

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
École Nationale Polytechnique



Département de Génie Industriel Mémoire de projet de fin d'études
Pour l'obtention du diplôme d'Ingénieur d'état en Génie Industriel
Option : Data Science et Intelligence Artificielle

Prévision des achats de pièces de rechange au
sein de Biopharm par l'application des méthodes
prédictives et le machine learning.

Réalisé par :

- M. BELKHOUMALI Ahmed Abdelmounaim

Encadré par :

- Mme. Bahia BOUCHAFAA
- M. Mohamed BELMOUHOU

Présenté et soutenu publiquement le 04 Juillet 2023, devant le jury composé de :

Présidente	Mme Latifa DEBBI	Professeur	ENP
Promotrice	Mme Bahia BOUCHAFAA	MCA	ENP
Examineur	M Oussama ARKI	MCB	ENP

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
École Nationale Polytechnique



Département de Génie Industriel Mémoire de projet de fin d'études
Pour l'obtention du diplôme d'Ingénieur d'état en Génie Industriel
Option : Data Science et Intelligence Artificielle

Prévision des achats de pièces de rechange au
sein de Biopharm par l'application des méthodes
prédictives et le machine learning.

Réalisé par :

- M. BELKHOUMALI Ahmed Abdelmounaim

Encadré par :

- Mme. Bahia BOUCHAFAA
- M. Mohamed BELMOUHOU

Présenté et soutenu publiquement le 04 Juillet 2023, devant le jury composé de :

Présidente	Mme Latifa DEBBI	Professeur	ENP
Promotrice	Mme Bahia BOUCHAFAA	MCA	ENP
Examineur	M Oussama ARKI	MCB	ENP

Dédicaces

Tout d'abord, à mes chers parents, qui ont été présents à chaque étape de mon parcours. Leurs sacrifices, leur amour et leur soutien indéfectible m'ont permis d'atteindre mes objectifs. Je leur suis infiniment reconnaissant.

Je souhaite également dédier ce travail à mon frère, Sidali, et ma sœur, Assia. Leurs présences et leurs encouragements ont été d'une grande valeur pour moi. Je suis fier d'avoir des frères et sœurs aussi merveilleux.

Dans mon entourage, j'ai également la chance d'avoir de nombreux amis qui ont partagé des moments précieux avec moi. Dans le premier groupe, je tiens à remercier Mohamed Guezdiya, Mimo Scarfysse et Ryad. Leur amitié et leur soutien m'ont aidé à traverser les hauts et les bas de la vie.

Dans le deuxième groupe, je souhaite exprimer ma gratitude envers Walid, Ishak, Adel, Islam, Abdelhak et tous mes camarades de classe. Leur présence dans ma vie a été une source de joie, de soutien et d'apprentissage mutuel.

Enfin, je tiens à mentionner mes amis du lycée, Youcef, Foued et Redouane toto, ainsi que mes amis des années de lycée. Leurs amitiés ont été précieuses et ont contribué à ma croissance personnelle et académique.

À tous ceux qui ont croisé ma route et m'ont apporté leur soutien, je vous suis profondément reconnaissant. C'est grâce à vous que je suis devenu la personne que je suis aujourd'hui.

Mounaim

Remerciement

"Je rends grâce à Dieu seul, clément et miséricordieux. Tout d'abord, j'exprime mes vifs remerciements à mon encadrante, Madame Bahia BOUCHAFA, qui m'a donné l'opportunité d'approfondir mes connaissances au sein de Biopharm.

Mes sincères remerciements vont ensuite à M. Mohammed Belmouhoub qui m'a assisté tout au long de la période du projet, avec son soutien moral et ses précieux conseils qui m'ont permis de mener à bien mon projet de fin d'études, dans l'espoir qu'il soit à la hauteur de ses attentes.

J'exprime ma gratitude à l'ensemble du personnel de Biopharm Industrie qui m'ont aidé et ont permis de travailler dans un cadre et une ambiance agréable.

Je saisis cette occasion pour adresser mes plus sincères remerciements à mes enseignants ainsi qu'aux membres du jury qui me font l'honneur d'évaluer mon travail .

Que toutes les personnes ayant participé de près ou de loin à la réalisation de mon projet trouvent ici mon témoignage de gratitude et de remerciements."

Mounaim

ملخص:

تتعامل صناعة الأدوية مع معوقات كبيرة في مجال الإشراف على المخزون وفي مجال اقتناء قطع الغيار التي تعتبر ضرورية لاستمرارية العمليات والإنتاج. في هذا الإطار، يسعى هذا المشروع إلى رفع مستوى فعالية عملية توقع الاحتياجات لقطع الغيار في شركة "بيوفارم"، وهي شركة دوائية جزائرية. لإنجاز هذا الهدف، تم استعراض تقنيات علوم البيانات، مع التركيز بشكل خاص على تحليل السلاسل الزمنية. السلاسل الزمنية هي مجموعة من البيانات التي يتم جمعها بفترات زمنية منتظمة، وتعتبر مهمة لفهم الاتجاهات والقيام بالتوقعات في عدة مجالات، بما في ذلك إدارة المخزون. تم تدريب مجموعة من نماذج التوقع باستخدام بيانات سابقة تتعلق بشراء قطع الغيار، وتم تقييم هذه النماذج وفقاً لدقتها وقوتها في القيام بتوقعات موثوقة. هذا أدى إلى تحديد النماذج الأكثر فعالية في التوقعات ضمن البيئة الخاصة بـ "بيوفارم". في ختام هذا المشروع، تم إنشاء واجهة رسومية للمستخدم (GUI). تمكن هذه الواجهة مستخدمي "بيوفارم" من إدخال البيانات بسهولة والحصول على توقعات حية لاحتياجات قطع الغيار. هذا يبسط عملية اتخاذ القرار في مجال الاقتناء ويسهم في تحسين إدارة المخزون، مما ينتج عنه خفض في التكاليف وتعزيز في الأداء التشغيلي.

