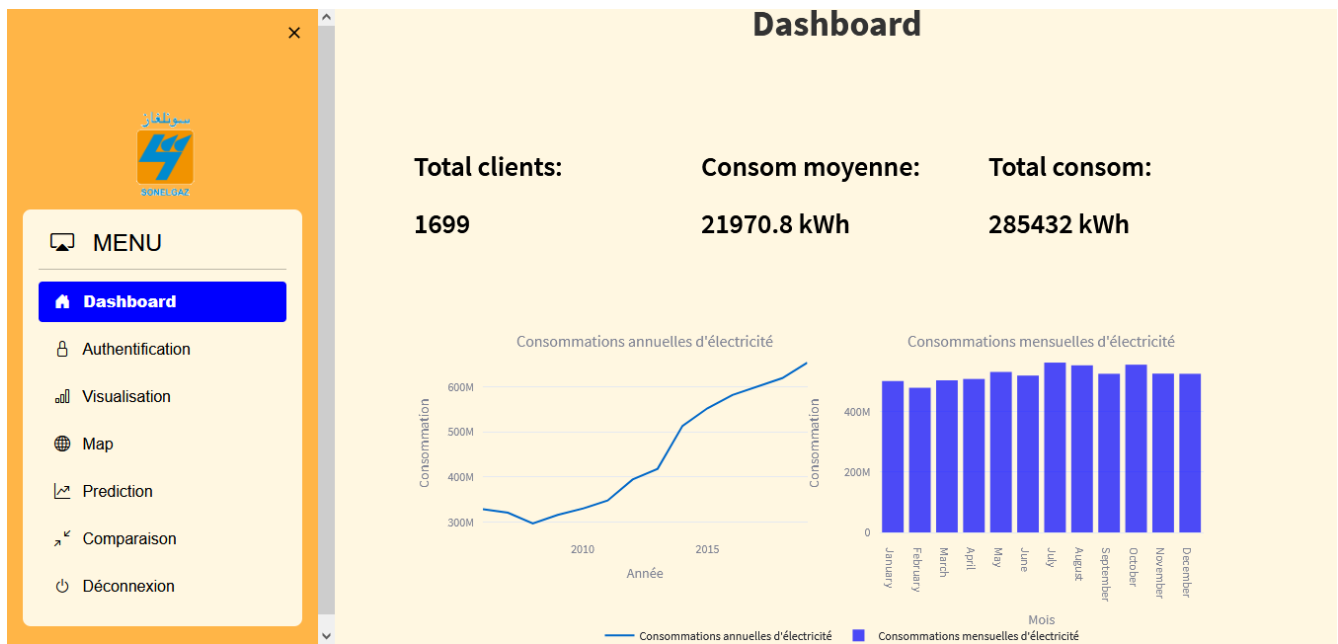


FIGURE 3.4 – Interface de tableau de bord

3.4.2.3 Interface Inscription

L'interface d'inscription, représentée dans la figure 3.5, est l'interface qui s'affiche lorsque les utilisateurs souhaitent créer un nouveau compte.

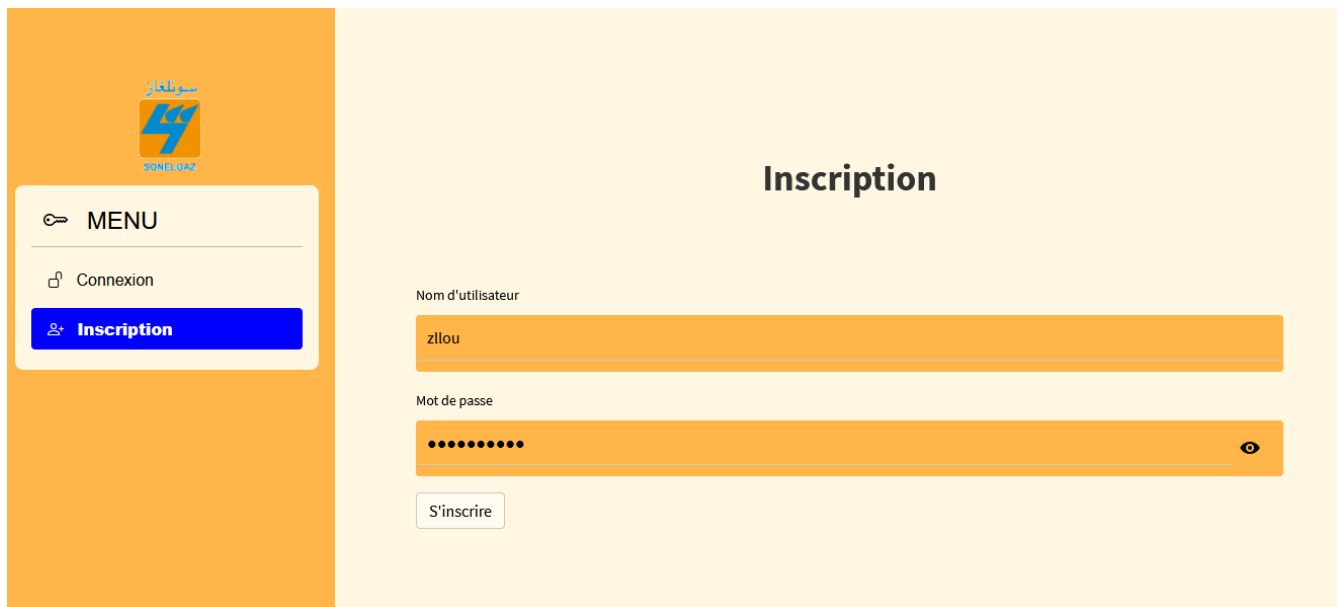


FIGURE 3.5 – Interface Inscription

3.4.2.4 Interface de visualisation générale des ventes

L'interface de visualisation générale des ventes, présentée dans les figures 3.6,3.7 offre aux utilisateurs une vue d'ensemble complète et intuitive de leurs données de vente, en kilowatts (KW).

