



UNIVERSITÉ CADI AYYAD
FACULTÉ POLY-DISCIPLINAIRE DE SAFI
DÉPARTEMENT DES SCIENCES
MATHÉMATIQUES ET INFORMATIQUE



Mémoire du projet de fin d'études

En vue de l'obtention du diplôme

Licence fondamentale

Sciences Mathématiques et Informatique

Sous le thème :

**Optimisation de la gestion des stocks
avec la Prédiction de la demande**

Réalisé par :

NMISSI NADIA

JAIFRA ICHRAK

Encadré par : Monsieur BOUDERBA SAIF ISLAM

Soutenu le 05/07/2024, devant les Membres du Jury :

- Pr. AFILAL Mounir
- Pr. ALIOUA Nawal

REMERCIEMENTS

En premier lieu, nous souhaitons exprimer notre gratitude envers ALLAH tout-puissant pour nous avoir donné la force, la patience et la détermination nécessaires pour mener à bien cette tâche.

Nous voulons également exprimer notre profonde reconnaissance envers le professeur BOUDERBA SAIF ISLAM pour l'honneur qu'il nous a accordé en acceptant de superviser ce travail. Ses conseils avisés, son soutien constant et sa direction précieuse ont été d'une aide inestimable tout au long de cette période.

Nous remercions chaleureusement les membres du jury, pour avoir accepté d'évaluer ce mémoire. Leurs compétences et leurs retours constructifs ont grandement contribué à l'amélioration de notre travail.

Nos remerciements vont également à tous nos enseignants qui ont façonné notre parcours depuis nos débuts à l'université. Leur engagement et leur enseignement ont été d'une importance inestimable tout au long de notre parcours académique.

Enfin, nous tenons à exprimer toute notre gratitude envers nos familles pour leur soutien indéfectible et leurs encouragements constants durant nos études. Leur amour et leur soutien sans conditions ont été une source constante d'inspiration et de motivation. Nous voulons exprimer notre gratitude à tous nos amis et à toutes les personnes qui ont participé de près ou de loin à l'accomplissement de cette tâche. Leur coopération et leur soutien ont été extrêmement précieux.

Ensemble, ces individus ont joué un rôle crucial dans la concrétisation.

RÉSUMÉ

Nous sommes tenues de réaliser un projet de fin d'étude en tant qu'étudiantes en Science Mathématiques et Informatiques. L'objectif de ce projet est de mettre en pratique les connaissances acquises au cours de notre formation en créant une outil pour optimisation de gestion des stocks avec la prédition de la demande.

Deux modules distincts font partie de notre outil de gestion des stocks. Nous les avons conçus pour répondre aux besoins spécifiques en termes de gestion des stocks et de prévision de la demande.

Le premier module est dédié à l'optimisation du système actuel d'approvisionnement régulier. Il permet une veille en temps réel des taux de stock, un contrôle efficace des commandes passées et son réapprovisionnement, ainsi qu'une analyse des tendances dans les demandes. Grâce à ce module, les entreprises peuvent mieux anticiper les besoins en stock, réduire les surstocks et/ou la rupture des approvisionnements tout en optimisant leur actifs engagés dans le stock.

Le deuxième module porte sur la prédition de la demande future pour les entreprises. En utilisant des techniques avancées de modélisation prédictive et d'analyse des données, qui s'appuient sur les données historiques des ventes, les tendances du marché, les événements saisonniers, et d'autres variables, ce module émet des prévisions fiables et précises. En résultat, les entreprises pourraient planifier leur production et les niveaux de stockage de manière plus efficace, réduisant le risque de surplus de stockage ou de dépassement de stockage.

Donc, notre outil constituera une solution complète sur la base des deux modules qui aideront les entreprises à satisfaire plus facilement les demandes du marché tout en opérationnelle. De plus, notre solution permettra aux entreprises de réduire les coûts grâce à une meilleure planification de leurs stocks.

TABLE DES MATIÈRES

1	<i>Introduction Générale</i>	7
2	La Gestion des Stocks	8
2.1	Introduction	8
2.2	Définitions des stocks et la gestion des stocks :	8
2.3	Les types des stocks	9
2.4	Importance de la gestion des stocks	11
2.5	Les problèmes de la gestion des stocks	11
2.6	Exemples des problèmes de la gestion des stocks	13
2.6.1	Le problème des stocks excédentaires de Nike	13
2.6.2	Le cauchemar des stocks de Noël de Best Buy	14
2.6.3	L'échec désastreux de l'expansion de Target au Canada	14
2.7	Solutions traditionnelles	15
2.8	Objectifs de l'étude	15
2.9	Conclusion	16
3	Contexte Générale	17
3.1	Introduction	17
3.2	Intelligence Artificielle	17
3.2.1	Définition	17
3.2.2	Les techniques de IA	18
3.3	L'apprentissage automatique (Machine learning)	22
3.3.1	Définition	22
3.3.2	Approches du Machine Learning	23
3.3.3	Avantages et inconvénients du Machine Learning	23
3.3.4	Types principales des modèles d'ensemble du machine learning	24
3.4	Conclusion	27
4	Méthodologie	28
4.1	Introduction	28
4.2	Collecte des données	28

4.2.1 Le dataset 1 :	29
4.2.2 Le dataset 2 :	30
4.3 Visualisation des données	31
4.4 Prétraitement des données	32
4.4.1 Le dataset 1 :	32
4.4.2 Le dataset 2 :	33
4.5 Matériel	35
4.5.1 Environnement de Développement	35
4.5.2 Langage de programmation	36
4.6 Conclusion	39
5 Résultats et discussion	40
5.1 Introduction	40
5.2 Prédiction de la demande	40
5.2.1 Modules de machine learning utilisés	41
5.2.2 Séparation des données en ensembles de formation, validation et test	41
5.2.3 L'entraînement et Évaluation des modèles de machine learning	42
5.3 Optimisation de la gestion des stocks	43
5.3.1 les méthodes et algorithmes	43
5.3.2 Optimisation de la gestion des stocks en pratique	45
5.4 Conclusion	48
6 Conclusion et Perspective	49

TABLE DES FIGURES

2.1 Ensemble des stocks	9
2.2 Gestion des stocks	9
2.3 Le déséquilibre entre l'offre et la demande	11
2.4 le sous-stockage	12
2.5 le sur-stockage	12
2.6 L'entreprise Nike	13
2.7 La boutique en ligne Best Buy	14
2.8 La marque Target	14
2.9 Objectif de l'étude	15
3.1 Les principaux sous-domaines de l'IA	18
3.2 Réseaux de neurones artificiels	18
3.3 Les principales étapes dans le processus d'un système de raisonnement par cas	19
3.4 Décision de catégorisation mentale	19
3.5 Schéma de principe d'un contrôleur Flou	20
3.6 Systèmes multi-agents	20
3.7 Domaines d'applications du NLP	21
3.8 Différentes méthodes d'optimisation	22
3.9 Machine Learning	22
3.10 prédictions du classificateur de vote dur	24
3.11 Moyenne	25
3.12 Empilage	25
3.13 Ensachage	26
3.14 Boostant	26
4.1 Extrait du dataset1	29
4.2 Les informations sur le dataset 1	29
4.3 Extrait du dataset 2	30
4.4 Les informations sur le dataset 2	31
4.5 Coût Total et Unités Vendues par Type d'Article de dataset1	31
4.6 Coût Total et la quantité de différents types de catégories de dataset2	31

4.7 Commandes par catégorie de dataset1	31
4.8 Commandes par catégorie de dataset2	31
4.9 Un bref aperçu des principales caractéristiques.	32
4.10 Extraction des caractéristiques temporelles	32
4.11 Normalisation des Variables Numériques	33
4.12 Suppression des colonnes inutiles et des colonnes datetime	33
4.13 Un aperçu rapide des principales caractéristiques des données	33
4.14 Extraction des caractéristiques temporelles	34
4.15 Encodage des Variables Catégorielles (garder les colonnes d'origine nécessaires)	34
4.16 Normalisation des Variables Numériques	34
4.17 Suppression des colonnes inutiles et des colonnes datetime	35
4.18 Colab	35
4.19 Anaconda	36
4.20 Jupyter Notebook	36
4.21 Python	37
4.22 NumPy	37
4.23 Pandas	38
4.24 Matplotlib	38
4.25 SKlearn	38
5.1 dataset 1	42
5.2 dataset 2	42
5.3 Comparaison résultat dataset 1	46
5.4 Comparaison résultat dataset 2	47

CHAPITRE 1

INTRODUCTION GÉNÉRALE

De nombreuses entreprises sont confrontées à un défi majeur : l'optimisation de la gestion des stocks. Cette tâche exige une planification précise et efficace pour minimiser les coûts tout en répondant à la demande des clients. La gestion des stocks est un élément crucial pour assurer la fluidité des opérations et la satisfaction des clients. Un stock mal géré peut entraîner des coûts supplémentaires importants et des pertes de ventes en raison de ruptures de stock ou d'un surstockage.

Dans ce contexte, l'utilisation de techniques d'intelligence artificielle (IA) pour prédire la demande est essentielle. Les méthodes traditionnelles de gestion des stocks, basées sur des prévisions manuelles et des approches statistiques simples, ne suffisent plus à répondre aux défis actuels. Les avancées technologiques et la disponibilité de vastes quantités de données ouvrent de nouvelles opportunités pour améliorer la précision des prévisions et l'efficacité des processus de gestion des stocks.

Ce projet se concentre sur l'exploration et la mise en œuvre de solutions innovantes pour optimiser la gestion des stocks. En utilisant des modèles prédictifs basés sur l'IA, nous cherchons à anticiper les besoins des clients et à réduire les risques de surstockage ou de rupture de stock. Pour y parvenir, nous analysons les tendances de la demande, les saisons, et d'autres facteurs qui influencent les achats des consommateurs. Nous combinons des algorithmes d'apprentissage automatique et des techniques de traitement avancé des données pour développer des modèles robustes et fiables.

L'objectif est de transformer la gestion des stocks en un processus plus adaptable, réactif et axé sur le marché. En intégrant ces informations dans les processus de planification des stocks, les entreprises peuvent ajuster leurs niveaux de stock en temps réel, améliorer leur efficacité opérationnelle et rester compétitives sur le marché. Cette approche vise à rendre les opérations plus rentables et plus efficaces, en réduisant les coûts de stockage et en augmentant la satisfaction des clients grâce à une meilleure disponibilité des produits.

En conclusion, ce projet vise à démontrer que l'intégration de l'IA dans la gestion des stocks peut révolutionner les pratiques actuelles. En adoptant ces nouvelles technologies, les entreprises peuvent non seulement améliorer leur gestion des stocks, mais aussi gagner un avantage concurrentiel significatif. La réussite de ce projet pourrait servir de modèle pour d'autres entreprises cherchant à optimiser leurs opérations et à s'adapter aux exigences d'un marché de plus en plus dynamique et complexe.

CHAPITRE 2

LA GESTION DES STOCKS

2.1 Introduction

Dans la nouvelle mondialité, la gestion des stocks est un abord grave de la gestion d'un commerce, qu'il sans large ou petite. Pour les bambins usines en particulier, une gérance épais des stocks peut avantagez à maximiser le destination des argent limitées et à tenter la rentabilité. En effet, une gestion des stocks bien planifiée permet de minimiser les coûts tout en garantissant la disponibilité des produits pour répondre à la demande des clients. Dans notre projet, nous allons explorer les raisons pour lesquelles la gestion des stocks est importante pour les petits commerces, ainsi que les outils et les meilleures pratiques qu'ils peuvent utiliser pour améliorer leur gestion des stocks.

2.2 Définitions des stocks et la gestion des stocks :

— Définition des stocks :

le stock est L'ensemble des marchandises, des matières, des produits semi-ouvrés et finis, etc..., appartenant à une entreprise, constitue son stock (1). Selon André RAMBAUX, le stock se définit comme un ensemble de biens ou d'articles accumulés pour une utilisation ultérieure (fig :2.1). Ce stock permet de répondre aux besoins des utilisateurs dès qu'ils en ont besoin, tout en les protégeant contre les contraintes des délais et les fluctuations des fournisseurs de livraison. Le stock se situe entre deux cycles, la fabrication et la vente, ou une autre opération de fabrication, et est modifié par chacun de ces cycles ainsi que parfois par ceux en amont et en aval; qui sont la propriété d'une entreprise (2).

— Définition de la gestion des stocks :

La gestion des stocks comprend les processus et pratiques utilisés par les entreprises pour contrôler et superviser efficacement le flux des stocks (fig :2.2), de l'achat à l'utilisation ou à la vente, dans le but d'optimiser les coûts tout en maintenant des niveaux de service satisfaisant les clients (3; 4; 5).

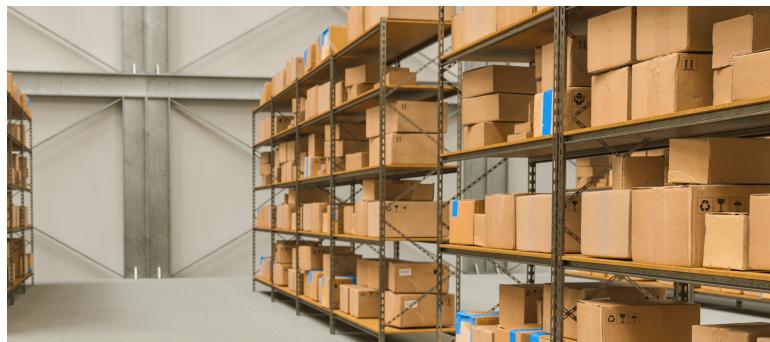


FIGURE 2.1 – Ensemble des stocks



FIGURE 2.2 – Gestion des stocks

2.3 Les types de stocks

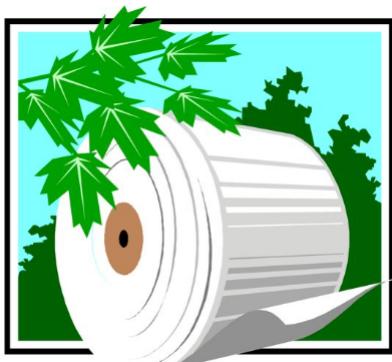
Les types de stocks peuvent varier selon l'industrie, les besoins spécifiques de l'entreprise, et la stratégie de gestion de la chaîne d'approvisionnement. D'après (6) voici les principaux types de stocks rencontrés dans la pratique de la logistique :

- Stock de matières premières :

Ce type de stock comprend les matériaux bruts nécessaires à la production. Il est essentiel pour les entreprises de fabrication qui dépendent de l'approvisionnement continu en matières premières pour maintenir leurs opérations de production.