الكلمات المفتاحية: توقعات الاقتناء، قطع الغيار، الصناعة الدوائية، بيوفارم، علوم البيانات، السلاسل الزمنية، نماذج التنبؤ، واجهة المستخدم الرسومية، إشراف على المخزون.

Abstract

The pharmaceutical landscape grapples with substantial hurdles in inventory control and sourcing of essential spare parts, pivotal for ongoing operations and manufacturing processes. Targeting this issue, the project focuses on enhancing the spare parts procurement prediction systems at Biopharm, a pharmaceutical firm based in Algeria. Employing Data Science methodologies with an emphasis on time series analytics, we sought to make accurate and efficient forecasts. Time series data, collected at uniform time intervals, serve as instrumental tools for trend analysis and predictive measures across diverse areas, including inventory management. We trained multiple forecast models using archival data on spare parts acquisitions, and assessed them on the basis of prediction accuracy and reliability. This exercise pinpointed the most effective models tailored for Biopharm's unique procurement landscape. Conclusively, we engineered a Graphical User Interface (GUI), allowing Biopharm staff to seamlessly input data and obtain instantaneous forecasts for spare part requirements. This simplifies procurement decision-making and contributes to optimized inventory control, thereby leading to cost savings and heightened operational efficiency.

Keywords: Procurement Prediction, Spare Parts, Pharmaceutical Sector, Biopharm, Data Science, Time Series Analysis, Forecasting Models, Graphical User Interface, Inventory Management

Résumé

Le domaine pharmaceutique est aux prises avec d'importants obstacles en matière de contrôle des stocks et de sourcing de pièces de rechange indispensables, éléments clés pour le maintien des opérations et des processus de fabrication. Ciblant cette problématique, ce projet a pour objectif d'améliorer les systèmes de prédiction d'achat de pièces de rechange chez Biopharm, une entreprise pharmaceutique algérienne. En utilisant des méthodologies en science des données, avec une attention particulière portée à l'analyse de séries chronologiques, nous avons cherché à réaliser des prévisions précises et efficaces. Les données de séries chronologiques, collectées à intervalles de temps uniformes, sont des outils essentiels pour l'analyse des tendances et les mesures prédictives dans divers domaines, y compris la gestion des stocks. Nous avons formé plusieurs modèles de prévision en utilisant des données archivées sur les acquisitions de pièces de rechange, et les avons évalués sur la base de la précision et de la fiabilité de leurs prédictions. Cet exercice a permis d'identifier les modèles les plus efficaces adaptés au contexte spécifique de Biopharm. En conclusion, nous avons développé une Interface Graphique Utilisateur (GUI), permettant au personnel de Biopharm de saisir facilement des données et d'obtenir des prévisions instantanées pour les besoins en pièces de rechange. Ceci simplifie la prise de décision en matière d'approvisionnement et contribue à l'optimisation du contrôle des stocks, entraînant ainsi des économies de coûts et une meilleure efficacité opérationnelle.

Mots-clés : Prédiction d'Achat, Pièces de Rechange, Secteur Pharmaceutique, Biopharm, Science des Données, Analyse de Séries Chronologiques, Modèles de Prévision, Interface Graphique Utilisateur, Gestion des Stocks

Sommaire

Liste des tableaux

Table des figures

Introduction générale	11
Chapitre 1 Etat des lieux	15
Introduction.....	15
1. Présentation du secteur d'activité.....	15
1.1. L'industrie pharmaceutique dans le monde.....	15
1.2. L'industrie pharmaceutique en Algérie	16
2. Présentation de l'entreprise	17
2.1. Présentation générale.....	17
2.2. Structure du groupe "BIOPHARM GROUP"	17
2.3. Structure interne de "BIOPHARM INDUSTRY"	18
3. Périmètre du projet	19
3.1. Département IT	19
3.2. Département Maintenance	20
3.3. Département DA&A.....	20
4. Diagnostic et problématique	21
4.1. Diagnostic externe.....	21
4.2. Diagnostic Interne	24
4.3. Synthèse du diagnostic	26
Énoncé de la problématique.....	28
L'énoncé complet de la problématique sera donc.....	28
Conclusion	29
Chapitre 2 Etat de l'art	31
Introduction.....	31
1. Termes et définitions	31
1.1 Prévision	31
1.2 Prédiction	31
1.3 Prévision vs Prédiction.....	31
1.4 Prévision d'achat des pièces de rechange dans le domaine de la pharma	32
1.5 Base de données	32
1.6 Série temporelle	32
1.7 Intelligence Artificielle (IA).....	33
1.8 Machine Learning (ML)	33
2. Concepts fondamentaux sur les séries temporelles	33
2.1 Types de séries temporelles	33
2.1.1 Séries temporelles univariées	33
2.1.2 Séries temporelles multivariées	33
2.2 Stationnarité.....	35
2.3 Retard (Lag).....	35
2.4 Autocorrélation.....	35
2.5 Transformations et différenciations :	35
2.6 Transformation Logarithmique	34