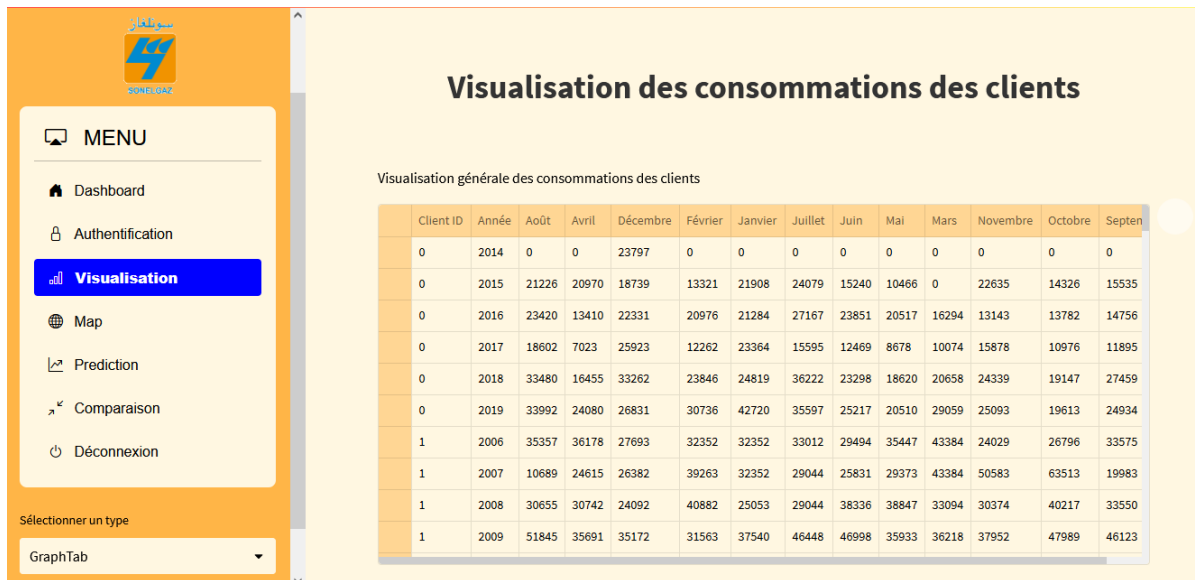


FIGURE 3.6 – Interface de visualisation générale des ventes-Tableau

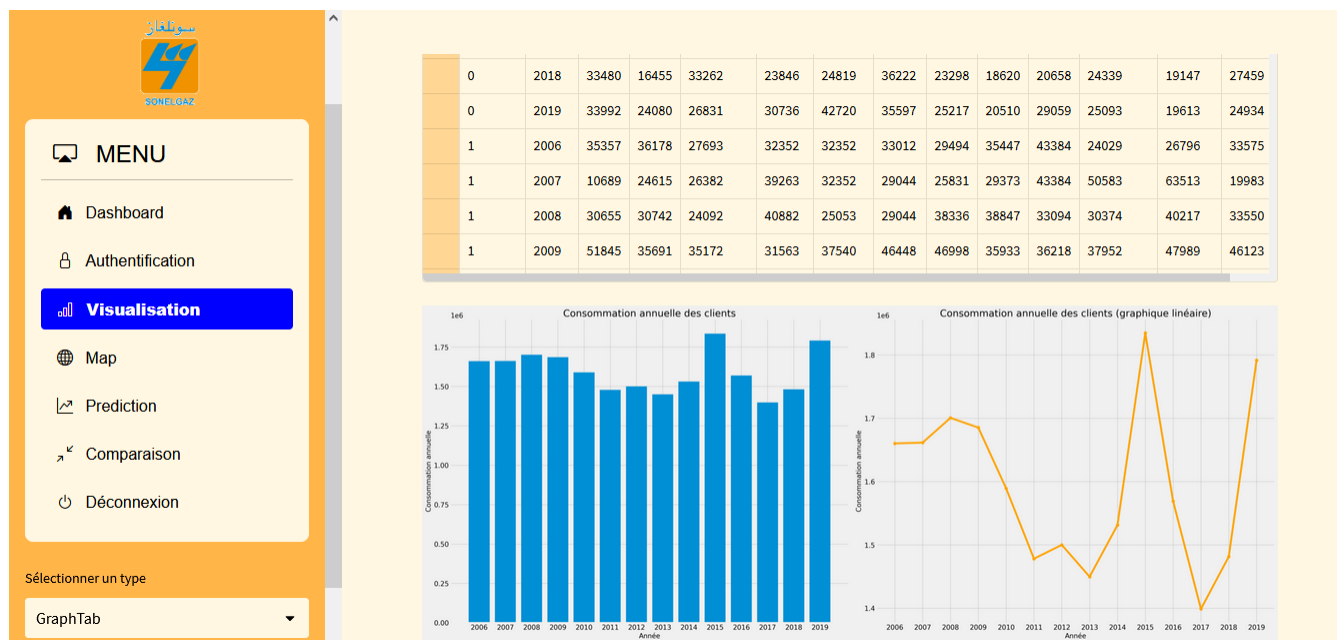


FIGURE 3.7 – Interface de visualisation générale des ventes-Graphe

3.4.2.5 Interface visualisation des ventes par client

L’interface de visualisation des ventes, illustrée dans la figure 3.8, offre aux utilisateurs une expérience interactive pour explorer les données de vente à travers des graphiques et des tableaux.