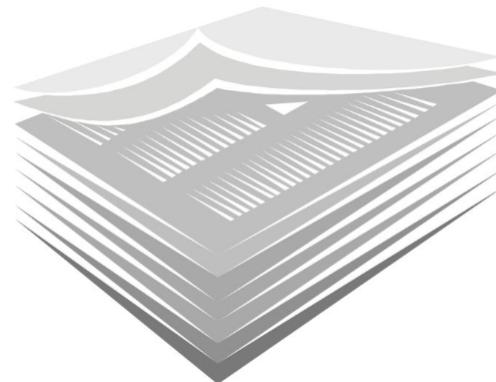
- Stock en cours de fabrication (ou en-cours) :



Ce stock inclut les matériaux et composants qui sont en train d'être transformés en produits finis. Il représente une étape intermédiaire dans le processus de production.

— Stock de produits finis :

Il s'agit de produits complètement manufacturés et prêts à être vendus ou distribués au client final. La gestion efficace de ce type de stock est nécessaire pour répondre rapidement à la demande des clients sans encourir des coûts de stockage excessifs.



— Stock de marchandises :



Pour les entreprises de distribution ou de revente, ce stock consiste en des produits finis acquis auprès de fournisseurs et destinés à être vendus directement aux consommateurs.

— Stock de sécurité (ou stock tampon) :

Ce stock est maintenu pour se prémunir contre les incertitudes de la demande et de l'approvisionnement. Il permet à l'entreprise de continuer à fonctionner et de satisfaire la demande des clients en cas de retard d'approvisionnement ou de hausse imprévue de la demande comme représenté dans (7).

— Stock saisonnier :

Certains produits ont une demande qui varie selon la saison. Le stock saisonnier est constitué en prévision de ces variations pour s'assurer que la demande peut être satisfaite.

— Stock spéculatif :

Les entreprises peuvent accumuler ce type de stock en prévision d'une augmentation des prix, de pénuries de matériaux ou de toute autre situation qui pourrait affecter négativement l'approvisionnement ou les coûts.

— Stock mort :



Il s'agit de stock qui n'est plus utilisable ou qui ne peut être vendu en raison de l'obsolescence, du dépassement de la date de péremption, ou de changements dans la demande du marché. La gestion efficace des stocks vise à minimiser ce type de stock pour réduire les pertes.

2.4 Importance de la gestion des stocks

On constate que Les stocks peuvent constituer l'actif le plus important d'une entreprise. La gestion des stocks est le point vers lequel convergent tous les éléments de la chaîne d'approvisionnement. Des stocks insuffisants peuvent entraîner le mécontentement des clients. Mais des stocks importants ont leurs propres inconvénients (8) :

- le coût de l'entreposage et de l'assurance, ainsi que le risque de perte, de vol et de détérioration. Les entreprises qui ont des chaînes d'approvisionnement et des procédés de fabrication complexes doivent trouver le bon équilibre entre excédents de stock et insuffisance de stocks ce qui montre la figure (fig :2.3).



FIGURE 2.3 – Le déséquilibre entre l'offre et la demande

2.5 Les problèmes de la gestion des stocks

Cette composante du commerce peut s'avérer très pénible surtout si l'on ne possède pas système de gestion automatisé. De nombreuses problématiques peuvent découler de la gestion des stocks. Il est donc essentiel de les anticiper afin de prévenir les risques d'erreur et maintenir ses stocks à jour (8).



Le sous-stockage :

Cette notion décrit le fait de ne pas stocker suffisamment, vous possédez alors une quantité de stock trop faible par rapport à la demande, celle-ci peut désorganiser totalement votre boutique. Certes, le sous-stockage permet d'avoir davantage de trésorerie mais vous ôte des opportunités de vente. Les conséquences peuvent être dramatique : manque de produits, impossibilité de faire face à la demande, insatisfaction des clients, baisse de notoriété, perte de clients et difficulté à lutter contre la concurrence. Les clients se déplacent en magasin pour acquérir le produit de suite, être en rupture de stock a donc un impact très négatif sur l'image de votre commerce et pousse les consommateurs à aller chez vos concurrents.



FIGURE 2.4 – le sous-stockage

Le sur-stockage

A l'inverse une quantité de stock trop élevée entraîne une forte augmentation des charges financières. En effet, le sur-stockage impacte les charges fixes avec une augmentation des coûts liés au locaux, rangement, ou encore aux machines ; et des charges variables notamment concernant les frais d'électricité, de loyer, de personnel ou encore d'entretien. À savoir, le sur-stockage peut de même, entraîner une dévaluation des stocks avec des risques de détérioration, vol, perte... Cette situation est souvent signe pour l'entreprise d'une mauvaise santé financière.



FIGURE 2.5 – le sur-stockage

Les erreurs d'inventaire

On peut expliquer les erreurs d'inventaire par les raisons suivantes : Les pertes ou vols de produits, pour cela, il est nécessaire de compatibiliser régulièrement les produits de votre boutique et les répertorier dans un fichier informatisé. Légalement vous devez mener un inventaire total minimum une fois par an, toutefois nous vous conseillons de faire fréquemment des inventaires partiels notamment sur les produits à risque et tenir ainsi votre inventaire à jour. La mauvaise gestion de la comptabilité des stocks, le mieux pour faire face à ce problème est de se munir d'un logiciel intuitif permettant de mettre à jour les stocks simplement.

2.6 Exemples des problèmes de la gestion des stocks

D'après [9], diverses entreprises ont connu des défis qui ont affecté la performance et la rentabilité de l'entreprise.

2.6.1 Le problème des stocks excédentaires de Nike

L'entreprise Nike a eu des difficultés avec ses stocks. Dans les années 2000, l'entreprise a perdu environ 100 millions de dollars de ventes à cause de problèmes de suivi des marchandises. Pour résoudre cela, ils ont utilisé un nouveau logiciel de gestion des stocks. Cependant, ce logiciel a eu des bugs et des erreurs de données, ce qui a conduit à des prévisions de demande incorrectes et à des pertes supplémentaires.

Ce problème montre l'importance de bien gérer les stocks et votre système de gestion des stocks. Lorsque vous choisissez un logiciel de gestion des stocks, il est important de s'assurer qu'il est précis, flexible et adapté à votre entreprise. Il doit pouvoir s'adapter aux changements de votre entreprise et de votre clientèle.

Nike continue de rencontrer des problèmes avec ses stocks. En 2016, la marge brute de Nike a diminué en raison de ventes à prix réduits causées par des problèmes de gestion des stocks. Ils essaient maintenant d'améliorer leur fabrication et d'utiliser de nouvelles technologies pour améliorer leurs stocks. Malgré ces défis, Nike reste un leader mondial en innovant et en explorant de nouveaux marchés.



FIGURE 2.6 – L'entreprise Nike

2.6.2 Le cauchemar des stocks de Noël de Best Buy

En décembre 2011, pendant la période de Noël, Best Buy a dit que beaucoup de gens voulaient acheter des produits sur leur site web. Cela a créé des problèmes pour eux car ils n'ont pas pu satisfaire toutes les commandes en ligne à temps. Ils ont dit aux clients désolé pour ça. Les clients étaient en colère parce que Best Buy a annulé les commandes au lieu de les envoyer plus tard. Cela était probablement parce qu'ils n'avaient pas assez de produits. Il semble que Best Buy a vendu beaucoup d'articles lors du Black Friday, mais ils n'avaient pas assez en stock pour tout le monde.



FIGURE 2.7 – La boutique en ligne Best Buy

Comme l'exemple de Best Buy l'a montré, acheter des articles en ligne au lieu de les acheter en magasin peut comporter des risques. Les clients ne sont pas toujours sûrs de recevoir leur produit comme prévu. Même s'ils reçoivent un reçu et des estimations de livraison, il reste toujours quelques incertitudes quant à la finalisation de leur achat à chaque fois. On peut penser que Best Buy a probablement perdu beaucoup de clients au profit d'Amazon après les problèmes rencontrés en 2011.

2.6.3 L'échec désastreux de l'expansion de Target au Canada

Target est une marque populaire aux États-Unis. Donc, quand elle a décidé de s'étendre au nord du Canada, cela semblait logique qu'elle soit aussi bien accueillie là-bas. Les responsables de Target devaient prendre une décision importante. Ils devaient trouver un moyen de surveiller leurs stocks. Ils ont choisi d'utiliser un tout nouveau système qui n'avait jamais été testé auparavant. Target Canada a alors découvert ce qui se passe lorsque des employés inexpérimentés doivent ouvrir un magasin en utilisant une technologie que personne, même au siège social aux États-Unis, ne connaissait.



FIGURE 2.8 – La marque Target

2.7 Solutions traditionnelles

1. **Rotation des Stocks** : Il est important de changer régulièrement les stocks pour éviter que les produits deviennent vieux et pour réduire les coûts de stockage. Utiliser le principe "Premier Entré, Premier Sorti" (PEPS) aide à utiliser les produits les plus anciens en premier. Il est aussi utile de vérifier régulièrement les produits qui ne se vendent pas bien et de faire des promotions pour les vendre plus vite.
2. **Automatisation des Processus** : Automatiser les commandes, le réapprovisionnement et le suivi des stocks aide à réduire les erreurs humaines et à rendre les opérations plus efficaces. Utiliser des logiciels de gestion des stocks, des scanners de codes-barres et des technologies comme la RFID permet de suivre les mouvements de stock de manière précise et rapide.
3. **Utilisation de la Technologie** : Investir dans des systèmes de gestion des stocks avancés est très important pour suivre les niveaux de stock en temps réel et améliorer les opérations. Ces systèmes donnent une vue complète de l'inventaire et permettent de faire des prévisions plus précises sur la demande. Des technologies comme l'intelligence artificielle peuvent aider à prévoir les tendances de consommation et à ajuster les niveaux de stock en conséquence.
4. **Réapprovisionnement automatique** : Mettre en place des seuils de réapprovisionnement pour les produits est une bonne pratique pour garder des niveaux de stock optimaux. Quand le stock atteint un niveau minimum, une commande d'approvisionnement est automatiquement lancée pour s'assurer que les articles nécessaires sont toujours disponibles. Cela aide à répondre rapidement aux changements de demande et à éviter les interruptions dans les opérations. Utiliser des systèmes de gestion des stocks pour gérer ces seuils peut aussi aider à optimiser les coûts et à améliorer la satisfaction des clients.

2.8 Objectifs de l'étude

Nous visons dans cette étude à améliorer l'efficacité et la rentabilité en anticipant les besoins des clients. En utilisant des modèles de machine learning, les entreprises peuvent analyser les tendances passées et actuelles pour prédire avec précision la demande future. Cela leur permet de mieux planifier leurs approvisionnements, de réduire les coûts liés aux stocks excédentaires et de répondre de manière plus efficace aux fluctuations du marché. En fin de compte, cette approche vise à garantir que les produits sont disponibles lorsque les clients en ont besoin, tout en minimisant les pertes et en maximisant les bénéfices.



FIGURE 2.9 – Objectif de l'étude

2.9 Conclusion

La gestion efficace des stocks est essentielle pour assurer le bon fonctionnement d'une entreprise. En comprenant les définitions et en reconnaissant l'importance de cette pratique, il est clair que les problèmes de gestion des stocks peuvent entraîner des conséquences financières désastreuses. Les exemples réels de Nike, Best Buy et Target soulignent l'impact significatif que de mauvaises décisions en matière de stocks peuvent avoir sur une entreprise. Alors que les solutions traditionnelles offrent une certaine aide, l'intégration de l'intelligence artificielle dans la gestion des stocks offre de nouvelles perspectives pour améliorer l'efficacité, réduire les coûts et prévenir les problèmes futurs. En combinant les connaissances établies avec les avancées technologiques, l'avenir de la gestion des stocks est prometteur grâce à l'intelligence artificielle.

CHAPITRE 3

CONTEXTE GÉNÉRALE

3.1 Introduction

Depuis des dizaines d'années, notre monde a connu de nombreuses avancées technologiques, mais une avancée nous intéresse particulièrement : L'intelligence artificielle. Des scientifiques, notamment Alan Turing, John McCarthy, Marvin Minsky, Frank Rosenblatt, sont parvenus à la créer. Il y a plusieurs domaines d'applications où l'intelligence artificielle apparaît, le plus souvent au sein des jeux vidéo mais aussi dans d'autres domaines tels que la médecine, l'armée, le transport.

3.2 Intelligence Artificielle

3.2.1 Définition

L'intelligence artificielle (IA) est un domaine qui regroupe un ensemble de théories et de techniques visant à créer des systèmes informatiques capables de reproduire les facultés intellectuelles humaines. Elle utilise des méthodes de résolution de problèmes à forte complexité, que ce soit d'un point de vue logique ou algorithmique. En termes plus accessibles, elle consiste à créer des programmes informatiques intelligents capables de s'adonner à des tâches qui demandent des processus mentaux de haut niveau, tels que l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique. Malgré les fantasmes, les inquiétudes et les débats philosophiques qu'elle suscite, l'IA continue d'évoluer et offre des solutions innovantes dans de nombreux domaines.

