



Département : Génie Logistique Et Transport

Mémoire de fin d'étude en vue de l'obtention du diplôme

D'INGÉNIEUR D'ÉTAT

-Filière-

INGÉNIERIE DES TRANSPORTS

-Spécialité -

INGÉNIERIE DE LA CHAÎNE LOGISTIQUE

- Thème -

Optimisation de la planification des opérateurs logistiques et d'allocation des quais dans un entrepôt

Cas Numilog, Bouira

Réalisé par

DOUZI Ilham & MESSOUS Romaiissa

Les membres de Jury :

SELLIDJ Wassila	Présidente
MOULAÏ Ratiba	Promotrice
REZGUI Wail	Co-Promoteur
BOUGHALEB Sadek El Amine	Examinateur
TOUMI Manel	Examinatrice

Alger, le 23/06/2024

Année universitaire 2023 –2024

Dédicace

À ma famille,

Les témoins privilégiés de mon parcours, dans mes succès comme dans mes échecs. Votre présence constante a été une source de réconfort et de soutien, me guidant à chaque étape de ma vie. Vos conseils avisés, votre orientation bienveillante et vos avertissements judicieux ont façonné la personne que je suis aujourd'hui. Grâce à vous, j'ai pu grandir et évoluer.

À mes chères sœurs, Hayet et Imane,

ainsi qu'à mes adorables cousines, Khanessa, Soumia et Narimane,

Je tiens à exprimer toute ma gratitude pour votre soutien inconditionnel et vos encouragements constants tout au long de ce parcours. Votre présence et vos mots réconfortants ont été une source de motivation et de force, me poussant à surmonter chaque défi avec détermination. Je vous dédie ce travail avec tout mon amour et ma reconnaissance.

À mes amies, Melissa, Ilhem, Katia et Selma,

Vous étiez là dans les moments de certitude, apportant votre soutien et vos encouragements. Dans les moments de doute, vous étiez présentes pour me motiver et me remonter le moral. Votre présence bienveillante a été un cadeau inestimable tout au long de ce chemin. Merci d'avoir toujours été là pour moi.

À Yasmine,

Ta présence constante et ton soutien inébranlable ont été une source de force et d'inspiration. Avec toi à mes côtés, j'ai su surmonter chaque défi avec détermination. Je suis profondément reconnaissante pour tout ce que tu as fait pour moi.

Ce projet est le résultat de mon engagement et témoigne des sacrifices et de la persévérance que j'ai investis. Il marque une étape cruciale dans mon développement intellectuel, m'aidant à évoluer vers une meilleure version de moi-même. Je suis profondément reconnaissante envers toutes les personnes qui m'ont soutenue et encouragée tout au long de cette expérience.

Romaïssa...

Dédicace

À mes parents DOUZI Abdennour et ABDELLI Souhila,
pour votre amour inconditionnel, votre soutien sans faille et votre confiance en moi. Vos conseils et vos encouragements ont été et resteront toujours ma force motrice. Vous avez toujours été là pour moi, dans les moments de doute et les moments de joie. Merci pour tous les sacrifices que vous avez faits pour me permettre de poursuivre mes rêves.

À mes chères sœurs, Yousra, Maroua, Malek et Douaa,
pour votre soutien constant et votre affection. Vos encouragements m'ont toujours poussé à aller de l'avant. Votre croyance en mes capacités et vos mots de réconfort m'ont donné la force de persévérer. Merci d'avoir été mes piliers et mes confidentes tout au long de ce parcours.

À mes amies, Sabrina, Sara et Romaïssa,
pour votre aide précieuse, votre amitié sincère et les moments de joie partagés pendant cette aventure académique. Votre soutien inébranlable et votre présence ont rendu ce voyage plus agréable et supportable. Merci pour votre soutien et vos encouragements.

À H.El-Khedim,
Pour ta présence constante et ton soutien inestimable qui ont été une source de réconfort tout au long de ce parcours. Merci d'avoir été là à chaque étape, pour m'encourager et me soutenir.

Ce travail est le fruit de mon parcours, de mes efforts, de mes sacrifices et de ma détermination. Je suis fière du chemin que j'ai parcouru et reconnaissante envers toutes les personnes qui m'ont soutenue et encouragée tout au long de cette aventure.

Ilham...

Remerciements

En tout premier lieu, nous remercions le « Bon Dieu » le tout puissant qui nous a mis sur le bon chemin et nous avoir donné la santé, le courage et la volonté pour accomplir ce modeste travail.

Nous tenons à exprimer notre gratitude la plus sincère envers nos encadrants, **Mme. Moulai Ratiba** et **Mr. Rezgui Wail**, pour leur bonne volonté d'accepter de nous encadrer, pour leurs conseils et leur suivi permanent. Nous leur sommes reconnaissants pour le temps précieux qu'ils nous ont accordé et pour leur soutien à la fois scientifique et moral, qui a été un pilier essentiel pour nous.

Nous souhaitons également exprimer notre gratitude envers l'ensemble du personnel de Numilog qui nous a accueillis et orientés tout au long de notre travail. Nous tenons particulièrement à remercier **Mr. Chebout Adel**, dont l'expertise et les conseils éclairés ont été d'une valeur inestimable. Sa confiance en nos capacités nous a encouragés à repousser nos limites et à atteindre des résultats remarquables.

Nous tenons à remercier sincèrement les membres du jury pour l'honneur qu'ils nous accordent en consacrant leur temps à lire et évaluer notre travail.

Enfin, nous exprimons notre profond respect et notre gratitude sincère envers tous les enseignants du département de Génie Logistique et Transport, qui ont toujours été disponibles et à l'écoute de tous les étudiants.

Table des matières

Dédicace	
Dédicace	
Remerciements	
Liste des figures	II
Liste des tableaux	III
Liste des abréviations	IV
Introduction générale	1
1 Présentation de l'entreprise et problématique	4
1.1 Introduction	4
1.2 L'entreprise Numilog et son environnement	4
1.2.1 Le groupe Cevital	4
1.2.2 Numilog	5
1.2.2.1 L'origine et le développement de Numilog	5
1.2.2.2 L'infrastructure de Numilog	6
1.2.2.3 Les activités de Numilog	6
1.2.2.4 La stratégie de Numilog	6
1.2.2.5 Clients de Numilog	7
1.2.3 La plateforme logistique de Bouira	7
1.2.3.1 Présentation générale	7
1.2.3.2 Activités principales	8
1.2.4 Présentation du cas d'étude	9
1.2.4.1 Système d'information CEVIFOOD	9
1.2.4.2 Les processus d'entreposage	11
1.3 Défis rencontrés chez CEVIFOOD	12
1.4 Conclusion	13
2 Etat de l'art	14
2.1 Introduction	14
2.2 La chaîne logistique et l'entreposage	14
2.2.1 La chaîne logistique	14
2.2.1.1 Les différents types de flux dans une chaîne logistique	14
2.2.1.2 L'externalisation logistique	15

2.2.2	L'entrepôt	15
2.2.2.1	Stratégies de conception d'entrepôt	16
2.2.2.2	Les Piliers de la logistique d'entreposage	16
2.2.2.3	Les processus logistiques dans un entrepôt	17
2.3	L'intelligence artificielle	18
2.4	Machine learning	19
2.4.1	Les différents types de l'apprentissage automatique	20
2.4.1.1	L'apprentissage supervisé :	20
2.4.1.2	L'apprentissage non supervisé :	21
2.4.1.3	L'apprentissage par renforcement :	21
2.5	Les méthodes de prévision	21
2.5.1	Méthodes classiques	22
2.5.1.1	Qu'est-ce qu'une série chronologique ?	22
2.5.1.2	Les composantes d'une série chronologique	22
2.5.1.3	Les modèles de décomposition d'une série chronologique	23
2.5.1.4	Prévision par la méthodologie de Box-Jenkis	23
2.5.1.5	Prévision par Lissage Exponentiel	28
2.5.1.6	Prévision par TBATS	29
2.5.2	Méthodes de prévision à base d'apprentissage automatique	31
2.5.2.1	Régression Linéaire	31
2.5.2.2	Random Forest	34
2.5.2.3	XGBoost	37
2.5.3	Evaluation de la qualité des prévisions	39
2.6	Distance de Minkowski	40
2.7	Conclusion	41
3	Implémentation des méthodes de prévision classiques	42
3.1	Introduction	42
3.2	Les outils	42
3.2.1	Logiciel R	42
3.2.2	Utilisation de Rstudio	43
3.2.3	Package FORECAST	43
3.2.4	Package ZOO	43
3.2.5	Package FTSA	43
3.3	Collecte et description des données	44
3.4	Analyse exploratoire des données (EDA)	44
3.4.1	Visualisation des données	44
3.4.2	Identification des anomalies	45
3.4.3	Analyse des tendances et des saisonnalités	47
3.4.4	Analyse de Modèles : Multiplicatif vs Additif pour les flux d'entrepôt	47
3.5	Modélisation et prévisions	47
3.5.1	Méthode de Box-Jenkins	48
3.5.1.1	La stationnarisation de la série	48
3.5.1.2	Le choix du modèle	48
3.5.1.3	Prévisions des flux et évaluation	49
3.5.2	Méthode de lissage exponentiel	50

3.5.3	Méthode de TBATS	51
3.6	Comparaison des modèles	52
3.7	Conclusion	53
4	Planification des opérateurs logistiques et l'allocation des quais	54
4.1	Introduction	54
4.2	Les outils	54
4.2.1	Python	54
4.2.2	Bibliothèques	55
4.3	Mise en œuvre des prévisions journalières des opérateurs logistiques	56
4.3.1	Collecte des données	56
4.3.2	Analyse et compréhension des données	58
4.3.3	Traitement des données	60
4.3.4	Division de données	62
4.3.5	Sélection du modèle	63
4.3.5.1	Utilisation du modèle Random Forest	64
4.3.5.2	Utilisation de XGBoost	65
4.3.5.3	Utilisation de la Régression Linéaire	66
4.3.6	Discussion des résultats	68
4.4	Optimisation de l'allocation des quais	70
4.4.1	Données relatives à l'affectation des quais	70
4.4.2	Implémentation de la solution	71
4.4.3	Discussion des résultats	73
4.5	Interface de l'utilisateur et fonctionnalités	74
4.6	Conclusion	78
	Conclusion générale	79
	Annexes	81
	Bibliographie	92
	Résumé	97

Table des figures

1.1	Logo Cevital	5
1.2	Logo Numilog	5
1.3	Clients Numilog	7
1.4	Structure de l'entrepôt Numilog Bouira	8
2.1	Les domaines de l'intelligence artificielle	19
2.2	Les types d'apprentissage automatique	20
2.3	Modèles de décomposition d'une série chronologique	23
2.4	Un modèle de régression linéaire simple	32
2.5	Un modèle de régression linéaire multiple	33
2.6	Exemple d'un arbre de décision	35
2.7	Schéma illustratif de principe de fonctionnement de Random Forest	37
2.8	Schéma illustratif de principe de fonctionnement de XGBoost	39
3.1	Logo de logiciel R	43
3.2	Logo de logiciel RStudio	43
3.3	Illustration du flux de marchandises	44
3.4	Les flux de l'entrepôt en fonction du temps	45
3.5	Détection des anomalies par diagramme en boîte	45
3.6	Diagramme en boîte des données après correction des anomalies	46
3.7	Les flux de l'entrepôt après correction des valeurs aberrantes	46
3.8	Analyse de la tendance et de la saisonnalité avec le test de Fisher	47
3.9	Résultat du test de Buys Ballot	47
3.10	Résultats générés par la fonction de minimisation de l'AIC	48
3.11	Les coefficients du modèle SARMA	49
4.1	Logo de Python	55
4.2	Logo de Numpy	55
4.3	Logo de Pandas	55
4.4	Logo de Matplotlib	56
4.5	Logo de Scikit-learn	56
4.6	Structure de la base de données	58
4.7	Description des données	58
4.8	Visualisation des lignes dupliquées	59
4.9	Visualisation des données manquantes	59
4.10	Visualisation des dates invalides	59
4.11	Visualisation des valeurs aberrantes	60
4.12	Visualisation des valeurs manquantes après le traitement de données	60
4.13	Conversion des dates	61

4.14	Transformation logarithmique de données	61
4.15	Normalisation de données	62
4.16	Visualisation des valeurs aberrantes après le traitement de données	62
4.17	test, training and validation sets	63
4.18	Comparaison des valeurs prédites avec les valeurs réelles Cas : Random Forest	64
4.19	Comparaison des valeurs prédites avec les valeurs réelles Cas : XGBoost	66
4.20	Comparaison des valeurs prédites avec les valeurs réelles Cas : Régression Linéaire	67
4.21	Données d'entrée dans les modèles d'apprentissage automatique	68
4.22	Comparaison des schémas de productivité	69
4.23	Structure des emplacements dans les cellules	70
4.24	Extrait d'une commande	71
4.25	Assignation des emplacements au quai le plus proche	72
4.26	Interface principale	75
4.27	Fenêtre secondaire affichée par le premier bouton	75
4.28	Résultat affiché par le premier bouton	76
4.29	Résultat affiché si la suggestion est refusée	76
4.30	Résultat affiché si la suggestion est acceptée	77
4.31	Fenêtre secondaire affichée par le deuxième bouton	77
4.32	Résultat affiché par le deuxième bouton	78
A.1	Implémentation du test de Fisher et de Buys-Ballot	81
A.2	Test de Dickey-Fuller	82

Liste des tableaux

1.1	Les quatres principales activités de Numilog Bouira	9
3.1	Résultat du modèle SARMA	49
3.2	Métrique d'évaluation du modèle SARMA	49
3.3	Résultat du modèle Holt-Winters	50
3.4	Métrique d'évaluation du modèle Holt-Winters	51
3.5	Résultat du modèle TBATS	51
3.6	Métrique d'évaluation du modèle TBATS	52
3.7	Tableau des métriques	52
4.1	Extrait du résultat du modèle Random Forest	64
4.2	Métrique d'évaluation du modèle Random Forest	65
4.3	Extrait du résultat du modèle XGBoost	65
4.4	Métrique d'évaluation du modèle XGBoost	66
4.5	Extrait du résultat du modèle Régression Linéaire	67
4.6	Métrique d'évaluation du modèle Régression Linéaire	67
4.7	Tableau récapitulatif des performances des modèles	69
4.8	Matrice des distances de Manhattan entre quais	72
4.9	Résultats obtenus pour le choix des quais	74
B.1	28 jours de prévisions du modèle SARMA	83
B.2	28 jours de prévisions du modèle Holt-Winters	84
B.3	28 jours de prévisions du modèle TBATS	85
C.1	Résultat du modèle Random Forest	86
C.2	28 jours de prévisions du modèle Random Forest	87
C.3	Résultat du modèle XGBoost	88
C.4	28 jours de prévisions du modèle XGBoost	89
C.5	Résultat du modèle Régression Linéaire	90
C.6	28 jours de prévisions du modèle Régression Linéaire	91

Liste des abréviations

ADM	Agents Administratifs
AIC	Critère d'Information d'Akaike
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
CLR	Centres Logistiques Régionaux
DS	Difference Stationary
FIFO	First In First Out
EDA	Exploratory Data Analysis
GBOOST	Gradient Boosting
IA	Intelligence Artificielle
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
OPL	Opérateurs Logistiques
PDA	Personal Digital Assistant
RMSE	Root Mean Square Error
SARIMA	Seasonal Autoregressive Integrated Moving Average
SARMA	Seasonal Autoregressive Moving Average
SPA	Société par Actions
SSR	Sum of Squared Residuals
TBATS	Trigonometric, Box-Cox transform, ARMA errors, Trend, and Seasonal components
TS	Trend Stationary
WMS	Warehouse Management System

XGBOOST Extreme Gradient Boosting

Introduction générale

La logistique représente un pilier central de l'industrie moderne, essentielle à la fluidité des échanges commerciaux et à la satisfaction des clients. Au cœur de cette chaîne, les entrepôts logistiques jouent un rôle crucial en assurant la réception, le stockage et la distribution efficace des marchandises. Cependant, pour rester compétitifs dans un environnement économique dynamique et exigeant, il est indispensable d'adopter des méthodes avancées visant à optimiser et à améliorer ces centres de distribution.

L'optimisation des entrepôts dépasse largement la simple maximisation de l'espace de stockage ou la réduction des coûts opérationnels. Il s'agit d'un processus stratégique visant à aligner les capacités logistiques sur les exigences du marché tout en garantissant une efficacité opérationnelle optimale. Des entrepôts bien optimisés peuvent significativement réduire les délais de livraison, et améliorer la flexibilité pour répondre aux fluctuations de la demande. De plus, ils permettent de réduire les risques associés aux pertes et aux erreurs.

Un aspect crucial de cette optimisation réside dans la capacité à anticiper et à répondre de manière proactive à la demande future, ce qui nécessite des prévisions précises des flux de marchandises pour ajuster les niveaux de stock et maintenir des délais de livraison compétitifs. De même, prévoir le nombre nécessaire d'opérateurs logistiques permet d'optimiser la répartition des tâches et d'assurer une gestion efficace des ressources humaines, adaptée aux variations saisonnières et aux fluctuations de la demande.

Pour cela, l'utilisation de méthodes de prévision adaptées est cruciale, telles que les méthodes classiques qui analysent méthodiquement les tendances passées pour anticiper les besoins futurs. Ainsi que les méthodes avancées de l'intelligence artificielle qui facilitent la prise de décisions éclairées dans la planification des opérations logistiques grâce à des analyses avancées des données historiques et en temps réel.

Un autre aspect crucial de l'optimisation des entrepôts réside dans l'allocation efficace des quais. Cette pratique revêt une importance capitale car elle influence directement la fluidité des opérations de réception et d'expédition des marchandises. Une gestion stratégique des quais permet de minimiser les distances parcourues pour la mise en stock et la préparation des commandes, réduisant ainsi leur temps de traitement. Cela contribue à la fois à la réduction des coûts opérationnels et à l'amélioration de l'efficacité globale de la chaîne logistique.

Dans cette perspective, notre étude se concentre sur l'entrepôt Numilog, qui joue un rôle central en tant que leader en Algérie dans le domaine de la logistique. L'entreprise

s'engage fermement à maintenir sa compétitivité en répondant précisément aux attentes de ses clients en termes de coûts, de qualité et de délais. Cependant, elle cherche à améliorer sa productivité, notamment en raison de l'indisponibilité des prévisions concernant le nombre d'opérateurs logistiques, ce qui nuit directement à l'efficacité opérationnelle. De plus, l'allocation arbitraire des quais de chargement entraîne des pertes de temps significatives et des inefficiences opérationnelles, nécessitant des distances supplémentaires. Ces inefficiences ont un impact direct sur la productivité de l'entreprise.

C'est dans ce contexte que s'inscrit notre travail de recherche, visant à répondre à la problématique suivante : **comment améliorer la productivité de l'entrepôt Numilog en exploitant de manière optimale les données disponibles ?**

Pour relever ce défis, nous avons développé des outils d'aide à la décision qui abordent trois sous problèmes :

- **Les prévisions de flux de marchandises :** Utilisation des méthodes de prévision classiques pour estimer les volumes journaliers de flux de marchandises.
- **Les prévisions du nombre d'opérateurs journaliers :** Application des algorithmes de machine learning sur les données historiques pour déterminer le nombre nécessaire d'opérateurs logistiques en fonction des prévisions de flux de marchandises.
- **Allocation des quais :** Mise en place d'un algorithme pour optimiser l'attribution des quais lors des opérations de chargement.

L'objectif principal de cette étude est d'explorer l'utilisation de diverses méthodes de prévision et d'adopter des concepts informatiques et mathématiques pour améliorer la productivité de l'entrepôt logistique de Numilog en Algérie. Pour ce faire, notre travail est organisé comme suit :

Chapitre 1 : "Étude de l'existant et problématique"

Ce chapitre vise à présenter l'étude de l'existant réalisée au sein de l'entreprise Numilog, qui constitue notre terrain de stage, et à identifier les défis auxquels elle est confrontée.

Chapitre 2 : "Etat de l'art"

Ce chapitre est dédié à la recherche dans la littérature. Nous avons abordé les notions de base liées au contexte général de notre projet de fin d'étude, notamment la chaîne logistique, plus précisément les entrepôts logistiques. Ensuite, nous avons mentionné des méthodologies et des outils que nous avons utilisés pour résoudre la problématique.

Chapitre 3 : "Implémentation des méthodes de prévision classiques"

Ce chapitre expose les différentes étapes impliquées dans la prévision des flux de marchandises en utilisant plusieurs modèles classiques, suivi d'une évaluation des résultats obtenus.

Chapitre 4 : "Planification des opérateurs logistiques et l'allocation des quais"

Ce chapitre a pour objectif de définir un modèle d'apprentissage automatique permettant de prédire le nombre d'opérateurs logistiques nécessaires en se basant sur les flux de marchandises antérieurs. Ensuite, il traite de la résolution du problème d'allocation des

quais et analyse les résultats obtenus. Enfin, il examine en détail l'intégration de ces résultats dans une interface utilisateur conçue pour relever le défis rencontré par l'entrepôt Numilog.

Enfin, nous résumons en conclusion l'ensemble de notre démarche et les résultats obtenus suivis par des perspectives pour de futurs travaux.

Chapitre 1

Présentation de l'entreprise et problématique

1.1 Introduction

Ce chapitre a pour objectif de présenter l'étude de l'existant effectuée au sein de Numilog. Cette présentation se décline en deux parties. Dans la première, nous introduisons le contexte général dans lequel notre projet de fin d'études est réalisé. Nous commençons par une vue d'ensemble de l'entreprise, incluant son historique, son infrastructure, ses principaux domaines d'activité, ainsi que l'activité spécifique sur laquelle nous nous concentrerons. Dans la deuxième partie, nous discuterons des contraintes auxquelles l'entreprise est confrontée, pour enfin déterminer notre problématique.

1.2 L'entreprise Numilog et son environnement

Avant de parler de Numilog, nous allons donner un bref aperçu de son origine, le groupe Cevital :

1.2.1 Le groupe Cevital

Créé en 1998 avec des fonds privés, Cevital est une entreprise familiale structurée en société par actions (SPA). Elle occupe la première place dans le secteur agroalimentaire en Algérie et se classe troisième en Afrique, avec un chiffre d'affaires dépassant les 4 milliards de dollars et 26 filiales dans divers secteurs, tels que l'agro-industrie, la distribution, l'automobile, la logistique et l'industrie manufacturière.

Cevital Agro-industrie est le leader de la production de sucre et d'huile végétale en Afrique et dans la région méditerranéenne. Depuis quelques années, l'entreprise développe également ses activités à l'international, notamment en Europe (France, Italie, Espagne), en Tunisie, au Maroc et au Brésil[1].



Figure 1.1: Logo Cevital
Source: [1]

1.2.2 Numilog

Dans ce qui suit, nous présentons l'entreprise d'accueil, sa naissance ainsi que son développement.

1.2.2.1 L'origine et le développement de Numilog

Avant la création de Numilog, le groupe Cevital faisait appel à des prestataires logistiques externes pour assurer le transport de ses marchandises. Chaque filiale du groupe disposait de sa propre structure logistique, entraînant des coûts élevés et une pression sur leur trésorerie. La création de Numilog visait principalement à centraliser les services logistiques pour réduire les coûts liés au transport et aux besoins logistiques, améliorant ainsi l'efficacité et la coordination au sein du groupe.

En 2007, Numilog a été fondée par le groupe Cevital pour renforcer le développement de ses activités en assurant un support logistique de qualité. Numilog joue un rôle crucial dans la chaîne logistique en soutenant diverses industries dans la consolidation des flux de marchandises, le développement des projets et l'amélioration des activités. Cela est réalisé grâce à une équipe d'experts en transport et en logistique, ainsi qu'à des infrastructures et des outils modernes, incluant la géolocalisation, la traçabilité et des technologies de pointe [2].

Numilog s'adresse à des clients soucieux de la qualité de service et se définit comme un partenaire durable pour les entreprises. Elle offre une prestation complète de Supply Chain, allant de l'enlèvement à la distribution des produits sur tout le territoire national, en passant par le stockage, avec une capacité de 150 000 palettes dont 35 000 sous température dirigée[3].



Figure 1.2: Logo Numilog
Source: [2]

1.2.2.2 L'infrastructure de Numilog

Le réseau de distribution de Numilog comprend quatre plateformes logistiques situées à l'est, au centre et à l'ouest du pays (Bouira, Oran, Constantine, Sétif), couvrant l'ensemble du territoire national. Il inclut également sept agences de transport (Bouira, Béjaïa, Constantine, Oran, Sétif, Tizi Ouzou, Alger) et 30 centres logistiques régionaux (CLR) répartis stratégiquement pour réduire les délais de livraison. Numilog dispose de 180 000 m² d'espace logistique et d'une flotte de 950 camions, incluant des camions de maraîchage, des plateaux, des porte-conteneurs, des camions frigorifiques et des camions-citernes[4].

1.2.2.3 Les activités de Numilog

Le transport et la logistique sont les deux moteurs de l'activité de Numilog. Ce sont deux métiers intrinsèquement liés.

L'activité de logistique:

Elle repose sur une gestion des flux de marchandises, incluant le stockage, la préparation des commandes et la logistique des produits réfrigérés. Cette approche s'appuie sur une traçabilité complète des mouvements de marchandises, assurant ainsi une visibilité sur les opérations. L'objectif principal est d'optimiser la gestion des stocks pour garantir une disponibilité constante des produits, tout en offrant un service adapté à chaque secteur d'activité.

L'activité de transport:

Elle offre une distribution répondant aux besoins et aux exigences des clients en matière de délais et de services. Cette approche repose sur la conception et la mise en place de nouveaux schémas logistiques, ainsi que sur l'organisation et l'optimisation des itinéraires [5].

Numilog organise ses activités en se concentrant sur trois missions principales [3] :

- Accompagner la croissance des activités du groupe Cevital au travers de prestations logistiques et de transport.
- Proposer aux acteurs économiques et industriels en Algérie des prestations de transport et/ou logistiques à travers tout le territoire.
- Proposer un accompagnement en conseil et solutions logistiques.

1.2.2.4 La stratégie de Numilog

Numilog oriente ses efforts en interne tout en explorant des opportunités pour une expansion à l'international.