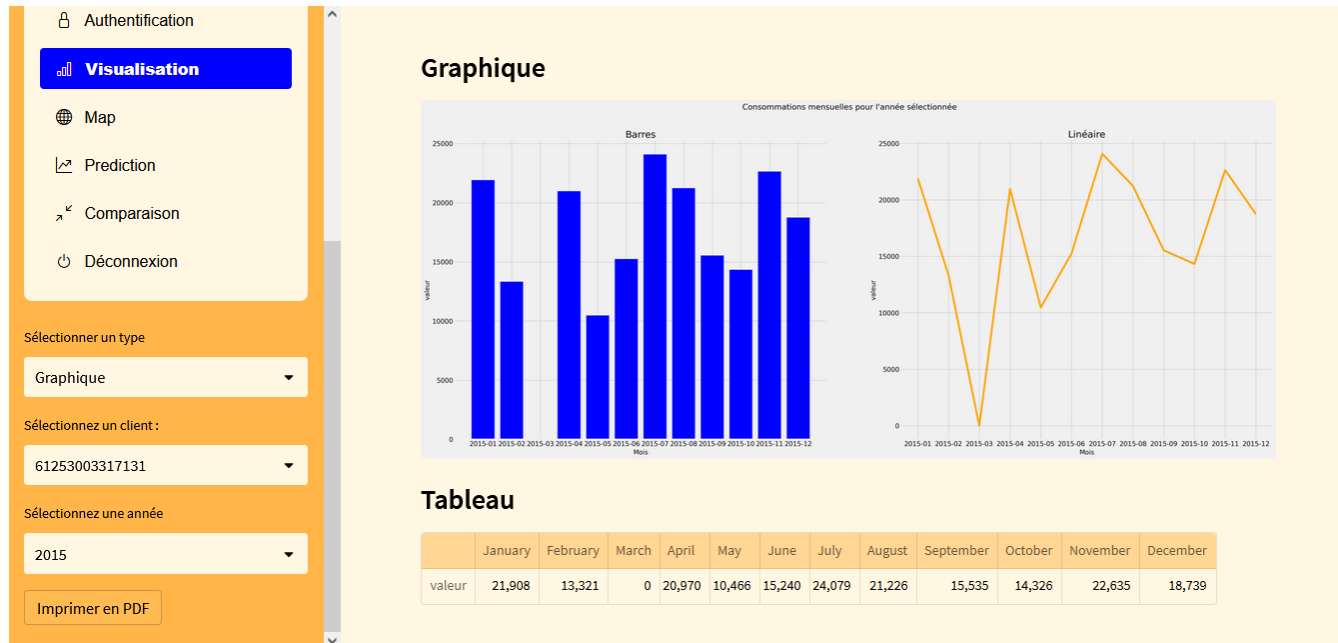


FIGURE 3.8 – Interface visualisation des ventes par client

3.4.2.6 Interface localisation des clients

L’interface de visualisation de la localisation des grands clients, illustrée dans la figure 3.9, présente sur une carte géographique de la wilaya de Bejaia les emplacements de quelques clients importants

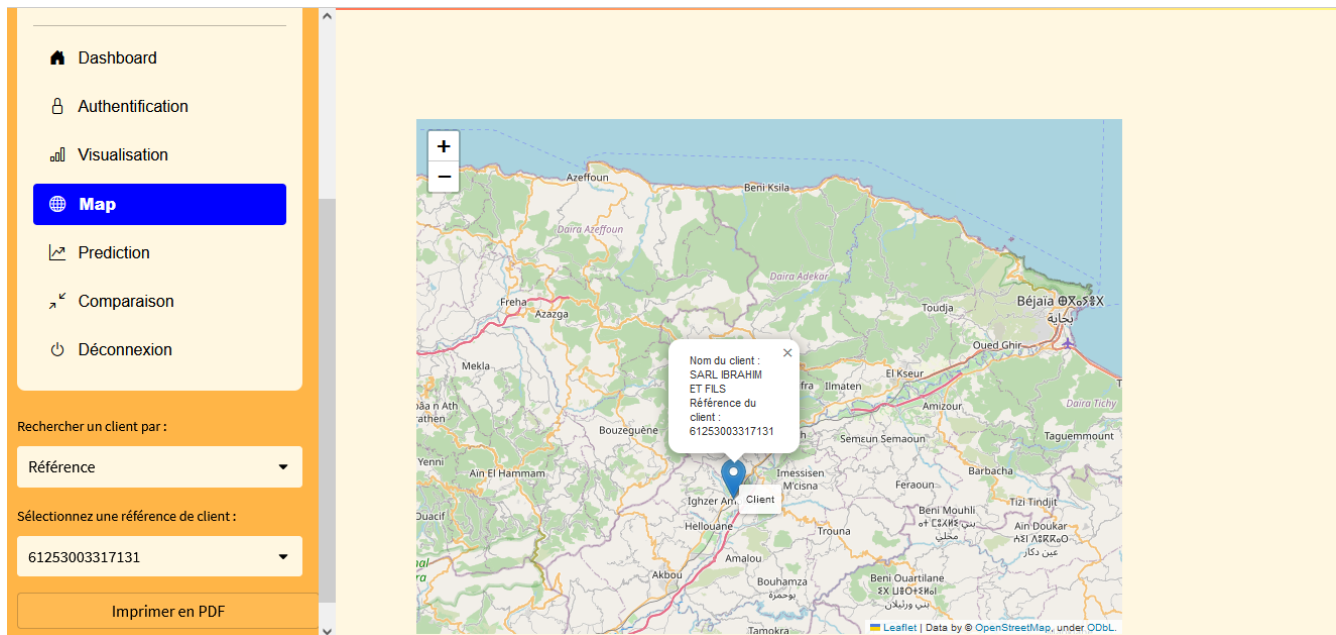


FIGURE 3.9 – Interface localisation des clients

3.4.2.7 Interface prédiction des ventes

L'interface de prédiction des ventes, offre aux utilisateurs deux méthodes de prédiction : SARIMA(Machine Learning) et LSTM(Deep Learning). Ces deux approches permettent d'estimer les ventes futures en se basant sur les données historiques.

- **Machine learning**

Grâce à cette interface, présentée dans les figures 3.10, 3.11, 3.12, les utilisateurs ont la possibilité d'afficher les prédictions de la consommation des onze principaux clients HTA en utilisant la méthode ARIMA.