2.7	Transformation de Box-Cox	36
2.8	Normalisation.....	36
2.9	Considérations pour l'Interprétation	37
2.10	Différenciation des séries temporelles	37
3.	Concepts fondamentaux sur le machine learning	37
3.1	Type de machine learning.....	37
3.1.1	Apprentissage supervisé (Supervised Learning).....	37
3.1.2	Apprentissage non supervisé (Unsupervised Learning)	38
3.1.3	Apprentissage semi-supervisé (Semi-Supervised Learning).....	38
3.2	Concepts basiques du machine learning.....	39
3.2.1	Données (data)	39
3.2.2	Caractéristiques (features)	39
3.2.3	Étiquettes (labels)	39
3.2.4	Données d'entraînement (training data).....	39
3.2.5	Données de test (test data)	39
3.2.6	Modèle (model).....	39
3.2.7	Entraînement (training)	40
3.2.8	Prédiction (prediction)	40
3.2.9	Fonction de perte (loss function)	40
3.2.10	Algorithme d'optimisation (optimization algorithm)	40
3.2.11	Métriques d'évaluation (evaluation metrics)	40
3.2.12	Généralisation (generalization)	41
3.3	Concepts avancés du machine learning	41
3.3.1	Trade-off bias-variance	41
3.3.2	Techniques de régularisation	41
3.3.3	Techniques de sélection e caractéristiques (features selection techniques).....	42
3.3.4	Apprentissage profond (Deep Learning).....	42
4.	Modèles de prédiction.....	42
4.1	Modèles de prédiction pour les séries temporelles	42
4.1.1	Modèle autorégressif (AR).....	42
4.1.2	Modèle à moyenne mobile (MA).....	44
4.1.3	Modèle autorégressif à moyenne mobile (ARMA)	44
4.1.4	Modèle autorégressif intégré à la moyenne mobile (ARIMA).....	44
4.2	Modèles de prédiction pour le machine learning	44
4.2.1	Régression linéaire	44
4.2.2	Régression par forêts aléatoires	45
4.2.3	Régression par Vecteurs de Support	45
4.2.4	Amplification Extrême du Gradient (XGBOOST).....	46
4.2.5	Modèles basés sur les réseaux de neurones	46
5.	Evaluation des modèles	47
5.1	Modèles de régression.....	47
5.1.1	Erreur quadratique moyenne (MSE).....	47
5.1.2	Erreur absolue moyenne (MAE)	47
5.1.3	Coefficient de détermination (R^2)	47
5.2	Modèles de séries temporelles	48
5.2.1	Erreur quadratique moyenne (RMSE).....	48

5.2.2 Erreur absolue en pourcentage (MAPE)	48
6. CRISP-DM.....	49
6.1 Compréhension du problème métier.....	49
6.3 Préparation des données	49
6.4 Modélisation des données	49
6.5 Évaluation.....	49
6.6 Déploiement	50
Conclusion.....	50
Introduction.....	53
1. Compréhension du problème métier.....	53
2. Compréhension des données	54
2.1 Données fournis	54
2.2 La Sélection des données nécessaires.....	57
3. Préparation des données	58
3.1 Valeurs nulles	58
3.2 Séparation des données	59
3.3 Éliminer certaines valeurs	61
3.4 Catégorisation de la demande	60
3.5 Visualisation temporelle des séries chronologiques pour chaque article	63
3.6 Complétion des données.....	65
4. Modélisation des données	66
5. Évaluation et résultats	67
5.1 Méthodologie.....	67
5.2 Résultats.....	67
5.3 Analyse des Résultats.....	69
6. Déploiement	70
6.1. Exploration de l'Interface Utilisateur : Présentation du GUI (Graphical User Interface) développé	70
6.2. Intégration Harmonieuse des Modèles de Prédiction au sein de l'Interface	73
1. Intégration du code des modèles	73
7. Perspectives et améliorations	75
7.1. Ajout des fournisseurs	75
7.2. Prédiction des niveaux de stock.....	75
7.3. Analyse des ruptures de stock.....	75

7.4. Déclenchement automatique des commandes	76
Conclusion	76
Conclusion Générale	77
Bibliographie	78
Webographie	81
Annexes	83

Liste des tableaux

Tableau 1 Analyse PESTEL	23
Tableau 2 Récapitulatif des modèles utilisés	66

Table des figures

Figure 1 Hausse des capacités de production et investissements dans l'industrie pharmaceutique française en 2023 [2]	16
Figure 2 Structure organisationnelle du groupe BIOPHARM	17
Figure 3 Structure organisationnelle de Biopharm Industrie	19
Figure 4 Analyse SWOT	20
Figure 5 Les composants d'une série temporelle [3]	34
Figure 6 Types d'apprentissage automatique [4]	38
Figure 7 Les étapes de la méthodologie CRISP-DM [5]	50
Figure 8 Récapitulatif des données	56
Figure 9 Résumé des caractéristiques des données (Fonction <code>resume_table</code>)	57
Figure 10 Résultats de la fonction (<code>resume_table</code>) sur notre jeu de données	58
Figure 11 Nos données statiques	59
Figure 12 Nos données dynamiques	60
Figure 13 Les 4 types de demandes	62
Figure 14 Répartition des types de demandes sur notre jeu de données	63
Figure 15 Visualisation des séries temporelles de notre jeu de données pour chaque article	64
Figure 16 Dernières dates pour chaque article	65
Figure 17 Une partie des résultats de notre étude (Prédictions des modèles de machine learning)	68
Figure 18 Une partie des résultats de notre étude (Les erreurs du modèle de deep learning)	68
Figure 19 Résultats des modèles sur notre jeu de données	69
Figure 20 Page 1 de l' interface (SBP decision helper)	72
Figure 21 Page 2 de l' interface (SBP decision helper)	72
Figure 22 Page 3 de l' interface (SBP decision helper)	73

Introduction générale

Le marché pharmaceutique mondial a atteint un chiffre d'affaires de 1284,6 milliards de dollars en 2021, avec une croissance annuelle d'au moins 3%. Le marché du médicament en Algérie, quant à lui, est en constante évolution avec un chiffre d'affaires de 360 milliards de dinars algériens en 2018, selon le rapport de la concurrence pharmaceutique. Cependant, ce marché est confronté à des défis majeurs tels que la traçabilité des marchandises, le contrôle des coûts et des matières premières, la maîtrise des stocks, les barrières à l'entrée du marché et le respect de la réglementation nationale et internationale.