La stratégie locale:

Numilog occupe une position importante dans le secteur de la logistique et du transport en Algérie, en offrant des services aux clients internes du Groupe Cevital ainsi qu'à

l'ensemble du marché. L'entreprise met l'accent sur l'optimisation de sa logistique opérationnelle pour contrôler les coûts, garantir un niveau de service satisfaisant et améliorer ses performances. De plus, elle accorde une attention particulière à la flexibilité et à la réactivité pour répondre aux attentes des clients dans un environnement concurrentiel.

La stratégie internationale

Numilog a développé de nouvelles ressources logistiques et de transport spécialement adaptées à ses activités internationales d'exportation. En France, la filiale CTLOG International a été établie par Numilog. Au Maroc, l'entreprise a loué un entrepôt de 5000 mètres carrés à Casablanca pour soutenir le développement de sa filiale, Brandt, sur le marché marocain. Numilog a également élargi son réseau de distribution en entrant sur le marché tunisien, et envisage également de s'implanter prochainement en Italie[6].

1.2.2.5 Clients de Numilog

Numilog s'appuie sur un portefeuille diversifié de clients, englobant une vaste gamme d'entreprises, allant des petites et moyennes entreprises aux grandes corporations. Cette diversité clientèle reflète l'engagement de Numilog à répondre aux besoins de divers secteurs d'activité, qu'il s'agisse de transport ou de logistique.



Figure 1.3: Clients Numilog

Source: Document interne de l'entreprise Numilog

1.2.3 La plateforme logistique de Bouira

Nous présentons dans cette partie le lieu de notre stage, la plateforme de Bouira.

1.2.3.1 Présentation générale

La plateforme logistique de Bouira est l'une des quatre plateformes logistiques exploitées par Numilog. La surface totale de stockage de la plateforme s'élève à 69 000 m², comprenant

15 cellules, dont 9 dédiées au stockage sec et 6 dédiées au stockage à température contrôlée. De plus, la plateforme dispose de 93 quais avec une capacité de stockage de 62 500 palettes [4].

La plateforme comporte 6 blocs administratifs :

- B0 - Bloc administratif dédié aux fonctions de gestion et de support.
- B1 - Bloc technique et maintenance.
- B2 - Administration de CEVIFOOD.
- B3 - Administration de l'UNODIS.
- B4 - Agence de transport.
- B5 - Administration de DANONE et SOBCO (Palmary).



Figure 1.4: Structure de l'entrepôt Numilog Bouira
Source: Image satellite de Google Maps

1.2.3.2 Activités principales

La plateforme de Bouira exerce deux activités principales, à savoir la logistique et le transport.

L'activité de logistique:

Dans son activité logistique, la plateforme de Bouira gère quatre activités principales :

Activité	Description
CEVIFOOD	Il constitue l'activité la plus importante en termes de flux physiques et informationnels. En outre, il joue le rôle d'une plateforme de distribution des produits Cevital, renforçant ainsi les liens avec la société mère Cevital.
UNODIS	Il est spécialement élaboré pour stocker les produits des clients d'Unodis, regroupant environ 30 000 à 40 000 articles, incluant des produits électroménagers, des boissons, des cosmétiques et d'autres catégories.
DANONE	Cet espace est spécialement conçu pour stocker les produits du client Danone tels que les biscuits céréales, Dannette, Activia, Oikos et autres. Il s'agit d'une activité récente qui utilise des surfaces à température régulée.
SOBCO	Il est dédié au stockage des produits Palmary, tels que les pâtes à tartiner, les biscuits Maxon et les biscuits Kool qui occupent une place de premier plan sur le marché des biscuits mosaïques. Cette marque possède un fort potentiel de croissance et l'entrepôt joue un rôle essentiel dans son développement.

Tableau 1.1: Les quatre principales activités de Numilog Bouira
Source: **Élaboration personnelle.**

L'activité de transport:

Disposant d'une flotte exploitable de 79 camions et de 118 chauffeurs, elle réalise le transport de marchandises pour les clients internes au groupe Cevital : CeviFood, Danone, Unodis, Sobco.

1.2.4 Présentation du cas d'étude

Au sein des multiples activités pris en charge par NUMILOG, notre étude se concentre spécifiquement sur CEVIFOOD. Cette entité joue un rôle central dans les activités de l'entreprise, étant la principale pierre angulaire de ses opérations à la plate-forme de Bouira. En raison de sa complexité et de sa diversité en matière de processus et de gestion des produits, CEVIFOOD constitue un sujet d'étude extrêmement enrichissant pour analyser les pratiques logistiques de l'entreprise.

1.2.4.1 Système d'information CEVIFOOD

Le système d'information de CEVIFOOD représente un élément vital de son infrastructure opérationnelle. En effet, il englobe non seulement les ressources humaines, mais également

les équipements matériels nécessaires à la manipulation et à l'informatisation des processus logistiques.

Les moyens humaines:

Au niveau opérationnel, il y a 28 opérateurs logistiques (OPL) répartis en trois équipes de deux niveaux différents :

- Les OPL de niveau 2 sont chargés de la manutention.
- Les OPL de niveau 3 sont chargés de l'ensemble des opérations logistiques.

Les moyens matériels :

Pour assurer une manipulation aisée et garantir le bon état des produits de l'entrepôt, plusieurs équipements sont mobilisés, parmi lesquels :

- Les palettes : Ces palettes en bois sont standardisées avec des dimensions de 800*1200 mm et 1000*1200 mm.
- Les chariots incluant :
 - Les chariots élévateurs, utilisés pour le stockage en rack.
 - Les chariots électriques, utilisés pour le chargement et le déchargement des camions.
 - Les chariots à pince, utilisés pour le stockage de masse.
- Les transpalettes à fourche: utilisés pour la préparation des commandes, l'accumulation, le chargement et le déchargement.

En parallèle avec le matériel de manutention, CEVIFOOD s'appuie sur une infrastructure informatique sophistiquée pour gérer efficacement ses opérations logistiques. Cela inclut notamment deux piliers essentiels : le système de gestion d'entrepôt (WMS) et les appareils mobiles de type PDA (Personal Digital Assistant).

- WMS : Le logiciel WMS est un outil interne visant à automatiser et optimiser les processus logistiques des entrepôts. Il offre de nombreuses fonctionnalités telles que la gestion des réceptions, des expéditions, des stocks et des préparations. En intégrant divers périphériques informatiques, ce système permet au personnel de l'entrepôt de bénéficier d'une meilleure gestion des stocks, de simplifier la préparation des commandes et d'assurer une meilleure traçabilité des marchandises.

- PDA : Le PDA est un ordinateur mobile pour l'entrepôt, existant sous forme d'un appareil seul ou embarqué sur le chariot élévateur. Il assiste les OPL dans la gestion des marchandises, permettant à l'opérateur de suivre les instructions insérées par le chef d'équipe pour effectuer les opérations et valider chaque étape. De plus, le PDA est connecté à l'appareil d'impression Zebra pour générer des étiquettes de support lors de la réception des marchandises.

1.2.4.2 Les processus d'entreposage

L'activité de CEVIFOOD se résume en trois processus :

Processus de réception :

Le processus de réception chez CEVIFOOD est parfaitement intégré via l'interface WMS, depuis l'arrivée des véhicules de fret jusqu'à l'enregistrement des marchandises dans le système informatique WMS et leur mise en stock. Généralement, le processus global de réception est divisé en plusieurs étapes, notamment :

- Présentation du conducteur à l'administration : Les documents de transport de la marchandise (bon de transfert) sont remis par le conducteur aux agents administratifs (ADM).
- Contrôle des documents de livraison : Les documents sont minutieusement examinés pour garantir leur exactitude et exhaustivité
- Affectation du chauffeur vers le quai de déchargement : un ordre de déchargement est remis au chauffeur pour le diriger vers le quai de déchargement approprié.
- Déchargement et vérification des marchandises : Les marchandises sont déchargées du véhicule par des OPL et vérifiées simultanément pour s'assurer de leur conformité en termes de qualité et de quantité.
- Validation de la réception : À l'aide du PDA, l'OPL valide la réception. Les produits validés sont ensuite intégrés dans WMS.

Note : Si la marchandise est non conforme, elle est placée dans la zone d'avarie. Cette zone est réservée pour stocker les marchandises endommagées, défectueuses ou non conforme.

Processus de stockage :

La mise en stock physique de la marchandise implique leur déplacement depuis la zone de réception et de contrôle vers des emplacements de stockage prédéfinis, tout en veillant à ce que les conditions de stockage soient appropriées. Cette opération nécessite une manipulation précise et efficace des produits, réalisée à l'aide de moyens de manutention appropriés tels que des transpalettes et des chariots élévateurs, ainsi qu'une répartition

stratégique dans les différentes zones de stockage selon la méthode ABC et le principe FIFO.

Processus de préparation de commande et d'expédition :

Cette phase nécessite une gestion minutieuse pour respecter les délais de livraison. Voici les étapes détaillées :

- Enregistrement de la commande : Les commandes sont intégrées automatiquement via une interface de communication WMS.
- Lancement des opérations de préparation : Le chef d'équipe logistique commence les opérations de préparation en se basant sur l'ordre de préparation de commande donnée par les ADM et attribue les tâches de préparation aux préparateurs, en tenant compte de l'ordre de priorité de chargement.
- Préparation physique : La préparation physique de la marchandise est lancée par les OPL qui commencent à prélever les articles en les validant à l'aide du PDA. Les palettes sont ensuite transportées vers la zone de préparation et chargées dans un quai approprié, conformément à l'ordre de chargement donné par les ADM. Cependant, l'affectation des quais se fait souvent de manière arbitraire, sans l'application d'un système structuré pour cette tâche.
- Clôture et contrôle de la préparation : Le chef d'équipe logistique assure la clôture et le contrôle de la préparation en vérifiant l'achèvement des tâches de préparation des commandes. Ensuite, l'OPL vérifie la conformité des produits préparés en termes de quantité et de qualité.
- Validation des marchandises : Une fois cette vérification effectuée, l'OPL valide la préparation à l'aide du PDA, s'assurant ainsi que les palettes sont prêtes pour l'expédition.
- Expédition et transport : La marchandise est chargée dans le camion, accompagnée d'un bon de livraison et une facture, remis au chauffeur pour la documentation de la cargaison avant son transport.

1.3 Défis rencontrés chez CEVIFOOD

Durant notre stage chez Numilog, nous avons relevé plusieurs problématiques liées à la gestion d'entrepôt. Parmi eux, nous avons constaté des difficultés significatives liées à la planification des opérateurs logistiques et à la sélection des quais pour les opérations de chargement et de déchargement. Ainsi, nous avons observé :

- Une absence de planification basée sur les données historiques : Cette lacune entraîne un déséquilibre dans les jours de travail, où certains jours peuvent être surchargés en personnel par rapport à la charge de travail, tandis que d'autres jours peuvent manquer de personnel nécessaire. Ce déséquilibre affecte directement la productivité globale de l'entrepôt. Les jours surchargés provoquent une inefficacité due à une coordination difficile et à des temps d'attente accrus, tandis que les jours sous-chargés ne permettent pas d'optimiser l'utilisation des ressources humaines.
- Une allocation aléatoire des quais de chargement : Cette pratique peut entraîner des retards importants dans les opérations de chargement, surtout lorsque les quais assignés sont éloignés des zones de prélèvement. Le chariot devra alors parcourir une distance supplémentaire, allongeant ainsi le temps de déplacement et impactant le temps nécessaire pour la préparation des commandes. Ces inefficacités dans l'organisation des quais se traduisent par des délais plus longs, affectant négativement la productivité de l'entrepôt.

1.4 Conclusion

Dans ce chapitre, nous avons présenté l'entreprise Numilog en mettant un accent particulier sur CEVIFOOD. Nous avons examiné son système d'information ainsi que son processus d'entreposage. Face au volume important de marchandises que CEVIFOOD traite quotidiennement, et en analysant de près le fonctionnement de ses mécanismes, nous avons identifié plusieurs défis qui influent sur sa performance logistique.

Afin de relever ces défis et de répondre de manière efficace à notre problématique, il est impératif d'adopter des outils de prévision avancés et de mettre en place des processus d'affectation des quais optimisés. Cette approche stratégique permettra à CEVIFOOD d'augmenter significativement sa productivité tout en répondant aux besoins des clients de manière plus rapide et efficiente.

Dans le prochain chapitre, nous aborderons les concepts et outils utilisés afin d'étudier le problème en profondeur et d'obtenir une solution qui améliore la productivité d'un entrepôt.

Chapitre 2

Etat de l'art

2.1 Introduction

Ce chapitre se concentre sur une revue approfondie de la littérature, structurée autour de trois axes principaux. Nous commençons par une analyse approfondie de la littérature portant sur les entrepôts, mettant en lumière les zones d'entrepôt, les piliers de la logistique d'entreposage, ainsi que les processus logistiques internes. Ensuite, nous nous pencherons sur la littérature concernant l'intelligence artificielle, en particulier le machine learning, en nous concentrant sur les types de modèles de régression. Cette approche découle de notre intérêt à développer des prévisions précises, en utilisant une variété de méthodes de prévision, qu'elles soient basées sur l'IA ou sur des approches plus classiques. Nous citerons également des méthodes pour évaluer la qualité des prévisions. Enfin, nous examinerons la distance de Minkowski, en nous appuyant sur la littérature existante pour explorer différents modèles de distance.

2.2 La chaîne logistique et l'entreposage

2.2.1 La chaîne logistique

La chaîne logistique, ou supply chain en anglais, englobe l'ensemble des étapes nécessaires pour satisfaire une demande client ; de la production à la livraison. Elle vise à intégrer les moyens internes et externes afin de répondre efficacement aux besoins des clients ; tout en réduisant les coûts, assurant un niveau de service élevé et raccourcissant les délais pour améliorer la compétitivité de l'entreprise.

2.2.1.1 Les différents types de flux dans une chaîne logistique

Les fondements de la réussite opérationnelle d'une entreprise reposent sur la synchronisation harmonieuse et la fluidité de ses flux ; ces flux sont essentiels à son bon fonctionnement en reliant chaque étape du processus. Ils représentent l'essence même de son dynamisme ; ces

flux peuvent être regroupés en trois catégories : les flux physiques, les flux d'informations et les flux monétaires.

- **Les flux physiques** : ils englobent toutes les entités matérielles qui circulent dans la chaîne logistique de l'amont vers l'aval ou de l'aval vers l'amont ; notamment les flux de matière première, d'encours et de produits finis.
- **Les flux d'information** : ils représentent les échanges d'informations entre les divers acteurs de la chaîne logistique. Ils permettent de transmettre la demande des clients, de surveiller les niveaux de stocks et de suivre l'avancement des travaux en cours ; cette fluidité dans la communication entre les intervenants favorise une prise de décision plus efficace.
- **Les flux financiers** : ils désignent les flux de trésorerie liés aux flux physiques. Ces flux parcourent principalement la chaîne d'aval en amont ; mais certaines opérations financières peuvent également se dérouler en sens inverse, comme les remboursements ou les paiements en cas de conflit.

2.2.1.2 L'externalisation logistique

L'externalisation, souvent désignée sous le terme Outsourcing, est le fait de confier la gestion d'une ou plusieurs activités à un fournisseur ou à un prestataire externe, plutôt que de les traiter en interne. Ce concept est crucial dans le secteur de la logistique, car il permet aux entreprises de se concentrer sur leurs connaissances clés tout en déléguant les tâches non essentielles à des spécialistes externes[7].

Le recours à l'externalisation est influencé par plusieurs facteurs. En premier lieu, la grande expertise des prestataires permet aux entreprises de répondre aux attentes élevées de leurs clients en termes de rapidité de livraison et de qualité de service ; renforçant ainsi leur compétitivité sur le marché. En deuxième lieu, l'externalisation permet aux entreprises de réduire leurs investissements dans les actifs physiques et les coûts de main-d'œuvre ; ce qui favorise une allocation plus judicieuse des budgets.

2.2.2 L'entrepôt

La notion d'entrepôt logistique est examinée sous divers angles dans la littérature :

En tant que composante de la chaîne logistique :

Un entrepôt est un bâtiment spécialement conçu pour le stockage de marchandises en vue de leur expédition ; qu'il s'agisse de clients internes ou externes à l'entreprise, en gros ou en détail. Il peut également servir à entreposer des produits en cours de fabrication. La propriété et la gestion de l'entrepôt peuvent être assumées directement par l'entreprise ou déléguées à un prestataire logistique externe [8].

En tant que centres logistiques :

Les entrepôts sont un élément important dans la distribution des marchandises et ne sont plus considérés comme un simple lieu de stockage ; ils deviennent plutôt des endroits dynamiques où les produits sont emballés et les commandes sont regroupées et préparées [9].

En tant que prestataire de service :

L'entrepôt peut être envisagé sous trois angles différents : les processus, les ressources et l'organisation. Les produits qui arrivent dans un entrepôt passent ensuite par plusieurs étapes appelées processus. Les ressources englobent tous les moyens, équipements et personnel nécessaires au fonctionnement d'un entrepôt. Enfin, l'organisation englobe toutes les procédures de planification et de contrôle utilisées pour son bon fonctionnement [9].

2.2.2.1 Stratégies de conception d'entrepôt

La conception d'un entrepôt implique une série de décisions coordonnées, prises à trois niveaux hiérarchiques : stratégique, tactique et opérationnel. Ces décisions visent à atteindre divers critères de performance prédéfinis [10].

Au niveau stratégique, les décisions ont un impact à long terme et portent principalement sur le choix de l'emplacement de l'entrepôt en fonction de la proximité avec les fournisseurs et les clients, ainsi que les dimensions de l'entrepôt pour anticiper les futurs besoins de stockage. Nous investissons également dans des technologies avancées telles que les systèmes de gestion d'entrepôt et les robots de stockage pour augmenter l'efficacité.

Au niveau tactique, nous planifions à moyen terme pour optimiser l'efficacité de l'entrepôt. En se basant sur les résultats des décisions stratégiques. Ces décisions impliquent généralement la disposition interne pour minimiser les déplacements et maximiser l'efficacité des opérations. De plus, nous ajustons les effectifs en fonction des prévisions de charge de travail anticipée.

Au niveau opérationnel, l'accent est mis sur la gestion quotidienne de l'entrepôt. Les principales décisions opérationnelles concernent le contrôle des ressources humaines et des équipements, ainsi que la maintenance et la sécurité de l'entrepôt pour maintenir un environnement de travail sûr et fonctionnel.

En ce qui suit, nous allons aborder la conception du point de vue tactique.

2.2.2.2 Les Piliers de la logistique d'entreposage

Dans un entrepôt, deux piliers essentiels sont à considérer :

Ressources humaines :

Selon Antoine de Saint-Exupéry : « Une organisation aussi parfaite soit-elle ne vaut que par la qualité des hommes qui la servent ».

Une quantification nette de l'ensemble du personnel opérationnel, administratif et d'encadrement est un élément très important pour un dimensionnement pertinent d'un entrepôt. Cette démarche vise à éviter à la fois la surcharge de travail et le surplus d'effectifs ; tout en permettant une optimisation des coûts liés à la main-d'œuvre. En

outre, cette approche favorise une meilleure productivité et efficacité en assignant les employés aux tâches appropriées selon les besoins du moment.

Ressources de manutention :

L'efficacité des engins de manutention est un élément clé dans le bon fonctionnement d'un entrepôt. Ils comprennent une variété de modèles disponibles, tels que le chariot élévateur, les transpalettes, etc., qui jouent un rôle essentiel dans le transport fluide des produits entre les différentes zones de l'entrepôt ; notamment le déchargement, le stockage, le chargement et la préparation de commande. L'un des principaux avantages de ces engins de manutention est leur capacité à remplacer le travail manuel ; permettant ainsi d'économiser un temps précieux et de réduire les risques de blessures liées à la manipulation manuelle de charges lourdes ou volumineuses. Ils contribuent ainsi à améliorer la productivité et la rentabilité des entrepôts, tout en offrant un environnement de travail plus sûr et plus confortable pour les employés.

2.2.2.3 Les processus logistiques dans un entrepôt

L'objectif de l'entrepôt est d'assurer le meilleur service possible en livrant en temps et en heure les produits demandés par le client. Ceci est réalisé à travers quatre points clés :

Le processus de réception :

Ce processus consiste à recevoir les marchandises entrant dans l'entrepôt et implique la réalisation de toutes les opérations nécessaires ; le contrôle quantitatif et qualitatif, l'enregistrement des informations pertinentes, ainsi que leur déchargement et leur placement dans l'entrepôt. L'objectif est d'identifier les procédures qui permettent de maximiser la productivité en réduisant les temps d'attente[11].

Le stockage :

Une fois les marchandises reçues, elles doivent être placées dans des emplacements adaptés à l'intérieur de l'entrepôt. Cela implique l'utilisation de divers systèmes de stockage, tels que les palettiers, qui sont des étagères métalliques conçues pour stocker des palettes. Chaque produit est étiqueté et adressé de manière précise afin de garantir une gestion optimale du stock.

L'adressage logistique détermine la manière d'organiser les produits ainsi que leur emplacement au sein de l'entrepôt. Il doit être défini de façon simple, efficace et logique afin d'assurer le bon fonctionnement des aires de stockage. Il existe deux grandes catégories d'adressage :

- **Fixe ou spécifique :** une position est préalablement affectée à chaque article entreposé. Cela facilite la localisation des produits et ne nécessite pas l'intervention d'un logiciel informatique.
- **Banalisé :** les références sont placées dans les espaces disponibles au fur et à mesure qu'elles sont reçues. Il est essentiel d'utiliser un logiciel de gestion d'entrepôt pour appliquer cette méthode.

Dans notre cas, le système de stockage utilisé est le dédié. Cela signifie que chaque produit est affecté à une zone spécifique de l'entrepôt en fonction de :

- La rangée : l'ensemble des travées consécutives le long d'une allée de circulation [12].
- La travée : l'ensemble des alvéoles superposées comprises entre deux échelles consécutives, qui sont les structures verticales supportant les barres horizontales d'un palettier [12].
- Le niveau : la hauteur spécifique dans la travée où le produit est placé [12].
- L'emplacement : la position exacte dans l'alvéole où le produit se trouve [12].

L'objectif est d'optimiser l'utilisation de l'espace dans l'entrepôt en adaptant une structure de stockage aux produits [13].

La préparation de commande :

Il s'agit de rassembler les produits spécifiques requis pour chaque commande client. Cela peut impliquer la localisation des articles dans l'entrepôt, leur prélèvement et leur regroupement dans un seul emplacement en vue de leur expédition. L'objectif est de garantir que la marchandise soit préparée et emballée à temps pour son expédition, tout en maximisant la productivité des préparateurs et en assurant la fiabilité du processus de préparation.

L'expédition :

Une fois les commandes préparées, les produits sont emballés, étiquetés et prêts à être expédiés aux clients. Cette étape peut impliquer la coordination des services de transport, la génération de documents d'expédition appropriés et l'organisation nécessaire pour optimiser les plans de tournées. L'objectif est de réduire les coûts de transport tout en garantissant des livraisons rapides et efficaces [11].

Après avoir présenté les principaux principes d'un entrepôt, nous allons explorer dans les sections suivantes des méthodes éprouvées pour l'optimiser. Cela inclut les techniques traditionnelles de prévision des flux de marchandises, les algorithmes d'intelligence artificielle pour la prévision du nombre d'OPL journalier, ainsi que la méthode de calcul des distances pour la sélection des quais.

2.3 L'intelligence artificielle

Au début du 19e siècle, Mary Shelley a souligné dans ses écrits que l'intelligence artificielle était un thème significatif de la science-fiction ; ce qui a contribué à populariser le sujet [14]. Par la suite, l'intelligence artificielle est devenue un domaine de recherche

scientifique, en particulier avec le développement des machines à calculer programmables dans les années 1950. Cette période a marqué les premiers questionnements sur la capacité des machines à penser [15].

En 1956, le terme Intelligence Artificielle (IA) a été utilisé pour la première fois lors d'une proposition pour un atelier de recherche estival à Dartmouth College. L'objectif initial de cet atelier était de créer des machines capables de simuler certains processus d'apprentissage de l'intelligence humaine ; de s'en inspirer et de les reproduire [16].

L'intelligence artificielle vise donc à imiter l'intelligence humaine à travers l'utilisation d'algorithmes exécutés dans un environnement informatique dynamique. Ce domaine englobe plusieurs sous-ensembles, tels que l'apprentissage automatique (machine learning) et l'apprentissage profond (deep learning), qui se concentrent sur le développement d'algorithmes et de modèles capables d'apprendre ; de faire des prédictions ou de prendre des décisions basées sur des données [17].

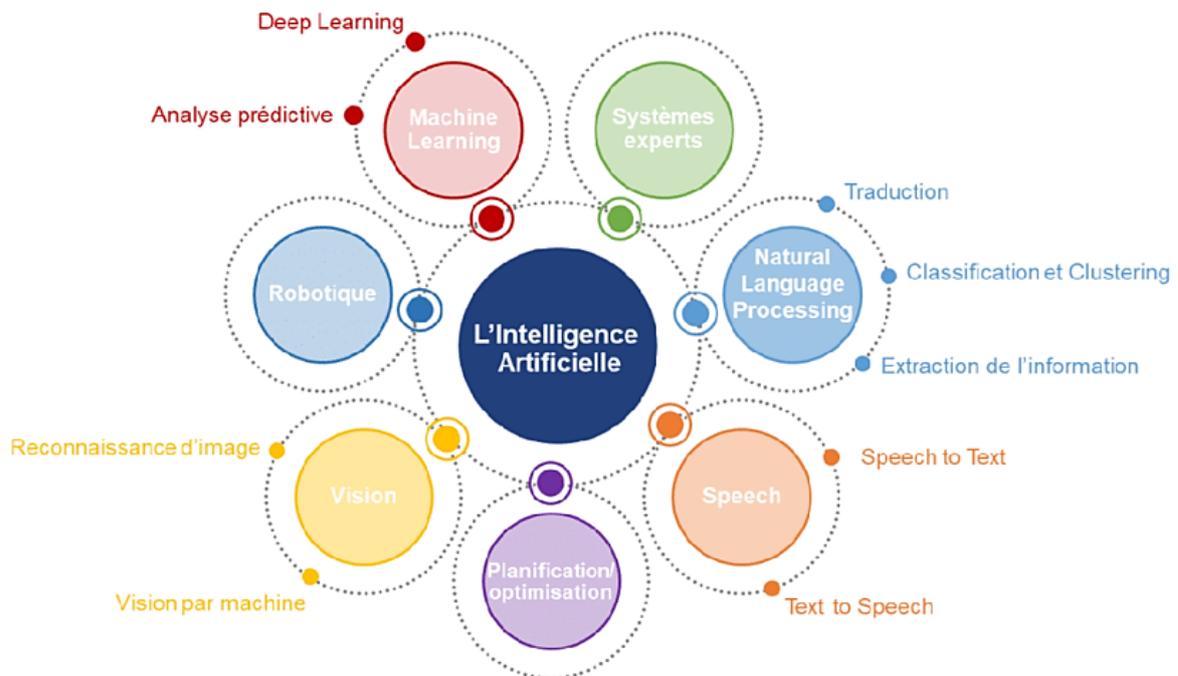


Figure 2.1: Les domaines de l'intelligence artificielle

Source: [18]

Dans notre étude, nous allons examiner et mettre en œuvre des méthodes d'apprentissage automatique. Cela implique l'utilisation de techniques et d'algorithmes spécialisés afin de résoudre efficacement notre problème de recherche.