Pour cela, ils doivent simplement saisir l'identifiant du client concerné. Par exemple, si l'utilisateur saisit l'identifiant du client 100, les prédictions correspondantes s'afficheront dans les figures suivantes :



FIGURE 3.10 – Interface prédiction avec ARIMA-Graphe

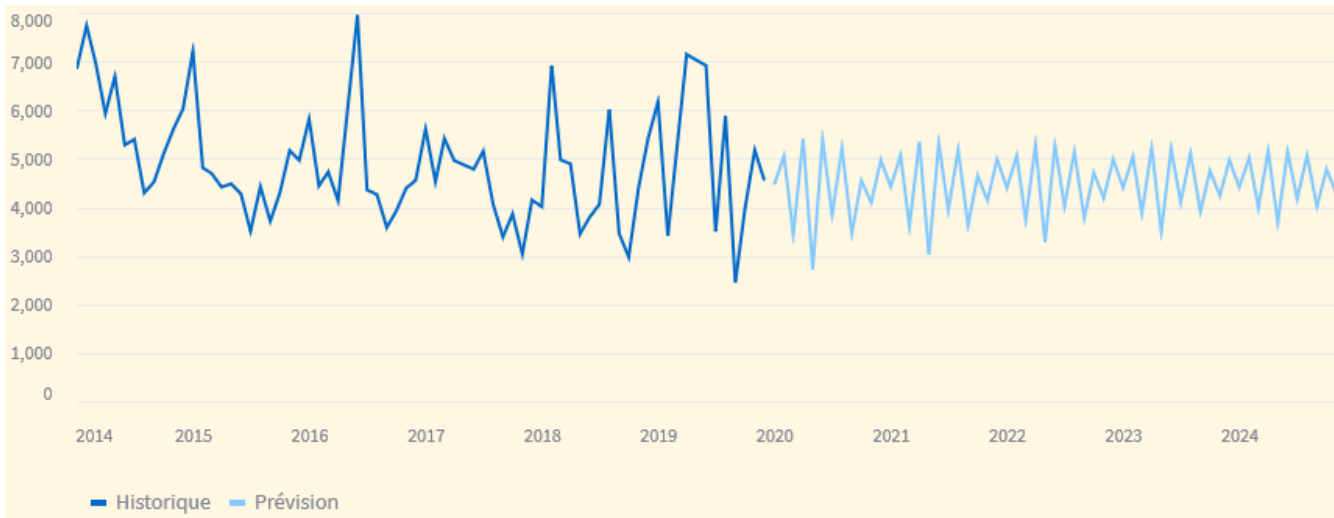


FIGURE 3.11 – Prédiction avec ARIMA-Graphe

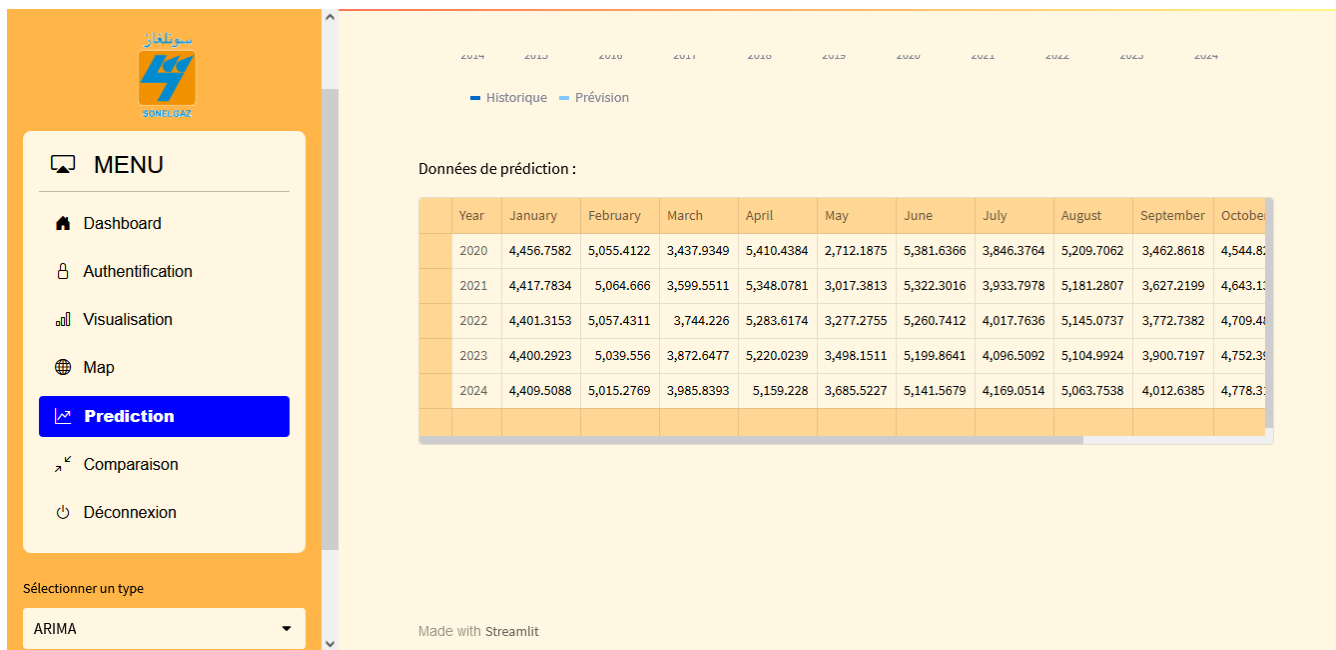


FIGURE 3.12 – Interface prédiction avec ARIMA-Tableau

Grâce à cette interface, présentée dans les figures 3.13, 3.14, 3.15, les utilisateurs ont la possibilité d’afficher les prédictions de la consommation des onze principaux clients HTA en utilisant la méthode SARIMA.