L'industrie pharmaceutique est un secteur en pleine croissance dans le monde entier, et l'Algérie n'échappe pas à cette tendance. Cependant, le marché pharmaceutique algérien a connu des pénuries ces dernières années, ce qui a perturbé les systèmes de santé et a mis en évidence une perte d'indépendance sanitaire préoccupante pour le pays. Dans ce contexte compétitif, les entreprises pharmaceutiques doivent chercher à améliorer leur réactivité et leur flexibilité afin d'augmenter leur chiffre d'affaires et de s'approprier les marchés de clients exigeants. Pour ce faire, il est essentiel de garantir une optimisation de la performance sur les différentes phases de production d'un médicament, tout en prenant en compte les spécificités de chaque entreprise. C'est dans ce contexte que Biopharm Industrie s'inscrit et vise à consolider sa place dans le marché algérien ainsi que favoriser sa croissance économique, en améliorant la performance opérationnelle des processus de ses départements et se munissant d'outils informatiques d'aide à la décision pour faciliter et optimiser ses flux, coûts, qualité et délais.

L'industrie pharmaceutique est confrontée à des défis constants, notamment dans le domaine des achats et de l'approvisionnement en pièces de rechange. Pour répondre à ces défis, Biopharm Industrie a lancé une étude complète intitulée "Prévision des achats de pièces de rechange au sein de Biopharm par l'application des méthodes prédictives et le machine learning " qui utilise des outils de data science pour prévoir les achats nécessaires dans les prochains mois et années et cela pour chaque pièce.

Cette étude se concentre principalement sur les pièces de rechange, qui représentent une part importante des achats de l'entreprise. En effet, la maintenance régulière des équipements de production est cruciale pour garantir la qualité et la sécurité des produits pharmaceutiques.

Néanmoins, la gestion des pièces de rechange peut s'avérer délicate et onéreuse, avec des défis tels que le suivi des stocks, la maîtrise des coûts et des délais, ainsi que le respect des réglementations relatives à la santé et à la sécurité. Pour aider à la prise de décision d'achat des pièces de rechange, Biopharm Industrie a adopté des techniques de data science pour affiner ses processus d'achat et de

réapprovisionnement. En exploitant des algorithmes de prévision basés sur l'analyse des données passées et des modèles mathématiques avancés, l'entreprise est en mesure d'anticiper les futurs besoins en pièces de rechange et d'ajuster efficacement ses stocks en fonction de ces prévisions.

Ce qui nous amène à nous poser cette question :

“Comment peut-on améliorer le processus d’achat et d’approvisionnement des produits importants pour le processus de production “**Pièces de rechanges**” en utilisant des techniques de la data science et de l’intelligence artificielle ? ”

Pour y répondre, nous proposons les hypothèses suivantes :

. Une révision du processus d’achats des pièces de rechanges ainsi que l’exploration du flux d’information de l’ERP de l’entreprise.

. L’implémentation d’un outil ou d’une interface interactive pourrait favoriser une gestion plus efficace de l’approvisionnement et faciliter la prise de décision lors de l’achat de pièces de rechange.

Pour assurer une approche rigoureuse et méthodique dans l’accomplissement de ce projet, nous avons choisi d’adopter la méthodologie scientifique **CRISP-DM** (Cross Industry Standard Process for Data Mining). Cette approche structurée nous permettra de suivre les étapes clés de la gestion de projet, de l’exploration des données à l’optimisation des achats et de l’approvisionnement des pièces de rechange. Ainsi, nous pourrions bénéficier d’une démarche cohérente et reproductible pour atteindre nos objectifs de manière efficace et fiable.

De plus, lors de la phase de déploiement, nous tirerons parti non seulement des outils de la DSIA, mais également des langages de développement logiciel.

Notre travail aura la structure suivante :

- Le premier chapitre de ce mémoire de fin d’étude aura pour objectif de présenter successivement le marché pharmaceutique mondial, le marché pharmaceutique algérien, ainsi que l’entreprise Biopharm Industrie. Nous procéderons à une analyse interne des processus d’achats et des entretiens, ainsi qu’à une analyse externe mettant en évidence les nombreux défis et enjeux de l’environnement global de l’entreprise en utilisant la méthode PESTEL. Enfin, ces étapes nous permettront d’identifier et de classer les dysfonctionnements afin de déterminer la problématique centrale.
- Le deuxième chapitre sera consacré à l’exposition et à l’explication des différents outils, notions et concepts que nous avons utilisés pour résoudre notre problématique.

- Dans le dernier chapitre, nous proposerons une résolution de la problématique en développant une démarche d'amélioration de la performance. Nous mettrons en place une interface graphique spécialement conçue pour faciliter la prise de décision lors de l'achat des pièces de rechange. Cette interface graphique fournira des outils interactifs et conviviaux permettant d'analyser les différentes options d'achat, d'évaluer leur pertinence en fonction de critères prédéfinis, et de générer des recommandations précises. Cette approche basée sur une interface graphique apportera une assistance efficace aux acteurs impliqués dans le processus d'achat des pièces de rechange, en leur offrant une meilleure visibilité et une prise de décision plus éclairée.

Nous concluons notre travail en proposant des perspectives prometteuses et des pistes concrètes pour améliorer la situation. Enfin, nous clôturerons notre étude par une synthèse et une conclusion finale.

Chapitre 1

État des lieux

Chapitre 1 Etat des lieux

Introduction

Dans ce chapitre, nous adopterons une approche en entonnoir en commençant par une présentation générale du secteur pharmaceutique avant de nous concentrer sur le cas spécifique de Biopharm Algérie. Nous procéderons ensuite à un diagnostic interne et externe qui comprendra une analyse de l'état actuel, des processus cibles et des contraintes, afin d'identifier les compétences et les besoins. Enfin, nous énonçons une problématique qui sera examinée dans les trois chapitres suivants.