2.4 Machine learning

Dans la littérature, l'apprentissage automatique est une subdivision de l'intelligence artificielle qui combine des notions statistiques, mathématiques et des techniques informatiques évoluées pour imiter l'intelligence humaine en assimilant des connaissances de l'environnement.

Ce domaine est considéré comme le moteur principal de la révolution des Big Data, et son utilisation s'étend à divers secteurs, notamment la reconnaissance des formes, la vision par ordinateur, la gestion d'entrepôt, ainsi que les applications biomédicales et médicales [19].

2.4.1 Les différents types de l'apprentissage automatique

Dans le vaste domaine du machine learning, une multitude d'approches sont exploitées pour analyser les données et résoudre des problèmes spécifiques. Parmi celles-ci se distinguent l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement. Chaque méthode propose différents modèles adaptés à des défis particuliers. Afin de simplifier cette diversité et d'en faciliter la compréhension, nous présentons une figure synthétique résumant les principes fondamentaux de chaque approche avec des explications succinctes ci-dessous.

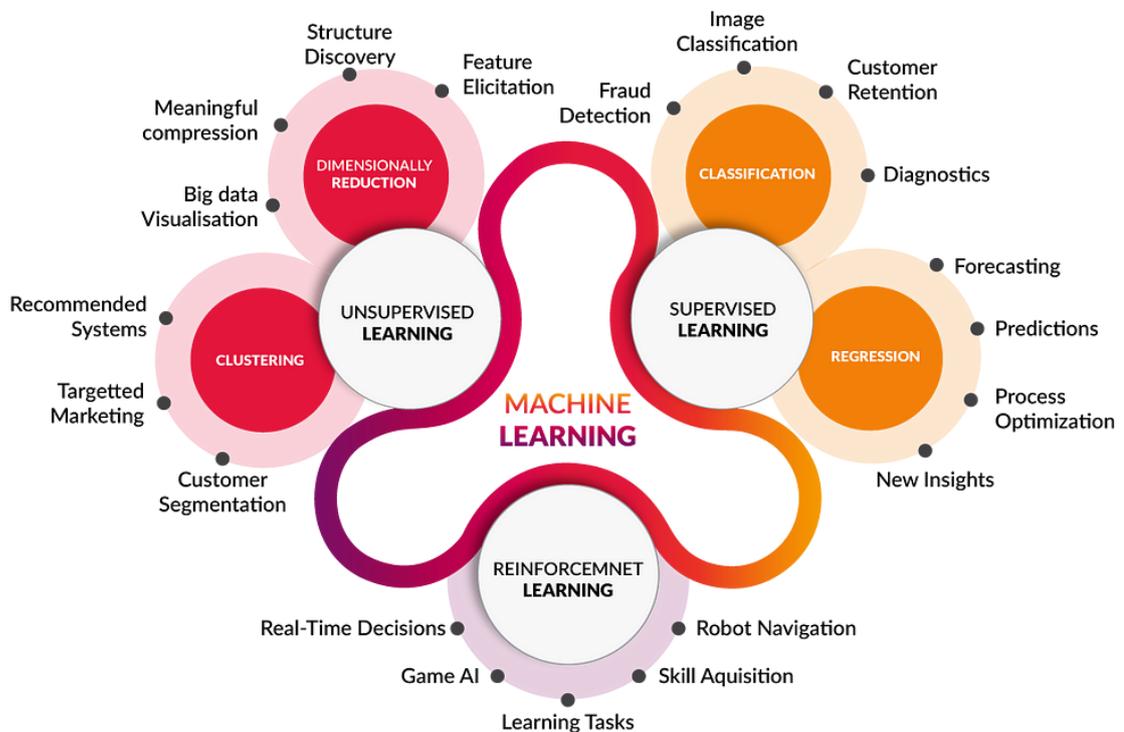


Figure 2.2: Les types d'apprentissage automatique
Source: [20]

2.4.1.1 L'apprentissage supervisé :

Dans l'apprentissage supervisé, un algorithme apprend à partir d'un ensemble de données étiquetées, où chaque exemple est associé à une sortie connue. Ces exemples sont utilisés pour construire des modèles prédictifs capables de généraliser et de faire des prévisions précises sur de nouvelles données non étiquetées [21].

Dans le domaine de l'apprentissage supervisé, on distingue deux principaux types d'algorithmes :

- **Les algorithmes de régression** : visent à anticiper des valeurs continues en modélisant la relation entre une variable dépendante et une ou plusieurs variables indépendantes. Leur objectif est de prédire ou d'estimer la valeur de cette variable dépendante
- **Les algorithmes de classification** : ils cherchent à prédire une classe/catégorie, ce qui implique la prédiction d'une variable discrète.

2.4.1.2 L'apprentissage non supervisé :

L'apprentissage non supervisé consiste généralement à trouver une structure cachée dans des collections de données non étiquetées [22]. Dans l'apprentissage non supervisé, le modèle doit découvrir les motifs et les structures des données par lui-même. En l'absence de sortie correcte et de superviseur pour guider le processus, l'algorithme apprend à partir des données d'entrée et regroupe les données en fonction de leurs similitudes[23] .

2.4.1.3 L'apprentissage par renforcement :

L'apprentissage par renforcement vise à créer des algorithmes capables d'apprendre et de s'adapter de manière autonome aux changements de l'environnement en se basant sur les expériences d'essais et erreurs. Dans ce processus, les bonnes décisions entraîneront une prime tandis que les mauvaises entraîneront une pénalité, dans le but ultime de maximiser les performances et d'optimiser les résultats. Ces caractéristiques d'auto-amélioration et d'apprentissage en ligne font de l'apprentissage par renforcement l'une des technologies de base des agents intelligents [24].

2.5 Les méthodes de prévision

La prévision joue un rôle crucial dans la chaîne logistique en établissant le fondement des plans opérationnels. Une prévision précise permet d'optimiser les niveaux de stock en anticipant les demandes futures et en allouant efficacement les opérateurs disponibles. Cependant, la préparation d'un plan de prévision est une tâche ardue pour les entreprises dans de nombreux secteurs, car les prévisions sont très souvent influencées par plusieurs facteurs, tels que le besoin des clients, la concurrence et les délais de livraison.

Dans ce contexte complexe, l'analyste doit prendre en compte les comportements des consommateurs, les effets météorologiques, les tendances du marché et l'évolution des prévisions. Il est essentiel de choisir les meilleures méthodes adaptées aux données disponibles pour obtenir des prévisions fiables et précises, tout en évitant une complexité excessive des outils de prévision.

Face à ce problème, de nombreuses solutions existent dans la littérature pour différents domaines d'application. Dans les travaux de (McCarthy et al., 2006), une enquête a été réalisée auprès des professionnels sur les techniques et les pratiques employées. Ils ont regroupé les techniques en deux niveaux : les techniques qualitatives sont principalement utilisées pour la prévision à moyen ou long terme, elles utilisent des données subjectives basées sur l'analyse des avis d'experts ou de clients. D'un autre côté, les techniques quantitatives s'appuient sur des données numériques. Parmi elles, on distingue deux grandes familles de méthodes de prévision : les méthodes de séries chronologiques et les méthodes causales. Les méthodes de séries chronologiques se basent sur une suite d'observations dans le temps prises à intervalles réguliers, permettant de prévoir la demande en fonction des données historiques. En revanche, les méthodes causales établissent des relations de cause à effet entre certaines variables de l'environnement (variables indépendantes) et la demande (variable dépendante), permettant ainsi de faire des prévisions associatives [25].

Dans les sections suivantes, nous mettrons en lumière trois méthodes classiques de prévision : la méthode de Box-Jenkins, le lissage exponentiel et la méthode TBATS. Ensuite, nous nous pencherons sur les méthodes de prévision basées sur l'intelligence artificielle, notamment la régression linéaire, Extreme Gradient Boosting (XGBoost) ainsi que les forêts aléatoires.

2.5.1 Méthodes classiques

La problématique de prévisions peut être traitée par une approche de modélisation des séries chronologiques. Ces techniques sont plus classiques en comparaison avec les méthodes d'apprentissage automatique, mais elles sont toujours très utilisées dans plusieurs domaines.

2.5.1.1 Qu'est-ce qu'une série chronologique ?

Une série chronologique (ou temporelle) est un processus stochastique. Elle représente une suite d'observations d'une grandeur aléatoire liée à un phénomène, habituellement observées à différentes dates, c'est-à-dire une suite de valeurs numériques indexées par le temps [26].

2.5.1.2 Les composantes d'une série chronologique

Une série chronologique peut être décomposée en trois principales composantes : la tendance, la saisonnalité et les résidus.

La tendance : notée f_t , désigne un mouvement de longue durée, continu et lent (à la hausse ou à la baisse), qui représente l'évolution fondamentale d'un phénomène [27].

La saisonnalité : notée S_t , correspond à un phénomène qui se répète à intervalles de temps approximativement réguliers à l'intérieur d'une année, d'un mois, d'une semaine, etc.

La composante résiduelle (ou bruit ou résidu) : notée ϵ_t , est aléatoire. Elle représente tout ce qui n'a pas été expliqué par la tendance et la saisonnalité. Elle n'est pas toujours visible sur le graphique sauf après élimination des composantes précédentes. Les variations résiduelles sont la composante des séries chronologiques qui nous intéressent le plus lorsque nous essayons de trouver le bon modèle pour prévoir la série [27].

2.5.1.3 Les modèles de décomposition d'une série chronologique

Les modèles suivants indiquent comment combiner les différentes composantes d'une série chronologique :

Modèle additif

Ce modèle stipule que la série est la juxtaposition additive des différentes composantes. Il prend la forme générale suivant [28] :

$$X_t = f_t + S_t + \epsilon_t \quad (2.1)$$

Modèle multiplicatif

Ce modèle prend la forme générale suivante [28]:

$$X_t = f_t \times S_t \times \epsilon_t \quad (2.2)$$

Il est à noter que le modèle additif traduit un mouvement saisonnier d'amplitude constante, tandis que le modèle multiplicatif traduit un mouvement saisonnier d'amplitude proportionnelle au niveau de la série.

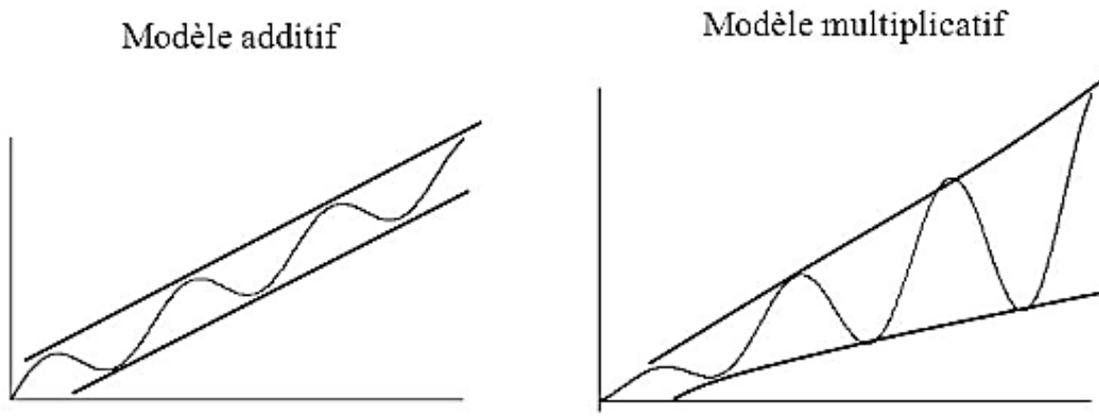


Figure 2.3: Modèles de décomposition d'une série chronologique
Source: [29]

2.5.1.4 Prévision par la méthodologie de Box-Jenkins

La méthodologie de Box-Jenkins est une approche itérative pour sélectionner et ajuster un modèle approprié pour la prévision des séries chronologiques. Nous supposons dans

cette procédure que nous avons préalablement supprimé la saisonnalité de la série (que nous allons bien sur réintégrer au niveau des prévisions), c'est-à-dire que la série n'est affectée éventuellement que d'une tendance. Ainsi, les étapes principales de la méthodologie de Box et Jenkins sont les suivantes :

a) **La stationnarisation de la série :**

Avant de recourir à la méthodologie de Box-Jenkins, il est essentiel de vérifier la stationnarité de la série chronologique en utilisant un test de stationnarité, le plus connu étant le test de Dickey-Fuller. Ce test permet de déterminer la présence d'une racine unitaire et de déterminer le type de non stationnarité en termes de tendance.

Il est important de distinguer deux types de non-stationnarité : DS (Difference Stationary) et TS (Trend Stationary).

- **DS (Difference Stationary) :** Si la série présente une racine unitaire dans sa partie autoregressive, elle est considérée comme non stationnaire de type DS. Pour rendre cette série stationnaire, il est nécessaire d'appliquer des différences successives. Cela signifie qu'il faut appliquer le filtre $(1 - B)^d$, où d représente le nombre de différenciations nécessaires.
- **TS (Trend Stationary) :** Si la série ne présente pas de racine unitaire mais contient une tendance déterministe, par exemple une tendance linéaire, alors elle est de type TS. Dans ce cas, il suffit de soustraire cette tendance de la série initiale pour la rendre stationnaire.

b) **Le choix du Modèle de prévision :**

Cette étape consiste à choisir le modèle de prévision le plus adéquat pour nos données. Il existe plusieurs modèles stationnaires et non stationnaires que nous allons présenter brièvement :

Modèles stationnaires :

- **Modèle autorégressif à moyenne mobile ARMA(p,q) :** Ce modèle est utilisé pour analyser et prévoir les séries chronologiques. Il combine deux composants :
 - **Terme autorégressif (AR) :**
 - * **Description :** Ce terme décrit la relation entre la variable à modéliser et ses valeurs précédentes.
 - * **Paramètre :** p indique le nombre de retards utilisés dans le modèle pour capturer cette relation.
 - * **Formule :**

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t \quad (2.3)$$

Tel que:

- X_t représente la série chronologique à l'instant t .
- $\phi_1, \phi_2, \dots, \phi_p$ sont les coefficients des termes autorégressifs.

- ϵ_t représente le bruit blanc à l'instant t .
 - **Terme moyenne mobile (MA) :**
 - * **Description :** Ce terme décrit la relation linéaire entre les résidus de la série.
 - * **Paramètre :** q indique le nombre de retards utilisés dans le modèle de moyenne mobile.
 - * **Formule :**

$$X_t = \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (2.4)$$
- Telque:
- $\theta_1, \theta_2, \dots, \theta_q$ sont les coefficients des termes de moyenne mobile.

La formule totale du modèle ARMA(p,q) devient :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (2.5)$$

- **Modèle Autorégressif et de Moyenne Mobile Saisonniers (p,0,q)(P,0,Q)m:**
Le modèle SARMA est une extension des modèles ARMA qui prend en compte la saisonnalité dans les séries chronologiques [30]. Il est composé des éléments suivants :
 - **Composantes tendance :**
 - * p : L'ordre de l'auto régression de la tendance.
 - * q : L'ordre de la moyenne mobile de la tendance.
 - **Composantes saisonnalité :**
 - * P : L'ordre de l'auto régression saisonnière.
 - * Q : L'ordre de la moyenne mobile saisonnière.
 - * m : Le nombre de pas de temps dans une période saisonnière unique.
 - * 0 : Indique l'absence de termes de différenciation saisonnière.

La formule totale du modèle SARMA(p,0,q)(P,0,Q) m devient :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \phi'_1 X_{t-m} + \phi'_2 X_{t-2m} + \dots + \phi'_P X_{t-mP} + \theta'_1 \epsilon_{t-m} + \theta'_2 \epsilon_{t-2m} + \dots + \theta'_Q \epsilon_{t-mQ} \quad (2.6)$$

Telque:

- $\phi'_1, \phi'_2, \dots, \phi'_P$ sont les coefficients des termes autorégressifs saisonniers.
- $\theta'_1, \theta'_2, \dots, \theta'_Q$ sont les coefficients des termes moyennes mobiles saisonniers.

Modèles non stationnaires :

- **Modèle autorégressif à moyenne mobile intégrée ARIMA(p,d,q) :** Ce modèle est une extension des modèles ARMA qui intègre une composante de différenciation pour rendre la série stationnaire. Voici comment il se différencie des modèles ARMA :

– **Terme de différenciation (I) :**

* **Description :** Ce terme est utilisé pour rendre la série temporelle stationnaire en différenciant les observations.

* **Paramètre d :** Indique le nombre de différenciations nécessaires pour rendre la série stationnaire.

* **Formule :**

$$X'_t = (1 - B)^d X_t \quad (2.7)$$

Telque:

· X'_t représente la série différenciée.

· B représente l'opérateur de décalage (retard).

· X_t représente la série chronologique originale.

La formule totale du modèle ARIMA(p,d,q) devient :

$$\Phi(B)(1 - B)^d X_t = \Theta(B)\epsilon_t \quad (2.8)$$

Telque:

– $\Phi(B)$ représente le polynôme autorégressif de l'ordre p :

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2.9)$$

– $\Theta(B)$ représente le polynôme de moyenne mobile de l'ordre q :

$$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \quad (2.10)$$

- **Modèle ARIMA Saisonnier SARIMA (p,d,q)(P,D,Q)m :** Ce modèle est une extension des modèles SARMA qui intègre une composante de différenciation pour rendre la série stationnaire, tout en prenant en compte la saisonnalité. Voici comment il se différencie des modèles SARMA :

– **Terme de différenciation (I) :**

* **Description :** Ce terme est utilisé pour rendre la série temporelle stationnaire en différenciant les observations.

* **Paramètre d :** Indique le nombre de différenciations nécessaires pour rendre la série stationnaire en termes de tendance.

* **Formule :**

$$X'_t = (1 - B)^d X_t \quad (2.11)$$

– **Terme de différenciation saisonnière (S) :**

* **Description :** Ce terme est utilisé pour rendre la série temporelle saisonnière stationnaire en différenciant les observations saisonnières.

* **Paramètre D :** Indique le nombre de différenciations saisonnières nécessaires pour rendre la série stationnaire en termes de saisonnalité.

* **Formule :**

$$X''_t = (1 - B^m)^D X_t \quad (2.12)$$

Tel que :

- X_t'' représente la série chronologique différenciée deux fois de manière saisonnière.
- X_t représente la série chronologique originale.
- m représente la périodicité saisonnière.

La formule totale du modèle SARIMA(p, d, q)(P, D, Q) m devient :

$$\Phi_p(B)\Phi_P(B^m)(1 - B^m)^D(1 - B)^d X_t = \Theta_q(B)\Theta_Q(B^m)\epsilon_t \quad (2.13)$$

Tel que :

- * $\Phi_P(B^m)$ représente le polynôme autorégressif saisonnier:

$$\Phi_P(B^m) = 1 - \phi'_1 B^m - \phi'_2 B^{2m} - \dots - \phi'_P B^{Pm} \quad (2.14)$$

- * $\Theta_Q(B^m)$ est le polynôme de moyenne mobile saisonnière :

$$\Theta_Q(B^m) = 1 + \theta'_1 B^m + \theta'_2 B^{2m} + \dots + \theta'_Q B^{Qm} \quad (2.15)$$

c) L'estimation des paramètres du modèle :

Une fois le modèle sélectionné, les paramètres du modèle sont estimés à l'aide d'algorithmes, tels que l'estimation par maximum de vraisemblance ou l'estimation par moindres carrés non linéaires.

d) Tester la validité du modèle :

la validation de la pertinence du modèle se fait en examinant les résidus. Différents tests sont alors utilisés, notamment :

- Le test d'absence d'autocorrélation des résidus : Il vérifie que les résidus ne présentent pas de autocorrélation significative.
- Le test d'homoscédasticité : Il évalue si la variance des résidus reste constante à travers le temps, ce qui est une hypothèse fondamentale des modèles linéaires.
- Le test de normalité : Il examine si les résidus suivent une distribution normale, ce qui est souvent une condition pour les inférences statistiques valides.

e) Les prévisions :

Une fois que le modèle est validé, il peut être utilisé pour générer des prévisions futures basées sur les données historiques. Il est à noter que la formule des prévisions reste identique pour toutes les formules du modèle en ajoutant un chapeau sur les valeurs prédites. Par exemple, la formule de prévision pour un modèle ARMA(p, q) pourrait être exprimée comme suit :

$$\hat{X}_T(h) = \hat{\phi}_1 \hat{X}_{T+h-1} + \hat{\phi}_2 \hat{X}_{T+h-2} + \dots + \hat{\phi}_p \hat{X}_{T+h-p} + \hat{\theta}_1 \hat{\epsilon}_{T+h-1} + \hat{\theta}_2 \hat{\epsilon}_{T+h-2} + \dots + \hat{\theta}_q \hat{\epsilon}_{T+h-q} \quad (2.16)$$

$$\hat{\epsilon}_{T+h-j} = \begin{cases} 0 & \text{si } h > j \\ \epsilon_{T+h-j} & \text{si } h \leq j \end{cases} \quad (2.17)$$

Notons $\hat{X}_T(h)$ la prévision à l'horizon h .

2.5.1.5 Prévision par Lissage Exponentiel

Introduite par Holt en 1958 puis par Winters en 1960, les méthodes de lissage constituent un ensemble de techniques empiriques de prévision qui accordent plus ou moins d'importance aux valeurs passées d'une série temporelle

Méthode de Holt-Winters

La méthode de Holt et Winters est une approche privilégiée pour la prévision des séries chronologiques qui présentent simultanément une tendance et une saisonnalité. Elle repose sur l'estimation de trois composantes fondamentales : le niveau (a_t), qui représente la moyenne des observations dans la série, la tendance (b_t) et la saisonnalité (s_t). Cette méthode propose des approches différentes, chacune adaptée à des types spécifiques de séries chronologiques : une pour les séries avec décomposition multiplicative et une pour les décompositions additives.

- **Lissage Exponentiel de Holt-Winters Additif :**

Ce modèle est approprié lorsque l'amplitude de la saisonnalité est constante au fil du temps[31]. La prévision est de la forme :

$$\hat{y}_{t+h} = \hat{a}_t + h\hat{b}_t + \hat{s}_{t-p+h} \quad (2.18)$$

où :

- \hat{a}_t représente le niveau ajusté à l'instant t :

$$\hat{a}_t = \alpha(\hat{y}_t - \hat{s}_{t-p}) + (1 - \alpha)(\hat{a}_{t-1} + \hat{b}_{t-1}) \quad (2.19)$$

- \hat{b}_t représente la tendance à l'instant t :

$$\hat{b}_t = \beta(\hat{a}_t - \hat{a}_{t-1}) + (1 - \beta)\hat{b}_{t-1} \quad (2.20)$$

- \hat{s}_t représente la composante saisonnière à l'instant t :

$$\hat{s}_t = \gamma(\hat{y}_t - \hat{a}_t) + (1 - \gamma)\hat{s}_{t-p} \quad (2.21)$$

- α , β , et γ sont les coefficients de lissage pour le niveau, la tendance et la saisonnalité respectivement.

- p représente la période saisonnière.

- h représente l'horizon de prédiction.

- **Lissage Exponentiel de Holt-Winters Multiplicatif :**

Ce modèle est approprié lorsque l'amplitude de la saisonnalité varie proportionnellement au niveau de la série[31]. La prévision est de la forme :

$$\hat{y}_{t+h} = (\hat{a}_t + h\hat{b}_t) \times \hat{s}_{t-p+h} \quad (2.22)$$

où les composantes sont définies comme suit :

$$\hat{a}_t = \alpha(\hat{y}_t / \hat{s}_{t-p}) + (1 - \alpha)(\hat{a}_{t-1} + \hat{b}_{t-1}) \quad (2.23)$$

$$\hat{b}_t = \beta(\hat{a}_t - \hat{a}_{t-1}) + (1 - \beta)\hat{b}_{t-1} \quad (2.24)$$

$$\hat{s}_t = \gamma \begin{pmatrix} \hat{y}_t \\ \hat{a}_t \end{pmatrix} + (1 - \gamma)\hat{s}_{t-p} \quad (2.25)$$

Il est à noter qu'il existe une multitude d'autres méthodes de lissage exponentiel, telles que le lissage exponentiel simple pour les séries chronologiques qui ne présentent ni tendance ni saisonnalité, et le lissage exponentiel double pour les séries chronologiques présentant une tendance linéaire mais pas de saisonnalité significative.

Cependant, le modèle traditionnel de lissage exponentiel ne peut pas décrire les modèles saisonniers multiples, les modèles saisonniers non entiers, ni les séries temporelles complexes avec des effets calendaires doubles. Pour répondre à ces limitations et mieux accommoder les séries temporelles multi-saisonnières, le modèle TBATS a été proposé [32].

2.5.1.6 Prévision par TBATS

TBATS (Trigonometric Seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal components), est une méthode de prévision permettant de modéliser les données de séries temporelles, dont l'objectif principal est de prévoir les séries temporelles avec des modèles saisonniers complexes en utilisant le lissage exponentiel.