Pour cela, ils doivent simplement saisir l’identifiant du client concerné. Par exemple, si l’utilisateur saisit l’identifiant du client 100, les prédictions correspondantes s’afficheront dans les figures suivantes :

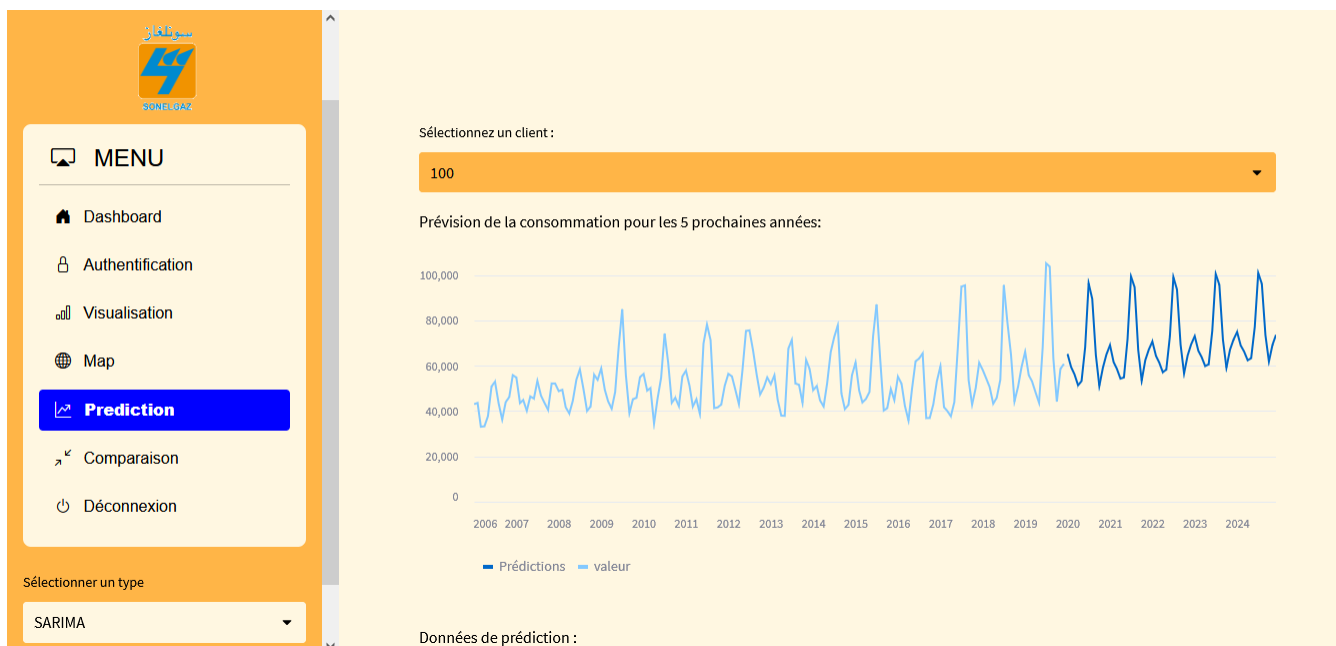


FIGURE 3.13 – Interface prédiction avec SARIMA-Graphe

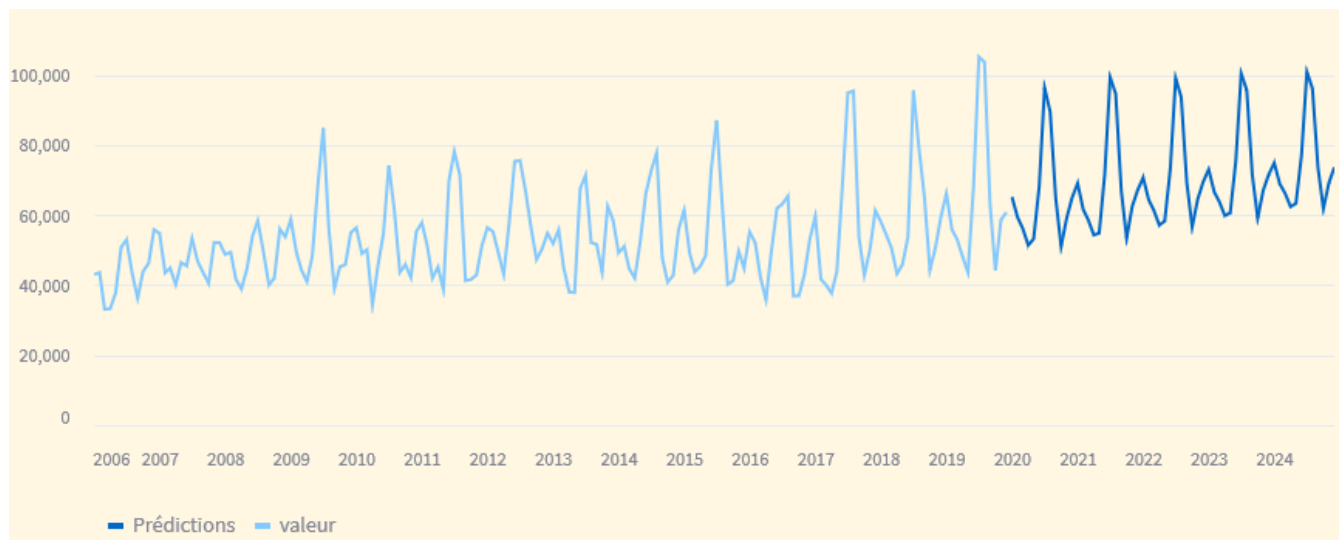


FIGURE 3.14 – Prédiction avec SARIMA-Graphe



FIGURE 3.15 – Interface prédiction avec SARIMA-Tableau

● Deep learning

Grâce à cette interface, présentée dans les figures 3.16, 3.17, 3.18, les utilisateurs ont la possibilité d'afficher les prédictions de la consommation des onze principaux clients HTA en utilisant la méthode LSTM.