Voici quelques-uns des principaux sous-domaines de l'IA :

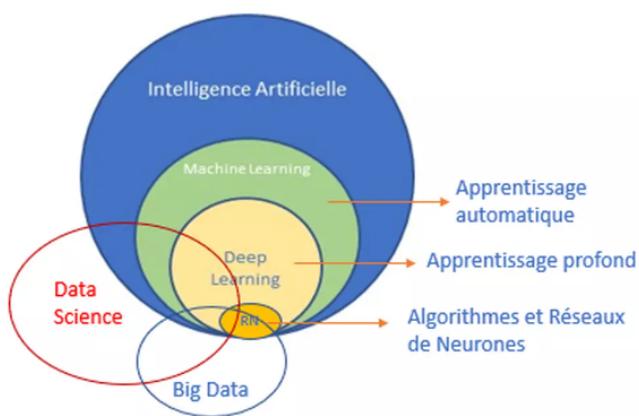


FIGURE 3.1 – Les principaux sous-domaines de l'IA

3.2.2 Les techniques de IA

- **Réseaux de neurones artificiels (Artificial neural networks)** : sont des modèles informatiques conçus en s'inspirant du fonctionnement du cerveau humain et sont employés dans des tâches telles que la reconnaissance de motifs, la classification, la prédiction, et bien d'autres encore.
Les réseaux de neurones artificiels sont constitués de couches nodales, comprenant une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie. Chaque nœud, ou neurone artificiel, se connecte à un autre et possède un poids et un seuil associés. Voici comment ils fonctionnent :
Entrée : Les données sont introduites dans la couche d'entrée.
Poids : Chaque connexion entre les neurones a un poids qui détermine l'importance de cette connexion.
Somme pondérée : Les entrées sont multipliées par leurs poids respectifs et additionnées.
Fonction d'activation : La sortie est transmise via une fonction d'activation qui détermine si le neurone est activé.
Propagation avant : La sortie d'un neurone devient l'entrée du neurone suivant.
Grâce à des techniques d'apprentissage supervisé ou non supervisé, les ANN peuvent apprendre à partir de données pour ajuster les poids des connexions entre les neurones et améliorer leurs performances sur des tâches spécifiques.

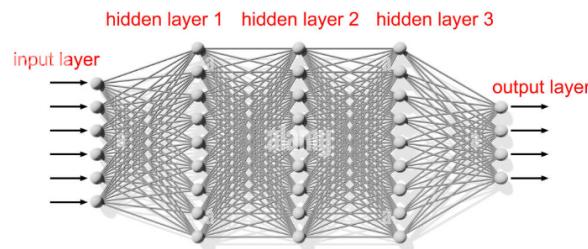


FIGURE 3.2 – Réseaux de neurones artificiels

- **Raisonnement basé sur des cas (Case-based reasoning)** : En fonction de (10)

Le raisonnement à partir de cas (RàPC), aussi appelé case-based reasoning (CBR) en anglais, est une forme de raisonnement qui s'inspire du processus cognitif humain. Il implique l'utilisation de l'expérience passée pour résoudre des problèmes rencontrés dans la vie quotidienne. Cette méthode consiste

à se souvenir de situations similaires déjà rencontrées et à les comparer à la situation actuelle afin de construire une nouvelle solution adaptée.

Un système RàPC dispose d'une base de cas. Chaque cas possède une description et une solution. Pour utiliser ces informations, un moteur est aussi présent. Celui-ci va retrouver les cas similaires au problème posé. Après analyse, le moteur fournit une solution adaptée qui doit être validée. Enfin le moteur ajoute le problème et sa solution dans la base de cas.

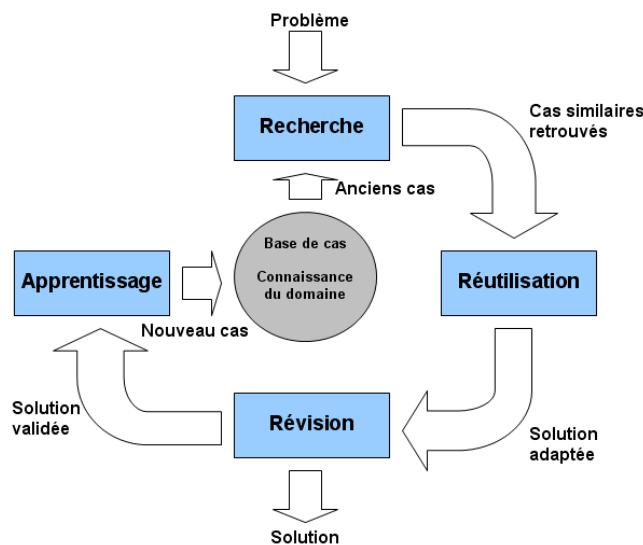


FIGURE 3.3 – Les principales étapes dans le processus d'un système de raisonnement par cas

- **Cartographie cognitive (Cognitive mapping) :** La cartographie cognitive, également connue sous le nom de carte mentale, est une technique qui permet de représenter graphiquement les connaissances, les perceptions et les relations mentales d'un individu sur un sujet donné. L'objectif principal de la cartographie cognitive est de visualiser les différentes idées, concepts et informations associés à un domaine spécifique, ainsi que les connexions entre eux.

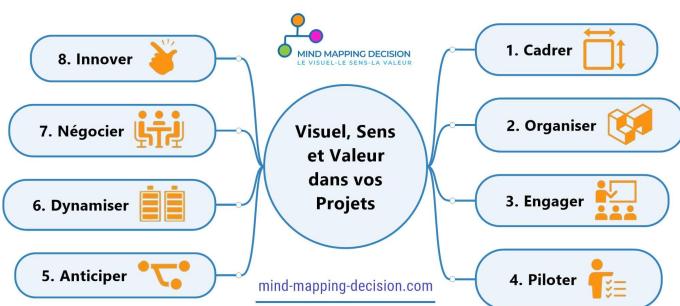


FIGURE 3.4 – Décision de catégraphie mentale

- **Logique floue (Fuzzy logic) :** est une sous-discipline de la logique qui vise à modéliser des concepts et des processus où les limites entre les catégories traditionnelles sont floues ou non définies. Contrairement à la logique binaire classique, qui considère les propositions comme étant soit vraies soit fausses, la logique floue permet de traiter des degrés de vérité, ce qui la rend mieux adaptée pour aborder des situations complexes et incertaines.

En logique floue, les variables peuvent prendre des valeurs sur un continuum, et les propositions peuvent être partiellement vraies ou partiellement fausses en fonction de la mesure dans laquelle elles satisfont certains critères. Cette flexibilité permet de mieux représenter des concepts tels que le flou linguistique, les approximations et les jugements subjectifs.

La logique floue trouve une large application dans de nombreux secteurs, notamment le contrôle des systèmes, la reconnaissance de formes, la prise de décision, la modélisation linguistique et l'intelligence artificielle. Grâce à sa capacité à gérer des données incomplètes, imprécises ou sujettes à des variations, elle offre une solution efficace pour résoudre une variété de problèmes.

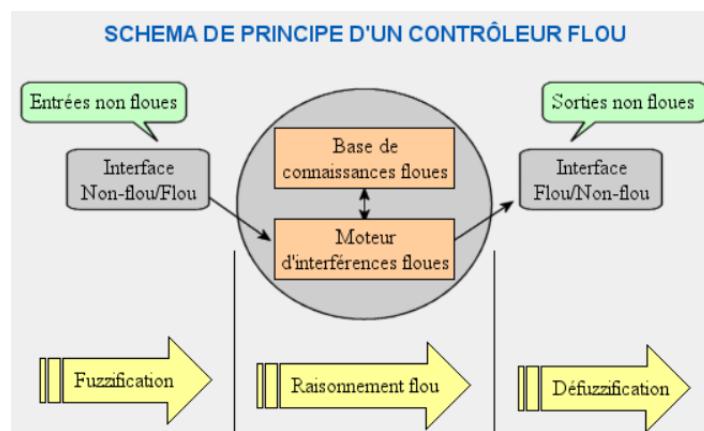


FIGURE 3.5 – Shéma de principe d'un contrôleur Flou

- **Systèmes multi-agents (Multi-agent systems)** : est un système composé d'un ensemble d'agents, telles que des processus, des robots, des êtres humains ou même des organismes biologiques comme les fourmis, qui sont actifs au sein d'un environnement donné et interagissent selon des règles prédéfinies. Chaque entité, ou agent, dans ce système est caractérisée par son autonomie, ce qui signifie qu'elle peut prendre des décisions de manière indépendante, sans nécessiter de contrôle centralisé sur l'ensemble du système.

Ces systèmes multi-agents représentent un domaine de recherche essentiel en intelligence artificielle distribuée, offrant une approche captivante pour modéliser les interactions complexes observées dans les sociétés humaines et animales. Leurs applications s'étendent à divers domaines, y compris les sciences humaines, où ils permettent la représentation et la simulation de phénomènes sociaux et comportementaux. Ce contenu est inspiré de (11).

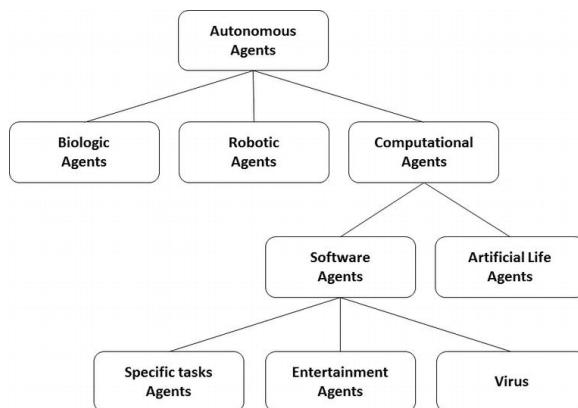


FIGURE 3.6 – Systèmes multi-agents

— **Traitement automatique du langage naturel (Neural language processing)** : Basé sur (12)

Le traitement automatique du langage naturel (TALN), aussi appelé neural language processing en anglais, qui se concentre sur la capacité des ordinateurs à comprendre, interpréter et générer un langage humain de manière naturelle. Ce domaine utilise des méthodes informatiques pour analyser et manipuler de vastes ensembles de données textuelles dans le but d'extraire des informations pertinentes et de réaliser diverses tâches telles que la traduction automatique, la génération de texte, la classification de documents, l'analyse des sentiments, entre autres.

Le traitement automatique du langage naturel (TALN) fréquemment recourt à des architectures basées sur des réseaux neuronaux, incluant les réseaux de neurones récurrents (RNN) et les transformateurs, afin d'analyser et de comprendre le langage humain de façon plus précise et élaborée.

Le TALN est utilisé dans divers domaines tels que la recherche d'informations, l'analyse des réseaux sociaux, l'assistance virtuelle, la compréhension des requêtes des utilisateurs, la génération de contenu, etc. Son utilisation croît avec les progrès de l'IA et de l'apprentissage automatique, ouvrant de nouvelles perspectives pour l'interaction homme-machine et l'automatisation des tâches linguistiques.

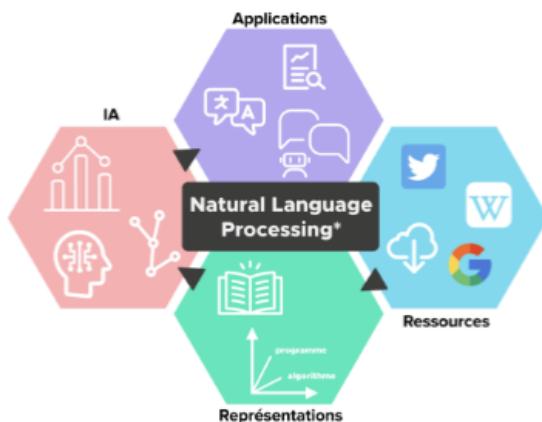


FIGURE 3.7 – Domaines d'applications du NLP

— **Algorithme génétique (Genetic algorithm)** : L'algorithme génétique, créé dans les années 70 par John H. Holland, est une méthode d'optimisation basée sur le principe de la sélection naturelle. Appartenant à la catégorie des algorithmes évolutionnistes, il utilise des algorithmes générés aléatoirement dans un environnement spécifié, où ils interagissent les uns avec les autres et avec cet environnement. Ils peuvent assimiler les informations présentes dans cet environnement et évoluer en conséquence, ou "muter".

Ces algorithmes sont essentiellement aléatoires et ne sont pas conçus pour exécuter des tâches spécifiques. Ils existent simplement et évoluent en fonction des éléments intégrés dans l'environnement. Ils sont ensuite soumis à une sélection basée sur des critères de performance. Seuls les algorithmes les mieux notés sont conservés et confrontés à de nouveaux algorithmes générés aléatoirement. Ce processus de mutation et de sélection est répété de génération en génération sur une période définie. Au fil du temps, grâce à ce processus de mutation-sélection, des algorithmes de plus en plus performants émergent de manière quasi spontanée (13; 14).

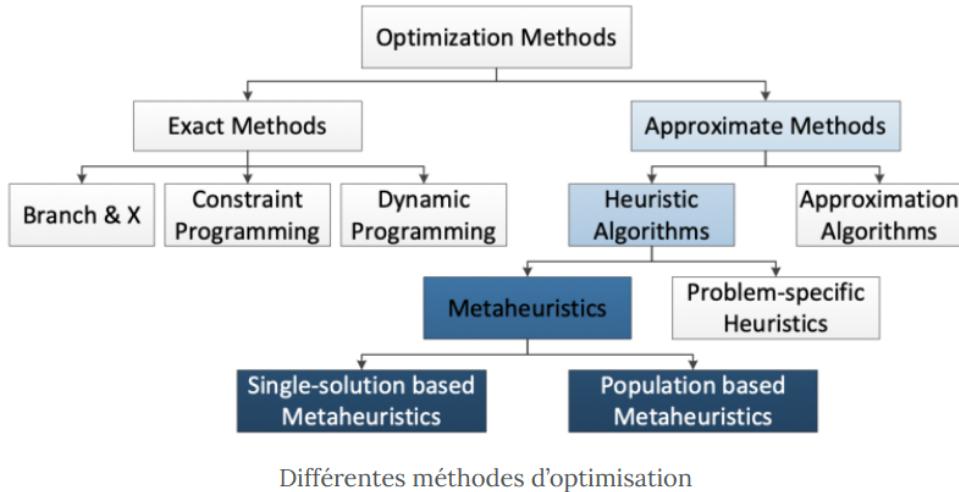


FIGURE 3.8 – Différentes méthodes d'optimisation

3.3 L'apprentissage automatique (Machine learning)

3.3.1 Définition

L'apprentissage automatique (machine learning en anglais) est un champ d'étude de l'IA qui vise à donner aux machines la capacité d'« apprendre » à partir de données, via des modèles mathématiques. Plus précisément, il s'agit du processus par lequel des informations pertinentes sont extraites d'un ensemble de données d'entraînement.