Ce modèle utilise une combinaison de termes de Fourier avec un modèle d'espace d'état de lissage exponentiel et une transformation de Box-Cox, de manière complètement automatisée [33].

TBATS présente les équations suivantes :

Transformation de Box-Cox :

La transformation de Box-Cox est une technique statistique utilisée pour stabiliser la variance des données de la série temporelle. Elle est particulièrement utile lorsque la variance des données change avec le temps, ce qui est courant dans de nombreuses séries chronologiques. La transformation de Box-Cox peut être ajustée automatiquement pour trouver le paramètre optimal qui stabilise la variance des données.

On définit une transformation de Box-Cox avec le paramètre ω comme suit [34][36] :

$$Y_t^{(\omega)} = \begin{cases} \frac{y_t^\omega - 1}{\omega} & \text{si } \omega \neq 0 \\ \log(y_t) & \text{si } \omega = 0 \end{cases} \quad (2.26)$$

Modèle saisonnier :

Le modèle saisonnier est donné par [35] :

$$Y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-m_i}^{(i)} + d_t \quad (2.27)$$

Tendances globales et locales :

Ces composantes modélisent la partie non saisonnière de la série temporelle en utilisant des équations de lissage qui mettent à jour dynamiquement le niveau et la tendance [36].

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t \quad (2.28)$$

$$b_t = \phi b_{t-1} + \beta d_t \quad (2.29)$$

Tel que :

- α est le paramètre de lissage pour le niveau.
- β est le paramètre de lissage pour la tendance.
- ϕ est le paramètre d'amortissement de la tendance.
- d_t représente l'erreur à l'instant t .

Composante Saisonnière :

Cette composante ajuste la valeur actuelle de la saisonnalité en fonction de sa valeur passée [36].

$$s_t^{(i)} = s_{t-m_i}^{(i)} + \gamma_i d_t \quad (2.30)$$

Tel que :

- $s_{t-m_i}^{(i)}$: Valeur de la composante saisonnière à l'instant $t - m_i$ ajustée en fonction du décalage saisonnier m_i .
- γ_i : Paramètre de lissage.
- d_t : Erreur à l'instant t .

L'erreur ARMA :

La composante ARMA modélise les erreurs résiduelles de la série temporelle à l'aide de termes auto-régressifs et de moyenne mobile, en capturant les dépendances temporelles et les fluctuations non saisonnières[36].

$$d_t = \sum_{i=1}^p \phi_i d_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t \quad (2.31)$$

Tel que :

- ϕ_i : Coefficients auto-régressifs.
- θ_i : Coefficients de moyenne mobile.
- ϵ_t : Bruit blanc à l'instant t .

Modèle de lissage trigonométrique exponentiel pour les données saisonnières: Il vise à capturer les motifs saisonniers complexes présents dans une série temporelle[36].

$$s_t^{(i)} = \sum_{j=1}^{k_i} a_{j,t}^{(i)} \cos(\lambda_j^{(i)} t) \quad (2.32)$$

$$a_{j,t}^{(i)} = a_{j,t-1}^{(i)} + k1^{(i)} d_t \quad (2.33)$$

$$\beta_{j,t}^{(i)} = \beta_{j,t-1}^{(i)} + k2^{(i)} d_t \quad (2.34)$$

Où $k1^{(i)}$ et $k2^{(i)}$ sont les paramètres de lissage, et $\lambda_j^{(i)} = \frac{2\pi j}{m_i}$ est la fréquence angulaire pour la saisonnalité i .

En intégrant ces composantes, le modèle TBATS s'adapte efficacement à une variété de données complexes et dynamiques, offrant ainsi une capacité de prévision robuste et flexible pour les séries temporelles.

2.5.2 Méthodes de prévision à base d'apprentissage automatique

L'apprentissage automatique offre une approche plus flexible pour la prévision de séries temporelles. Les modèles d'apprentissage automatique sont capables de capturer des modèles non linéaires et de modéliser des relations complexes entre les différentes variables.

Dans notre approche, nous avons délibérément mis l'accent sur les méthodes de régression, en cohérence avec nos objectifs spécifiques. Cette décision découle de la nature de nos données et de notre intention de privilégier les modèles de régression pour obtenir des prévisions précises et fiables. Ainsi, nous présenterons dans cette partie les algorithmes de machine learning les plus adaptés à notre contexte.

2.5.2.1 Régression Linéaire

La régression linéaire est une méthode statistique fondamentale en machine learning utilisée pour modéliser et analyser les relations entre une variable dépendante et une ou plusieurs variables explicatives. Elle permet de prédire la valeur de la variable dépendante en fonction des valeurs des variables explicatives. La régression linéaire peut être divisée en deux types principaux : la régression linéaire simple et la régression linéaire multiple.

Régression linéaire simple

La régression linéaire simple modélise la relation entre une seule variable explicative et une variable dépendante. Cette relation est représentée par une droite dans un plan 2D. Le modèle se réduit à la meilleure droite qui correspond aux données, déterminée de manière à minimiser l'erreur de prévision [37][38].

L'équation de la droite de régression linéaire simple est :

$$y = ax + b \quad (2.35)$$

où :

- y est la variable dépendante.
- x est la variable explicative.
- a est le coefficient de régression (pente de la droite).
- b est l'ordonnée à l'origine (intercept).

La droite est déterminée du sorte que la somme des carrés des erreurs de prédiction est minimisée.

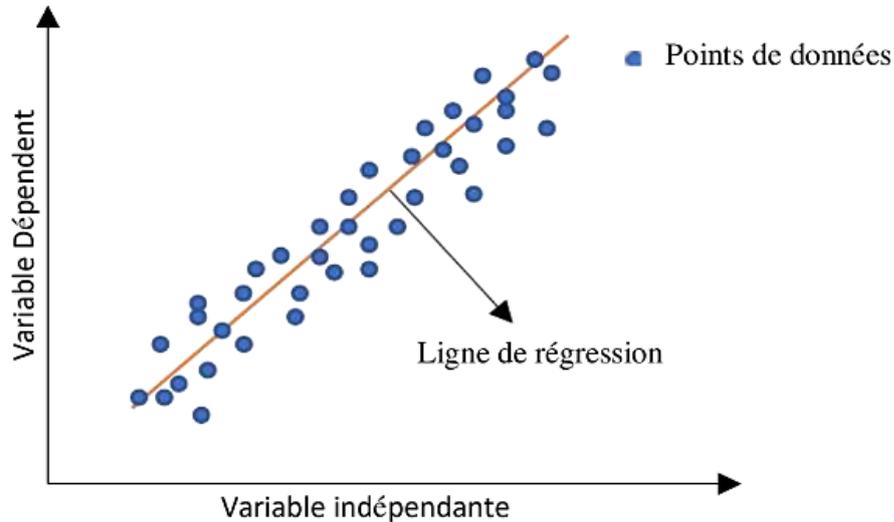


Figure 2.4: Un modèle de régression linéaire simple
Source: [38]

Régression linéaire multiple

Lorsque plusieurs variables explicatives sont présentes, on utilise la régression linéaire multiple pour modéliser la relation entre la variable dépendante y et les variables explicatives x_1, x_2, \dots, x_p . Ce modèle permet de prendre en compte l'influence de plusieurs variables sur la variable dépendante [37].

L'équation de la régression linéaire multiple avec deux variables explicatives est :

$$y = a_1x_1 + a_2x_2 + b \quad (2.36)$$

Le modèle peut être étendu à p variables explicatives :

$$y = a_1x_1 + a_2x_2 + \dots + a_px_p + b \quad (2.37)$$

Dans ce cas, le modèle de régression linéaire multiple peut être visualisé sous la forme d'un hyperplan dans un espace à $p + 1$ dimensions (où p est le nombre de variables explicatives). Par exemple, avec deux variables explicatives, le résultat est un plan de régression dans un espace 3D. Avec plus de deux variables explicatives, nous avons un hyperplan dans un espace de dimensions supérieures, ce qui est difficile à visualiser concrètement mais repose sur le même principe mathématique [37].

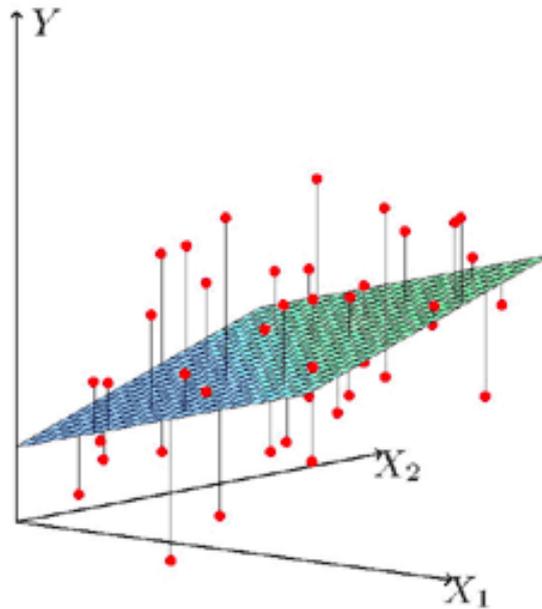


Figure 2.5: Un modèle de régression linéaire multiple

Source: [39]

Pour déterminer les valeurs des paramètres a_1, a_2, \dots, a_p et b , diverses méthodes d'estimation peuvent être utilisées. Parmi celles-ci, la descente de gradient est une technique couramment employée. Elle trouve les coefficients optimaux du modèle en mettant à jour les coefficients à chaque itération de manière à réduire la somme des erreurs entre le modèle ajusté et les valeurs réelles des données d'apprentissage. Par plusieurs itérations, elle trouve le minimum local en se déplaçant dans le sens du gradient négatif[38]. L'algorithme est le suivant:

Algorithme 1 : Régression Linéaire

Entrées :

- X : Matrice des features de dimension $m \times n$ (m exemples, n features)
- y : Vecteur des valeurs cibles de dimension $m \times 1$
- α : Taux d'apprentissage
- max_iter: Nombre maximum d'itérations

Sorties :

- Coefficients finaux du modèle

- 1 **Debut :**
 - 2 Initialiser les coefficients du modèle $\theta = (a_1, a_2, \dots, a_n, b)$ à des valeurs aléatoires ;
 - 3 Définir le modèle linéaire $F(X) = X\theta$;
 - 4 **pour** i de 1 à max_iter **faire**
 - 5 Calculer les prédictions: $y_{\text{pred}} = X\theta$;
 - 6 Calculer l'erreur: $erreur = y_{\text{pred}} - y$;
 - 7 Calculer le gradient: $gradient = \frac{1}{m} X^T erreur$;
 - 8 Mettre à jour les coefficients du modèle: $\theta = \theta - \alpha \times gradient$;
 - 9 **fin**
 - 10 **Fin**
-

Quelle est la différence entre la régression linéaire dans la statistique et dans machine learning ?

La différence principale réside dans l'approche de modélisation et d'estimation des paramètres. La régression linéaire statique suit une approche plus traditionnelle et rigide, tandis que le machine learning offre une plus grande flexibilité et peut capturer des relations non linéaires entre les variables.

2.5.2.2 Random Forest

Dans le domaine de l'apprentissage automatique, les forêts aléatoires (Random Forests) sont largement reconnues pour leur efficacité et leur robustesse. Avant d'explorer cette méthode avancée, il est crucial de comprendre les fondements sur lesquels elle repose, à savoir les arbres de décision.

Les arbres de décision sont des modèles d'apprentissage supervisé utilisés pour la classification et la régression. Ils fonctionnent en établissant une séquence de tests sur les valeurs d'un ensemble de variables, appelées attributs. À partir des résultats de ces tests, le modèle génère une valeur numérique ou choisi parmi un ensemble discret de conclusions. L'objectif principal des arbres de décision est de diviser les données en groupes homogènes en fonction de la variable cible à prédire[40]. Pour ce faire, ils segmentent les données en sous-groupes homogènes basés sur les caractéristiques. Chaque nœud de l'arbre représente une caractéristique, chaque branche est une décision basée sur cette caractéristique, et chaque feuille est une prévision de la variable cible.

La segmentation des données est guidée par une mesure d'erreur, la somme des carrés des résidus (SSR), qui quantifie la différence entre les valeurs observées et prédites par le modèle pour chaque observation[40].

Cette mesure est définie par la formule:

$$SSR = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.38)$$

Tel que :

- y_i représente la valeur observée de la variable dépendante pour la i -ème observation.
- \hat{y}_i représente la valeur prédite par le modèle pour la i -ème observation.
- n est le nombre total d'observations dans l'échantillon.

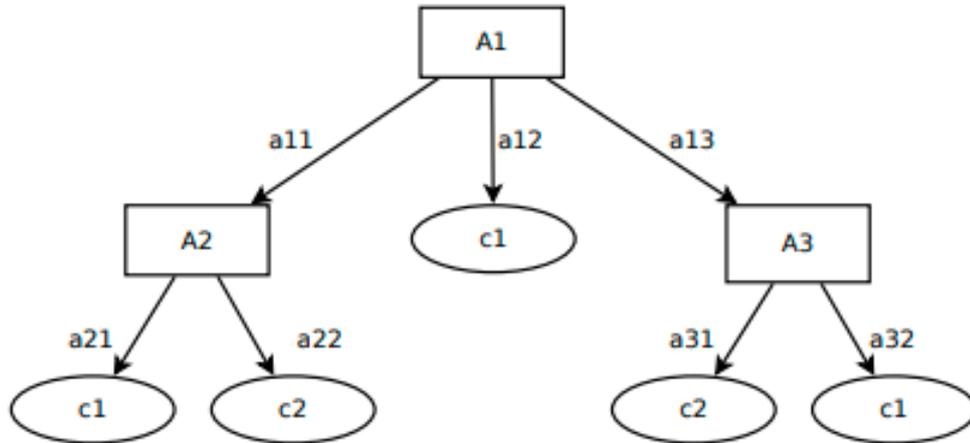


Figure 2.6: Exemple d'un arbre de décision

Source:[41]

Ces modèles sont populaires en raison de leur simplicité et de leur capacité à être visualisés facilement, ce qui les rend interprétables. Cependant, ils sont susceptibles de surapprentissage, ce qui signifie qu'ils peuvent s'ajuster trop étroitement aux données d'entraînement et perdre en capacité de généralisation.

Pour remédier à cette limitation, Leo Breiman a introduit en 2001 les forêts aléatoires, une technique qui améliore les performances des arbres de décision en les combinant dans un ensemble. Une forêt aléatoire est constituée de nombreux arbres de décision, chacun étant construit sur un échantillon bootstrap des données d'entraînement et utilisant un sous-ensemble aléatoire des caractéristiques pour chaque division de nœud. Cette approche introduit deux sources de variabilité : l'échantillonnage aléatoire des données (bagging) et la sélection aléatoire des caractéristiques[42]. Cette diversité réduit la variance du modèle et améliore sa robustesse globale.

Les prévisions des forêts aléatoires sont obtenues en agrégeant les résultats des arbres individuels, soit par vote majoritaire pour la classification, soit par moyenne pour la régression. En plus de leur robustesse, les forêts aléatoires sont particulièrement appréciées pour leur capacité à gérer les données manquantes et à réduire les biais, tout en évitant le surapprentissage grâce à leur construction diversifiée. L'algorithme est le suivant:

Algorithme 2 : Random Forest

Entrées :

X : Matrice des features de dimension $m \times n$ (m exemples, n features)

y : Vecteur des valeurs cibles de dimension $m \times 1$

B : Nombre d'arbres dans le modèle

min_node_size : Taille minimale des nœuds

p : Nombre de variables prédictives

Sorties :

Ensemble d'arbres

```
1 Debut :
2 pour  $i = 1$  à  $B$  faire
3   Tirer un échantillon bootstrap de taille  $m$  à partir des données
   d'entraînement;
4   Construire un arbre de décision  $T_i$  avec l'échantillon bootstrap;
5   tant que  $\text{node size} \neq \text{min\_node\_size}$  faire
6     Sélectionner aléatoirement un sous-ensemble de  $p$  variables prédictives
     parmi les  $n$  totales ;
7     pour  $j = 1$  à  $p$  faire
8       Évaluer le critère de division pour le  $j$ -ème prédicteur;
9       si le  $j$ -ème prédicteur optimise le critère de division alors
10        Diviser le nœud interne en deux nœuds enfants;
11        Arrêter la sélection pour ce nœud;
12      fin
13    fin
14  fin
15 fin
16 Fin
```

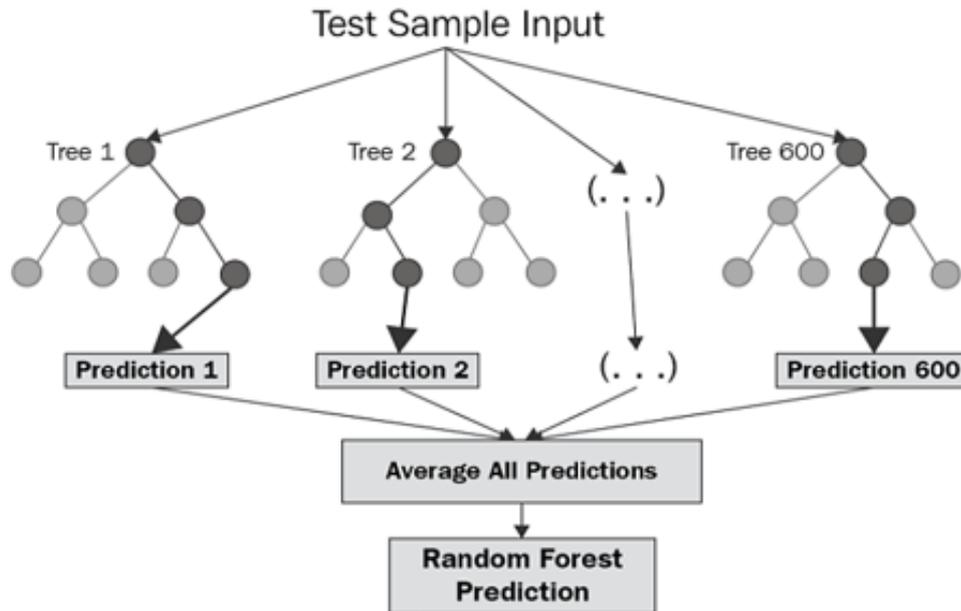


Figure 2.7: Schéma illustratif de principe de fonctionnement de Random Forest
Source: [43]

2.5.2.3 XGBoost

Le développement de XGBoost a été précédé par Gradient Boosting (GBoost), une méthode d'apprentissage ensembliste qui construit séquentiellement plusieurs modèles faibles pour améliorer les prédictions en ajustant itérativement les résidus des prévisions précédentes. Contrairement au Random Forest, où chaque arbre est construit de manière indépendante, GBoost construit chaque arbre en tenant compte des erreurs des arbres précédents. Cela permet à chaque nouvel arbre de se concentrer sur les erreurs restantes et de corriger les faiblesses des modèles précédents.

Gradient Boosting a été largement salué pour sa capacité à capturer des modèles complexes et a été utilisé avec succès dans de nombreux domaines tels que la finance, la biostatistique, et le marketing prédictif.

Cependant, l'introduction de XGBoost en 2016 a marqué une avancée significative dans cette approche. XGBoost a été conçu pour être hautement performant et capable de gérer efficacement de grands ensembles de données complexes grâce à plusieurs améliorations et fonctionnalités avancées [44], notamment :

- **Amélioration de la vitesse d'exécution :** Utilisation de techniques de parallélisation des calculs et une exploitation efficace des ressources matérielles (CPU et mémoire).
- **Gestion automatique des valeurs manquantes :** Permet aux arbres de décision de trouver les meilleures divisions même en présence de données incomplètes.
- **Pruning (élagage) des arbres :** Suppression des nœuds des arbres de décision qui n'améliorent pas la classification des feuilles, permettant de construire des modèles plus simples et plus robustes.

- **Intégration de termes de régularisation** : Pénalisation de la complexité des modèles pour éviter le surapprentissage.

Algorithme 3 : XGBoost

Entrées :

X : Matrice des features de dimension $m \times n$ (m exemples, n features)
 y : Vecteur des valeurs cibles de dimension $m \times 1$
 M : Nombre d'arbres
 η : Taux d'apprentissage
 λ et γ : Termes de régularisation

Sorties :

Modèle final

1 Debut :

2 pour $m = 1$ *to* M **faire**

3 Calculer les résidus $r_i^{(m)} = -\frac{\partial L(y_i, \hat{y}_i^{(m-1)})}{\partial \hat{y}_i^{(m-1)}}$;

4 Construire un nouvel arbre T_m des résidus;

5 **pour** *chaque feuille* j *de l'arbre* T_m **faire**

6 Calculer la somme des gradients $G_j = \sum_{i \in R_j} r_i^{(m)}$;

7 Calculer la somme des hessians $H_j = \sum_{i \in R_j} h_i^{(m)}$ avec $h_i^{(m)} = \frac{\partial^2 L(y_i, \hat{y}_i^{(m-1)})}{\partial (\hat{y}_i^{(m-1)})^2}$;

8 Calculer le score de gain pour chaque split potentiel :

$$\text{Gain} = \frac{1}{2} \left(\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{G^2}{H + \lambda} \right) - \gamma$$

;

9 G_L, H_L, G_R, H_R : Sommes des gradients et hessians pour les sous-nœuds gauche et droit respectivement. ;

10 Sélectionner le split avec le gain le plus élevé qui dépasse un certain seuil γ

;

11 **fin**

12 Mettre à jour les prédictions $\hat{y}_i^{(m)} = \hat{y}_i^{(m-1)} + \eta \cdot T_m(x_i)$;

13 fin

14 Fin

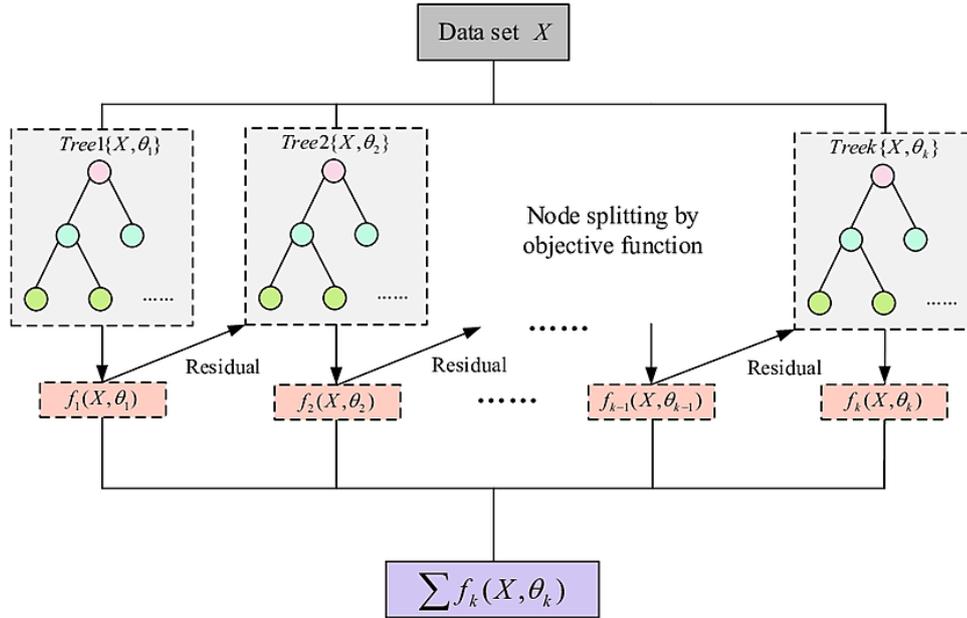


Figure 2.8: Schéma illustratif de principe de fonctionnement de XGBoost
Source: [45]

2.5.3 Evaluation de la qualité des prévisions

Pour évaluer la qualité d'un modèle, il est crucial de comprendre les différentes métriques qui permettent de comparer les méthodes de prévision. Ces mesures nous aident à choisir la méthode la mieux adaptée à nos données. Les principales métriques utilisées pour cette évaluation sont :

MAE (Mean Absolute Error) : l'erreur absolue moyenne mesure la moyenne des erreurs absolues entre les prévisions et les valeurs réelles. Elle fournit une indication de l'exactitude globale des prévisions. En d'autres termes, le mae mesure la différence absolue entre chaque prédiction et sa valeur réelle.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.39)$$

RMSE (Root Mean Squared Error) : La racine carrée de l'erreur quadratique moyenne est une mesure qui accorde plus de poids aux grandes erreurs. Le RMSE est calculé en prenant la racine de la moyenne des carrés des différences entre les valeurs prédites et les valeurs réelles.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.40)$$

MAPE (Mean Absolute Percentage Error) : L'erreur de pourcentage absolue moyenne est une mesure d'erreur utilisée pour évaluer la précision d'un modèle de prévision en pourcentage. Elle calcule la moyenne des valeurs absolues des erreurs relatives entre les valeurs observées et prédites.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (2.41)$$

Tel que:

- y_i représente la valeur observée de la variable dépendante pour la i -ème observation.
- \hat{y}_i représente la valeur prédite par le modèle pour la i -ème observation.
- n est le nombre total d'observations dans l'échantillon.

Ces métriques sont généralement appliquées sur des données de test, c'est-à-dire des données qui n'ont pas participé à la construction du modèle de prévision. En utilisant ces mesures, nous pourrions comparer les différentes méthodes de prévision et sélectionner celle qui offre la meilleure précision.

Pour clôturer ce chapitre nous allons présenter dans la section suivante une mesure de distance qui est largement utilisée en pratique pour quantifier la similarité et la dissimilarité entre les points.