1. Présentation du secteur d'activité

1.1. L'industrie pharmaceutique dans le monde

L'industrie pharmaceutique est responsable du développement, de la production et de la commercialisation des médicaments et d'autres produits pharmaceutiques. Son importance en tant que secteur global est indiscutable. Le chiffre d'affaires mondial de ce secteur a dépassé le cap du millier de milliards de dollars des États-Unis en 2014, un chiffre qui, d'ailleurs, a été généré de moitié par les États-Unis. Le deuxième marché le plus important est le marché européen, avec environ 200 milliards de dollars de recettes dans la vente de produits pharmaceutiques [1].

Selon Xerfi, la production pharmaceutique en France devrait connaître une légère augmentation de 1% en 2022, malgré une baisse en début d'année due à la politique de déstockage des fabricants de médicaments et des baisses de prix sur le marché national. En revanche, les prévisions pour 2023 sont plus encourageantes, avec une croissance de 1,5% de la production de l'industrie pharmaceutique, soutenue par des investissements importants annoncés lors du sommet "Choose France" début 2022. Des entreprises telles que GSK, Merck KGaA et Seqens ont annoncé des plans d'investissement significatifs dans leurs usines françaises, ce qui devrait stimuler la croissance de l'industrie pharmaceutique en 2023.

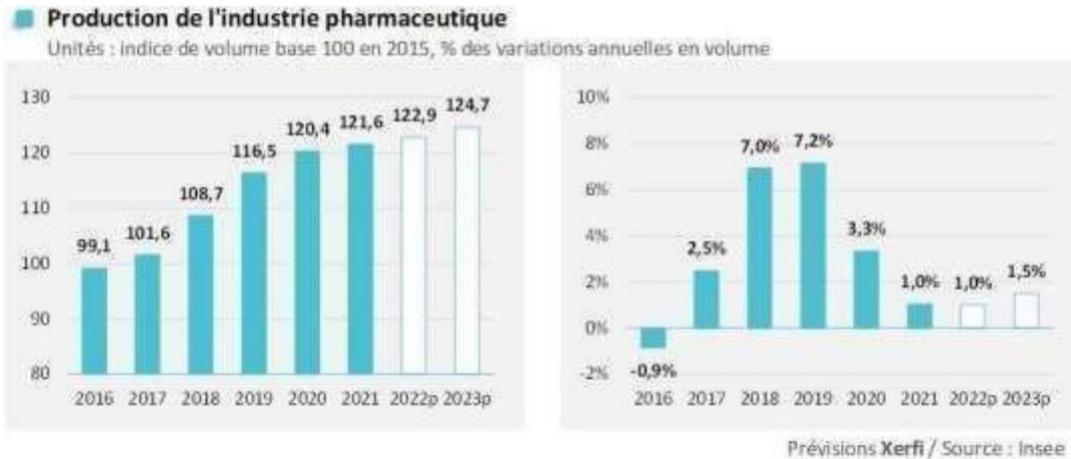


Figure 1 Hausse des capacités de production et investissements dans l'industrie pharmaceutique française en 2023 [2]

1.2. L'industrie pharmaceutique en Algérie

En quelques décennies, l'Algérie s'est hissée parmi les principaux producteurs pharmaceutiques d'Afrique. Plus de 2,5 milliards d'euros de médicaments ont été produits l'an dernier. Quelque 200 unités de production couvrent près de 70 % des besoins du pays en générique [1].

Un exemple avec les laboratoires Frater-Razes qui, en pleine pandémie, ont mis sur le marché un anticoagulant qui a été prescrit aux malades du Covid-19. C'est la première fois en Algérie qu'étaient produits des médicaments biosimilaires, c'est-à-dire fait à partir d'organismes vivants.

Plusieurs réformes ont permis un changement de paradigme dans le pays, où les produits de santé qui étaient une enveloppe budgétaire sont devenus un levier de croissance. Cet autre laboratoire privé, Biopharm, commercialise depuis peu des produits d'oncologies, des anticancéreux, dont l'élaboration nécessite des infrastructures complexes et des investissements massifs.

Cependant, le secteur pharmaceutique algérien doit encore relever plusieurs défis, notamment la nécessité de moderniser ses installations de production, d'améliorer la qualité de ses produits et de renforcer la recherche et le développement pour proposer des médicaments innovants. Malgré cela, l'industrie pharmaceutique en Algérie offre de nombreuses opportunités d'investissement pour les entreprises locales et internationales souhaitant participer à son développement.

2. Présentation de l'entreprise

2.1. Présentation générale

BIOPHARM est un groupe industriel et commercial algérien qui opère dans le secteur pharmaceutique. Créé dans les années 1990, il s'est développé au fil des ans pour devenir une entreprise dotée d'une unité de production aux normes internationales et d'un réseau de distribution auprès de grossistes et de pharmacies.

Le groupe BIOPHARM s'est récemment restructuré autour de différents métiers, notamment la production de médicaments via BIOPHARM, la distribution en gros de produits pharmaceutiques via BIOPHARM DISTRIBUTION, la répartition aux officines via BIOPURE, la promotion et l'information médicales via HHI (Human Health Information), ainsi que la logistique pour l'industrie pharmaceutique via BIOPHARM LOGISTIC.

Pour soutenir son développement, le groupe a également ouvert son capital à un consortium international composé de ADPI, un fonds d'investissement spécialisé en Afrique, DEG, une filiale de la KfW spécialisée dans le financement de projets à long terme, et FMC COFARES, un véhicule d'investissement détenu par deux fonds institutionnels spécialisés sur l'Afrique du Nord. Cette alliance vise à tirer profit des réseaux du consortium pour poursuivre le plan de développement de ses activités en amont et en aval, assurer un transfert de technologie, renforcer l'encadrement managérial et développer ses marchés à l'exportation en Afrique et en Europe.