Pour cela, ils doivent simplement saisir l'identifiant du client concerné. Par exemple, si l'utilisateur saisit l'identifiant du client 100, les prédictions correspondantes s'afficheront dans les figures suivantes :



FIGURE 3.16 – Interface prédiction avec LSTM-Graphe

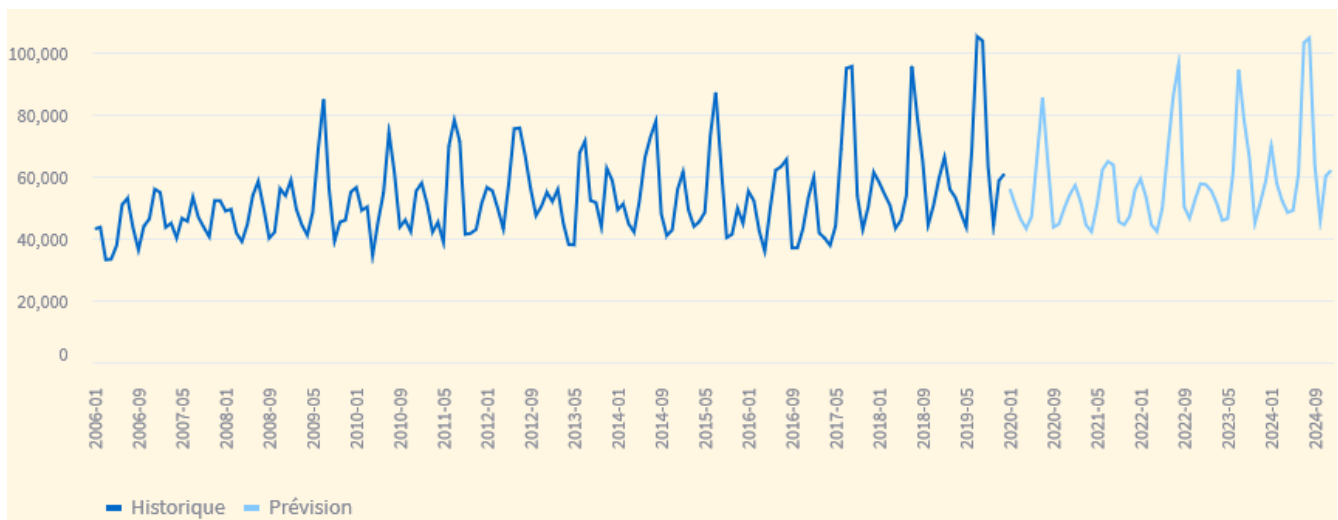


FIGURE 3.17 – Prédiction avec LSTM-Graphe



FIGURE 3.18 – Interface prédiction avec LSTM-Tableau

3.4.2.8 Interface comparaison des ventes

L’interface de comparaison des ventes, présentée dans la figure 3.19, permet aux utilisateurs de comparer les consommations en sélectionnant deux mois spécifiques.



FIGURE 3.19 – Interface comparaison des ventes

3.4.2.9 Interface comparaison des ventes (Consommation-Prédiction(SARIMA))

L'interface de comparaison des ventes, présentée dans la figure 3.20, permet aux utilisateurs de comparer les ventes réelles avec les prédictions(SARIMA) en sélectionnant deux mois spécifiques.



FIGURE 3.20 – Interface comparaison des ventes (Consommation-Prédiction(SARIMA))

3.5 La comparaison entre les méthodes ARIMA, SARIMA et LSTM

En analysant les réalisations et les prédictions obtenues avec les méthodes SARIMA, ARIMA et LSTM, nous avons observé des différences significatives dans leurs performances.

Pour commencer, en ce qui concerne les prédictions de ARIMA, nous avons constaté que celles-ci ne sont pas aussi précises et fiables par rapport à celles de SARIMA et de LSTM. Les prédictions de ARIMA ont montré une tendance à sous-estimer ou surestimer les valeurs réelles de la série temporelle. Cela peut être dû aux limitations de ARIMA dans la modélisation de structures plus complexes et dans la capture des dépendances à long terme.

En revanche, les prédictions de SARIMA ont présenté des résultats plus satisfaisants et comparables à ceux de LSTM. SARIMA, en incluant les termes saisonniers, a pu mieux capturer les variations saisonnières et les tendances de la série temporelle. Cela a conduit à des prédictions plus précises et cohérentes par rapport à ARIMA.

Quant à LSTM, ses prédictions ont été remarquablement bonnes et professionnelles. Grâce

à sa capacité à capturer les motifs temporels complexes et les dépendances à long terme, LSTM a surpassé les autres méthodes dans la précision et la qualité des prédictions. Il a démontré une capacité à apprendre et à modéliser les caractéristiques subtiles de la série temporelle, ce qui lui a permis de générer des prédictions plus fiables et adaptées aux données.

3.5.1 Discussion des résultats

Après avoir comparé les performances des trois méthodes (LSTM, ARIMA et SARIMA) en utilisant la métrique MAPE, voici les résultats obtenus :

Le modèle LSTM a donné le plus faible MAPE, ce qui indique une meilleure précision de prédiction par rapport aux modèles ARIMA et SARIMA.

| Méthode | MAPE (%) |
|---------|----------|
| ARIMA | 12.5 |
| SARIMA | 11.8 |
| LSTM | 10.2 |

TABLEAU 3.1 – Comparaison des méthodes ARIMA, SARIMA et LSTM selon la métrique MAPE.

ARIMA a montré des prédictions moins performantes que SARIMA et LSTM.

SARIMA a réussi à capturer efficacement les tendances saisonnières et les variations de la série temporelle.

Tandis que le modèle LSTM s'est démarqué en fournissant des prédictions exceptionnelles grâce à sa capacité à modéliser les motifs complexes et les dépendances à long terme.

En résumé, Chacune des méthodes - ARIMA, SARIMA et LSTM - présente ses propres caractéristiques et avantages :

ARIMA et SARIMA sont des approches classiques qui fonctionnent bien pour modéliser des séries temporelles simples et bien structurées.

LSTM, en revanche, est une méthode plus avancée qui peut capturer des dépendances complexes à long terme et s'adapte aux séries temporelles non stationnaires.

3.6 Conclusion

Dans ce chapitre, nous avons abordé l'aspect pratique de notre projet. Nous avons commencé par décrire les environnements matériels et logiciels que nous avons utilisés.

Ensuite, nous avons mis l'accent sur les différentes interfaces de notre application et expliqué comment elles fonctionnent dans la pratique.