2.6 Distance de Minkowski

La distance de Minkowski est l'une des principales mesures de distance car elle généralise une large gamme d'autres distances, telles que la distance de Manhattan, la distance euclidienne, ainsi que la distance de Chebyshev. Elle est définie par la formule suivante [46]:

$$D(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}} \quad (2.42)$$

où \mathbf{x} et \mathbf{y} sont deux points de n dimensions, et p est un paramètre positif qui détermine le type de distance calculée. En ajustant le paramètre p , la distance de Minkowski peut représenter plusieurs autres distances bien connues :

- **Distance de Manhattan** : pour $p = 1$

Cette distance mesure la somme des valeurs absolues des différences entre les coordonnées correspondantes des points.

$$D(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n |x_i - y_i| \quad (2.43)$$

- **Distance Euclidienne** : pour $p = 2$

C'est la distance courante mesurant la longueur du segment de ligne droite entre deux points.

$$D(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^n (x_i - y_i)^2 \right)^{\frac{1}{2}} \quad (2.44)$$

- **Distance de Chebyshev** : pour $p \rightarrow \infty$

Cette distance correspond à la plus grande différence entre les coordonnées des deux points.

$$D(\mathbf{x}, \mathbf{y}) = \max_i |x_i - y_i| \quad (2.45)$$

2.7 Conclusion

Dans ce chapitre, nous avons abordé la partie théorique de la résolution de notre problème en commençant par définir la chaîne logistique, en mettant un accent particulier sur les entrepôts. Nous avons examiné en détail les différents aspects des entrepôts, en mettant en lumière les stratégies de conception d'entrepôt, ainsi que les piliers de la logistique d'entreposage et les processus logistiques lesquels influent sur la productivité de l'entrepôt. Nous avons également discuté des outils nécessaires, incluant l'utilisation des méthodes classiques de prévision, l'intégration de l'intelligence artificielle, et la maîtrise des concepts mathématiques fondamentaux pour une analyse approfondie.

Dans les prochains chapitres, nous ferons la transition de la théorie à la pratique en mettant en œuvre nos solutions pour résoudre les problèmes identifiés et évaluer l'impact des méthodes d'optimisation sur la productivité.

Chapitre 3

Implémentation des méthodes de prévision classiques

3.1 Introduction

Dans le chapitre précédent, nous avons exploré en détail les méthodologies nécessaires à la création de prévisions. Ce chapitre se concentre sur l'application de modèles classiques pour prédire avec précision les flux journaliers. Ces prévisions revêtent une importance cruciale pour anticiper l'avenir et faciliter la planification stratégique.

Ce chapitre a pour objectif d'évaluer plusieurs modèles de prédiction afin de choisir celui qui offre les prévisions les plus précises et adaptées à nos données spécifiques. Ces prévisions seront utilisées comme données d'entrée essentielles pour la planification des opérations logistiques. Cette étude sera examinée en détail dans le prochain chapitre.

3.2 Les outils

3.2.1 Logiciel R

Le logiciel R est un outil statistique créé par Ross Ihaka et Robert Gentleman. Il s'agit à la fois d'un langage de programmation et d'un environnement de travail. Les commandes sont exécutées par des instructions codées dans un langage relativement simple, avec des résultats affichés sous forme de texte et des graphiques visualisés directement dans une fenêtre dédiée. R est largement utilisé dans le domaine de la statistique en raison de sa flexibilité, de sa puissance et de la vaste gamme de packages disponibles pour diverses analyses.



Figure 3.1: Logo de logiciel R
Source: [47]

3.2.2 Utilisation de Rstudio

RStudio est un outil récent qui vient combler un vide dans la gamme des outils associés à R. Il s'agit d'un environnement de développement intégré fonctionnel, libre, gratuit et multiplateforme. RStudio facilite la saisie, l'exécution de code, la visualisation des résultats, etc. Il est disponible sur Windows, Mac OS X et Linux, RStudio offre une flexibilité d'utilisation étendue.



Figure 3.2: Logo de logiciel RStudio
Source: [48]

3.2.3 Package FORECAST

Le package "forecast" de R est une bibliothèque qui offre des outils pour effectuer des prévisions statistiques et des analyses de séries chronologiques. Il fournit des fonctionnalités pour modéliser et prédire des données temporelles. Il est particulièrement apprécié pour sa facilité d'utilisation et ses capacités robustes en matière de prévision.

3.2.4 Package ZOO

Le package "zoo" de R est une bibliothèque largement utilisée pour manipuler et analyser des séries temporelles irrégulières ou régulières. Il offre une gamme d'outils pour travailler avec des données temporelles, y compris la gestion des données manquantes et la fusion de séries temporelles. "zoo" est particulièrement apprécié pour sa flexibilité et sa facilité d'utilisation, ce qui en fait un choix populaire parmi les chercheurs et les analystes travaillant avec des séries temporelles dans le langage de programmation R.

3.2.5 Package FTSA

Le package "FTSA" de R est une bibliothèque spécialisée dans l'analyse et la modélisation des séries chronologiques, offrant une approche flexible pour l'analyse des données temporelles

dans le langage de programmation R. Il permet également le calcul des erreurs de prévision telles que MAE, RMSE et MAPE, permettant ainsi d'évaluer la performance des modèles de séries chronologiques.

3.3 Collecte et description des données

Pour notre étude, nous avons recueilli des données quantitatives provenant de l'entrepôt logistique Numilog, spécifiquement du dossier Cevifood. Ces données s'étendent sur la période allant du 1er janvier 2023 au 11 mai 2024 et comprennent la somme des flux de marchandise entrants et sortants.

La figure suivante illustre une partie de notre série chronologique :

Date	Flux
2023-01-01	1583
2023-01-02	1756
2023-01-03	2057
2023-01-04	1808
2023-01-05	1775
2023-01-06	1475
2023-01-07	2321
2023-01-08	2204
2023-01-09	2532
2023-01-10	1905
2023-01-11	2416
2023-01-12	2185
2023-01-13	1364
2023-01-14	2157
2023-01-15	2065
2023-01-16	2310
2023-01-17	2584
2023-01-18	2573

Figure 3.3: Illustration du flux de marchandises

3.4 Analyse exploratoire des données (EDA)

3.4.1 Visualisation des données

La visualisation des données revêt d'une importance cruciale dans notre processus d'analyse exploratoire, car elle nous permet de saisir rapidement les tendances et la saisonnalité. Nous avons choisi de représenter ces données sous forme d'une série chronologique journalière, avec une période de 7 jours. En d'autres termes, nous avons créé une série chronologique sous R en indiquant des données journalières avec une fréquence de 7 jours pour pouvoir capter la saisonnalité à l'intérieur d'une semaine.

La représentation graphique illustrée dans la figure ci-dessous montre que notre série est affectée d'une tendance à la baisse et d'une saisonnalité, car il y a des pics et des creux qui se répètent régulièrement chaque semaine. Nous avons observé un flux plus important les jeudis, vendredis et samedis. Nous remarquons également la présence de pics significatifs, c'est-à-dire des flux qui peuvent avoir des valeurs soit très élevées soit très basses par rapport à l'ensemble des données. La prochaine section traitera de ces points aberrants ou extrêmes.

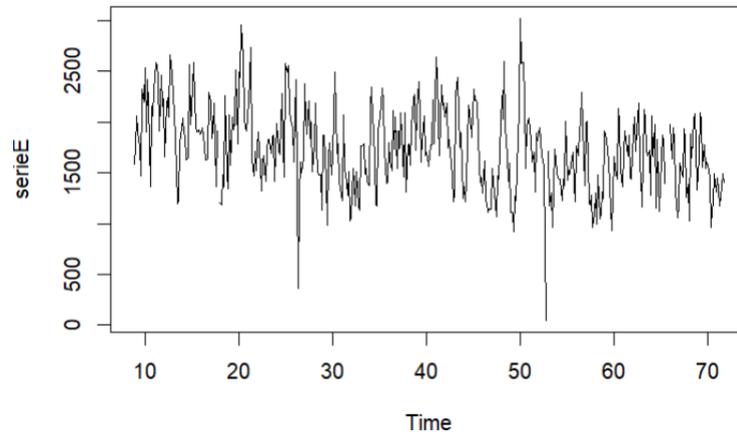


Figure 3.4: Les flux de l'entrepôt en fonction du temps

3.4.2 Identification des anomalies

Dans la section Identification des anomalies, nous avons mis en œuvre un processus de détection des valeurs aberrantes dans nos données de flux d'entrepôt. Cette étape est cruciale pour repérer les observations qui s'écartent significativement des schémas attendus et qui pourraient influencer notre analyse. Nous avons utilisé une approche basée sur un test statistique simplifié et visualiser les données à l'aide d'un diagramme en boîte pour identifier les valeurs extrêmes.

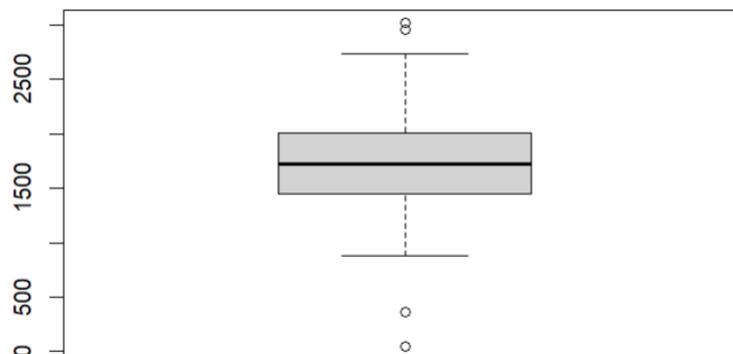


Figure 3.5: Détection des anomalies par diagramme en boîte

Le diagramme de la figure 3.5 indique la présence dans notre série de valeurs aberrantes représentés par les petits points

Nous avons remplacé ces valeurs aberrantes par des valeurs manquantes (NA) et utilisé l'interpolation linéaire pour les imputer, c'est-à-dire les remplacer par des valeurs plus en adéquation avec notre série . Cette méthode nous permet de nettoyer nos données et de les préparer pour des analyses plus approfondies.

Après avoir procédé à l'imputation, nous obtenons le diagramme en boîte suivant qui indique que notre série est nettoyée des valeurs aberrantes :

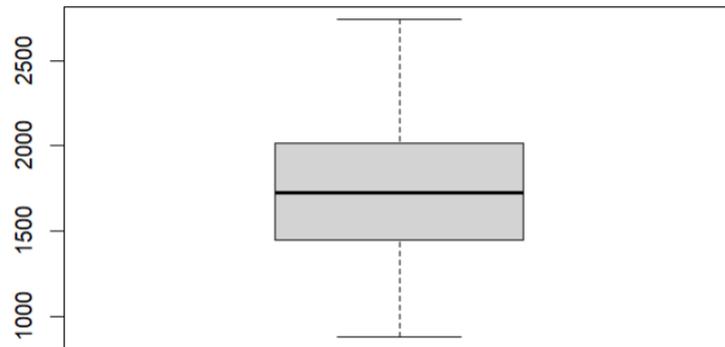


Figure 3.6: Diagramme en boîte des données après correction des anomalies

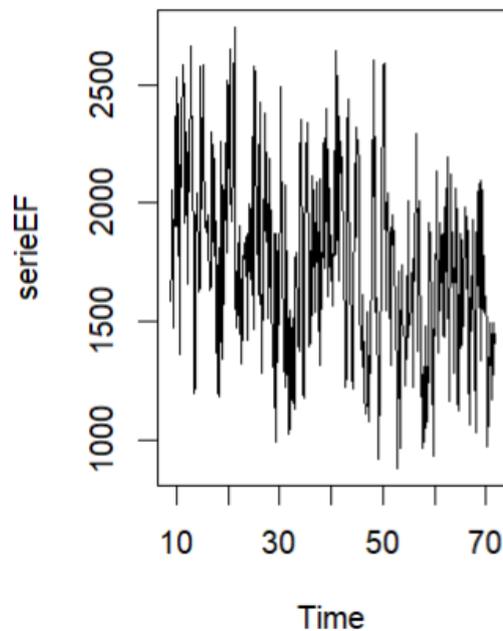


Figure 3.7: Les flux de l'entrepôt en fonction du temps après correction des valeurs aberrantes

3.4.3 Analyse des tendances et des saisonnalités

La représentation graphique de notre série nous a indiqué la présence de tendance et de saisonnalité. Pour confirmer ces observations, nous avons utilisé le test de Fisher, comme décrit en détail dans l'annexe A.1. Ce test statistique nous a permis de valider de manière robuste la présence de tendances et de saisonnalités dans nos données.

```
$Test_Fisher
      Fisher      pvalue
saisonnalité 9.083540 3.074326e-09
tendance      5.590354 0.000000e+00

$décision
[1] "série avec saisonnalité" "série avec tendance"
```

Figure 3.8: Analyse de la tendance et de la saisonnalité avec le test de Fisher

3.4.4 Analyse de Modèles : Multiplicatif vs Additif pour les flux d'entrepôt

Notre étude a impliqué une analyse approfondie afin de déterminer si nos données présentent une structure additive ou multiplicative. Cette distinction revêt une importance capitale pour la précision de nos prévisions futures. Pour ce faire, nous avons fait appel au test de Buys Ballot (voir annexe A.1).

Nos résultats indiquent un modèle multiplicatif, suggérant que les variations saisonnières sont proportionnelles au niveau de la série. Cette reconnaissance de la nature multiplicative de nos données est essentielle pour le paramétrage des modèles de prévision.

```
$Schéma_décomposition
[1] 0.01175497

$décision
[1] "schéma_multiplicatif"
```

Figure 3.9: Résultat du test de Buys Ballot

3.5 Modélisation et prévisions

Dans cette partie, nous procéderons à une évaluation comparative de trois modèles de prévision spécifiques afin de déterminer celui qui offre les prévisions les plus précises. Pour ce faire, nous avons découpé notre série en deux parties : les données d'entraînement et les données de test. Sur les données d'entraînement, nous allons construire nos modèles, et sur les données de test (qui correspondent aux 28 derniers jours de la série), nous évaluerons

la qualité prédictive de nos modèles. Cette approche nous permettra d'identifier le modèle final qui générera les prévisions les plus précises à intégrer dans le processus de planification quotidienne des opérations logistiques.

3.5.1 Méthode de Box-Jenkins

3.5.1.1 La stationnarisation de la série

Suite à notre analyse des données, il est apparu que notre série chronologique présente des tendances et des saisonnalités, ce qui indique que nos données ne sont pas stationnaires. Afin d'appliquer le modèle Box-Jenkins, il est impératif de stationnariser nos données. Pour ce faire, nous avons commencé par éliminer la saisonnalité. Ensuite, nous avons procédé au test de Dickey-Fuller. Les résultats ont confirmé la non-stationnarité de nos données, indiquant un modèle de type "ts", ce qui nécessite également l'élimination de la tendance. Enfin, nous avons répété le test de Dickey-Fuller pour confirmer la stationnarité de nos données transformées (voir l'annexe A.2).

3.5.1.2 Le choix du modèle

Étant donné que notre modèle était de type "ts", nous n'avions pas besoin de le différencier, donc il s'agissait d'un modèle SARMA ou ARMA. Nous avons supposé que le modèle était SARMA. Nous avons donc procédé à la recherche des paramètres optimaux pour ce modèle. Nous avons créé une fonction spécifique pour rechercher les paramètres (p, d, q) et (P, D, Q) qui minimisent le critère d'information d'Akaike (AIC). Les résultats ont montré que le modèle SARMA optimal avait des paramètres (5, 0, 7)(2, 0, 0). Cette configuration confirme qu'il ne s'agit pas d'un modèle ARMA, car les paramètres de la composante saisonnière ne sont pas nuls.

```
$`ARIMA(5,0,7)(2,0,2)`  
[1] 3718.641  
  
$`ARIMA(6,0,7)(2,0,2)`  
[1] 3722.956  
  
$`ARIMA(7,0,7)(2,0,2)`  
[1] 3724.41  
  
> # Trouver le modèle avec le plus petit AIC  
> best_model_name <- names(which.min(unlist(all_models_aic)))  
> best_model_name  
[1] "ARIMA(5,0,7)(2,0,0)"  
> best_model_aic <- min(unlist(all_models_aic))  
> best_model_aic  
[1] 3709.1
```

Figure 3.10: Résultats générés par la fonction de minimisation de l'AIC

Les coefficients du modèle SARMA optimal sont présentés dans la figure ci-jointe:

```

Coefficients:
  ar1      ar2      ar3      ar4      ar5      ma1      ma2      ma3      ma4      ma5      ma6
s.e.  0.9746  1.0717 -1.0427 -0.3517  0.3383  0.9987 -0.0554 -0.0553 -0.1124 -0.1108 -0.8319
      ma7      sar1      sar2      intercept
s.e.      NaN  0.0734  0.0546   9.0830
    
```

Figure 3.11: Les coefficients du modèle SARMA

3.5.1.3 Prévisions des flux et évaluation

En utilisant le modèle SARMA avec les paramètres $(5, 0, 7)(2, 0, 0)$, nous avons généré des prévisions pour les 28 prochains jours (données test) en utilisant la fonction `forecast` dans R. Ensuite, en réintégrant la tendance et la saisonnalité préalablement supprimées, nous avons comparé ces prévisions avec les observations réelles en calculant des métriques d'évaluation telles que MAE, RMSE et MAPE. Ces mesures nous ont fourni une indication de la précision de nos prévisions et nous ont permis d'évaluer l'efficacité de notre modèle SARMA.

Date	Réelles	Prévisions
14/04/2024	1385	1876.239
15/04/2024	1032	1918.668
16/04/2024	1877	1819.141
17/04/2024	1653	1721.004
18/04/2024	2016	1656.073
19/04/2024	2082	1645.249
20/04/2024	1777	1449.56
21/04/2024	1337	1761.34
22/04/2024	1654	1768.567
23/04/2024	2091	1653.863
24/04/2024	2027	1550.146
25/04/2024	1555	1515.33
26/04/2024	1774	1541.77
27/04/2024	1542	1384.314
28/04/2024	1606	1732.762
29/04/2024	1591	1755.078
30/04/2024	1454	1654.638

Tableau 3.1: Résultat du modèle SARMA

Métrique	Valeur
MAE	285
RMSE	345.41
MAPE	20.54

Tableau 3.2: Métrique d'évaluation du modèle SARMA

Les mesures d'erreur fournies révèlent des informations essentielles sur la performance

du modèle de prévision. La valeur de MAE est de 275.69 indique qu'en moyenne, les prévisions diffèrent de plus ou moins 275.69 unités des vraies valeurs. Le RMSE de 345.41 indique une dispersion plus importante des erreurs par rapport à la réalité. Enfin, le MAPE de 20.54% exprime une différence moyenne de 20.54% entre les prévisions et les valeurs réelles.

Bien que le modèle fournisse des prévisions utiles, il est crucial d'explorer d'autres méthodes de modélisation pour améliorer la précision des prévisions.

3.5.2 Méthode de lissage exponentiel

Suite à notre analyse, il est apparu que notre série chronologique présente à la fois une tendance et une saisonnalité. Pour modéliser efficacement ces caractéristiques, nous avons opté pour le modèle de Holt-Winters qui est reconnu pour sa capacité à capturer les tendances et les variations saisonnières dans les données temporelles. Le processus d'ajustement des paramètres du modèle implique de trouver les valeurs optimales pour ces paramètres, ce que nous avons accompli en utilisant la fonction Holt-Winters dans R.

Ensuite, pour générer nos prévisions, nous avons employé la fonction forecast disponible dans R, qui utilise les paramètres optimisés du modèle Holt-Winters pour produire des projections futures. Cette approche nous permet de mieux anticiper les tendances et les variations saisonnières, fournissant ainsi des estimations plus fiables pour nos prévisions à venir.

Date	Réelles	Prévisions
14/04/2024	1385	1404.241
15/04/2024	1032	1769.181
16/04/2024	1877	1760.352
17/04/2024	1653	1741.794
18/04/2024	2016	1615.572
19/04/2024	2082	1616.784
20/04/2024	1777	13/9.398
21/04/2024	1337	1406.472
22/04/2024	1654	1771.993
23/04/2024	2091	1763.148
24/04/2024	2027	1744.561
25/04/2024	1555	1618.137
26/04/2024	1774	1619.35
27/04/2024	1542	1381.587
28/04/2024	1606	1408.704
29/04/2024	1591	1774.804
30/04/2024	1454	1765.945

Tableau 3.3: Résultat du modèle Holt-Winters

Métrique	Valeur
MAE	275.69
RMSE	342.02
MAPE	20.16

Tableau 3.4: Métrique d'évaluation du modèle Holt-Winters

Le modèle de lissage exponentiel a montré une légère amélioration par rapport au modèle Box-Jenkins. Cependant, malgré cette amélioration, il est conseillé d'explorer une troisième méthode de prévision pour obtenir des résultats plus significatifs et fiables.

3.5.3 Méthode de TBATS

Nous avons également choisi d'utiliser le modèle TBATS, reconnu pour sa capacité à capturer les tendances complexes et les variations saisonnières dans les données temporelles. L'ajustement des paramètres de ce modèle consiste à déterminer les valeurs optimales pour les différents paramètres du modèle, ce que nous avons accompli en utilisant la fonction `Tbats` dans R.

Pour la génération de nos prévisions, nous avons utilisé la fonction `forecast` disponible dans R, exploitant ainsi les paramètres optimisés du modèle TBATS pour produire des projections futures.

Date	Réelles	Prévisions
14/04/2024	1385	1465.345
15/04/2024	1032	1581.515
16/04/2024	1877	1508.159
17/04/2024	1653	1485.802
18/04/2024	2016	1640.993
19/04/2024	2082	1670.865
20/04/2024	1777	1523.712
21/04/2024	1337	1571.055
22/04/2024	1654	1672.581
23/04/2024	2091	1581.343
24/04/2024	2027	1541.076
25/04/2024	1555	1637.731
26/04/2024	1774	1624.074
27/04/2024	1542	1548.949
28/04/2024	1606	1597.434
29/04/2024	1591	1637.281
30/04/2024	1454	1576.189

Tableau 3.5: Résultat du modèle TBATS

Métrique	Valeur
MAE	252.67
RMSE	308.20
MAPE	17.95

Tableau 3.6: Métrique d'évaluation du modèle TBATS

Ces résultats révèlent une nette amélioration des performances par rapport aux modèles Box-Jenkins et de lissage exponentiel. Cette amélioration découle principalement de la capacité du modèle TBATS à capturer efficacement les aspects complexes de notre série temporelle. MAPE est de 17,95%, ce qui suggère que les prévisions du modèle sont bonnes et acceptables, avec une précision d'environ 82,05% par rapport à la réalité.

3.6 Comparaison des modèles

Après avoir formé et évalué plusieurs modèles, nous avons comparé leurs performances en utilisant diverses métriques, notamment l'erreur RMSE, MAE et MAPE. En outre, nous avons évalué les performances des prévisions en calculant les métriques MAE, RMSE et MAPE par rapport aux valeurs prévisionnelles de l'entreprise. Cela nous a offert une perspective supplémentaire sur la fiabilité de nos propres prévisions.

Modèle	MAE	RMSE	MAPE(%)
Box-Jenkins	285	345.41	20.54
Lissage Exponentiel	275.69	342.02	20.16
TBATS	252.67	308.20	17.95
Entreprise	514.97	614.1	32.21

Tableau 3.7: Tableau des métriques

L'analyse comparative des performances des divers modèles de prévision, ainsi que des prévisions émises par l'entreprise, fournit des perspectives essentielles pour éclairer les décisions prises dans le cadre de nos études. Les résultats obtenus démontrent de manière convaincante que la méthode TBATS surpasse nettement les autres approches en termes de précision et de fiabilité des prévisions.

La méthode TBATS se distingue par des mesures d'erreur considérablement plus faibles que celles des méthodes de lissage exponentiel et de Box-Jenkins, confirmant l'efficacité supérieure de la méthode TBATS pour capturer et prédire les tendances des flux d'entrepôt.

Notamment, les prévisions de l'entreprise présentent des performances encore moins satisfaisantes, avec des valeurs plus élevées de MAE, RMSE et MAPE par rapport aux modèles examinés. Cette comparaison directe met en évidence l'avantage distinct que représente l'adoption de la méthode TBATS par rapport aux pratiques de prévision actuelles de l'entreprise.

En conclusion, les résultats de cette analyse soutiennent fermement la recommandation d'utiliser la méthode TBATS pour l'application de prévision des flux d'entrepôt. Cette approche se distingue par sa précision supérieure, offrant ainsi une base solide pour notre étude ultérieure.

3.7 Conclusion

Dans ce chapitre, nous avons fourni une description détaillée des outils utilisés pour prédire les flux journaliers. Nous avons commencé par mettre en lumière les bibliothèques spécifiques que nous avons utilisées pour développer notre solution. Ensuite, nous avons souligné l'importance des étapes clés de la science des données : la collecte des données, le prétraitement incluant la gestion des valeurs manquantes et aberrantes, ainsi que l'analyse des données.

Enfin, nous avons appliqué les méthodologies des modèles traditionnels en utilisant des données réelles de l'entreprise pour tester nos modèles. Les résultats ont été examinés et interprétés afin de sélectionner le modèle le plus performant, en l'occurrence le modèle TBATS. Les prévisions des flux de marchandises générées par ce modèle seront utilisées dans le prochain chapitre pour estimer le nombre d'opérateurs nécessaire journaliers à l'aide de méthodes de l'apprentissage automatique.

Chapitre 4

Planification des opérateurs logistiques et l'allocation des quais

4.1 Introduction

Dans le chapitre précédent, nous avons exploré en détail toutes les étapes essentielles pour obtenir les prévisions des flux. Ce chapitre constitue une continuation de ce travail, en se concentrant sur l'application de modèles de machine learning pour prédire le nombre d'opérateurs logistiques journaliers en utilisant les prévisions des flux déterminées précédemment. Pour cela, nous avons évalué plusieurs modèles de prévision afin de choisir celui qui offre les prévisions des opérateurs logistiques les plus précises et adaptées à nos données spécifiques.

Nous avons également pour objectif dans ce chapitre de trouver le meilleur quai de chargement parmi l'ensemble des quais disponibles dans l'entrepôt. Pour cela, nous avons développé un algorithme permettant de choisir le meilleur quai pour les opérations de chargement. Nous avons ensuite analysé les résultats obtenus après avoir appliqué notre algorithme à un cas réel. Enfin, nous avons également créé une interface utilisateur qui centralisera et facilitera l'accès aux résultats obtenus dans ce chapitre.

4.2 Les outils

4.2.1 Python

Python est un langage de programmation puissant et facile à apprendre. Il dispose de structures de données efficaces de haut niveau. La syntaxe élégante et le typage dynamique de Python, ainsi que sa nature interprétée, en font un langage idéal pour le scripting et le développement rapide d'applications dans de nombreux domaines sur la plupart des plates-formes [49].