Le Groupe BIOPHARM a pour ambition de réaliser un programme de développement ambitieux tout en assurant la sécurité financière et en garantissant à ses partenaires une forte croissance de ses activités.

2.2. Structure du groupe "BIOPHARM GROUP"

La structure organisationnelle du groupe BIOPHARM se compose de quatre filiales qui reflètent les différentes activités du groupe dans le secteur pharmaceutique :

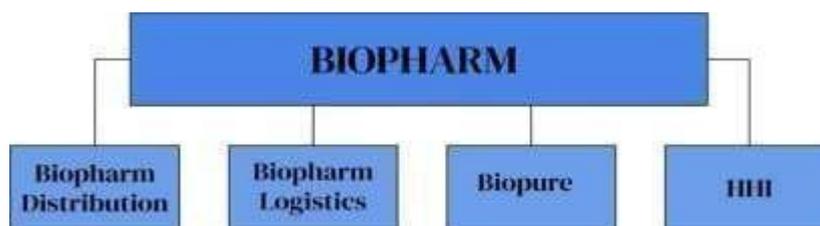


Figure 2 Structure organisationnelle du groupe BIOPHARM

BIOPHARM : cette filiale est le noyau central du groupe et se concentre sur la production de médicaments. Elle dispose d'une unité de production aux normes internationales et est responsable de la fabrication des produits pharmaceutiques du groupe BIOPHARM.

BIOPHARM DISTRIBUTION : cette filiale s'occupe de la distribution en gros de produits pharmaceutiques auprès des grossistes et des pharmacies. Elle dispose d'un réseau de distribution bien établi et travaille en étroite collaboration avec les autres filiales du groupe.

BIOPURE : cette filiale est spécialisée dans la répartition de produits pharmaceutiques aux officines. Elle assure la logistique de la chaîne d'approvisionnement en s'occupant de la gestion des stocks, de la préparation des commandes et de la livraison aux pharmacies.

HHI (Human Health Information) : cette filiale est responsable de la promotion et de l'information médicales. Elle a pour mission d'informer les professionnels de la santé sur les produits pharmaceutiques du groupe BIOPHARM, ainsi que sur les bonnes pratiques médicales.

En résumé, ces quatre filiales du groupe BIOPHARM couvrent l'ensemble de la chaîne de valeur du secteur pharmaceutique, depuis la production jusqu'à la distribution, en passant par la répartition et la promotion médicale. Cette structure organisationnelle permet au groupe de coordonner efficacement ses activités et de maximiser ses performances commerciales.

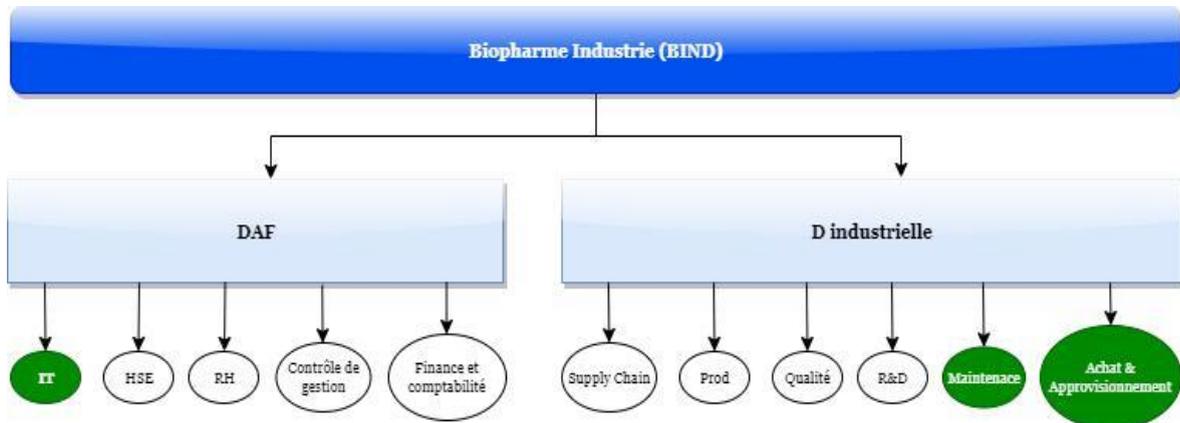
2.3. Structure interne de “BIOPHARM INDUSTRY”

Biopharm Industrie est la société mère du groupe, spécialisée dans la production de médicaments. Pour répondre aux différents métiers, l'entreprise a adapté sa structure organisationnelle et est composée de deux directions principales :

La direction industrielle : cette direction contient les départements et services opérationnels de l'entreprise et est entièrement axée sur la production, le contrôle de qualité des produits et des ressources, ainsi que sur le bon déroulement des processus de la chaîne d'approvisionnement (supply chain).

La direction administrative et financière (DAF) : cette direction est responsable des affaires administratives, de la comptabilité et de la gestion des ressources humaines. Elle réunit la majorité des processus support de l'entreprise et comprend différents départements et services.

Nous avons présenté la structure interne de l'entreprise dans l'organigramme suivant :



3. Périmètre du projet

Notre étude portera sur 3 départements impliqués dans le processus achat. Afin de délimiter le périmètre du projet ainsi qu'effectuer le diagnostic de l'entreprise nous avons élaboré un questionnaire et nous nous sommes entretenus avec les membres des différents départements (IT, Maintenance, DA&A).

Suite à nos entretiens avec les membres des différents départements impliqués dans le processus achat, nous avons identifié que le département IT est considéré comme le **maître d'œuvre** dans ce projet, c'est-à-dire qu'il sera en charge de la mise en place de la solution technique. En revanche, les départements Maintenance et DA&A seront considérés comme **maîtres d'ouvrage**, c'est-à-dire qu'ils seront responsables de la définition des besoins et des spécifications fonctionnelles de la solution. Cette distinction est importante pour assurer une bonne collaboration entre les différents acteurs du projet et garantir la réussite de celui-ci. Nous allons donc travailler en étroite collaboration avec les membres de ces départements pour élaborer une solution adaptée à leurs besoins et à leurs contraintes.