Le but de cette phase est l'obtention des paramètres d'un modèle qui atteindront les meilleures performances, notamment lors de la réalisation de la tâche attribuée au modèle. Une fois l'apprentissage réalisé, le modèle pourra ensuite être déployé en production. En basée sur (15; 16).

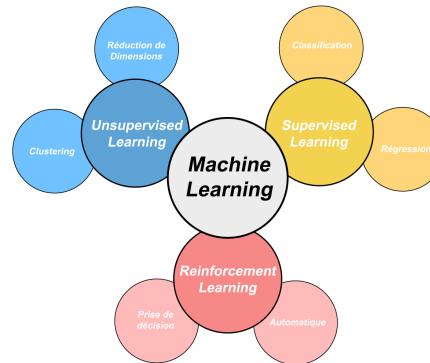


FIGURE 3.9 – Machine Learning

3.3.2 Approches du Machine Learning

— Apprentissage supervisé :

L'apprentissage supervisé, parfois désigné sous le terme d'apprentissage automatique supervisé, est l'utilisation d'ensembles de données étiquetées pour entraîner des algorithmes qui permettent de classer des données ou de prédire des résultats avec précision. Lorsque de nouvelles données d'entrée sont introduites dans le modèle, celui-ci ajuste ses paramètres jusqu'à ce qu'il soit correctement adapté aux données fournies. Cette opération se produit dans le cadre du processus de validation croisée, afin que le modèle évite le sur-ajustement ou le sous-ajustement.

— Apprentissage non supervisé :

L'apprentissage non supervisé, ou apprentissage automatique non supervisé, utilise des algorithmes d'apprentissage automatique pour analyser et regrouper des jeux de données non étiquetées. Ces algorithmes découvrent des modèles cachés ou des regroupements de données sans nécessiter d'intervention humaine. La capacité de l'apprentissage automatique non supervisé à détecter des similarités et des différences dans les données en fait une solution privilégiée pour des tâches telles que l'analyse exploratoire des données, la recommandation de produits complémentaires, la segmentation de la clientèle et la reconnaissance de motifs visuels. En outre, il contribue à simplifier les modèles en réduisant le nombre de caractéristiques grâce à des techniques de réduction de la dimensionnalité.

— Apprentissage par renforcement :

L'apprentissage automatique par renforcement, semblable à l'apprentissage supervisé mais se distinguant par le fait que l'algorithme n'est pas formé sur des échantillons de données, repose sur un processus itératif d'essais et d'erreurs. Au fil du temps, il apprend en évaluant les résultats de ses actions. Les séquences de résultats favorables sont renforcées afin de déterminer la meilleure recommandation ou stratégie pour résoudre un problème spécifique.

3.3.3 Avantages et inconvénients du Machine Learning

Avantages du Machine Learning :

Adaptabilité : Le ML permet aux modèles d'apprendre à partir de données et de s'adapter à de nouvelles informations sans être explicitement programmés. Il peut gérer des tâches complexes et des données non structurées.

Précision : Les modèles ML peuvent fournir des prédictions plus précises que les méthodes traditionnelles. Ils détectent des modèles subtils et des tendances non évidentes.

Automatisation : Le ML automatise des tâches répétitives et fastidieuses, ce qui permet aux humains de se concentrer sur des tâches plus créatives.

Applications diverses : Le ML est utilisé dans la reconnaissance d'images, la traduction automatique, la prédiction financière, etc.

Inconvénients du Machine Learning :

Besoin de données de haute qualité : Les modèles ML dépendent de données d'entraînement de haute qualité. Des données bruitées ou biaisées peuvent affecter les performances.

Sur-apprentissage et sous-apprentissage : Les modèles peuvent apprendre à partir de données spécifiques et ne pas généraliser correctement.

Le surapprentissage (overfitting) et le sous-apprentissage (underfitting) sont des problèmes courants.

Interprétabilité des résultats : Certains modèles ML, comme les réseaux de neurones profonds, sont difficiles à interpréter.

Comprendre pourquoi un modèle prend une décision peut être complexe.

Sélection appropriée des caractéristiques : Le choix des caractéristiques (features) influencera les performances du modèle.

Une mauvaise sélection peut entraîner des résultats médiocres.

3.3.4 Types principales des modèles d'ensemble du machine learning

D'après (17) Les modèles d'ensemble représentent une approche puissante en apprentissage automatique, exploitant la diversité des prédictions de plusieurs modèles pour produire des résultats plus fiables et précis. Cette méthode repose sur le principe selon lequel la combinaison de multiples modèles peut compenser les faiblesses individuelles de chaque modèle, aboutissant ainsi à une performance globale améliorée. Selon les techniques utilisées pour créer et fusionner ces modèles.

les modèles d'ensemble peuvent être classés en cinq types principaux. Chacun de ces types présente des caractéristiques distinctes, offrant une gamme variée d'approches pour résoudre différents problèmes d'apprentissage automatique avec efficacité et robustesse :

1. **Majorité (Majority)** : Un cas particulier d'ensachage est celui où les modèles sont utilisés pour la classification plutôt que pour la régression. Par exemple, si vous avez trois modèles qui prédisent s'il pleuvra à Paris par oui ou par non, vous pouvez prendre leur vote majoritaire et obtenir oui comme prédition finale. C'est ce qu'on appelle également un ensemble majoritaire ou un ensemble pluraliste.

L'avantage du vote majoritaire est qu'il réduit le taux d'erreur des prédictions, ce qui signifie qu'elles ont plus de chances d'être exactes. L'inconvénient est qu'il ne tient pas compte de la confiance ou de la probabilité de chaque prédition, ce qui signifie qu'il peut ignorer certaines informations utiles.

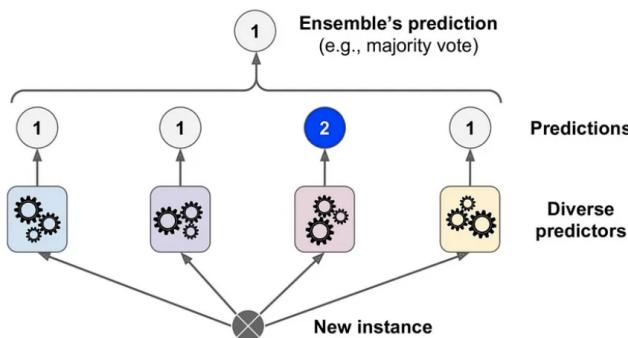


FIGURE 3.10 – prédictions du classificateur de vote dur

2. **Moyenne (Average)** : Le moyen le plus simple de combiner plusieurs modèles est de prendre leur moyenne.

Par exemple, si vous disposez de trois modèles qui prédisent la température à Paris à 15°C, 18°C et 20°C, vous pouvez prendre leur moyenne et obtenir 17,67°C comme prévision finale. C'est ce qu'on appelle aussi un ensemble moyen.

L'avantage de la moyenne est qu'elle réduit la variance des prédictions, ce qui signifie qu'elles sont moins susceptibles d'être éloignées de la vraie valeur. L'inconvénient est que cela réduit également le biais des prédictions, ce qui signifie qu'elles sont moins susceptibles d'être proches de la vraie valeur. En d'autres termes, la moyenne rend les prévisions plus cohérentes mais aussi plus conservatrices.

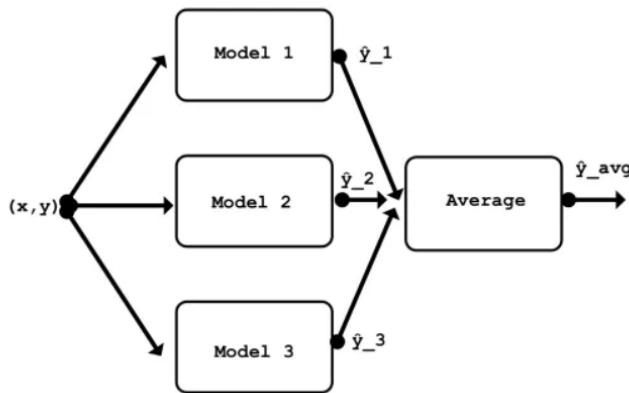


FIGURE 3.11 – Moyenne

3. **Empilage (Stacking)** : Une autre façon de combiner plusieurs modèles consiste à les utiliser comme entrées pour un autre modèle. Par exemple, si vous disposez de trois modèles qui prédisent la température à Paris à 15°C, 18°C et 20°C, vous pouvez utiliser leurs prévisions comme fonctionnalités pour un quatrième modèle qui apprend à les peser et à faire une prédition finale. . C'est ce qu'on appelle également un méta-apprenant ou un apprenant de deuxième niveau.

L'avantage du stacking est qu'il permet d'apprendre des forces et des faiblesses de chaque modèle et de faire une prédition plus précise. L'inconvénient est qu'il peut être plus complexe et sujet au surapprentissage, ce qui signifie qu'il peut fonctionner correctement sur les données d'entraînement mais peu sur les nouvelles données.

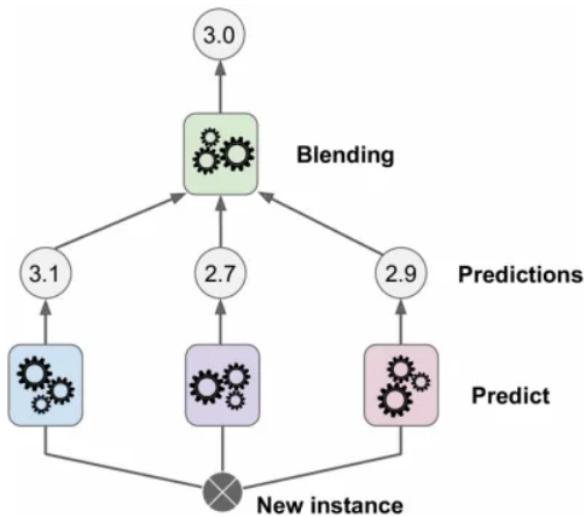


FIGURE 3.12 – Empilage

4. **Ensachage (Bagging)** : Une quatrième façon de combiner plusieurs modèles consiste à utiliser différents sous-ensembles de données pour les entraîner. Par exemple, si vous disposez d'un ensemble de données de 1 000 observations, vous pouvez échantillonner aléatoirement 500 observations avec remplacements (ce qui signifie que certaines observations peuvent être répétées) et les utiliser pour entraîner un modèle. Vous pouvez répéter ce processus plusieurs fois et entraîner différents modèles sur différents

sous-ensembles de données. Ceci est également appelé agrégation bootstrap ou bootstrapping.

L'avantage du bagging est qu'il réduit la variance des prédictions, ce qui signifie qu'elles sont moins susceptibles d'être éloignées de la vraie valeur. L'inconvénient est que cela ne réduit pas le biais des prédictions, ce qui signifie qu'elles sont toujours susceptibles d'être proches de la valeur réelle. En d'autres termes, l'ensachage rend les prédictions plus cohérentes mais pas plus précises.

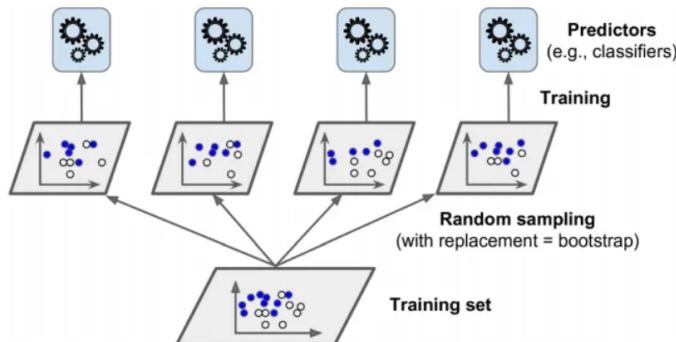


FIGURE 3.13 – Ensachage

5. **Boostant (Boosting)** : La dernière façon de combiner plusieurs modèles consiste à les utiliser de manière séquentielle et itérative. Par exemple, si vous disposez d'un modèle qui prédit une température de 15°C à Paris, vous pouvez utiliser ses erreurs ou ses résidus comme entrées pour un autre modèle qui tente de les corriger et de faire une meilleure prévision. Vous pouvez répéter ce processus plusieurs fois et obtenir différents modèles qui apprennent des erreurs de chacun. Ceci est également appelé boost adaptatif ou AdaBoost.

L'avantage du boosting est qu'il réduit à la fois la variance et le biais des prédictions, ce qui signifie qu'elles sont plus susceptibles d'être proches et précises de la valeur réelle. L'inconvénient est qu'il peut être plus sensible aux valeurs aberrantes et au bruit, ce qui signifie qu'il peut sur-ajuster ou sous-ajuster les données.

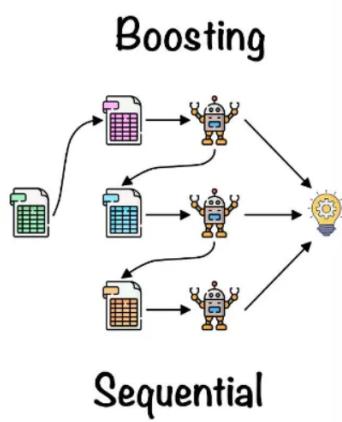


FIGURE 3.14 – Boostant

3.4 Conclusion

En somme, ce chapitre a mis en lumière les fondements de l'IA et de l'apprentissage automatique, soulignant leur importance croissante dans divers secteurs, y compris la gestion des stocks. L'IA, avec ses capacités d'analyse de données avancées et ses algorithmes prédictifs, offre des solutions innovantes et efficaces pour optimiser les processus de gestion des stocks. En comprenant ces concepts clés, nous sommes mieux équipés pour appliquer des techniques d'IA afin d'améliorer la précision des prévisions de demande, réduire les coûts opérationnels et augmenter la satisfaction client. Cette compréhension est essentielle pour notre étude, qui se concentre sur l'application de l'IA pour optimiser la gestion des stocks, démontrant ainsi le potentiel transformateur de ces technologies dans le monde des affaires.