Enfin, nous nous sommes approfondis dans les détails de la construction et de la modélisation des prédictions de ventes. Nous avons utilisé les puissants modèles de prévision de séries temporelles, tels que SARIMA et LSTM

Conclusion et perspectives

Pour conclure, notre projet consistait à concevoir et mettre en place une application web de prévision et d'analyse des ventes d'électricité pour les clients HTA de Sonelgaz. Nous avons développé une application web d'analyse et de prévision des ventes, axée sur le thème : "Application des techniques d'intelligence artificielle pour l'analyse prédictive des ventes" de la gestion des ventes électriques, pour répondre aux besoins du Responsable Commercial Sonelgaz de BEJAIA afin d'améliorer la prise de décision dans leur service grands comptes.

Notre approche a porté sur la résolution des problèmes suivants rencontrés par la direction commerciale Sonelgaz de BEJAIA : le déséquilibre entre production et besoins entraînant des surplus ou des déficits d'électricité, la détection des fraudes liées aux compteurs d'énergie chez les clients qui est difficile, entraînant des évaluations de consommation erronées et des facturations inexactes, ainsi que l'utilisation de méthodes d'analyse de ventes traditionnelles et l'absence de systèmes d'aide à la décision avancés.

L'adoption de la méthode Scrum nous a permis de gérer efficacement le projet en découpant les objectifs en itérations (sprints) et en favorisant une collaboration étroite avec le client. Le product backlog a joué un rôle central dans la définition des besoins fonctionnels, offrant une vision claire des tâches à accomplir et assurant la satisfaction des attentes du client.

Le langage UML a été utilisé avec succès pour modéliser les différentes fonctionnalités de notre application de prédiction. Cela nous a permis de mieux comprendre et de communiquer efficacement les concepts et les processus au sein de l'équipe de développement.

En intégrant deux modèles d'apprentissage automatique et profond, nous avons renforcé les capacités de prédiction de notre application. Ces modèles, soigneusement sélectionnés en fonction des spécificités du problème et des données disponibles, ont apporté une plus grande précision et ont permis d'obtenir des résultats significatifs dans la prédiction des événements futurs.

Ce projet a été une opportunité d'apprentissage précieuse, nous permettant de nous familiariser avec les nouvelles technologies, en particulier l'intelligence artificielle, l'apprentissage automatique et le deep learning. Nous avons pu exploiter les avantages offerts par ces approches innovantes pour améliorer nos compétences en programmation et approfondir notre compréhension des concepts clés de l'IA.

Outre les aspects techniques, ce projet nous a également confrontés aux réalités de la vie

professionnelle et aux défis inhérents à la gestion d'un projet d'envergure. Nous avons développé des compétences essentielles en matière de planification, de collaboration et de résolution de problèmes, qui nous seront bénéfiques dans notre future carrière.

Ce projet constitue une étape cruciale dans notre parcours académique, nous permettant d'acquérir des compétences précieuses et de nous préparer à relever les défis futurs dans le domaine de la programmation et de l'intelligence artificielle.

Pour les perspectives d'avenir, nous prévoyons de développer de nouvelles fonctionnalités, telles que la détection d'anomalies, afin de mieux équilibrer la production et les besoins en électricité, ainsi que de mettre en place des mécanismes de détection des fraudes liées aux compteurs d'énergie chez les clients. Nous chercherons également à étendre notre système pour analyser et prédire les ventes d'autres catégories de clients. En parallèle, nous continuerons à améliorer nos connaissances et compétences dans ce domaine en constante évolution, en explorant des méthodes d'analyse de ventes avancées et en exploitant les avancées de l'intelligence artificielle pour soutenir la prise de décision et améliorer les performances commerciales.

Bibliographie

- [1] Python. <https://www.python.org/>, Consulté le 1 Avril 2023.
- [2] Keras. <https://keras.io/>, Consulté le 10 Avril 2023.
- [3] Tensorflow. <https://www.tensorflow.org/?hl=fr>, Consulté le 11 Avril 2023.
- [4] Scikit-learn. <https://scikit-learn.org/>, Consulté le 12 Avril 2023.
- [5] Folium. <https://www.foliumofficiel.com/>, Consulté le 13 Avril 2023.
- [6] Gestion de projet agile. <https://le-consultant-digital.com/gestion-de-projet/gestion-de-projet-agile>, Consulté le 14 Février 2023.
- [7] Anaconda. <https://www.anaconda.com/>, Consulté le 17 Avril 2023.
- [8] Css basics. <https://CSS-basics/>, Consulté le 19 Avril 2023.
- [9] Pycharm. <https://www.jetbrains.com/fr-fr/pycharm/>, Consulté le 2 Avril 2023.
- [10] Streamlit. <https://streamlit.io/>, Consulté le 25 Mars 2023.
- [11] Visual paradigm. <https://www.visual-paradigm.com/>, Consulté le 3 Avril 2023.
- [12] Matplotlib. <https://matplotlib.org/>, Consulté le 4 Avril 2023.
- [13] Pandas. <https://pandas.pydata.org/>, Consulté le 5 Avril 2023.
- [14] Jupyter. <https://jupyter.org/>, Consulté le 6 Avril 2023.
- [15] Numpy. <https://numpy.org/>, Consulté le 7 Avril 2023.
- [16] Statsmodels. <https://www.statsmodels.org/>, Consulté le 9 Avril 2023.
- [17] Gestion de projet. <https://www.ipag.edu/blog/gestion-projet?>, Consulté le 9 Mai 2023.
- [18] Hamla Abdenour and Chebout Lamia. *Reconnaissance automatique de gestes manuels en langue des signes avec le Deep learning*. PhD thesis in Computer Science, Université Mohamed Larbi Ben M'hidi Oum El bouaghi, 2021-2022.