Dans notre étude, nous avons utilisé PyCharm comme principal outil de développement en Python. Ses fonctionnalités avancées, notamment la complétion automatique et la détection d'erreurs en temps réel, se sont avérées extrêmement utiles pour simplifier le processus d'écriture du code, en garantissant une fluidité et une efficacité optimales[50].

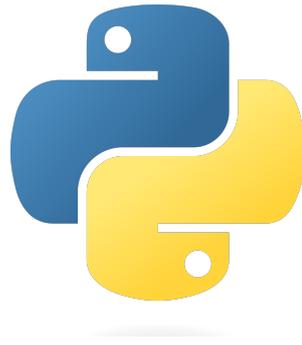


Figure 4.1: Logo de Python
Source: [51]

4.2.2 Bibliothèques

NumPy : Une bibliothèque essentielle pour le calcul numérique en Python, elle offre des opérations rapides sur des tableaux N-dimensionnels, des fonctions mathématiques complètes, des routines d'algèbre linéaire, des générateurs de nombres aléatoires, et bien plus encore[52].



Figure 4.2: Logo de Numpy
Source: [52]

Pandas : C'est une bibliothèque qui fournit des structures de données riches et des fonctions qui rendent la manipulation des données structurées rapide, facile et expressive. Elle est également facile à apprendre et permet de lire des données dans différents formats, tels que les fichiers CSV, texte, Microsoft Excel et les bases de données SQL [53].



Figure 4.3: Logo de Pandas
Source: [54]

Matplotlib : Elle est l'une des bibliothèques de visualisation de données les plus populaires en Python. Elle est largement utilisée pour créer des graphiques statiques,

animés et interactifs. Matplotlib offre une grande flexibilité pour produire une variété de visualisations, allant des simples tracés linéaires aux graphiques complexes [55].

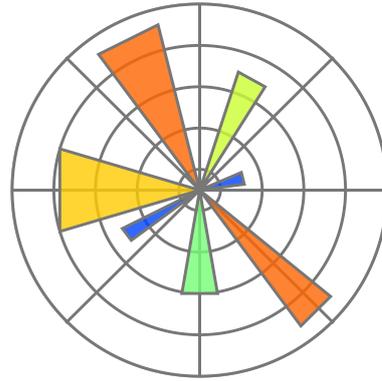


Figure 4.4: Logo de Matplotlib
Source: [56]

Scikit-learn : Cette bibliothèque est devenue l'une des bibliothèques d'apprentissage automatique les plus populaires. Elle fournit une vaste collection d'algorithmes pour les tâches d'apprentissage automatique, comprenant la régression, la classification, le clustering, et plus encore. De plus, scikit-learn propose des modules pour le prétraitement des fonctionnalités, l'extraction de données, l'optimisation des hyperparamètres et l'évaluation des modèles[57].



Figure 4.5: Logo de Scikit-learn
Source: [58]

4.3 Mise en œuvre des prévisions journalières des opérateurs logistiques

4.3.1 Collecte des données

Les données recueillies par l'entreprise pour la période allant du 1er janvier 2023 au 11 mai 2024 nous ont permis de compiler un historique détaillé des flux de marchandises, des données relatives à la productivité ainsi que le nombre d'OPL journaliers.

Voici une synthèse organisée de ces informations :

Flux de marchandise journaliers : Ces données comprennent les dates des flux historiques, ainsi que les quantités de marchandises traitées (réception et expédition combinées) journalières. Les flux historiques offrent une vision globale des volumes d'activité sur une période donnée, permettant ainsi de détecter les tendances et les variations dans le traitement des marchandises.

Productivité journalière : La productivité est un indicateur clé pour évaluer l'efficacité opérationnelle de l'entrepôt. Cet indicateur mesure le nombre d'unités traitées par heure travaillée, avec une pondération de 70% pour les flux de mise en stock et de préparation de commande, et de 30% pour les activités de chargement et de déchargement. Il est important de noter que cette règle de pondération est spécifiquement utilisée par l'entreprise pour refléter l'importance relative de chaque type d'activité dans l'évaluation de la productivité.

$$\text{Productivité} = \frac{0.3 \times \text{NPSP} + 0.7 \times \text{NPCD}}{\text{NHT}} \quad (4.1)$$

Tel que :

- NPSP : Nombre de palettes stockées et préparées
- NPCD : Nombre de palettes chargées et déchargées
- NHT : Nombre d'heures travaillées, il est calculé comme suit :

$$\text{NHT} = 7.5 \times \text{nombre d'OPL} \quad (4.2)$$

Cette règle indique que chaque opérateur travaille 7,5 heures par jour, et le total des heures travaillées est obtenu en multipliant ce chiffre par le nombre total d'OPL présents chaque jour.

Nombre d'OPL journaliers : Nous avons également extrait les données sur les OPL journaliers, qui indiquent le nombre d'opérateurs présents chaque jour. Ces informations sont cruciales pour calculer les heures travaillées et, par conséquent, pour évaluer la productivité quotidienne.

Pour analyser et modéliser l'ensemble de ces données, nous avons mis en place une matrice de données structurée de manière rigoureuse, offrant une vue complète et détaillée de toutes les variables pertinentes :

date	flux	productivité	OPL
01/01/2023	1583	14,5	17
02/01/2023	1756	16,3	15
03/01/2023	2057	20,7	18
04/01/2023	1808	16,8	16
05/01/2023	1775	16,2	15
06/01/2023	1475	17,2	13
07/01/2023	2321	18,6	19
08/01/2023	2204	17,5	19
09/01/2023	2532	17,6	16
10/01/2023	1905	19,1	18
11/01/2023	2416	18,2	20
12/01/2023	2185	16,8	16
13/01/2023	1364	18,9	14

Figure 4.6: Structure de la base de données

4.3.2 Analyse et compréhension des données

Dans cette partie, nous avons procédé à une analyse minutieuse de la matrice afin d'identifier toute anomalie, lacune ou incohérence qui pourrait compromettre nos analyses ultérieures. Voici nos principales actions :

- **Description des données** : L'analyse descriptive de nos données a été réalisée en utilisant la fonctionnalité de statistiques descriptives. Cette approche nous fournit des informations telles que la moyenne, l'écart type, les valeurs minimales et maximales, les quartiles de nos données ce qui nous a permis de comprendre rapidement les caractéristiques fondamentales de nos données

```
data.describe()
      flux  productivité  OPL
count  440.000000    440.000000  440.000000
mean   1742.718182    19.626136  14.593182
std    407.071560     2.855195   3.376921
min    368.000000    13.900000   4.000000
25%   1447.750000    17.600000  12.000000
50%   1724.500000    19.000000  14.000000
75%   2018.000000    20.800000  17.000000
max   3020.000000    34.500000  24.000000
```

Figure 4.7: Description des données

- **Détection des doublons** : nous avons utilisé des fonctions dédiées pour repérer les enregistrements dupliqués, permettant ainsi de nettoyer les données et d'éliminer toute redondance susceptible de biaiser nos résultats.

```
data[data.duplicated()]
Empty DataFrame
Columns: [date, flux, productivité , OPL ]
Index: []
```

Figure 4.8: Visualisation des lignes dupliquées

- **Identification des valeurs manquantes** : À l'aide de fonctions spécifiques, nous avons repéré les valeurs manquantes dans chaque colonne, puis nous avons entrepris une enquête supplémentaire pour comprendre leur origine et leur impact potentiel sur nos analyses.

```
data.isna().sum()
date          0
flux          1
productivité  1
OPL           1
dtype: int64
```

Figure 4.9: Visualisation des données manquantes

- **Vérification de la validité des dates** : Nous avons vérifié que les dates se situent dans une plage temporelle attendue, afin d'identifier et de rectifier toute erreur de saisie

```
data[data['date'].isna()]
Empty DataFrame
Columns: [date, flux, productivité , OPL ]
Index: []
```

Figure 4.10: Visualisation des dates invalides

- **Détection des valeurs aberrantes** : En utilisant des méthodes statistiques, nous avons repéré les valeurs aberrantes qui s'écartent significativement de la distribution normale des données, ce qui nous a permis de les traiter de manière appropriée.

```
data.loc[(data[colonne] < lower_bound)
 | (data[colonne] > upper_bound)]
30      0.778194
52      0.892922
54      0.947763
267     0.853568
269     0.742514
276     1.000000
361     0.758515
422     0.831493
437     0.903415
438     0.766429
440     0.766429
Name: productivité , dtype: float64
```

Figure 4.11: Visualisation des valeurs aberrantes

4.3.3 Traitement des données

Dans cette phase, plusieurs étapes ont été entreprises pour rendre les données prêtes à être utilisées dans un modèle d'apprentissage automatique. Voici les actions entreprises :

- **Remplacement des valeurs manquantes :** Nous avons gérées les valeurs manquantes en les remplaçant par la moyenne des valeurs disponibles dans la même colonne. Cela assure la continuité des données tout en minimisant l'impact sur les analyses.

```
fillna(data[colonne].mean())
Nombre de valeurs manquantes :
date          0
flux          0
productivité  0
OPL           0
dtype: int64
```

Figure 4.12: Visualisation des valeurs manquantes après le traitement de données

- **Conversion des dates :** Nous avons converti les dates dans un format approprié pour l'analyse, ce qui est essentiel pour que le modèle d'apprentissage automatique puisse exploiter efficacement ces données temporelles.

```
(data['date'] - pd.Timestamp('1970-01-01'))
// pd.Timedelta('1D')
      date  date_numeric
0  2023-01-01      19358
1  2023-01-02      19359
2  2023-01-03      19360
3  2023-01-04      19361
4  2023-01-05      19362
..      ...      ...
436 2024-05-07      19850
437 2024-05-08      19851
438 2024-05-09      19852
439 2024-05-10      19853
440 2024-05-11      19854
```

Figure 4.13: Conversion des dates

- **Transformation logarithmique :** Nous avons utilisé cette transformation pour stabiliser la variance.

```
np.log(data[colonne])
      flux  productivité  OPL  date_numeric
0  7.367077      2.674149  2.833213      9.870861
1  7.470794      2.791165  2.708050      9.870913
2  7.629004      3.030134  2.890372      9.870964
3  7.499977      2.821379  2.772589      9.871016
4  7.481556      2.785011  2.708050      9.871068
..      ...      ...      ...      ...
436 7.199678      3.005683  2.484907      9.895959
437 7.066467      3.453157  2.484907      9.896010
438 7.222566      3.328627  2.484907      9.896060
439 7.309881      3.169686  2.397895      9.896110
440 7.248504      3.328627  2.302585      9.896161
```

Figure 4.14: Transformation logarithmique de données

- **Normalisation de Données :** Nous avons procédé à la normalisation des données en utilisant la méthode Min-Max Scaling, qui consiste à redimensionner les valeurs des caractéristiques pour les faire tenir dans une plage spécifique entre 0 et 1.

```

scaler.fit_transform(data[[colonne]]
      flux      productivité      OPL      date_numeric
0      0.693132      0.046487      0.807541      0.000000
1      0.742405      0.175208      0.737686      0.002042
2      0.817567      0.438079      0.839442      0.004083
3      0.756269      0.208444      0.773706      0.006125
4      0.747518      0.168438      0.737686      0.008167
..      ...      ...      ...      ...
436     0.613605      0.411182      0.613147      0.992036
437     0.550320      0.903415      0.613147      0.994027
438     0.624479      0.766429      0.613147      0.996018
439     0.665960      0.591590      0.564585      0.998009
440     0.636801      0.766429      0.511392      1.000000

```

Figure 4.15: Normalisation de données

- **Remplacement des valeurs aberrantes :** Après avoir identifié les valeurs aberrantes, nous les avons remplacées par des valeurs NaN. Ensuite, ces valeurs NaN ont été remplacées par la moyenne de chaque colonne. Cette étape garantit l'exactitude des prédictions en traitant les données aberrantes de manière appropriée.

```

data.loc[(data[colonne] < lower_bound)
         | (data[colonne] > upper_bound)]
Series([], Name: productivité , dtype: float64)

```

Figure 4.16: Visualisation des valeurs aberrantes après le traitement de données

4.3.4 Division de données

La division de l'ensemble de données est une étape cruciale dans le développement et l'évaluation des modèles d'apprentissage automatique. Cette pratique permet de s'assurer que le modèle peut se généraliser sur des données non vues et d'éviter le surapprentissage (overfitting). Typiquement, l'ensemble de données est divisé en trois sous-ensembles : l'ensemble d'entraînement, l'ensemble de validation, et l'ensemble de test. Voici comment ces sous-ensembles sont généralement définis et utilisés :

Ensemble d'entraînement (Training Set) : Cet ensemble représente environ 70-80% des données totales et est utilisé pour former le modèle. Le modèle apprend les relations entre les caractéristiques et la variable cible, ce qui lui permet de faire des prévisions précises sur de nouvelles données.

Ensemble de validation (Validation Set) : Cet ensemble représente environ 10-15% des données. Il est utilisé pour évaluer les performances du modèle pendant la phase de formation. Cela permet d'ajuster les hyperparamètres et d'éviter le surapprentissage.

Ensemble de test (Test Set) : Cet ensemble, représentant environ 10-15% des données, est utilisé pour évaluer les performances finales du modèle après l'entraînement.

Il n'est jamais utilisé durant la phase de formation ou de validation, offrant une indication de la capacité du modèle à généraliser sur des données non vues.

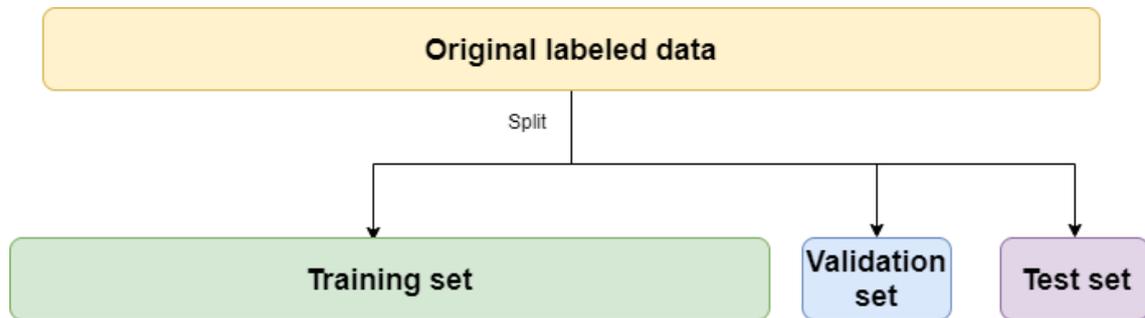


Figure 4.17: test, training and validation sets

Source: [59]

Dans notre approche, nous avons utilisé la validation croisée intégrée dans Grid Search CV pour sélectionner les meilleurs hyperparamètres et évaluer les performances du modèle. Cette méthode permet de diviser automatiquement l'ensemble de données d'entraînement en plusieurs sous-ensembles pour effectuer une validation croisée, ce qui élimine ainsi la nécessité d'un ensemble de validation distinct.

4.3.5 Sélection du modèle

Lors de la sélection d'un modèle pour la prévision, il est essentiel de considérer divers facteurs propres à notre problème, tels que la nature des données (classification, régression, etc.), la taille de l'ensemble de données, les exigences en matière d'interprétabilité et les objectifs spécifiques de notre étude. En tenant compte de ces éléments, nous serons en mesure de choisir le modèle le plus approprié pour nos besoins.

Pour atteindre nos objectifs de prévision, nous avons opté pour des méthodes éprouvées dans la résolution de problèmes de régression. Ainsi, nous avons choisi d'utiliser trois méthodes spécifiques : la régression linéaire, la forêt aléatoire (Random Forest) et XGBoost. Ces méthodes sont bien connues pour leur efficacité et leurs performances dans divers contextes de prévision.

Notre objectif est de trouver le meilleur modèle à utiliser pour calculer les prévisions futures. Pour cela, nous allons nous concentrer sur deux critères :

Évaluation des erreurs : Nous évaluerons les modèles en calculant l'erreur d'apprentissage sur les prédictions (y_{pred}) et en la comparant avec celle sur les données de test (y_{test}). Cela nous permettra de déterminer quel modèle a le mieux appris à générer les prévisions les plus précises et fiables.

Productivité : Pour renforcer le premier critère, nous utiliserons les prévisions générées par les trois modèles. Ces prévisions seront ensuite utilisées dans la formule de productivité pour identifier le modèle qui répond le mieux à l'objectif de l'entreprise d'atteindre une productivité de 20 ou plus.

Dans la partie qui suit, nous allons nous baser sur ces deux critères pour identifier le meilleur modèle. Pour ce faire, nous commencerons par retirer les 28 derniers jours

pour calculer les erreurs de prévisions, puis nous utiliserons ces prévisions pour calculer la productivité. La méthodologie employée et les résultats obtenus sont présentés ci-après.

4.3.5.1 Utilisation du modèle Random Forest

Le principe de Random Forest repose sur l'utilisation de plusieurs arbres de décision en parallèle. Chaque arbre est formé sur un sous-ensemble des données et fait sa propre prévision. Le modèle final fait ensuite la moyenne de toutes ces prévisions. Dans notre cas, nous avons sélectionné 100 arbres en utilisant la méthode de validation croisée pour optimiser sa performance, garantissant ainsi une robustesse et une précision accrues.

En utilisant ce modèle entraîné nous avons obtenue les résultats suivant :

y_test	y_pred
22	17
16	15
13	12
17	17
15	12
13	12
19	17
14	15
17	16
18	17

Tableau 4.1: Extrait du résultat du modèle Random Forest

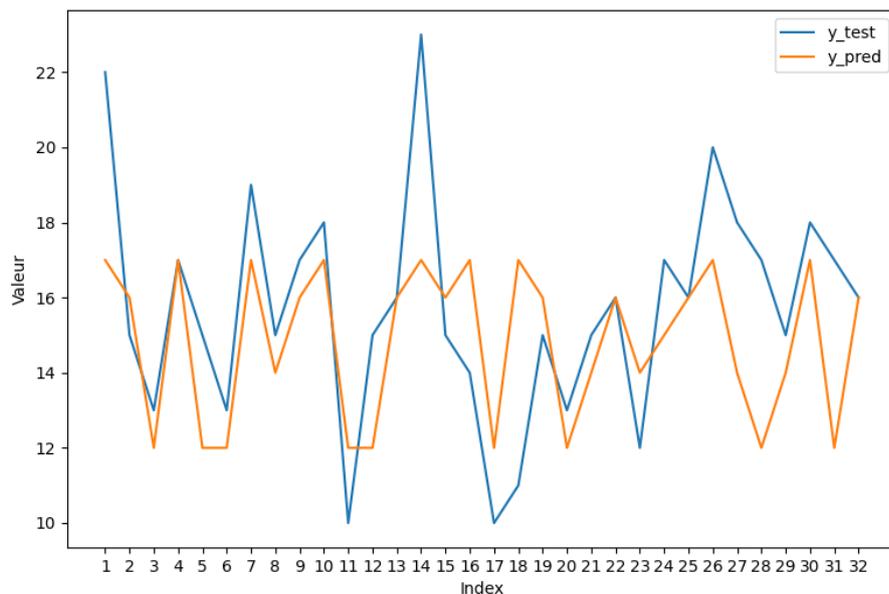


Figure 4.18: Comparaison des valeurs prédites avec les valeurs réelles
Cas : Random Forest

Métrique	Valeur
MAE	2.03
RMSE	2.69
MAPE	13.12

Tableau 4.2: Métrique d'évaluation du modèle Random Forest

L'analyse des erreurs de notre modèle de prédiction révèle des performances qui pourraient être considérées comme acceptables pour notre cas d'utilisation. Avec MAE de 2.68, notre modèle présente en moyenne une différence de 2.68 unités entre ses prédictions et les valeurs réelles. Le RMSE est de 2.03 indique une dispersion modérée des erreurs, tandis que le MAPE de 13.12% suggère des variations modestes entre les prédictions et les valeurs réelles. Ces résultats fournissent une base de référence pour évaluer d'autres modèles potentiels dans notre analyse.

4.3.5.2 Utilisation de XGBoost

Le principe de XGBoost est d'entraîner successivement plusieurs modèles relativement faibles, demandant à chaque modèle d'essayer de corriger les erreurs faites par son prédécesseur, de manière séquentielle (en série). Cela contraste avec l'approche de Random Forest, où les modèles sont développés en parallèle.

Dans notre démarche, nous avons opté pour la méthode de validation croisée afin de déterminer le meilleur nombre d'estimations d'arbres qui a été fixé à 100 arbres. Cette approche a été essentielle pour améliorer la performance du modèle, assurant ainsi une précision et une fiabilité renforcées dans nos prévisions. Les résultats de notre évaluation sont les suivants :

y_test	y_pred
22	21
15	15
13	11
17	16
15	12
13	12
19	17
15	15
17	16
18	16

Tableau 4.3: Extrait du résultat du modèle XGBoost

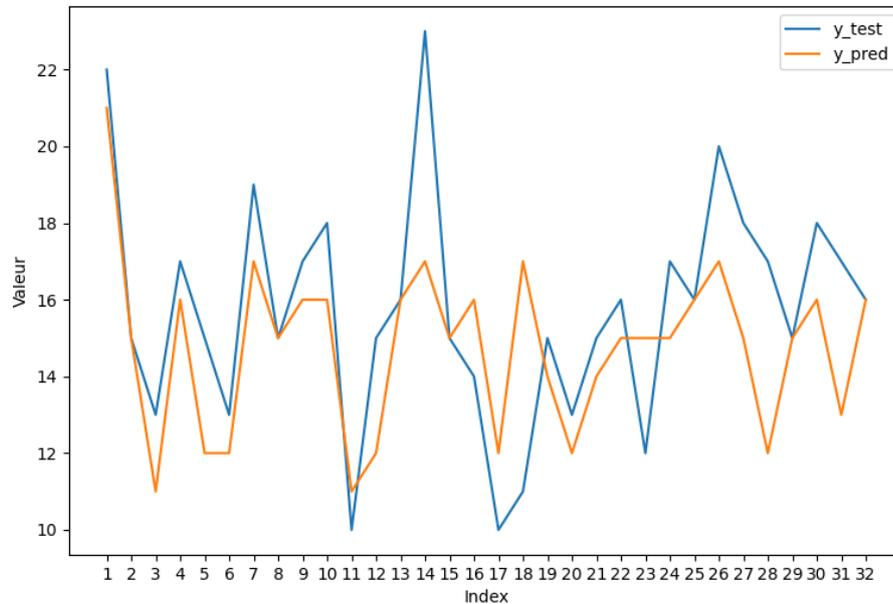


Figure 4.19: Comparaison des valeurs prédites avec les valeurs réelles
Cas : XGBoost

Métrique	Valeur
MAE	1.84
RMSE	2.47
MAPE	12.05

Tableau 4.4: Métrique d'évaluation du modèle XGBoost

Les résultats obtenus indiquent une nette amélioration par rapport au modèle de Random Forest. MAE, RMSE ainsi que MAPE ont tous diminué par rapport aux mesures précédentes. Cela suggère que le modèle XGBoost est plus précis dans ses prévisions et offre de meilleures performances en termes d'estimation du nombre d'OPL journalier.

4.3.5.3 Utilisation de la Régression Linéaire

Dans le cadre de la régression linéaire, nous visons à modéliser la relation entre une variable dépendante et une ou plusieurs variables indépendantes sous forme d'une relation linéaire. Afin de minimiser l'écart entre les valeurs prédites et observées de la variable dépendante, nous avons cherché à estimer les coefficients qui influent sur chaque variable indépendante. Cette estimation des coefficients est effectuée à l'aide de la fonction `fit()` de la régression linéaire, qui ajuste le modèle aux données d'entraînement, en optimisant les coefficients pour mieux correspondre aux valeurs observées.

Après avoir entraîné ce modèle sur nos données, nous avons obtenu les résultats suivants :

y_test	y_pred
22	19
15	15
13	12
17	16
15	13
13	12
19	16
15	14
17	16
18	17

Tableau 4.5: Extrait du résultat du modèle Régression Linéaire

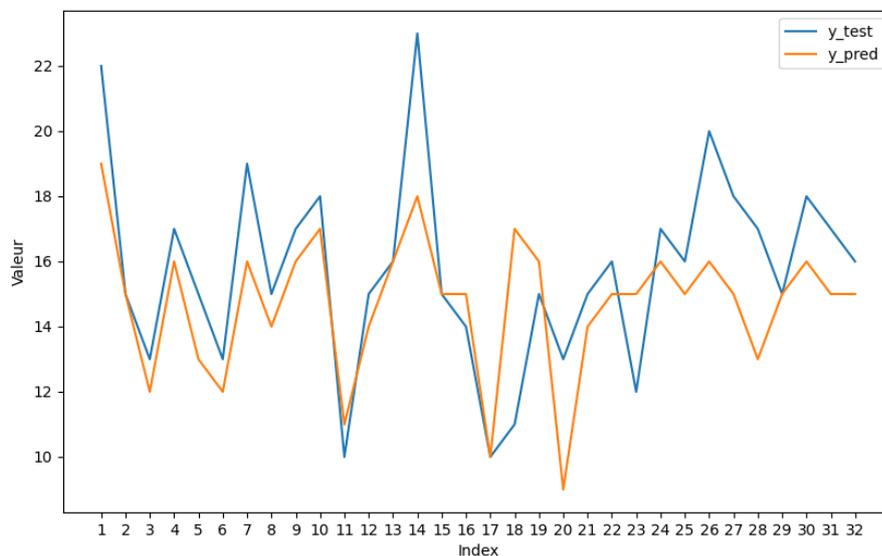


Figure 4.20: Comparaison des valeurs prédites avec les valeurs réelles
Cas : Régression Linéaire

Métrique	Valeur
MAE	1.75
RMSE	2.32
MAPE	11.22

Tableau 4.6: Métrique d'évaluation du modèle Régression Linéaire

Ces résultats indiquent une amélioration par rapport aux modèles précédents, avec des performances relativement meilleures pour le modèle Régression linéaire sur notre série. MAE et RMSE et Le MAPE ont diminué par rapport aux modèles précédents, ce qui suggère une réduction de la variance entre les prédictions et les vraies valeurs.