CHAPITRE 4

MÉTHODOLOGIE

4.1 Introduction

Dans le domaine de l'apprentissage automatique et de l'analyse de données, plusieurs étapes principales sont réalisées pour assurer la qualité et l'efficacité des résultats obtenus. Tout d'abord, la collecte des données brutes est effectuée, puis ces données sont prétraitées pour les rendre plus exploitables. De plus, une fois que cela est fait, il est crucial de sélectionner le matériel approprié, y compris le langage de programmation, les technologies et les bibliothèques, ainsi que les modèles de machine learning. En fait, les types de modèles utilisés influencent directement la qualité et la pertinence des résultats reçus. Dans cette section, nous nous concentrerons sur ces étapes principales, mettant en évidence l'importance de chacune pour le domaine spécifique de l'analyse de données et de l'apprentissage automatique.

4.2 Collecte des données

La collecte de données est une étape importante dans notre étude, mais parfois les données ne sont pas toujours disponibles, car beaucoup d'entreprises ne les partagent pas publiquement. Les données disponibles peuvent être de mauvaise qualité, avec des erreurs ou des valeurs manquantes, nécessitant un nettoyage. Trouver des données couvrant différents secteurs est difficile, tout comme gérer de grandes quantités de données qui demandent des ressources informatiques importantes. Les formats non standardisés des données nécessitent une conversion. Certaines données peuvent être sensibles ou confidentielles, posant des défis de sécurité. Enfin, de nombreuses entreprises sont réticentes à partager leurs données de ventes pour des raisons de confidentialité et de concurrence.

Pour Les ensembles de données que nous avons trouvés comprennent deux types :

- Un données générés via une logique aléatoire dans VBA, Il ne s'agit pas de données de ventes réelles.
 Pris à partir de <https://excelbianalytics.com/> (18).
- Un données réelles contient des informations complètes sur les ventes de divers produits dans différentes régions, avec des détails sur les clients et les transactions. Il est utile pour analyser les performances de vente, les tendances du marché, et les comportements des clients.

Pris à partir de (19).

4.2.1 Le dataset 1 :

La (fig :4.1) représente un extrait du dataset 1 :

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit
0	Australia and Oceania	Australia	Meat	Online	C	4/4/2011	451691138	5/23/2011	4300	421.89	364.69	1814127.00	1568167.00	245960.00
1	Asia	Tajikistan	Personal Care	Online	L	7/12/2018	144177377	8/1/2018	4145	81.73	56.67	338770.85	234897.15	103873.70
2	Sub-Saharan Africa	Mozambique	Cosmetics	Online	H	7/6/2011	982716166	7/17/2011	6407	437.20	263.33	2801140.40	1687155.31	1113985.09
3	Central America and the Caribbean	Panama	Personal Care	Offline	L	5/1/2011	784543836	5/11/2011	2810	81.73	56.67	229661.30	159242.70	70418.60
4	North America	Canada	Fruits	Online	H	11/15/2013	137209212	12/29/2013	2110	9.33	6.92	19686.30	14601.20	5085.10

FIGURE 4.1 – Extrait du dataset1

Chaque ligne représente une étude spécifique sur la vente comprenant 14 paramètres, qui sont représentés dans le tableau suivant (tab :4.1) :

Nom de colonne	Description
Region	La région géographique où la vente a eu lieu
Country	Le pays où la vente a eu lieu
Item Type	Le type de produit vendu
Sales Channel	Le canal de vente utilisé pour la transaction (en ligne ou hors ligne)
Order Priority	La priorité de la commande
Order Date	La date à laquelle la commande a été passée
Order ID	L'identifiant de la commande
Ship Date	La date à laquelle la commande a été expédiée
Units Sold	Le nombre d'unités du produit vendues
Unit Price	Le prix unitaire de chaque unité vendue
Unit Cost	Le coût unitaire de chaque unité vendue
Total Revenue	Le montant total généré par la vente (nombre d'unités vendues multiplié par le prix unitaire).
Total Cost	Le coût total de l'achat (nombre d'unités vendues multiplié par le coût unitaire)
Total Profit	Le profit total généré par la vente (total revenue moins total cost)

TABLE 4.1 – Description de dataset 1

```
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500000 entries, 0 to 499999
Data columns (total 14 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   Region      500000 non-null   object 
 1   Country     500000 non-null   object 
 2   Item Type   500000 non-null   object 
 3   Sales Channel 500000 non-null   object 
 4   Order Priority 500000 non-null   object 
 5   Order Date   500000 non-null   object 
 6   Order ID     500000 non-null   int64  
 7   Ship Date    500000 non-null   object 
 8   Units Sold   500000 non-null   int64  
 9   Unit Price   500000 non-null   float64 
 10  Unit Cost    500000 non-null   float64 
 11  Total Revenue 500000 non-null   float64 
 12  Total Cost   500000 non-null   float64 
 13  Total Profit 500000 non-null   float64 
dtypes: float64(5), int64(2), object(7)
memory usage: 53.4+ MB
```

FIGURE 4.2 – Les informations sur le dataset 1

4.2.2 Le dataset 2 :

La (fig :4.3) représente un extrait du dataset 2 :

	Order ID	Order Date	Country	City	Branch	Lat	Lng	Customer First Name	Customer Last Name	Email	Phone Number	Category	Sub Category	Categ
0	1.0	1/1/2023	Syria	homs	hs01	34.7326	36.7136	lina	alrrashid	LINA.ALRRASHID@MAIL.COM	Tel: 5634962795	Tablet	Apple Ipad	
1	2.0	1/1/2023	Saudi Arabia	riyadh	rd01	24.7136	46.6753	omar	eurul	OMAR.EURUL@MAIL.COM	Tel: 5311455624	Smartphone	Samsung Galaxy	
2	3.0	1/1/2023	Saudi Arabia	riyadh	rd03	24.7743	46.7386	iman	iismaeil	IMAN.IISMAEIL@MAIL.COM	Tel: 5272740777	Digital Camera	Panasonic Lumix	
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4	4.0	1/1/2023	United Arab Emirates	abu dhabi	ad01	24.4539	54.3773	ahmad	rihan	AHMAD.RIHAN@MAIL.COM	Tel: 5783343440	Tablet	Samsung Galaxy	

FIGURE 4.3 – Extrait du dataset 2

Chaque ligne correspondant à une transaction individuelle comprenant 19 paramètres, qui sont représentés dans le tableau suivant (fig :4.4) :

Nom de colonne	Description
Order ID	Identifiant unique pour chaque commande
Order Date	Date à laquelle la commande a été passée
Country	Pays où la commande a été passée
City	Ville où la commande a été passée
Branch	Code de la succursale où la commande a été traitée
Lat	Latitude de la ville de la succursale
Lng	Longitude de la ville de la succursale
Customer First Name	Prénom du client
Customer Last Name	Nom de famille du client
Email	Adresse e-mail du client
Phone Number	Numéro de téléphone du client
Category	Catégorie du produit acheté (ex : Tablet, Smartphone, etc.)
Sub Category	Sous-catégorie du produit
Item	Nom ou modèle spécifique du produit
SalesPerson ID	Identifiant de la personne ayant réalisé la vente
Quantity	Quantité de produits achetés
Unit Price	Prix unitaire du produit
Discount	Réduction appliquée sur le produit, si applicable
Status	Statut de la commande (ex : True pour complet, False pour en attente ou annulé)

TABLE 4.2 – Description de dataset 2

Data columns (total 17 columns):			
#	Column	Non-Null Count	Dtype
0	Order ID	20000 non-null	int64
1	Order Date	20000 non-null	object
2	Day	20000 non-null	object
3	Country	20000 non-null	object
4	City	20000 non-null	object
5	Lat	20000 non-null	float64
6	Lng	20000 non-null	float64
7	Full Name	20000 non-null	object
8	Category	20000 non-null	object
9	Sub Category	20000 non-null	object
10	Item	20000 non-null	object
11	SalesPerson ID	20000 non-null	object
12	Quantity	20000 non-null	int64
13	Unit Price	20000 non-null	int64
14	Discount	20000 non-null	float64
15	Total Cost	20000 non-null	float64
16	Status	20000 non-null	bool

FIGURE 4.4 – Les informations sur le dataset 2

4.3 Visualisation des données

La visualisation des données est une étape importante dans le processus d'analyse des données. Elle permet de comprendre les motifs, les tendances et les relations dans les données. Voici quelques types courants de visualisations de données :

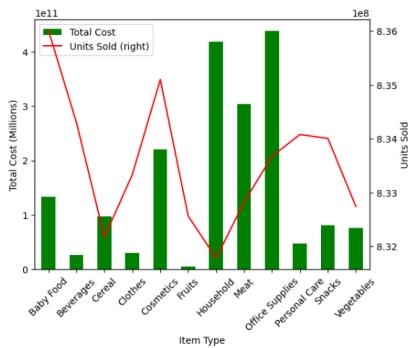


FIGURE 4.5 – Coût Total et Unités Vendues par Type d'Article de dataset1

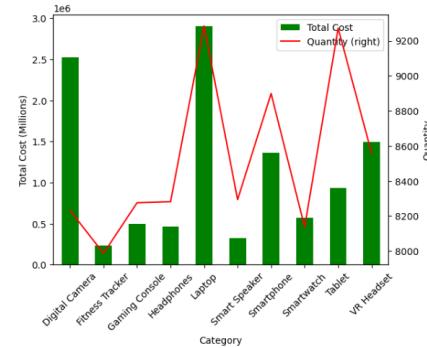


FIGURE 4.6 – Coût Total et la quantité de différents types de catégories de dataset2

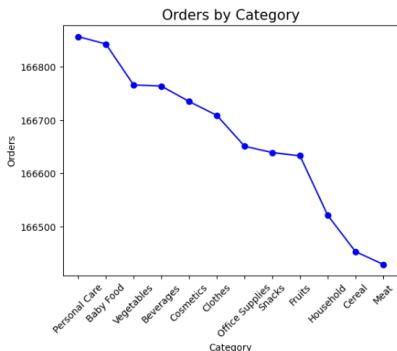


FIGURE 4.7 – Commandes par catégorie de dataset1

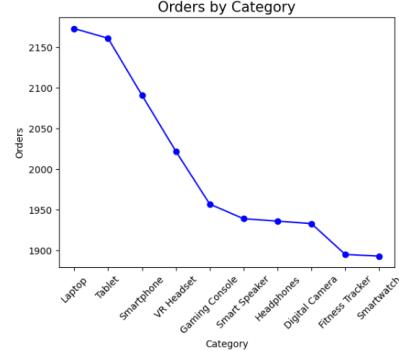


FIGURE 4.8 – Commandes par catégorie de dataset2

4.4 Prétraitement des données

Dans un algorithme d'apprentissage automatique, les données ne peuvent pas être utilisées telles quelles, c'est-à-dire dans leur forme brute. Elles doivent être préalablement préparées avant d'être intégrées dans les modèles d'apprentissage automatique. Cette étape est essentielle pour résoudre des problèmes qui ne sont pas encore connus par l'extracteur de connaissances. Par exemple, si vous avez un problème avec votre voiture et que vous voulez savoir combien d'huile il faut pour la conduire, vous pouvez demander au conducteur de la voiture de vous indiquer la quantité exacte d'huile nécessaire. Cela s'appelle le prétraitement des données. Le but du prétraitement est de déterminer quel type d'informations est nécessaire pour prendre des décisions concernant l'utilisation de la voiture. Des données correctement formatées et nettoyées sont essentielles pour le prétraitement. Le prétraitement des données inclut les activités suivantes :

4.4.1 Le dataset 1 :

- Fournir un résumé statistique des colonnes numériques en valeurs, moyenne, écart type, valeurs minimales et maximales et quartiles. Cela donne un aperçu rapide des principales caractéristiques des données(fig :4.9).

data.describe()							
	Order ID	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit
count	5.000000e+05	500000.000000	500000.000000	500000.000000	5.000000e+05	5.000000e+05	5.000000e+05
mean	5.501319e+08	4999.136110	266.036748	187.528564	1.330096e+06	9.376163e+05	3.924800e+05
std	2.599605e+08	2884.023907	216.937600	175.623982	1.468090e+06	1.148684e+06	3.787517e+05
min	1.000029e+08	1.000000	9.330000	6.920000	9.330000e+00	6.920000e+00	2.410000e+00
25%	3.251814e+08	2502.000000	81.730000	56.670000	2.783059e+05	1.620597e+05	9.538506e+04
50%	5.491843e+08	4999.000000	154.060000	97.440000	7.862426e+05	4.677120e+05	2.817492e+05
75%	7.756291e+08	7497.000000	421.890000	263.330000	1.824236e+06	1.198736e+06	5.653923e+05
max	9.999995e+08	10000.000000	668.270000	524.960000	6.682700e+06	5.249600e+06	1.738700e+06

FIGURE 4.9 – Un bref aperçu des principales caractéristiques.

- Extrait et crée de nouvelles colonnes à partir de la colonne "Order Date". Il récupère l'année, le mois, le jour et le jour de la semaine (où lundi est 0 et dimanche est 6) et les stocke dans de nouvelles colonnes nommées "Year", "Month", "Day" et "Weekday"(fig :4.10).