- [19] Pranshu Aggarwal. Sarima using python forecast seasonal data. <https://www.wisdomgeek.com/development/machine-learning/sarima-forecast-seasonal-data-using-python/>, 20 Juillet 2020, Consulté le 15 Février 2023.
- [20] Laurent Audibert. Uml 2. <https://laurent-audibert.developpez.com/Cours-UML>, Consulté le 15 Mai 2023.
- [21] Chloé-Agathe Azencott. *Introduction au Machine Learning*. Dunod, Paris, 2nd edition, 2 Février 2022.
- [22] Nada Belaidi. Modèles de machine learning les plus populaires. <https://blent.ai/modele-machine-learning-populaires/>, Le 24 mai 2022, Consulté le 07 Mars 2023.
- [23] Eugene Charniak. *Introduction au Deep Learning*. PhD thesis in Computer Science, Brown (à Providence, Rhode Island), Janvier 2021.
- [24] Yan Girard. *Séries chronologiques à une et plusieurs variables*. PhD thesis in Computer Science, University of Quebec, May 2011.
- [25] Romain Herault and Clement Chatelain. Initiez-vous au deep learning : Découvrez les cellules à mémoire interne : les lstm. <https://openclassrooms.com/fr/courses/5801891-initiez-vous-au-deep-learning/5814656-decouvrez-les-cellules-a-memoire-interne-les-lstm>, Consulté le 15 Mai 2023.
- [26] Jean-Claude Heudin. *Comprendre le Deep Learning : Une introduction aux réseaux de neurones*. Paris, Novembre 2016.
- [27] Charles. E Holt. Forecasting seasonals and trends. *Exponentially weighted averages(O.N.R. Memorandum No. 52)*, 1957.
- [28] Christophe Hurlin and Valérie Mignon. *Statistiques et probabilités en économie-gestion*. Dunod, Paris, 2015.
- [29] Kassidy Kelley and Gaétan Raoul. Machine learning : les 9 types d’algorithmes les plus pertinents en entreprise. <https://www.lemagit.fr/conseil/Machine-Learning-les-9-types-dalgorithmes-les-plus-pertinents-en-entreprise>, 8 juin 2020, Consulté le 07 Mars 2023.
- [30] Mahamoudou Kiemtoré. *Développement d’une approche de management des megaprojets d’investissement favorisant la performance*. PhD thesis in Computer Science, Université du Québec à Chicoutimi, 2016.
- [31] Agnès LAGNOUX. Séries chronologiques. *Prévision par lissage exponentiel*, 67, 1999 à 2004.
- [32] Véronique Messenger. *Gestion de projet agile, avec Scrum, Lean, Extreme Programming...* Eyrolles, 21 Février 2013.

- [33] V. Monbet. *Modélisation des séries temporelles*. PhD thesis in Computer Science, University of Rennes, 2016-2017.
- [34] Gourav Singh. Understanding architecture of lstm. <https://www.analyticsvidhya.com/blog/2021/01/understanding-architecture-of-lstm>, Published On January 21, 2021 , Consulté le 14 Mai 2023.
- [35] Groupe Sonelgaz. Sonelgaz. <https://www.sonelgaz.dz/fr>, 1969, Consulté le 15 Mars 2023.
- [36] Bruno Stuner. *Cohorte de réseaux de neurones récurrents pour la reconnaissance de l'écriture*. PhD thesis in Computer Science, Université de Rouen Normandie, 2018.
- [37] Denis Thuillier. Principes et application des réseaux de neurones. *Revue région et développement n05-1997*, 28, 1997.
- [38] M.-C. Viano and A. Philippe. Maîtrise d'économétrie. *Cours de Séries Temporelles*, 67, 1999 à 2004.
- [39] Gasmi Zahia. *Prévision par méthodes de lissage exponentiel (Application sur la série Taux brut de naissance en Algérie 1967-2000)*. Master in Computer Science, Université Kasdi Merbah de Ouargla, July 2019.

RÉSUMÉ

Ces dernières années, l'intelligence artificielle a connu des progrès significatifs dans les domaines de l'apprentissage automatique ou Machine Learning (ML) et de l'apprentissage profond ou Deep learning (DL). Les entreprises voient en ces technologies une opportunité d'améliorer leurs opérations et de relever les défis liés à la demande croissante en énergie. Ce projet propose une application web 3-tier utilisant l'intelligence artificielle pour analyser et prévoir les ventes au sein de la direction commerciale de Sonelgaz de BEJAIA. Les modèles LSTM, ARIMA et SARIMA sont utilisés pour la visualisation et la prédiction des ventes, et leur précision est comparée à celle du modèle ARIMA en utilisant la métrique MAPE. Les résultats montrent que ARIMA obtient un MAPE de 12.5%, SARIMA obtient un MAPE de 11.8%, et LSTM obtient le meilleur résultat avec un MAPE de 10.2%, démontrant ainsi son efficacité dans la prédiction des ventes. Pour mener à bien ce projet, la méthodologie Scrum, une approche agile de développement et de gestion de projet, a été adoptée. Des tests d'application ont été réalisés à partir des données de près de 1699 clients, couvrant une période de 14 ans de consommation mensuelle d'électricité à BEJAIA.

Mots clés : Série temporelle, Méthode Agile, Intelligence Artificielle, Machine Learning, Deep Learning, Analyse et Prédiction des ventes.

ABSTRACT

In recent years, artificial intelligence has made significant progress in the fields of machine learning (ML) and deep learning. Machine Learning (ML) and Deep Learning (DL). Companies see these technologies as an opportunity to improve their operations and meet the challenges of growing energy demand. This project proposes a 3-tier web application using artificial intelligence to analyze and forecast sales within Sonelgaz's sales department in BEJAIA. The LSTM, ARIMA and SARIMA models are used for sales visualization and prediction, and their accuracy is compared with that of the ARIMA model using the MAPE metric. The results show that ARIMA achieves a MAPE of 12.5%, SARIMA achieves a MAPE of 11.8%, and LSTM achieves the best result with a MAPE of 10.2%, demonstrating its effectiveness in sales prediction. The Scrum methodology, an agile approach to development and project management, was adopted for this project. Application tests were carried out using data from some 1,699 customers, covering a 14 year period of monthly electricity consumption in BEJAIA.

Key words : Time series, Agile method, Artificial Intelligence, Machine Learning, Deep Learning, Energy sales analysis and prediction.