4.3.6 Discussion des résultats

Après avoir calculé les erreurs pour chaque modèle, nous avons utilisé ces modèles pour générer des prévisions des opérateurs logistiques sur une période de 28 jours. Pour ce faire, nous avons utilisé les caractéristiques (date, flux, productivité) comme données d'entrée dans chaque modèle pour estimer le nombre prévisionnel d'OPL requis en sortie.

date	flux	productivité
14/04/2024	1465	20
15/04/2024	1582	20
16/04/2024	1508	20
17/04/2024	1486	20
18/04/2024	1641	20
19/04/2024	1671	20
20/04/2024	1524	20
21/04/2024	1571	20
22/04/2024	1673	20
23/04/2024	1581	20
24/04/2024	1541	20
25/04/2024	1638	20
26/04/2024	1624	20
27/04/2024	1549	20
28/04/2024	1597	20
29/04/2024	1637	20
30/04/2024	1576	20

Figure 4.21: Données d'entrée dans les modèles d'apprentissage automatique

La colonne Flux représente les prévisions du flux de marchandises calculées à partir du modèle TBATS décrit dans le chapitre précédent. tandis que la colonne Productivité représente notre objectif minimal à atteindre.

En intégrant les prévisions générées par les trois modèles, nous avons calculé la productivité sur une période de 28 jours. Ensuite, nous avons calculé le taux de réussite qui mesure la fréquence à laquelle chaque modèle atteint ou dépasse l'objectif de productivité. Pour ce faire, nous avons déterminé le nombre de jours où chaque modèle a atteint l'objectif et nous avons divisé ce nombre par le nombre total de jours évalués.

Ainsi, nous avons rajouté les prévisions d'OPL de l'entreprise dans notre évaluation. Cette approche nous permet de comparer tous les modèles, y compris celui de l'entreprise, afin de déterminer le plus adapté à nos besoins

Ci-dessous est présenté un tableau récapitulatif de ces éléments, incluant les erreurs calculées ainsi que le taux de réussite de l'objectif de productivité. Ce tableau nous

permet de comparer les modèles non seulement en termes de précision des prévisions mais également en termes de leur capacité à atteindre les objectifs de productivité définis par l'entreprise.

Modèle	MAE	RMSE	MAPE(%)	Taux de réussite(%)
Random Forest	2.03	2.69	13.12	68
XGBoost	1.84	2.47	12.05	82
Régression Linéaire	1.75	2.32	11.22	64
Entreprise	3.07	3.60	28.08	29

Tableau 4.7: Tableau récapitulatif des performances des modèles

En examinant le tableau, la régression linéaire se distingue par une meilleure précision en termes d'erreur ; tandis que XGBoost affiche une supériorité notable en matière de productivité. La régression linéaire surpasse XGBoost avec une faible erreur en moyenne de 0.83%, ce qui est négligeable comparé à l'avantage de XGBoost en termes de productivité, surpassant la régression linéaire de 18%.

De plus, on remarque bien que les prévisions générées par XGBoost ont montré une amélioration significative de la productivité par rapport aux prévisions générées par l'entreprise

Vous trouverez ci-joint un schéma visualisant la productivité estimée par XGBoost par rapport à la productivité estimée par l'entreprise :

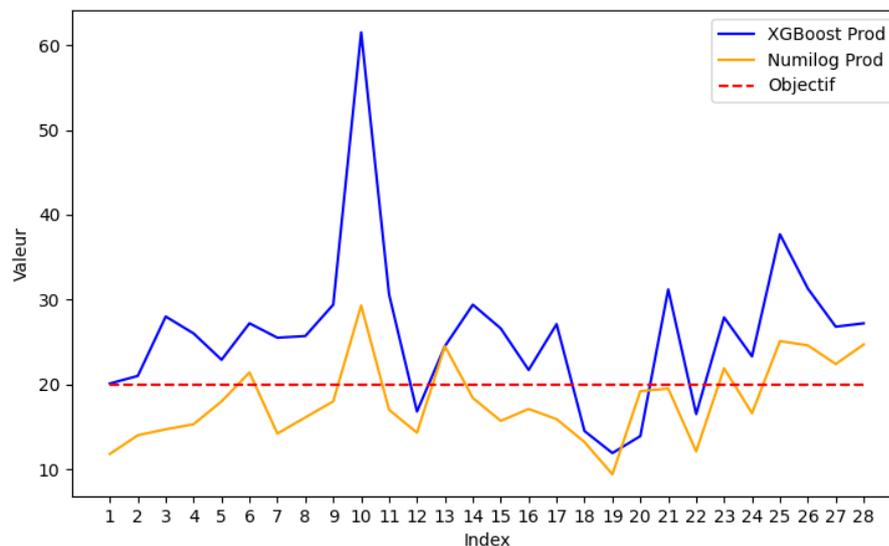


Figure 4.22: Comparaison des schémas de productivité

Sur la base de ces résultats, nous recommandons l'utilisation de XGBoost comme modèle de prévision pour trouver le nombre d'OPL nécessaires, en raison de sa capacité supérieure à atteindre les objectifs de productivité fixés par l'entreprise et de l'analyse des erreurs de prévisions qui indique des performances acceptables pour notre cas d'utilisation.

4.4 Optimisation de l'allocation des quais

4.4.1 Données relatives à l'affectation des quais

Dans notre stratégie de gestion des données pour l'affectation des quais, nous avons adopté une approche rigoureuse en construisant une matrice détaillée qui intègre les informations spécifiques obtenues de chaque emplacement, comprenant les rangées, travées, niveaux et emplacements; ces informations étant séparées, chaque tableau contient les données d'une cellule spécifique. Nous avons ainsi pu attribuer à chaque emplacement le quai le plus proche, puis agréger tous les tableaux dans une matrice finale contenant les emplacements de toutes les cellules ainsi que leur quai correspondant.

Rangée	Travée	Niveau	Emplacement	Quai
A08	001	00	1	24
A08	001	00	2	24
A08	001	00	3	24
A08	001	10	1	24
A08	001	10	2	24
A08	001	10	3	24
A08	001	20	1	24
A08	001	20	2	24
A08	001	20	3	24
A08	001	30	1	24
A08	001	30	2	24
A08	001	30	3	24
A08	001	40	1	24
A08	001	40	2	24
A08	001	40	3	24

Figure 4.23: Structure des emplacements dans les cellules

Dans la partie suivante, nous allons détailler les étapes d'implémentation de notre solution, qui consistent à sélectionner le quai approprié pour chaque préparation de commande.

Rangée	Travée	Niveau	Emplacement
D08	004	00	3
D08	006	20	3
D08	007	10	1
D08	007	20	2
H08	007	30	2
H08	009	30	2
H08	014	40	1
H08	014	40	2
H08	015	40	1
H08	015	40	2
H08	016	40	1
H08	016	40	2
W06	010	10	3
P06	002	00	2
L06	007	10	3

Figure 4.24: Extrait d'une commande

4.4.2 Implémentation de la solution

Dans cette partie de notre processus nous avons abordé la sélection du quai le plus proche pour chaque commande. L'objectif est de déterminer le meilleur quai afin de minimiser le temps nécessaire en réduisant la distance parcourue pour la préparation des commandes. Pour ce faire, nous avons créé un algorithme qui prend en compte la disponibilité des quais dans les cellules et les distances parcourues.

Tout d'abord, nous avons extrait les données d'emplacement des quais existants dans toutes les cellules ainsi que les détails des commandes. Notre objectif est désormais de trouver un seul quai pour chaque commande. Pour ce faire, nous avons ajouté une colonne "quai" dans la matrice des commandes, afin de déterminer le quai le plus proche pour chaque emplacement mentionné dans la commande. Cette approche a été choisie afin de travailler directement avec les quais plutôt qu'avec les emplacements eux-mêmes.



Figure 4.25: Assignment des emplacements au quai le plus proche
Source: [60]

Ensuite, nous avons une matrice où les lignes représentent les quais associés aux emplacements des commandes, et les colonnes représentent les quais disponibles dans l'entrepôt ; les valeurs dans cette matrice représentent les distances entre les quais de lignes et de colonnes en utilisant la formule de distance de Manhattan.

$$D(\text{quai}_i, \text{quai}_j) = |\text{quai}_i - \text{quai}_j|$$

	quai ₁	quai ₂	...	quai ₈₃
quai ₃	D(quai ₃ , quai ₁)	D(quai ₃ , quai ₂)	...	D(quai ₃ , quai ₈₃)
quai ₄	D(quai ₄ , quai ₁)	D(quai ₄ , quai ₂)	...	D(quai ₄ , quai ₈₃)
⋮	⋮	⋮	⋮	⋮
quai ₃₁	D(quai ₃₁ , quai ₁)	D(quai ₃₁ , quai ₂)	...	D(quai ₃₁ , quai ₈₃)
quai ₃₂	D(quai ₃₂ , quai ₁)	D(quai ₃₂ , quai ₂)	...	D(quai ₃₂ , quai ₈₃)
⋮	⋮	⋮	⋮	⋮
quai ₈₃	D(quai ₈₃ , quai ₁)	D(quai ₈₃ , quai ₂)	...	D(quai ₈₃ , quai ₈₃)
Somme	$\sum_{i=3}^{84} D(\text{quai}_i, \text{quai}_1)$	$\sum_{i=3}^{84} D(\text{quai}_i, \text{quai}_2)$...	$\sum_{i=3}^{84} D(\text{quai}_i, \text{quai}_{83})$

Tableau 4.8: Matrice des distances de Manhattan entre quais

Ensuite, nous avons cherché le quai de l'entrepôt qui minimise la somme des distances, tout en respectant la contrainte de disponibilité du quai choisi.

Nous avons construit un algorithme récapitulant notre rapproche :

Algorithme 4 : Sélection de quai

Entrées :

Fichier contenant les emplacements et les quais correspondants

Fichier contenant la commande

Liste `quaisCHOIX` contenant les quais disponibles pour le choix

Sorties :

Le quai approprié

```
1 Debut :
2 Créer un dictionnaire a partir du fichier contenant les emplacements et les quais
  correspondants où la clé est les détails de l'emplacement et la valeur est le quai
  correspondant ;
3 pour chaque ligne dans le fichier de commande faire
4 | Utiliser le dictionnaire pour trouver le quai correspondant ;
5 fin
6 Créer une liste quais à partir de la nouvelle colonne quai des données de
  commande ;
7 Définir une matrice de dimensions (taille de quais, taille de quaisCHOIX) ;
8 pour chaque quai dans quais faire
9 | pour chaque quai_choix dans quaisCHOIX faire
10 | | Calculer la distance entre le quai et le quai_choix ;
11 | fin
12 fin
13 Calculer la somme de chaque colonne de la matrice pour obtenir les distances
  totales par quaisCHOIX ;
14 Créer une liste de tuples (quai, somme des distances) ;
15 Trier la liste par la somme des distances en ordre croissant ;
16 pour chaque tuple (quai, distance) dans la liste faire
17 | Afficher le quai ;
18 | Demander à l'utilisateur si le quai est disponible ;
19 | si le quai est disponible alors
20 | | Afficher le quai comme optimal ;
21 | | Quitter la boucle ;
22 | fin
23 fin
24 si aucun quai n'est disponible alors
25 | Afficher un message indiquant qu'aucun quai n'est disponible et de patienter ;
26 fin
27 Fin
```

4.4.3 Discussion des résultats

Maintenant que nous avons détaillé notre méthode de sélection de quai de chargement, nous allons l'illustrer avec un exemple concret basé sur une commande datée du 14 mai 2024. Nous utiliserons notre programme pour déterminer le meilleur quai disponible dans

l'entrepôt ; ensuite, nous comparerons la distance minimale obtenue par notre méthode avec celle qui aurait été obtenue si nous avions choisi le quai recommandé par l'entreprise.

Après avoir exécuté notre méthode de résolution, nous avons obtenu les résultats suivants :

	Quai de chargement choisi	Distance (unité de distance)
Résultat obtenu par notre résolution	21	45
Résultat obtenu par l'entreprise	19	73

Tableau 4.9: Résultats obtenus pour le choix des quais

Ces résultats démontrent que notre méthode a identifié le quai 21 avec une distance de 45 unités. En revanche, l'entreprise, en choisissant arbitrairement le quai 19, obtient une distance de 73 unités. Cette différence significative souligne l'efficacité de notre approche. En minimisant la distance, notre méthode permet de réduire considérablement les temps de déplacement et, par conséquent, d'améliorer la productivité.

Ainsi, les résultats obtenus illustrent clairement l'efficacité de notre algorithme dans l'optimisation des processus logistiques. L'algorithme permet de prendre des décisions éclairées basées sur des données précises, contrairement à l'approche arbitraire actuelle. En optimisant les choix de quais, l'algorithme réduit les distances parcourues, ce qui se traduit par une amélioration des délais de traitement et une réduction des coûts opérationnels.

Cette approche analytique renforce également la capacité de l'entreprise à s'adapter aux variations de la demande. En intégrant ces données de manière systématique, nous assurons une efficacité opérationnelle accrue et une réponse plus rapide aux besoins des clients.

4.5 Interface de l'utilisateur et fonctionnalités

Pour récapituler et intégrer toutes nos solutions dans un format facile à utiliser par l'utilisateur, nous avons développé une interface utilisateur. Cette interface a pour but de simplifier l'accès aux différentes fonctionnalités de notre solution et de rendre l'expérience utilisateur plus intuitive et interactive.

Pour créer notre interface, nous avons utilisé la bibliothèque Tkinter, qui est la bibliothèque standard de Python pour la création d'interfaces graphiques. Tkinter permet de développer des applications de bureau avec une variété de widgets, tels que des boutons, des champs de texte, des étiquettes, et bien d'autres.

Grâce à sa simplicité et à sa flexibilité, Tkinter est un choix populaire pour le développement d'interfaces utilisateur intuitives et interactives.

Notre interface est présentée comme suit :



Figure 4.26: Interface principale

La fenêtre principale de notre application est configurée comme illustré dans la figure et comporte deux boutons principaux :

- **Bouton "Sélection de Quai" :**

- Ce bouton ouvre une fenêtre secondaire permettant à l'utilisateur de charger un fichier CSV contenant les emplacements de commande et de trouver le quai optimal pour la commande en fonction de la distance.

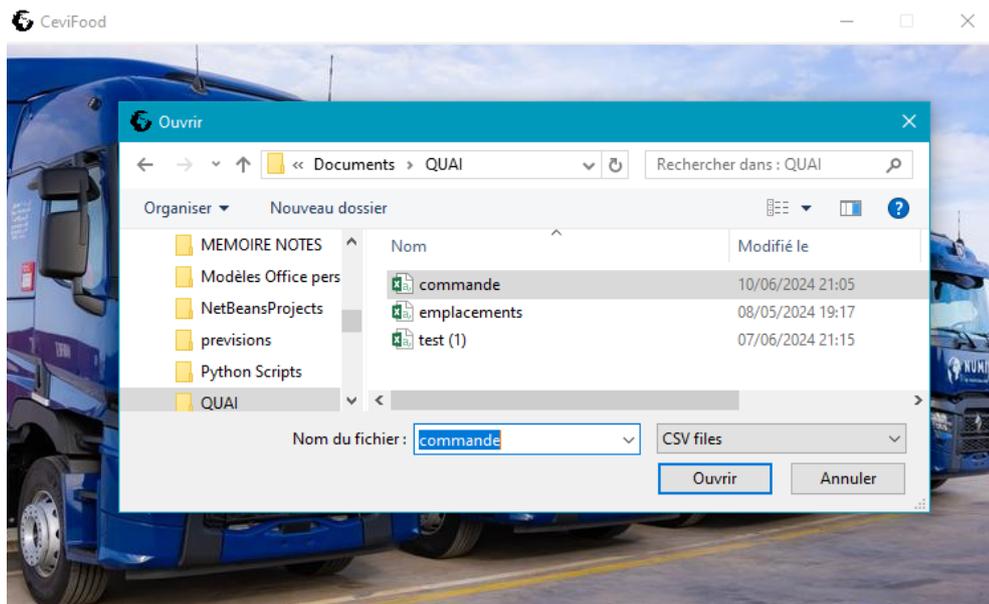


Figure 4.27: Fenêtre secondaire affichée par le premier bouton

- L'interface affiche les résultats de la sélection de quai optimal et offre des options pour accepter ou refuser les suggestions .



Figure 4.28: Résultat affiché par le premier bouton



Figure 4.29: Résultat affiché si la suggestion est refusée

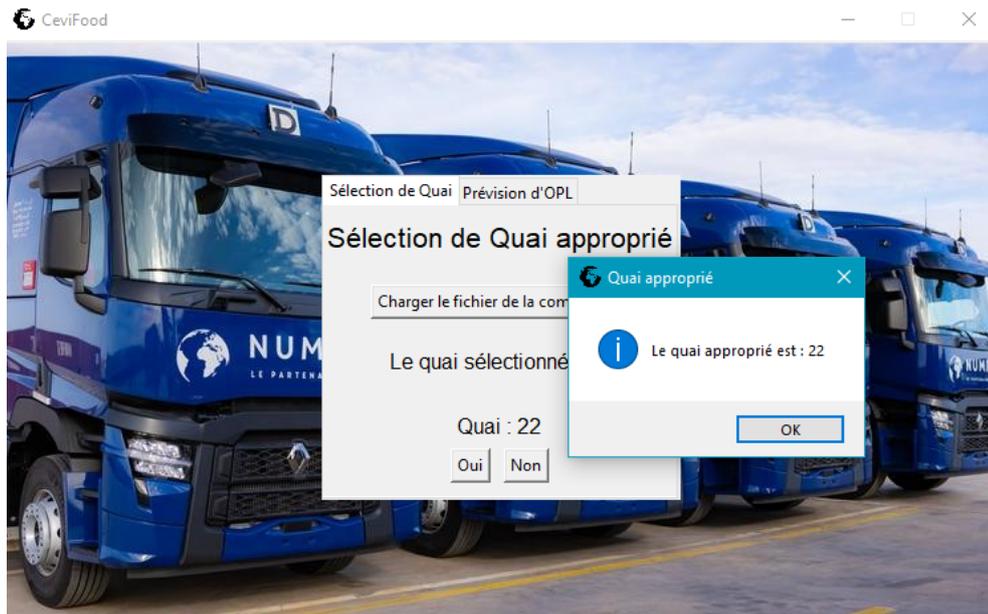


Figure 4.30: Résultat affiché si la suggestion est acceptée

- **Bouton "Prévision d'OPL" :**

- Ce bouton conduit à une fenêtre secondaire dédiée à la prévision d'OPL journaliers. L'utilisateur peut charger un fichier Excel contenant les données d'entrée pour la période à prévoir. Ces données sont ensuite traitées et utilisées pour générer des prévisions à l'aide d'un modèle de régression XGBoost.

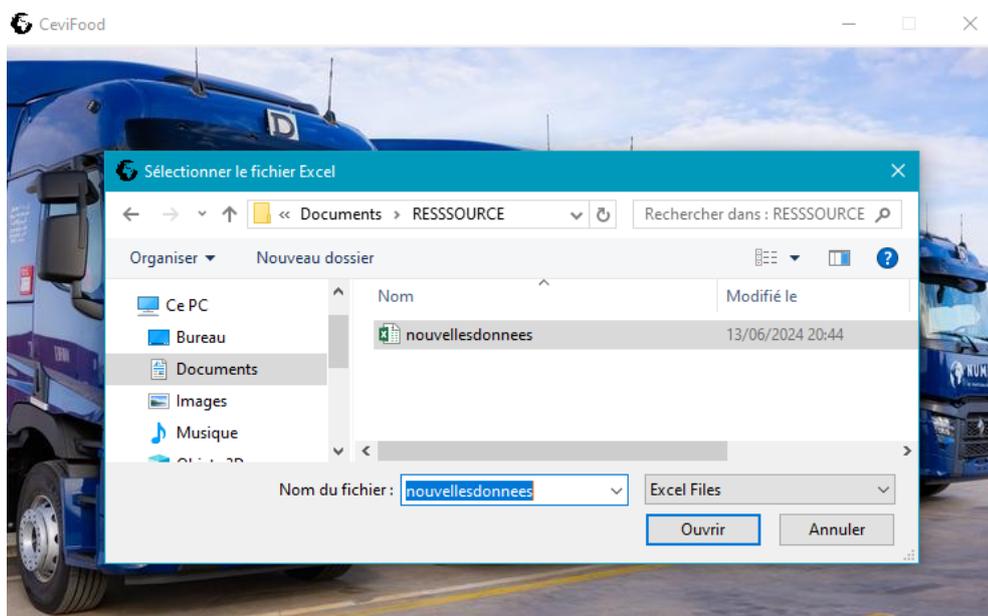


Figure 4.31: Fenêtre secondaire affichée par le deuxième bouton

- Les prévisions sont affichées de manière claire et concise, permettant aux utilisateurs de planifier efficacement le nombre d'OPL nécessaire.

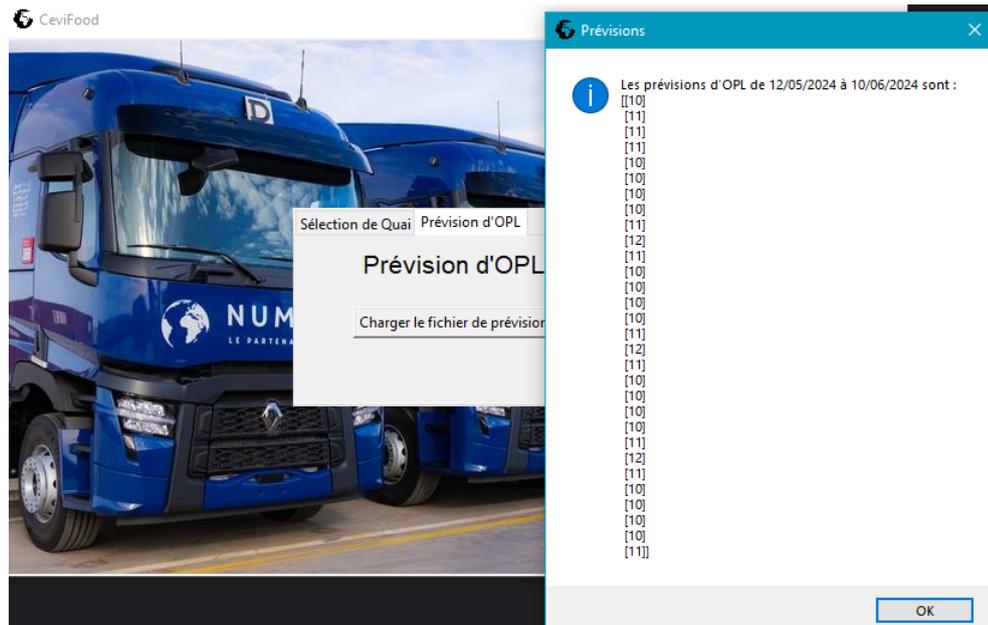


Figure 4.32: Résultat affiché par le deuxième bouton

4.6 Conclusion

Dans ce chapitre, nous avons fourni une description détaillée des outils utilisés pour prédire le nombre d'OPL nécessaires journaliers et pour choisir le meilleur quai dans l'entrepôt pour les opérations de chargement. Pour cela, nous avons commencé par mettre en lumière les bibliothèques spécifiques que nous avons utilisées pour développer notre solution. Ensuite, nous avons analysé et traité nos données provenant d'une entreprise afin de tester nos outils et trouver des solutions à nos problèmes.

Pour le problème de prévision des opérateurs, nous avons testé trois méthodes de machine learning : la régression linéaire, la forêt aléatoire (Random Forest) et XGBoost. Nous avons constaté qu'en termes d'erreur, la régression linéaire donnait des résultats plus précis, tandis qu'en termes de productivité, XGBoost était plus performant. Étant donné que la différence d'erreur entre la régression linéaire et XGBoost n'était pas significative, mais que XGBoost offrait une bien meilleure productivité, nous avons choisi d'utiliser la méthode XGBoost pour notre résolution.

Pour le choix du quai, nous avons créé un algorithme en Python visant à trouver le meilleur quai en termes de distance et de disponibilité. Enfin, nous avons détaillé l'utilité de notre interface utilisateur créée pour centraliser et faciliter l'accès aux résultats obtenus.

Conclusion générale

L'amélioration continue dans la logistique de distribution est devenue essentielle pour la survie des entreprises sur le marché. Elle constitue un axe de profitabilité et de réduction des coûts, tout en satisfaisant la demande des clients rapidement et avec une meilleure qualité. Dans ce cadre, nous avons eu l'opportunité d'appliquer nos connaissances théoriques et pratiques acquises durant notre cursus universitaire pour améliorer la productivité de l'entrepôt Numilog cas Cevifood.

Nous avons structuré notre travail en plusieurs étapes méthodiques. Tout d'abord, nous avons dressé une vue d'ensemble détaillée de l'entreprise pour mieux comprendre son environnement et son fonctionnement sur le terrain de notre stage. Cette première étape nous a permis d'acquérir une compréhension approfondie de l'organisation. Ensuite, nous avons identifié et analysé les principaux défis auxquels fait face l'entrepôt, incluant les prévisions de flux de marchandises, les prévisions du nombre d'OPL journaliers ainsi que l'allocation des quais de chargement. Cette analyse approfondie nous a conduit à formuler la problématique centrale de notre étude, orientée spécifiquement vers l'amélioration de la productivité au sein de l'entrepôt.

Par la suite, après avoir identifié ces défis, nous avons approfondi notre étude en nous appuyant sur la théorie. Nous avons exploré les concepts essentiels de la gestion d'entrepôt pour mieux appréhender le cadre d'application. En examinant diverses méthodes de résolution, nous avons étudié les meilleures pratiques et les stratégies efficaces utilisées pour obtenir des résultats performants. Cette exploration nous a permis d'élargir notre compréhension des principes fondamentaux nécessaires pour résoudre de manière efficace les défis spécifiques rencontrés par l'entrepôt.

Pour résoudre le problème de prévision de flux de marchandises, nous avons initié une analyse approfondie de nos données, mettant en lumière l'importance capitale de comprendre et de traiter nos données afin d'obtenir des prévisions précises. Nous avons exploré divers modèles traditionnels comme TBATS, le lissage exponentiel et Box-Jenkins, et après avoir évalué les mesures d'erreur appropriées, nous avons conclu que le modèle TBATS était le plus adapté pour calculer les prévisions de flux de marchandises.