<pre> data['Year'] = data['Order Date'].dt.year data['Month'] = data['Order Date'].dt.month data['Day'] = data['Order Date'].dt.day data['Weekday'] = data['Order Date'].dt.weekday data.head(3) </pre>																		
	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit	Year	Month	Day	Weekday
0	Sub-Saharan Africa	Namibia	Household	Offline	M	2015-08-31	897751939	10/12/2015	3604	668.27	502.54	2408445.08	1811154.16	597290.92	2015	8	31	0
1	Europe	Iceland	Baby Food	Online	H	2010-11-20	599480426	1/9/2011	8435	255.28	159.42	2153286.80	1344707.70	808579.10	2010	11	20	5
2	Europe	Russia	Meat	Online	L	2017-06-22	538911855	6/25/2017	4848	421.89	364.69	2045322.72	1768017.12	277305.60	2017	6	22	3

FIGURE 4.10 – Extraction des caractéristiques temporelles

- Applique une mise à échelle standard à certaines colonnes numériques . Il utilise la classe StandardScaler de scikit-learn pour transformer les valeurs de ces colonnes en ayant une moyenne de 0 et un écart type de 1. Cela peut améliorer les performances de nombreux algorithmes d'apprentissage automatique(fig :4.11).

```
numeric_cols = ['Unit Price', 'Unit Cost', 'Total Revenue', 'Total Cost', 'Total Profit']
scaler = StandardScaler()
data[numeric_cols] = scaler.fit_transform(data[numeric_cols])
data.head(3)
```

	order	Date	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit	Year	Month	Day	Weekday
0	2015-08-31	897751939	10/12/2015		3604	1.856116	1.795305	0.740998	0.766450	0.547265	2015	8	31	0
1	2010-11-20	599480426	1/9/2011		8435	-0.047813	-0.158917	0.566694	0.359264	1.106591	2010	11	20	5
2	2017-06-22	538911855	6/25/2017		4848	0.720277	1.010187	0.492942	0.728794	-0.299806	2017	6	22	3

FIGURE 4.11 – Normalisation des Variables Numériques

- Supprimer les colonnes inutiles pour l'analyse des ventes(fig :4.12).

```
data = data.drop(columns=['Order Date', 'Ship Date', 'Order ID'])
data.head(3)
```

	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit	Year	Month	Day	Weekday
0	3604	1.856116	1.795305	0.740998	0.766450	0.547265	2015	8	31	0
1	8435	-0.047813	-0.158917	0.566694	0.359264	1.106591	2010	11	20	5
2	4848	0.720277	1.010187	0.492942	0.728794	-0.299806	2017	6	22	3

FIGURE 4.12 – Suppression des colonnes inutiles et des colonnes datetime

4.4.2 Le dataset 2 :

- Fournir un résumé statistique des colonnes numériques en valeurs, moyenne, écart type, valeurs minimales et maximales et quartiles. Cela donne un aperçu rapide des principales caractéristiques des données (fig :4.13).

```
df1.describe()
```

	Order ID	Lat	Lng	Quantity	Unit Price	Discount
count	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000	18889.000000
mean	10000.500000	30.461665	11.339318	1.936750	609.292000	65.021528
std	5773.647028	8.126746	53.194139	2.288736	617.976051	145.481646
min	1.000000	18.231100	-115.149200	1.000000	29.000000	0.290000
25%	5000.750000	24.453900	2.352200	1.000000	179.000000	9.800000
50%	10000.500000	27.180900	37.134300	1.000000	399.000000	23.980000
75%	15000.250000	36.202100	42.500400	2.000000	899.000000	64.900000
max	20000.000000	48.856600	56.324200	13.000000	3999.000000	3999.000000

FIGURE 4.13 – Un aperçu rapide des principales caractéristiques des données

- Extrait et crée de nouvelles colonnes à partir de la colonne "Order Date". Il récupère l'année, le mois, le jour et le jour de la semaine (où lundi est 0 et dimanche est 6) et les stocke dans de nouvelles colonnes nommées "Year", "Month", "Day" et "Weekday"(fig :4.14).

```
# Extraction des caractéristiques temporelles
data['Year'] = data['Order Date'].dt.year
data['Month'] = data['Order Date'].dt.month
data['Day'] = data['Order Date'].dt.day
data['Weekday'] = data['Order Date'].dt.weekday
data.head(3)
```

	Order ID	Order Date	Day	Country	City	Lat	Lng	Full Name	Category	Sub Category	Item	SalesPerson ID	Quantity	Unit Price	Discount	Total Cost	Status	Year	Month	Weekday
0	1	2023-01-01	1	Syria	homs	34.7326	36.7136	Lina Alrrashid	Tablet	Apple iPad	iPad Pro 12.9"	N498	4	999	38.3616	891.9072	False	2023	1	6
1	2	2023-01-01	1	Saudi Arabia	riyadh	24.7136	46.6753	Omar Eurul	Smartphone	Samsung Galaxy	Galaxy S21 Ultra	X918	3	1199	517.9680	302.1480	True	2023	1	6
2	3	2023-01-01	1	Saudi Arabia	riyadh	24.7743	46.7386	Iman Iismaeil	Digital Camera	Panasonic Lumix	Panasonic Lumix GH5	I036	4	1299	883.3200	831.3600	True	2023	1	6

FIGURE 4.14 – Extraction des caractéristiques temporelles

- Convertit les données catégorielles en représentations numériques que les algorithmes d'apprentissage automatique peuvent comprendre(fig :4.15).

```
categorical_cols = ['Country', 'City', 'Full Name', 'Category', 'Sub Category', 'Item', 'status']
encoder = OneHotEncoder(sparse=False, drop='first')
encoded_cols = pd.DataFrame(encoder.fit_transform(data[categorical_cols]), columns=encoder.get_feature_names_out(categorical_cols))
data = data.drop(columns=categorical_cols)
data.head(3)
```

	Order ID	Order Date	Day	Lat	Lng	SalesPerson ID	Quantity	Unit Price	Discount	Total Cost	Year	Month	Weekday
0	1	2023-01-01	1	34.7326	36.7136	N498	4	999	38.3616	891.9072	2023	1	6
1	2	2023-01-01	1	24.7136	46.6753	X918	3	1199	517.9680	302.1480	2023	1	6
2	3	2023-01-01	1	24.7743	46.7386	I036	4	1299	883.3200	831.3600	2023	1	6

FIGURE 4.15 – Encodage des Variables Catégorielles (garder les colonnes d'origine nécessaires)

- Applique une mise à échelle standard à certaines colonnes numériques . Il utilise la classe StandardScaler de scikit-learn pour transformer les valeurs de ces colonnes en ayant une moyenne de 0 et un écart type de 1. Cela peut améliorer les performances de nombreux algorithmes d'apprentissage automatique(fig :4.16)

```
# Création de Nouvelles Variables
data['Total Sales'] = data['Quantity'] * data['Unit Price']
data['Discounted Sales'] = data['Total Sales'] * (1 - data['Discount'])
data['Region'] = data['Lat'].astype(str) + '_' + data['Lng'].astype(str)
# Normalisation des Variables Numériques
numeric_cols = ['Unit Price', 'Discount', 'Total Cost', 'Total Sales', 'Discounted Sales']
scaler = StandardScaler()
data[numeric_cols] = scaler.fit_transform(data[numeric_cols])
data.head(3)
```

	Order ID	Order Date	Day	Lat	Lng	SalesPerson ID	Quantity	Unit Price	Discount	Total Cost	Year	Month	Weekday	Total Sales	Discounted Sales	Region
0	1	2023-01-01	1	34.7326	36.7136	N498	4	0.630636	-0.408400	0.252358	2023	1	6	0.486025	0.219978	34.7326_36.7136
1	2	2023-01-01	1	24.7136	46.6753	X918	3	0.954281	-0.037213	-0.203340	2023	1	6	0.347755	0.086434	24.7136_46.6753
2	3	2023-01-01	1	24.7743	46.7386	I036	4	1.116104	0.245548	0.205574	2023	1	6	0.901876	-0.126347	24.7743_46.7386

FIGURE 4.16 – Normalisation des Variables Numériques

- Supprimer les colonnes inutiles pour l'analyse des ventes(fig :4.17).

	Order ID	Day	Quantity	Unit Price	Discount	Total Cost	Year	Month	Weekday	Total Sales	Discounted Sales
0	1	1	4	0.630636	-0.408400	0.252358	2023	1	6	0.486025	0.219978
1	2	1	3	0.954281	-0.037213	-0.203340	2023	1	6	0.347755	0.086434
2	3	1	4	1.116104	0.245548	0.205574	2023	1	6	0.901876	-0.126347

FIGURE 4.17 – Suppression des colonnes inutiles et des colonnes datetime

4.5 Matériel

Le choix du matériel sur lequel les modèles de machine learning sont développés et déployés est importante pour assurer des performances optimales.

4.5.1 Environnement de Développement

Colab

Google Colab est un service d'apprentissage automatique en ligne fonctionnant sur le cloud. Il permet aux utilisateurs de créer et d'exécuter des notebooks Jupyter en utilisant des ressources informatiques avancées. Ce service est gratuit mais cette configuration est payant, basé sur un navigateur, compatible avec de nombreux langages de programmation et offre une configuration comprenant 51 Go de RAM et 201.2 Go de stockage. De plus, il s'intègre facilement avec les données de Google Drive.



FIGURE 4.18 – Colab

Anaconda

La plateforme Anaconda représente un outil open source et gratuit pour Python, équipé de diverses bibliothèques essentielles au service des domaines des sciences des données et du calcul scientifique. Cette distribution a été concoctée par Continuum Analytics pour simplifier l'installation et la gestion des packages Python indispensables aux applications de science des données. En outre, il introduit un environnement de développement intégré nommé Spyder, en plus des outils Jupyter Notebook et JupyterLab pour créer et exécuter des blocs-notes interactifs. Anaconda englobe un large éventail de packages scientifiques renommés, notamment NumPy, Pandas, SciPy, Matplotlib, Scikit-learn et TensorFlow, entre autres..



FIGURE 4.19 – Anaconda

Jupyter Notebook

Le Jupiter Notebook constitue un environnement interactif de développement employé pour diverses tâches belles que l'analyse de données, la visualisation et l'apprentissage automatise. Il offre la possibilité de créer et de participer des documents incluant du code, des graphites, des explications textuelles, des aides que des formules mathématiques.

Ce logique est l'argument adopté par les professionnels tels que les scientifiques des données, les ingénieurs, les chercheurs, aides que les étudiants. Fonctionnant sur le coypu Python, il S'orienter et s'adapte à plusieurs langages de programmation, incant Python, R, Julia, Ruby, Haskell, et Lien d'autres encore.

Les utilisateurs empêchent d'écrire et exécuter du code Bans ces langages directement à l'intérieur du notebook, avec les résultats affichés sous avant de graphites, tableaux et azures types de visualisations de données.



FIGURE 4.20 – Jupyter Notebook

4.5.2 Langage de programmation

Python

Python est un langage interprété, polyvalent et adapté à différentes plateformes. Il encourage la programmation nécessaire, organisée, fonctionnelle et axée sur l'objet. Son typage dynamique est puissant, sa gestion de la mémoire par ramasse-miettes est automatique et son système de gestion d'exceptions est similaire à Perl, Ruby, Scheme et Smalltalk et Tcl. Python est un langage libre similaire à la licence BSD5 et peut être utilisé sur la plupart des plateformes informatiques, allant des smartphones aux ordinateurs centraux6, de Windows à Unix, y compris GNU/Linux, macOS, Android, iOS, et peut également être traduit en Java ou.NET. Son objectif est d'améliorer la productivité des programmeurs en proposant des outils de qualité supérieure.Son

objectif est d'améliorer la productivité des programmeurs en proposant des outils de qualité supérieure et une syntaxe facile à comprendre. Certains enseignants y trouvent aussi un langage où la syntaxe, bien distincte des mécanismes de base, permet une initiation facile aux concepts fondamentaux de la programmation.

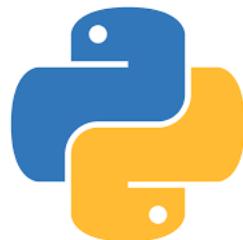


FIGURE 4.21 – Python

Les bibliothèque python utilisées

NumPy

NumPy est un package open source pour le langage de programmation Python, qui Fournit des structures de données pour manipuler des tableaux et des matrices multidimensionnels, ainsi que des fonctions mathématiques pour manipuler ces tableaux. Les tableaux NumPy sont similaires aux listes Python, mais ils sont plus efficaces des manières suivantes : Aspects de stockage et de calcul, tels qu'ils sont implémentés en C et Fortran. Simulation numérique Des fonctions mathématiques de base pour travailler avec des tableaux sont également fournies. Par exemple, calculez la moyenne, la somme, la variance et l'écart type, et Fonctions trigonométriques, logarithmiques et exponentielle.



FIGURE 4.22 – NumPy

Pandas

Pandas est une bibliothèque de manipulation et d'analyse de données écrite pour le langage de programmation Python. En particulier, il fournit des structures de données et des opérations pour travailler avec des tableaux numériques et des séries chronologiques. Pandas est un logiciel libre sous licence BSD2. Le nom vient du terme données de panel, qui est un terme économétrique désignant des ensembles de données comprenant des observations des mêmes personnes sur plusieurs périodes. Son nom est également un jeu de mots sur l'expression « Python Data Analysis ».



FIGURE 4.23 – Pandas

Matplotlib

Matplotlib est une puissante bibliothèque de traçage Python permettant de créer des visualisations statiques, animées et interactives. L'objectif principal de Matplotlib est de fournir aux utilisateurs des outils et des capacités pour les données graphiques, les rendant plus faciles à analyser et à comprendre. Il a été initialement développé par John D. Hunter en 2003 et est désormais maintenu par une large communauté de développeurs.



FIGURE 4.24 – Matplotlib

SKlearn

Scikit-learn est une bibliothèque d'apprentissage automatique Python gratuite. Il a été développé par de nombreux contributeurs 2 du monde académique, notamment des universités françaises et d'Inria 3. Dans ce cadre, nous fournissons une large bibliothèque d'algorithmes pouvant être implémentés clé en main. Ces bibliothèques sont particulièrement utiles pour les data scientists. Cela inclut les fonctions d'estimation de forêt aléatoire, la régression logistique, les algorithmes de classification et les machines à vecteurs de support, entre autres. Il est conçu pour fonctionner avec d'autres bibliothèques Python gratuites telles que NumPy et SciPy.



FIGURE 4.25 – SKlearn

4.6 Conclusion

En conclusion, le processus d'apprentissage automatique et d'analyse de données est une aventure complexe, mais non négligeable, pour extraire les informations de signification des données brutes. Des premières étapes de collecte de données à la sélection soignée du matériel et de modèles, chaque étape est importante pour la qualité d'avancement du projet.