3.1. Département IT

Le département IT joue un rôle clé dans notre projet. Ce département est responsable de collecter, stocker et gérer les données pertinentes pour l'analyse et la prédiction des besoins en pièces de rechange. L'utilisation de solutions informatiques performantes permettra d'optimiser les processus d'achat et

d'approvisionnement, de mieux suivre les stocks de pièces de rechange et de faciliter la prise de décision. Le système d'information intégré permettra également de centraliser les données sur les achats et les approvisionnements pour une gestion plus efficace.

L'utilisation de solutions informatiques performantes : Pour optimiser les processus d'achat et d'approvisionnement, le projet nécessite l'utilisation de solutions informatiques performantes. Les solutions logicielles modernes permettent une gestion plus efficace des processus d'achat et d'approvisionnement en automatisant les tâches répétitives et en fournissant des données en temps réel pour faciliter la prise de décision. Le système d'information intégré permettra également de centraliser les données sur les achats et les approvisionnements, ce qui facilitera la gestion et le suivi de l'ensemble du processus.

- Suivi des performances en temps réel : Le système d'information intégré permettra également de suivre les performances en temps réel, ce qui facilitera la prise de décision et la mise en place de plans d'action. Les indicateurs de performance clés tels que les délais d'attente pour l'obtention des pièces de rechange, les taux de rotation des stocks et les coûts d'achat et de stockage seront disponibles en temps réel pour permettre une gestion proactive des achats et des approvisionnements.

3.2. Département Maintenance

Le département Maintenance est essentiel dans le cadre de ce projet, car il est chargé de veiller au bon fonctionnement et à l'efficacité des équipements et des machines de l'entreprise. Son rôle dans ce projet comprend la définition des besoins en pièces de rechange et l'établissement des spécifications techniques.

Ce département possède l'expertise technique requise pour déterminer quels composants sont susceptibles de s'utiliser plus vite et nécessitent donc un remplacement régulier. Ces informations sont vitales pour s'assurer que le système d'information mis en œuvre par le département IT puisse prévoir avec exactitude les besoins en pièces de rechange.

Il est également impératif que le département Maintenance collabore étroitement avec le département IT. Cette collaboration permettra de garantir que les solutions techniques développées correspondent aux besoins spécifiques du département Maintenance et contribuent à optimiser les processus d'achat de pièces de rechange.

3.3. Département DA&A

Le département DA&A, pour Demande, Achat et Approvisionnement, est au cœur de la gestion de la chaîne d'approvisionnement et de l'optimisation des

achats. Il veille à ce que l'entreprise dispose des pièces de rechange nécessaires, au bon moment et au meilleur coût.

Une fonction cruciale de ce département est la gestion des demandes, qui implique d'anticiper les besoins en pièces de rechange en travaillant étroitement avec le département Maintenance pour comprendre quels composants pourraient nécessiter un remplacement et de planifier en conséquence.

Il est également chargé de la sélection des fournisseurs de pièces de rechange, ce qui nécessite d'analyser différents fournisseurs en termes de coûts, de qualité et de fiabilité, et de négocier les meilleures conditions possibles.

La gestion des stocks est un autre aspect essentiel du rôle de DA&A. Cela inclut le suivi des niveaux de stock de pièces de rechange, de s'assurer que les commandes sont passées en temps opportun et d'éviter que l'entreprise ne se retrouve en rupture de stock ou avec un excès d'inventaire.

En outre, le département DA&A doit collaborer avec le département IT pour s'assurer que le système d'information répond aux besoins de gestion des demandes, des achats et des approvisionnements, et facilite l'optimisation de ces processus.

Globalement, le département DA&A est indispensable pour veiller à ce que l'entreprise gère efficacement sa chaîne d'approvisionnement en s'assurant qu'elle a les pièces de rechange nécessaires, tout en optimisant les coûts et l'efficacité.

4. Diagnostic et problématique

Dans ce qui va suivre, nous allons procéder en deux temps :

1. Nous allons procéder à une première analyse externe de BIOPHARM en utilisant l'analyse SWOT-PESTEL afin de mieux comprendre son positionnement par rapport à son environnement.
2. Le deuxième diagnostic portera sur le processus d'achat de pièces de rechange chez BIOPHARM, en mettant l'accent sur la manière dont les besoins en pièces de rechange sont identifiés et planifiés pour une période donnée.

4.1. Diagnostic externe

Il est important de noter que l'analyse SWOT-PESTEL ne doit pas être considérée comme une méthode exhaustive pour évaluer l'entreprise. D'autres méthodes et outils d'analyse peuvent également être utilisés pour obtenir une compréhension plus approfondie de l'entreprise. Cependant, cette approche est souvent utilisée car elle fournit une vue d'ensemble de l'environnement dans

lequel l'entreprise opère et de son positionnement actuel sur le marché. Maintenant, sans plus tarder, commençons par l'analyse SWOT-PESTEL de BIOPHARM pour mieux comprendre cette entreprise.

Analyse SWOT

Dans cette analyse interne et externe, nous allons pouvoir détecter et énumérer les forces et faiblesses de l'entreprise en interne et les opportunités et les menaces en externe. Les résultats de l'analyse sont présentés sur la figure.