Ensuite, pour anticiper le nombre nécessaire d'OPL, nous avons élaboré des solutions en commençant par une analyse approfondie et un traitement rigoureux de nos données. En utilisant les prévisions obtenues par TBATS nous avons appliqué des techniques d'apprentissage automatique notamment, la Régression Linéaire, XGBoost et Random Forest. Après avoir comparé leurs performances en termes d'erreur, nous avons observé que la Régression Linéaire montrait une légère amélioration, tandis que XGBoost offrait une productivité supérieure. En raison de cette efficacité accrue, nous avons choisi

d'adopter XGBoost pour notre résolution.

Nous avons également créé un algorithme pour améliorer l'allocation des quais de chargement, démontrant ces améliorations à travers un exemple concret d'étude de cas comparatif avec les pratiques actuelles de l'entreprise.

Enfin, après avoir relevé ces défis, la productivité de l'entrepôt a significativement progressé comparativement à l'approche précédente de l'entreprise. Cela nous a permis de présenter notre interface ainsi que ses fonctionnalités, expliquant comment l'entrepôt pourra en tirer parti.

Malgré ces progrès, des limites subsistent. Tout d'abord, la qualité et la quantité des données disponibles sont cruciales pour la précision des prévisions. De plus, l'adaptation de nos modèles aux changements imprévus de la demande ou aux conditions du marché constitue un défi majeur. Enfin, l'implémentation de modèles avancés nécessite des ressources technologiques et humaines significatives, ce qui peut être un obstacle pour certaines entreprises.

Cette recherche ouvre cependant la voie à de futures études prometteuses dans le domaine de la logistique de distribution. Ces études pourraient se concentrer sur plusieurs axes d'amélioration et d'approfondissement :

- **Gestion avancée des ressources matérielles :** Approfondir l'étude de la gestion proactive des ressources matérielles en utilisant les méthodes développées dans cette recherche. Par exemple, explorer comment prédire les pannes des machines à l'aide de modèles prédictifs basés sur les données historiques et en temps réel, afin d'optimiser la maintenance préventive et de réduire les temps d'arrêt imprévus.
- **Intégration de l'interface développée dans d'autres secteurs de Numilog:** Partager l'interface développée pour qu'elle soit utilisée dans d'autres entrepôts de Numilog en Algérie. Cette initiative vise à uniformiser les pratiques et à optimiser les processus logistiques. En adoptant ces solutions innovantes, Numilog pourrait améliorer la productivité et la qualité des services dans tous ses entrepôts en Algérie.

Annexes

A Tests utilisés dans les méthodes classiques

A.1 Test de Fisher et test de Buys-Ballot

Étant donné que ces tests ne sont pas disponibles sous R, nous avons dû créer une fonction personnalisée pour les effectuer.

```
p=7; N=59 #nombre de semaines
Fisher=function(x,N,p,alpha){
  m=matrix(x,N,p,byrow=T)
  mp=apply(m,2,mean)
  ma=apply(m,1,mean)
  sa= apply(m,1,sd)
  ss=summary(lm(sa~ma))
  pvalueB =coef(ss)[[2,4]]
  mg=mean(x)
  sp=(sum((mp-mg)^2))*N
  sa=(sum((ma-mg)^2))*p
  st=sum((m-mg)^2)
  sr=st-(sp+sa)
  vp=sp/(p-1)
  va=sa/(N-1)
  vr=sr/((p-1)*(N-1))
  fis=vp/vr
  fite=va/vr
  v1=p-1
  v2=(p-1)*(N-1)
  v3=N-1
  pvalues=1-pf(fis, v1, v2)
  pvaluet=1-pf(fite, v3, v2)
  m=matrix(c(fis,fite,pvalues,pvaluet),2,2)
  colnames(m)=c("Fisher","pvalue")
  rownames(m)=c("saisonnalité","tendance")
  if(pvalueB<alpha) déc="schéma multiplicatif" else déc="additif"
  if(pvalues<alpha)décision1=" série avec saisonnalité" else décision1="pas de
saisonnalité"
  if(pvaluet<alpha)décision2="série avec tendance" else décision2="pas de
tendance"
  list( Schéma_décomposition=pvalueB, décision=déc, Test_Fisher=m,
décision=c(décision1,décision2))
}
Fisher(serieEFF,59,7,0.05)
```

Figure A.1: Implémentation du test de Fisher et de Buys-Ballot

A.2 Test de Dickey-Fuller

Étant donné que notre modèle était de type série temporelle (ts), nous avons supprimé la tendance et la saisonnalité. Ensuite, nous avons réappliqué le test de Dickey-Fuller pour vérifier si la série était devenue stationnaire. Voici la démarche que nous avons suivie pour appliquer ce test :

```
#refaire le test ADF sur la série csv2 pour vérifier qu'elle est bien stationnaire
#Estimation du modèle 1
layout(matrix(1:2,1,2))
plot(csv2)
acf(csv2)
di2=diff(csv2)
pacf(di2)

#refaire le test ADF sur la série csv2 pour vérifier qu'elle est bien stationnaire
p2=19
#Estimation de modèle 3
library(CADFTest)
adft03 =CADFTest(csv2, max.lag.y = p2 , type = "trend")
summary(adft03)
# on a pvalue=0.594234 qui sup à 0.05 donc on rejette pas H0 donc on passe à l'estimation du modèle 2

#Estimation du modèle 2
adft2 =CADFTest(csv2, max.lag.y = p2, type = "drift")
summary(adft2)
#pvalue=0.856867 qui sup à 0.05 on ne rejette pas (c=0), donc on passe à l'estimation du modèle 1

#Estimation du modèle 1
adft1 =CADFTest(csv2, max.lag.y = p2, type = "none")
summary(adft1)
#pvalue=1.73e-05 qui inf à 0.05 donc on rejette H0 alors la série adft2 est bien stationnaire
```

Figure A.2: Test de Dickey-Fuller

B Prévisions de flux de marchandises en utilisant les méthodes classiques

B.1 SARMA

Date	Réelles	Prévisions
14/04/2024	1385	1875.768
15/04/2024	1032	1916.761
16/04/2024	1877	1816.051
17/04/2024	1653	1716.047
18/04/2024	2016	1649.252
19/04/2024	2082	1636.251
20/04/2024	1777	1438.202
21/04/2024	1337	1748.337
22/04/2024	1654	1753.212
23/04/2024	2091	1636.863
24/04/2024	2027	1531.848
25/04/2024	1555	1496.42
26/04/2024	1774	1522.71
27/04/2024	1542	1365.402
28/04/2024	1606	1714.578
29/04/2024	1591	1737.405
30/04/2024	1454	1637.753
01/05/2024	969	1538.974
02/05/2024	1131	1504.289
03/05/2024	1208	1523.295
04/05/2024	1493	1358.226
05/05/2024	1315	1700.965
06/05/2024	1450	1720.455
07/05/2024	1339	1618.357
08/05/2024	1172	1518.515
09/05/2024	1370	1487.868
10/05/2024	1495	1512.038
11/05/2024	1406	1352.31

Tableau B.1: 28 jours de prévisions du modèle SARMA

B.2 Lissage Exponentiel

Date	Réelles	Prévisions
14/04/2024	1385	1404.241
15/04/2024	1032	1769.181
16/04/2024	1877	1760.352
17/04/2024	1653	1741.794
18/04/2024	2016	1615.572
19/04/2024	2082	1616.784
20/04/2024	1777	13/9.398
21/04/2024	1337	1406.472
22/04/2024	1654	1771.993
23/04/2024	2091	1763.148
24/04/2024	2027	1744.561
25/04/2024	1555	1618.137
26/04/2024	1774	1619.35
27/04/2024	1542	1381.587
28/04/2024	1606	1408.704
29/04/2024	1591	1774.804
30/04/2024	1454	1765.945
01/05/2024	969	1747.327
02/05/2024	1131	1620.703
03/05/2024	1208	1621.917
04/05/2024	1493	1383.777
05/05/2024	1315	1410.936
06/05/2024	1450	1777.615
07/05/2024	1339	1768.742
08/05/2024	1172	1750.094
09/05/2024	1370	1623.268
10/05/2024	1495	1624.484
11/05/2024	1406	1385.966

Tableau B.2: 28 jours de prévisions du modèle Holt-Winters

B.3 TBATS

Date	Réelles	Prévisions
14/04/2024	1385	1465.345
15/04/2024	1032	1581.515
16/04/2024	1877	1508.159
17/04/2024	1653	1485.802
18/04/2024	2016	1640.993
19/04/2024	2082	1670.865
20/04/2024	1777	1523.712
21/04/2024	1337	1571.055
22/04/2024	1654	1672.581
23/04/2024	2091	1581.343
24/04/2024	2027	1541.076
25/04/2024	1555	1637.731
26/04/2024	1774	1624.074
27/04/2024	1542	1548.949
28/04/2024	1606	1597.434
29/04/2024	1591	1637.281
30/04/2024	1454	1576.189
01/05/2024	969	1572.805
02/05/2024	1131	1626.147
03/05/2024	1208	1602.530
04/05/2024	1493	1569.112
05/05/2024	1315	1605.539
06/05/2024	1450	1615.876
07/05/2024	1339	1579.838
08/05/2024	1172	1588.926
09/05/2024	1370	1614.963
10/05/2024	1495	1594.324
11/05/2024	1406	1582.523

Tableau B.3: 28 jours de prévisions du modèle TBATS

C Prévisions de nombre d'OPL en utilisant les méthodes de machine learning

C.1 Random Forest

y_test	y_pred
22	17
16	15
13	12
17	17
15	12
13	12
19	17
14	15
17	16
18	17
10	12
15	12
16	16
23	17
15	16
14	17
10	12
11	17
15	16
13	12
15	14
16	16
12	14
17	15
16	16
20	17
18	14
17	12
15	14
18	17
17	12
16	16

Tableau C.1: Résultat du modèle Random Forest

Date	Prévisions
14/04/2024	12
15/04/2024	12
16/04/2024	12
17/04/2024	12
18/04/2024	12
19/04/2024	12
20/04/2024	12
21/04/2024	12
22/04/2024	12
23/04/2024	12
24/04/2024	12
25/04/2024	12
26/04/2024	12
27/04/2024	12
28/04/2024	12
29/04/2024	12
30/04/2024	12
01/05/2024	12
02/05/2024	12
03/05/2024	12
04/05/2024	12
05/05/2024	12
06/05/2024	12
07/05/2024	12
08/05/2024	12
09/05/2024	12
10/05/2024	12
11/05/2024	12

Tableau C.2: 28 jours de prévisions du modèle Random Forest

C.2 XGBoost

<u>y_test</u>	<u>y_pred</u>
22	21
15	15
13	11
17	16
15	12
13	12
19	17
15	15
17	16
18	16
10	11
15	12
16	16
23	17
15	15
14	16
10	12
11	17
15	14
13	12
15	14
16	15
12	15
17	15
16	16
20	17
18	15
17	12
15	15
18	16
17	13
16	16

Tableau C.3: Résultat du modèle XGBoost

Date	Prévisions
14/04/2024	10]
15/04/2024	10
16/04/2024	10
17/04/2024	10
18/04/2024	11
19/04/2024	11
20/04/2024	10
21/04/2024	10
22/04/2024	11
23/04/2024	10
24/04/2024	10
25/04/2024	11
26/04/2024	11
27/04/2024	10
28/04/2024	10
29/04/2024	11
30/04/2024	10
01/05/2024	10
02/05/2024	11
03/05/2024	11
04/05/2024	10
05/05/2024	11
06/05/2024	11
07/05/2024	10
08/05/2024	10
09/05/2024	11
10/05/2024	10
11/05/2024	10

Tableau C.4: 28 jours de prévisions du modèle XGBoost

C.3 Régression Linéaire

<u>y_test</u>	<u>y_pred</u>
22	19
15	15
13	12
17	16
15	13
13	12
19	16
15	14
17	16
18	17
10	11
15	14
16	16
23	18
15	15
14	15
10	10
11	17
15	16
13	9
15	14
16	15
12	15
17	16
16	15
20	16
18	15
17	13
15	15
18	16
17	15
16	15

Tableau C.5: Résultat du modèle Régression Linéaire

Date	Prévisions
14/04/2024	12
15/04/2024	12
16/04/2024	12
17/04/2024	12
18/04/2024	13
19/04/2024	13
20/04/2024	12
21/04/2024	12
22/04/2024	13
23/04/2024	12
24/04/2024	12
25/04/2024	13
26/04/2024	13
27/04/2024	12
28/04/2024	12
29/04/2024	13
30/04/2024	12
01/05/2024	12
02/05/2024	13
03/05/2024	13
04/05/2024	12
05/05/2024	13
06/05/2024	13
07/05/2024	12
08/05/2024	12
09/05/2024	13
10/05/2024	12
11/05/2024	12

Tableau C.6: 28 jours de prévisions du modèle Régression Linéaire

Bibliographie

- [1] L'histoire du groupe. (s. d.). Cevital. Consulté 19 juin 2024, à l'adresse <https://www.cevital.com/lhistoire-du-groupe/>
- [2] Numilog | Entreprise Algérienne de transport et de logistique. (s. d.). Numilog: entreprise algérienne de transport et de logistique. Consulté 12 juin 2024, à l'adresse <https://numilog.dz/a-propos/>
- [3] NUMILOG. (s. d.). Cevital. Consulté 11 juin 2024, à l'adresse <https://www.cevital.com/numilog/>
- [4] Numilog. (2023). Présentation Numilog [Présentation PowerPoint]. Document interne.
- [5] Solutions | Numilog: Entreprise algérienne de transport. (s. d.). Numilog: entreprise algérienne de transport et de logistique. Consulté 12 juin 2024, à l'adresse <https://numilog.dz/solutions/>
- [6] Cevital déploie sa logistique de part et d'autre de la Méditerranée—Jeune Afrique.com. (s. d.). JeuneAfrique.com. Consulté 12 juin 2024, à l'adresse <https://www.jeuneafrique.com/mag/312084/economie-entreprises/cevital-deploie-logistique-de-part-dautre-de-mediterranee/>
- [7] Arif, J. (2016). L'externalisation des activités logistiques, vers la mise au point d'un outil d'aide à la décision [Phdthesis, Université Sidi Mohamed Ben Abdellah Fès (Maroc)].
- [8] Commissariat général au développement durable. (2012, juillet). Les entrepôts et leur activité en 2010 (n° 334).
- [9] Mohamed, R., Radi, B., Okar, C. (2016). Critical Success Factors for Warehousing Performance Improvement in Moroccan Companies. *International Journal of Business and Management Invention*, 5, 32-40.
- [10] Tompkins, J. A., White, J. A., Bozer, Y. A., Tanchoco, J. M. A. (2010). *Facilities Planning*. John Wiley Sons.
- [11] Carrera, S. (2010). Planification et ordonnancement des plateformes logistiques [Phdthesis, Institut National Polytechnique de Lorraine - INPL].

-
- [12] Roux, M., Liu, T. (2010). Optimisez votre plate-forme logistique: Exercices corrigés calcul des dimensions, des temps, des coûts CD-Rom pour l'audit (4e éd). Éd. d'Organisation.
- [13] Maréchal, A. (2020). Optimisation et réorganisation des espaces de stockage de la société Mery-Bois. Travail de fin d'études présenté en vue de l'obtention du grade de Bachelier en management de la logistique. Haute Ecole de la Province de Liège (HEPL), Seraing, Belgique.
- [14] Cazals, F., Cazals, C. (2019). Intelligence artificielle: L'intelligence amplifiée par la technologie. De Boeck Supérieur.
- [15] Turing, A. M. (2009). Computing Machinery and Intelligence. In R. Epstein, G. Roberts, G. Beber (Éds.), Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer (p. 23-65). Springer Netherlands.
- [16] Dilhac, M.-A., Abrassart, C., Voarino, N. (2018). Rapport de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle.
- [17] A Review of Machine Learning and Deep Learning Applications | IEEE Conference Publication | IEEE Xplore. (s. d.). Consulté 21 juin 2024, à l'adresse <https://ieeexplore.ieee.org/abstract/document/8697857>
- [18] Romero, M., Aloui, H., Heiser, L., Galindo, L., Lepage, A. (2021). Un bref parcours sur les ressources, pratiques et acteurs en IA et éducation. GTnum Scol_ia.
- [19] El Naqa, I., Murphy, M. J. (2015). What Is Machine Learning? In I. El Naqa, R. Li, M. J. Murphy (Éds.), Machine Learning in Radiation Oncology: Theory and Applications (p. 3-11). Springer International Publishing.
- [20] Coding Deep Learning For Beginners—Types of Machine Learning. (2019, février 15).<https://resources.experfy.com/ai-ml/coding-deep-learning-for-beginners-types-of-machine-learning/>
- [21] Ciaburro, G. (2017). MATLAB for Machine Learning. Packt Publishing Ltd.
- [22] Sutton, R. S., Barto, A. G. (s. d.). Reinforcement Learning: An Introduction.
- [23] Hiran, K. K., Jain, R. K., Lakhwani, D. K., Doshi, D. R. (2021). Machine Learning: Master Supervised and Unsupervised Learning Algorithms with Real Examples (English Edition). BPB Publications.
- [24] Qiang, W., Zhongli, Z. (2011). Reinforcement learning model, algorithms and its application. 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC), 1143-1146.
- [25] Aye, G. C., Balcilar, M., Gupta, R., Majumdar, A. (2015). Forecasting aggregate retail sales: The case of South Africa. International Journal of Production Economics, 160, 66-79.

-
- [26] Lagnoux, A. (s.d.). Renforcement statistique : Séries chronologiques. ISMAG, Université de Toulouse. MASTER 1 - MI0B246X. Récupéré de lagnoux@univ-tlse2.fr
- [27] Coutrot, B., & Droesbeke, J.-J. (1990). Les Méthodes de prévision. FeniXX.
- [28] Palm, R., & Brostaux, Y. (2009). Etude des séries chronologiques par les méthodes de décomposition.
- [29] Commission Inter-IREM. (s.d.). Statistique et probabilités : Brochure APMEP n° 156. ISBN 2-912846-32-3.
- [30] Khan, S., & Alghulaiakh, H. (2020). ARIMA Model for Accurate Time Series Stocks Forecasting. *International Journal of Advanced Computer Science and Applications*, 11(7).
- [31] Kalekar, P. S. (2004, 6 décembre). Time series forecasting using Holt-Winters exponential smoothing. Kanwal Rekhi School of Information Technology.
- [32] Time series analysis and prediction of scarlet fever incidence trends in Jiangsu Province, China : Using ARIMA and TBATS models | Research Square. (s. d.). Consulté 21 juin 2024, à l'adresse <https://www.researchsquare.com/article/rs-2259096/v1>
- [33] De Livera, A. M. (2010). Forecasting time series with complex seasonal patterns using exponential smoothing. Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia.
- [34] Mishra, P., Ray, S., Al Khatib, A., Abotaleb, M., Tiwari, S., Badr, A., & Balloo, R. (2021). Estimation of Fish Production in India using ARIMA, Holt's Linear, BATS and TBATS Models. *Indian Journal of Ecology*, 48, 1254-1261.
- [35] Yonar, H., Yonar, A., Mishra, P., Abotaleb, M., Khatib, A. M. G. A., Makarovskikh, T., & Cam, M. (2022). Modeling and forecasting of milk production in different breeds in Turkey.
- [36] Al Khatib, A. M. G., Yonar, H., Abotaleb, M., Mishra, P., Yonar, A., Karakaya, K., Badr, A., & Dhaka, V. (2021). Modeling and forecasting of egg production in India using time series models.
- [37] Borcard, D. (2006). Régression multiple - corrélation multiple et partielle. Département de sciences biologiques, Université de Montréal. Récupéré de via Google Scholar.
- [38] Saint-Cirgue, G. (2019). Apprendre Le Machine Learning en une semaine.
- [39] H, P. N. (2023, décembre 18). Intricacies of Multiple Linear Regression. Medium.
- [40] Caron, S. (s. d.). Une introduction aux arbres de décision.
- [41] Mathéi, A., Geurts, P., Marée, R., & Droogenbroeck, M. V. (s. d.). Classification automatique d'images biologiques par variantes d'arbres de décision.

-
- [42] Genuer, R., & Poggi, J.-M. (2017). Arbres CART et Forêts aléatoires, Importance et sélection de variables.
- [43] Random Forest. (s. d.). Corporate Finance Institute. Consulté 21 juin 2024, à l'adresse <https://corporatefinanceinstitute.com/resources/data-science/random-forest/>
- [44] Ali, Z. A., Abduljabbar, Z. H., Taher, H. A., Sallow, A. B., & Almufti, S. M. (2023). Exploring the Power of eXtreme Gradient Boosting Algorithm in Machine Learning : A Review. *Academic Journal of Nawroz University*, 12(2), Article 2.
- [45] Guo, R., Zhao, Z., Wang, T., Liu, G., Zhao, J., & Gao, D. (2020). Degradation state recognition of piston pump based on ICEEMDAN and XGBoost. *Applied Sciences*, 10, 6593.
- [46] Merigó, J. M., & Casanovas, M. (2011). A New Minkowski Distance Based on Induced Aggregation Operators. *International Journal of Computational Intelligence Systems*, 4(2), 123-133.
- [47] R: The R Project for Statistical Computing. (s. d.). Consulté 21 juin 2024, à l'adresse <https://www.r-project.org/>
- [48] Fichier:RStudio Logo.png. (2018). In Wikipédia. Consulté 21 juin 2024, à l'adresse https://fr.wikipedia.org/w/index.php?title=Fichier:RStudio_Logo.png&oldid=144122579
- [49] van Rossum, G. (2003). *An Introduction to Python* (F. L. Drake, Jr., Éd.). Network Theory Limited.
- [50] PyCharm : L'IDE Python pour la science des données et le développement web. (s. d.). Consulté 21 juin 2024, à l'adresse <https://www.jetbrains.com/fr-fr/pycharm/>
- [51] Welcome to Python.org. (2024, juin 14). Python.Org. Consulté 21 juin 2024, à l'adresse <https://www.python.org/>
- [52] NumPy -. (s. d.). Consulté 21 juin 2024, à l'adresse <https://numpy.org/>
- [53] McKinney, W. (2013). *Python for Data Analysis*. O'Reilly Media, Inc.
- [54] pandas—Python Data Analysis Library. (s. d.). Consulté 21 juin 2024, à l'adresse <https://pandas.pydata.org/>
- [55] Nelli, F. (2015). *Python Data Analytics : Data Analysis and Science using pandas, matplotlib and the Python Programming Language*.
- [56] Matplotlib—Visualization with Python. (s. d.). Consulté 21 juin 2024, à l'adresse <https://matplotlib.org/>
- [57] Hackeling, G. (2017). *Mastering Machine Learning with scikit-learn*. Packt Publishing Ltd.

-
- [58] Scikit-learn : Machine learning in Python—Scikit-learn 1.5.0 documentation. (s. d.). Consulté 21 juin 2024, à l'adresse <https://scikit-learn.org/stable/>
- [59] Test, training and validation sets. (2020, janvier 28). BrainsToBytes. Consulté 21 juin 2024, à l'adresse <https://www.brainstobytes.com/test-training-and-validation-sets/>
- [60] Mecalux. (s. d.). Équipements et technologies pour les quais de chargement. Consulté 23 juin 2024, à l'adresse <https://www.mecalux.fr/blog/technologies-quais-chargeement-camions>

ملخص

تهدف هذه الأطروحة إلى تحسين إنتاجية شركة نوميلوج، مقدمة الخدمات اللوجستية والنقل، من خلال محورين رئيسيين. تتضمن الخطوة الأولى استخدام أساليب التنبؤ التقليدية لتقدير تدفق البضائع. تشكل هذه التوقعات أساساً لتطبيق تقنيات الذكاء الاصطناعي لتوقع العدد الأمثل من العمال اللوجستيين اللازمين يوميًا.

الخطوة الثانية تركز على تطوير خوارزمية مخصصة لتحسين تخصيص الأرصفة داخل المستودع. الهدف هو اختيار الرصيف الأقرب لكل عملية إعداد الطلبات ووضع المخزون.

النتائج التي تم الحصول عليها تظهر تحسناً كبيراً في دقة التنبؤات بالتدفقات، وكذلك في التنبؤ بعدد المشغلين اللازمين. بالإضافة إلى ذلك، فقد قللت خوارزمية تخصيص الأرصفة من المسافة المقطوعة، مما حسن الكفاءة اللوجستية.

الكلمات الرئيسية: الذكاء الاصطناعي، التنبؤات، تخصيص الأرصفة، المستودع، الإنتاجية، بايثون، آر.

Résumé

Ce mémoire vise à améliorer la productivité de Numilog, prestataire logistique et de transport, à travers deux axes principaux. La première étape implique l'utilisation de méthodes classiques de prévision pour estimer les flux de marchandises. Ces prévisions servent de base pour l'application de techniques d'intelligence artificielle afin de prédire le nombre optimal d'opérateurs logistiques nécessaires chaque jour.

La seconde étape consiste au développement d'un algorithme dédié à l'optimisation de l'allocation des quais dans l'entrepôt. L'objectif est de sélectionner le quai le plus proche pour chaque opération de préparation de commande et de mise en stock.

Les résultats obtenus montrent une amélioration significative de la précision des prévisions de flux, ainsi que des prévisions du nombre d'opérateurs nécessaires. De plus, l'algorithme d'allocation des quais a réduit la distance parcourue, améliorant ainsi l'efficacité logistique.

Mots-clés : Intelligence artificielle, Prévisions, Allocation des quais, Entrepôt, Productivité, Python, R.

Abstract

This thesis aims to enhance the productivity of Numilog, a logistics and transportation service provider, through two main objectives. The first step involves using traditional forecasting methods to estimate goods flow. These forecasts form the basis for applying artificial intelligence techniques to predict the optimal number of logistics operators needed daily.

The second step focuses on developing an algorithm dedicated to optimizing dock allocation within the warehouse. The goal is to select the nearest dock for each order preparation and stock placement operation.

The results obtained show a significant improvement in the accuracy of flow forecasts, as well as in the prediction of the number of necessary operators. Additionally, the dock allocation algorithm has reduced the distance traveled, thereby improving logistical efficiency.

Keywords: Artificial Intelligence, Forecasts, Dock Allocation, Warehouse, Productivity, Python, R.