CHAPITRE 5

RÉSULTATS ET DISCUSSION

5.1 Introduction

Dans cette section, nous nous concentrons sur la prédiction de la demande et l'optimisation de la gestion des stocks. La prédiction de la demande est essentielle pour anticiper les besoins des clients et planifier efficacement la production. Pour cela, nous utilisons divers modules de machine learning, nous séparons les données en ensembles de formation, validation et test, puis nous appliquons et évaluons les modèles de machine learning. Ensuite, nous passons à l'optimisation de la gestion des stocks, où nous présentons différentes méthodes et algorithmes pour améliorer l'efficacité des stocks, comme la normalisation des données, la sélection des caractéristiques, la réduction de la dimensionnalité, l'optimisation des hyperparamètres et l'ensemble learning. Enfin, nous examinons l'application pratique de ces techniques pour la gestion des stocks.

5.2 Prédiction de la demande

La prédiction de la demande est une composante essentielle de la gestion des stocks. Elle permet aux entreprises d'anticiper les besoins futurs des clients et de planifier adéquatement leur approvisionnement et production. En prévoyant la demande, les entreprises peuvent éviter les ruptures de stock et réduire les excès de stock, ce qui optimise leurs coûts.

Pour prédire la demande, plusieurs étapes et techniques sont utilisées. D'abord, les données historiques de ventes sont collectées et analysées. Ensuite, ces données sont utilisées pour identifier des tendances et des schémas dans les comportements des clients. Les entreprises peuvent aussi utiliser des modèles mathématiques et des algorithmes d'apprentissage automatique pour améliorer la précision de leurs prévisions. Enfin, les prévisions sont ajustées en fonction des événements à venir, comme les promotions ou les changements saisonniers.

Grâce à ces techniques, les entreprises peuvent mieux gérer leurs stocks et répondre plus efficacement aux besoins de leurs clients.

5.2.1 Modules de machine learning utilisés

Pour prédire la demande, nous utilisons plusieurs modules de machine learning. Les principaux modules comprennent :

RandomForestRegressor

- Un modèle d'ensemble basé sur plusieurs arbres de décision. Il est robuste aux données bruyantes et réduit le risque de surapprentissage (overfitting).
- Approprié pour les problèmes de régression complexes avec de nombreuses variables.

GradientBoostingRegressor

- Un autre modèle d'ensemble qui construit des arbres de décision de manière séquentielle. Chaque arbre corrige les erreurs des arbres précédents.
- Efficace pour les données où les relations non linéaires sont présentes.

DecisionTreeRegressor

- Un modèle de régression qui utilise un arbre de décision pour prédire la valeur cible en apprenant des règles de décision simples dérivées des caractéristiques des données.
- Simple à interpréter, mais peut souffrir de surapprentissage.

ExtraTreesRegressor

- Similaire au RandomForest, mais construit chaque arbre avec un sous-échantillon aléatoire des données.
- Améliore la précision en réduisant la variance.

XGBRegressor (XGBoost)

- Un modèle d'ensemble puissant basé sur le boosting qui optimise les arbres de décision avec des techniques avancées de régularisation.
- Très performant pour de nombreux types de problèmes de régression.

5.2.2 Séparation des données en ensembles de formation, validation et test

La séparation des données est une étape cruciale pour garantir que les modèles de machine learning sont bien évalués et généralisent correctement sur des données nouvelles. Les données sont divisées comme suit :

Ensemble de formation (Training set) :

- C'est l'ensemble de données sur lequel le modèle est entraîné.
- Le modèle utilise ces données pour apprendre les relations entre les caractéristiques (features) et la variable cible.
- Un bon modèle doit être capable de bien s'ajuster aux données d'entraînement sans sur-apprendre (overfitting), c'est-à-dire sans simplement mémoriser les données mais en comprenant leurs structures sous-jacentes.

Ensemble de validation (Validation set) :

- Après l'entraînement sur l'ensemble de formation, le modèle est évalué sur l'ensemble de validation.
- Cela permet de mesurer la performance du modèle sur des données qu'il n'a pas vues pendant l'entraînement.
- L'ensemble de validation est utilisé pour ajuster les hyperparamètres du modèle et pour comparer différentes configurations de modèles.
- C'est une étape cruciale pour choisir le meilleur modèle et éviter le sur-ajustement en vérifiant la performance sur des données indépendantes.

Ensemble de test (Test set) :

- Une fois que le modèle a été sélectionné et que tous les ajustements des hyperparamètres ont été effectués sur l'ensemble de validation, le modèle final est évalué sur l'ensemble de test.
- L'ensemble de test est utilisé pour estimer la performance réelle du modèle sur des données totalement nouvelles et non utilisées pendant l'entraînement ou la validation.
- Cela donne une estimation honnête de la capacité du modèle à généraliser sur de nouvelles données.

Séparer les données en ensembles de formation, validation et test à partir du DataFrame :

```
Taille de l'ensemble d'entraînement (X_train, y_train) : (1600000, 11), (1600000,)
Taille de l'ensemble de validation (X_valid, y_valid) : (200000, 11), (200000,)
Taille de l'ensemble de test (X_test, y_test) : (200000, 11), (200000,)
```

FIGURE 5.1 – dataset 1

```
Taille de l'ensemble d'entraînement (X_train, y_train) : (16000, 7), (16000,)
Taille de l'ensemble de validation (X_valid, y_valid) : (2000, 7), (2000,)
Taille de l'ensemble de test (X_test, y_test) : (2000, 7), (2000,)
```

FIGURE 5.2 – dataset 2

5.2.3 L'entraînement et Évaluation des modèles de machine learning

Une fois les données prétraitées, elles sont divisées en ensembles d'entraînement et de test. Les données d'entraînement sont utilisées pour ajuster les paramètres du modèle, tandis que les données de test sont utilisées pour évaluer ses performances sur des données invisibles durant l'entraînement.

Pour évaluer les modèles de machine learning, on utilise plusieurs métriques, chacune ayant ses propres avantages et domaines d'application. Voici quelques-unes des métriques les plus couramment utilisées (20) :

- **MSE** :(Mean Squared Error) mesure la moyenne des carrés des erreurs, c'est-à-dire la moyenne des différences au carré entre les valeurs prédites et les valeurs réelles. Il est particulièrement sensible aux grandes erreurs.
- **MAE** :(Mean Absolute Error) mesure la moyenne des erreurs absolues entre les valeurs prédites et les valeurs réelles. Contrairement à MSE, il n'est pas aussi sensible aux grandes erreurs, car il traite toutes les erreurs de la même manière.
- **R²** : (R-squared) coefficient de détermination, mesure la proportion de la variance de la variable dépendante qui est prévisible à partir des variables indépendantes. Il varie entre 0 et 1, où 1 indique une prédition parfaite.

Voici les résultats obtenus pour les deux datasets :

	Random Forest	Gradient Boosting	Decision Tree	XGBoost	ExtraTreesRegressor
MAE	30.105	1094.8	20.477	172.91	14.385
MSE	31199	19800	75655	10293	13930
R ²	0.9962	0.7608	0.9908	0.9875	0.9983

TABLE 5.1 – Résultat obtenu pour dataset 1

	Random Forest	Gradient Boosting	Decision Tree	XGBoost	ExtraTreesRegressor
MAE	0.2026	0.7962	0.0840	0.4814	0.3652
MSE	0.1919	1.1620	0.1980	0.6940	0.3451
R ²	0.9045	0.4220	0.9015	0.6548	0.8283

TABLE 5.2 – Résultat obtenu pour dataset 2

Les résultats montrent que les modèles Random Forest et Extra Trees sont les plus performants sur les deux datasets, avec des erreurs moyennes significativement plus faibles comparées aux autres modèles. Cela suggère que ces méthodes sont bien adaptées pour la prédition dans le contexte des données étudiées. En revanche, Gradient Boosting montre une performance variable : excellente sur le premier dataset mais moins satisfaisante sur le deuxième, indiquant une sensibilité aux caractéristiques spécifiques des données. Ces observations soulignent l'importance de choisir un modèle adapté à la structure et aux caractéristiques des données pour obtenir les meilleures performances prédictives en gestion des stocks.

5.3 Optimisation de la gestion des stocks

L'optimisation de la gestion des stocks vise à améliorer l'efficacité des systèmes d'approvisionnement et de stockage, réduisant ainsi les coûts et les risques de surstocks ou de ruptures.

5.3.1 les méthodes et algorithmes

1. Normalisation des données :

La normalisation est importante pour garantir que toutes les variables utilisées dans les modèles d'apprentissage automatique sont sur la même échelle. Sans normalisation, des variables avec des échelles différentes pourraient influencer de manière disproportionnée le modèle, ce qui peut conduire à une mauvaise convergence des algorithmes. Par exemple, si une variable est mesurée en milliers et une autre en unités simples, la variable mesurée en milliers pourrait dominer le processus d'apprentissage en raison de ses valeurs plus élevées. En normalisant les données, chaque variable est transformée pour avoir une moyenne de zéro et une variance d'un, assurant ainsi une distribution uniforme et facilitant l'apprentissage efficace des modèles.

- Méthodes courantes :

- **Min-Max Scaling** : Transforme les données pour qu'elles se situent entre 0 et 1.

- **Standard Scaling** : Centre les données autour de zéro avec un écart-type de un.

2. Sélection des caractéristiques :

La sélection des caractéristiques vise à identifier les variables les plus pertinentes pour prédire correctement la demande ou toute autre variable cible. Des techniques comme la méthode des forêts aléatoires et l'analyse des corrélations sont utilisées pour évaluer l'importance de chaque variable par rapport à la variable cible. Cette étape est cruciale pour réduire le bruit dans les données et améliorer la précision des modèles en se concentrant uniquement sur les caractéristiques les plus informatives.

- *Méthodes courantes :*

- **Filtrage (Filter Methods)** : Utilisation de tests statistiques pour sélectionner les meilleures caractéristiques (e.g., test du chi-carré, ANOVA).
- **Méthodes par enveloppe (Wrapper Methods)** : Utilisation de modèles pour évaluer la pertinence des caractéristiques (e.g., Recursive Feature Elimination - RFE).
- **Méthodes intrinsèques (Embedded Methods)** : Sélection des caractéristiques au cours de l'entraînement du modèle (e.g., Lasso Regression, Random Forest importance).

3. Réduction de la dimensionnalité :

La réduction de la dimensionnalité consiste à réduire le nombre de variables explicatives tout en conservant le maximum d'informations utiles. Cela aide à simplifier les modèles et à prévenir le surapprentissage, où un modèle apprend trop précisément à partir des données d'entraînement mais ne généralise pas bien sur de nouvelles données. L'analyse en composantes principales (PCA) est l'une des techniques les plus courantes utilisées pour réduire la dimensionnalité. Elle transforme les variables initiales en un ensemble de variables non corrélées (composantes principales) qui expliquent la variance maximale des données tout en réduisant le nombre de dimensions.

- *Méthodes courantes :*

- **Analyse en Composantes Principales (PCA)** : Transforme les données en un nouvel ensemble de variables non corrélées (composantes principales).
- **Auto-encodeurs** : Réseaux de neurones utilisés pour apprendre une représentation compressée des données.

4. Optimisation des hyperparamètres :

Les hyperparamètres contrôlent le comportement des algorithmes d'apprentissage automatique et leur permettent de s'adapter aux données spécifiques sur lesquelles ils sont formés. L'optimisation des hyperparamètres est le processus de recherche des valeurs les plus performantes pour ces paramètres. Des techniques telles que la recherche par grille (Grid Search) et la recherche bayésienne sont utilisées pour explorer de manière systématique l'espace des hyperparamètres et trouver la combinaison optimale qui maximise les performances du modèle sur les données de validation ou de test.

- *Méthodes courantes :*

- **Recherche par grille (Grid Search)** : Exploration exhaustive d'une grille de combinaisons de paramètres.
- **Recherche aléatoire (Random Search)** : Exploration aléatoire de l'espace des hyperparamètres.

5. Ensemble learning :

L'ensemble learning est une approche où plusieurs modèles d'apprentissage sont combinés pour améliorer la précision prédictive et la robustesse générale du système. En utilisant différents algorithmes ou en formant chaque modèle sur des sous-ensembles différents des données, les modèles d'ensemble peuvent compenser les faiblesses individuelles de chaque modèle et produire des prédictions plus fiables. Les techniques d'ensemble learning incluent le bagging, le boosting, et les méthodes de vote.

- *Méthodes courantes :*

- **Bagging** : Combinaison de modèles formés sur des échantillons différents du jeu de données (e.g., Random Forest).
- **Boosting** : Entraînement séquentiel de modèles où chaque modèle tente de corriger les erreurs des modèles précédents (e.g., XGBoost, AdaBoost).
- **Stacking** : Combinaison des prédictions de plusieurs modèles à l'aide d'un modèle de niveau supérieur.

5.3.2 Optimisation de la gestion des stocks en pratique

Pour notre étude, nous avons utilisé deux méthodes principales pour améliorer les résultats de prédiction :

- **Normalisation des données** : Nous avons normalisé les variables numériques telles que 'Unit Price', 'Discount', 'Total Cost', 'Total Sales', et 'Discounted Sales' pour assurer une contribution équitable de chaque caractéristique à l'apprentissage du modèle. Cette normalisation aide à stabiliser et à accélérer le processus d'apprentissage.

- **Optimisation des hyperparamètres** : Nous avons optimisé les hyperparamètres de certains modèles en utilisant RandomizedSearchCV. Cette méthode nous a permis de trouver les meilleures combinaisons d'hyperparamètres pour chaque modèle, améliorant ainsi leurs performances prédictives.

Grâce à ces techniques, nous avons pu obtenir des résultats de prédiction suivantes.

	Random Forest	Gradient Boosting	Decision Tree	XGBoost	ExtraTreesRegressor
MAE	30.609	45.744	31.758	67.273	25.140
MSE	32221	34079	75522	36083	20589
R ²	0.9961	0.9958	0.9908	0.9956	0.9975

TABLE 5.3 – Résultat obtenu pour dataset 1

	Random Forest	Gradient Boosting	Decision Tree	XGBoost	ExtraTreesRegressor
MAE	0.2017	0.0825	0.1292	0.4802	0.4216
MSE	0.2020	0.0273	0.2199	0.6707	0.4395
R ²	0.8994	0.9863	0.8906	0.6663	0.7813

TABLE 5.4 – Résultat obtenu pour dataset 2

Les résultats après l'optimisation des modèles montrent une amélioration générale des performances par rapport aux modèles non optimisés. Sur le premier dataset, les modèles Random Forest et Extra Trees continuent de présenter les meilleures performances en termes de MSE, MAE et R2, confirmant leur robustesse pour la prédiction de la gestion des stocks. Pour le deuxième dataset, on observe que Gradient Boosting a significativement amélioré ses résultats, montrant une meilleure capacité à modéliser les relations complexes entre les variables. Ces résultats soulignent l'importance de l'optimisation des hyperparamètres pour adapter les modèles aux spécificités des données, augmentant ainsi leur précision prédictive en gestion des stocks.

La (fig :5.3)et(fig :5.4) montrent la la comparaison des performances des modèles avant et après l'optimisation pour les deux datasets.

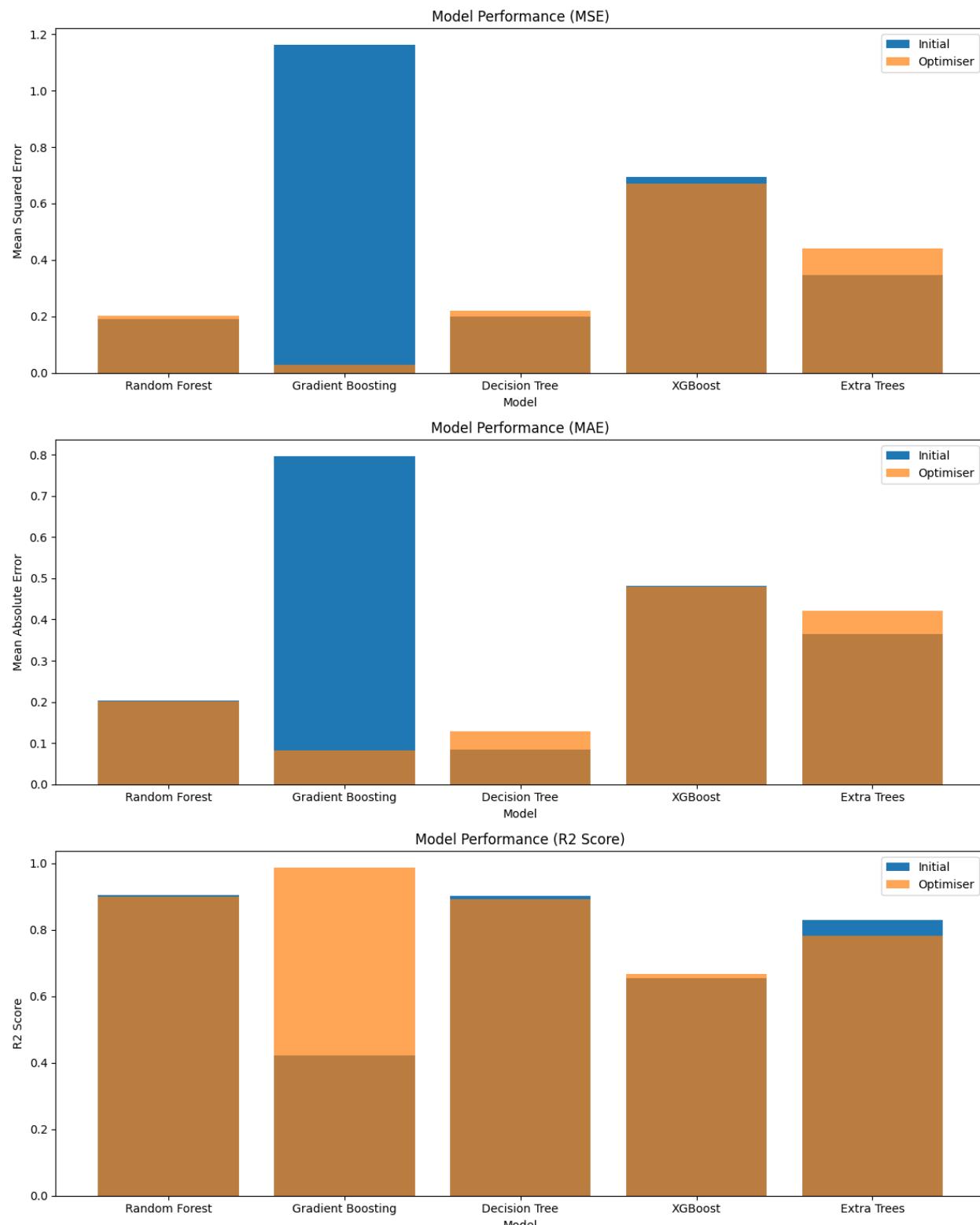


FIGURE 5.3 – Comparaison résultat dataset 1

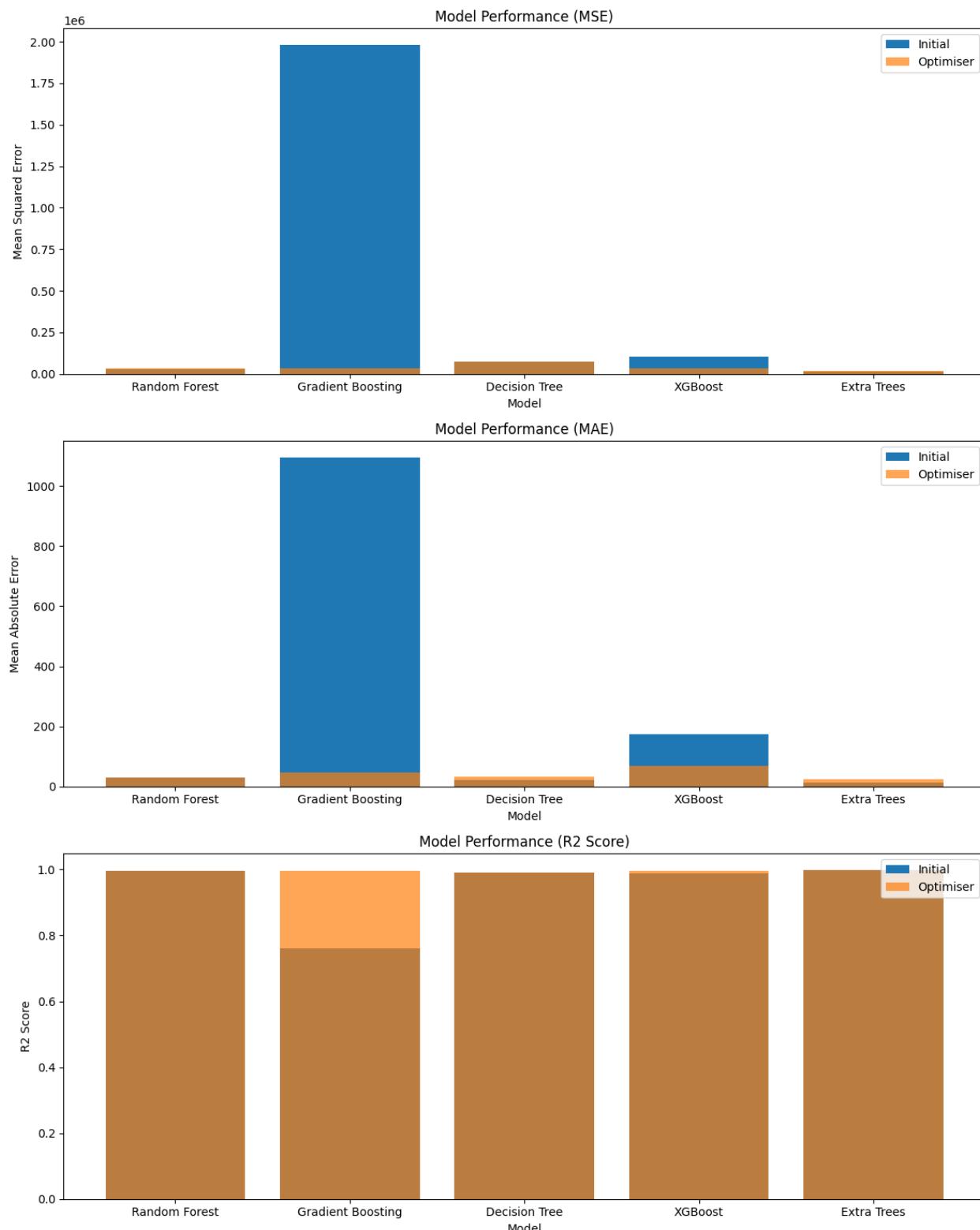


FIGURE 5.4 – Comparaison résultat dataset 2

5.4 Conclusion

En résumé, cette section a montré l'importance de la prédition de la demande et de l'optimisation de la gestion des stocks dans une entreprise. L'utilisation des techniques de machine learning permet de prévoir les besoins futurs avec précision, ce qui aide à éviter les surplus ou les pénuries de stocks. Par ailleurs, les méthodes et algorithmes d'optimisation présentés permettent d'améliorer la gestion des stocks en pratique, en maximisant l'efficacité et en minimisant les coûts. Ces approches combinées offrent une solution complète pour une gestion optimale des stocks, répondant aux défis du marché de manière proactive et efficace.

CHAPITRE 6

CONCLUSION ET PERSPECTIVE

Ce projet de fin d'étude a permis d'explorer en profondeur la gestion des stocks et son optimisation grâce à la prédiction de la demande à l'aide des modèles de machine learning. La gestion des stocks est un aspect crucial pour toute entreprise cherchant à minimiser les coûts tout en maximisant la satisfaction client. Notre étude a commencé par définir les concepts fondamentaux de la gestion des stocks, la gestion des entrepôts et les différents types de stocks.

Nous avons également examiné les problèmes courants de la gestion des stocks, en prenant des exemples concrets comme les excédents de Nike, les défis rencontrés par Best Buy pendant la période de Noël, et l'échec de l'expansion de Target au Canada. Ces études de cas ont mis en lumière les conséquences des mauvaises prévisions de la demande et de l'inefficacité de la gestion des stocks.

Ensuite, nous avons plongé dans le domaine de l'intelligence artificielle et du machine learning, en définissant leurs concepts et en discutant des techniques courantes. Nous avons analysé les approches du machine learning, les avantages et les inconvénients de ces technologies, et les principaux modèles utilisés pour la prédiction. La collecte et le prétraitement des données ont été des étapes essentielles pour assurer la qualité des prévisions.

Nous avons appliqué différents modèles de machine learning, comme les forêts aléatoires, les réseaux de neurones, et les modèles de régression, pour prédire la demande future. La séparation des données en ensembles de formation, de validation et de test a permis de vérifier la performance et la précision des modèles. Les résultats obtenus ont montré que le machine learning peut considérablement améliorer les prévisions de la demande.

Pour l'optimisation de la gestion des stocks, nous avons discuté de diverses méthodes et algorithmes, comme les algorithmes génétiques et les techniques de programmation linéaire. Nous avons démontré comment ces méthodes peuvent être mises en pratique pour optimiser les niveaux de stocks, réduire les coûts de stockage et éviter les ruptures de stock.

En termes de perspectives, plusieurs axes de développement peuvent être envisagés. Intégrer des techniques plus avancées, comme le deep learning, pourrait encore améliorer la précision des prévisions. Élargir la base de données en incorporant plus de variables externes, comme les tendances économiques ou les comportements des consommateurs, pourrait également améliorer les résultats.

Enfin, une collaboration étroite avec des entreprises permettrait de tester et de valider les modèles dans des contextes réels, offrant ainsi des retours d'expérience précieux et des opportunités d'amélioration continue.

Ce projet ouvre la voie à des recherches futures et à des applications pratiques qui peuvent transformer la gestion des stocks grâce à l'intelligence artificielle et au machine learning.

BIBLIOGRAPHIE

- [1] D'après le dictionnaire « LE PETIT LAROUSSE » (1997),
- [2] <https://www.scribd.com/document/470394531/Evaluation-De-La-Gestion-Des-Stocks-D-une-Entreprise>
- [3] <https://www.l-expert-comptable.com/a/51975-la-gestion-des-stocks-definition-enjeux-optimisation.html>
- [4] <https://www.ibm.com/fr-fr/topics/inventory-management>
- [5] <https://infonet.fr/lexique/definitions/gestion-de-stocks/>
- [6] <https://www.nomadia-group.com/ressources/blog/gestion-des-stocks/>
- [7] <https://www.ar-racking.com/fr/blog/types-de-stock-et-inventaires-dans-un-entrepot/>
- [8] <https://www.chrono-caisse.com/articles/5cf11022c32c2/1%20importance-de-la-gestion-des-stocks-enjeux-et-opportunites>
- [9] <https://www.assetpanda.com/resource-center/blog/catastrophic-inventory-mistakes-by-huge-brands-and-how-to-avoid-them/:text=Nike's%20Excess%20Inventory%20Problemtext=The%20software%20promised%20to%20help,led%20to%20millions%20more%20lost.>
- [10] https://fr.wikipedia.org/wiki/Raisonnement_%C3%A0_partir_de_cas
- [11] https://www.wikiwand.com/fr/Syst%C3%A8me_multiagents
- [12] <https://datascientest.com/introduction-au-nlp-natural-language-processing>
- [13] <https://www.journaldunet.fr/intelligence-artificielle/guide-de-l-intelligence-artificielle/1501301-algorithme-genetique-definition-et-apport-en-machine-learning/>
- [14] <https://ledatascientist.com/algorithme-genetique/>
- [15] <https://www.ibm.com/fr-fr/topics/machine-learning>
- [16] <https://www.redhat.com/fr/topics/ai/what-is-machine-learning>

- [17] <https://aryanbajaj13.medium.com/ensemble-models-how-to-make-better-predictions-by-combining-multiple-models-with-python-codes-6ac54403414e>
- [18] <https://excelbianalytics.com/wp/downloads-18-sample-csv-files-data-sets-for-testing-sales/>
- [19] <https://www.youtube.com/watch?v=0GHSXwPHV6U&list=PLof3yw6ZFPFg5W3Z3sv6GzY6WvvXBqtnDinde>
- [20] <https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e>
- [21] <https://openai.com/chatgpt/>
- [22] <https://scholar.google.com/>