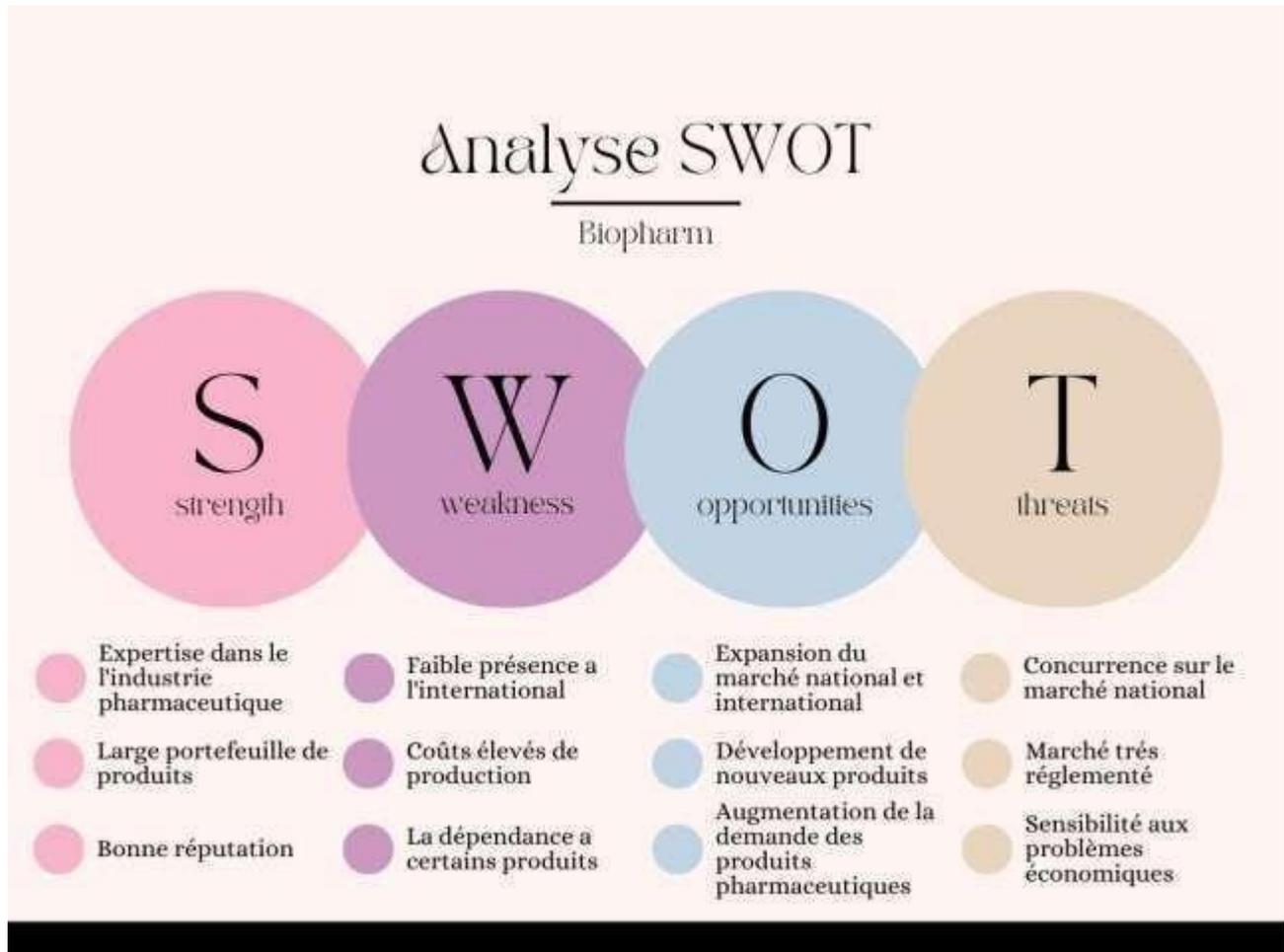


Figure 4 Analyse SWOT

Analyse PESTEL

Nous Nous examinerons les facteurs PESTEL (politiques, économiques, sociaux, technologiques, environnementaux et législatifs) impactant l'entreprise. Cette analyse nous aidera à identifier les opportunités et les menaces, et est résumée dans la figure ci-dessous.

Tableau 1 Analyse PESTEL

Facteur	Points clés	Chapitre 1. Etat des
Politique	- Réglementations gouvernementales sur le secteur pharmaceutique	
	- Politiques d'importation et d'exportation de médicaments	
Économique	- Taux de croissance économique du pays	
	- Tendances du pouvoir d'achat et de la consommation	
Socio-culturel	- Profil démographique de la population (âge, genre, etc.)	
	- Attitudes envers la santé et les médicaments	
Technologique	- Avancées technologiques dans la recherche médicale et pharmaceutique	
	- Adoption de nouvelles technologies dans la production et la distribution de médicaments	
Environnemental	- Réglementations environnementales liées à la production et à l'élimination des déchets	
	- Sensibilisation croissante à l'écologie et à la durabilité	
Légal	- Loi sur la propriété intellectuelle et les brevets	
	- Réglementations des essais cliniques et approbation médicaments.	

Pourquoi avons-nous choisi PESTEL ?

Politique : Les changements politiques peuvent avoir un impact sur la chaîne d'approvisionnement de Biopharm, notamment en ce qui concerne les réglementations relatives à l'importation et l'exportation de pièces de rechange.

Économique : Les changements économiques, tels que les fluctuations du taux de change, peuvent affecter les coûts d'achat et d'approvisionnement des pièces

de rechange.

Socioculturel : Les tendances socioculturelles, comme l'accent mis sur les entreprises respectueuses de l'environnement, peuvent influencer les choix de Biopharm en matière de fournisseurs de pièces de rechange.

Technologique : Les avancées technologiques, comme l'utilisation de l'intelligence artificielle pour prévoir les besoins en pièces de rechange, peuvent aider Biopharm à optimiser son processus d'approvisionnement.

Environnemental : Les préoccupations environnementales peuvent avoir un impact sur les choix de Biopharm en matière de fournisseurs de pièces de rechange.

Légal : Les changements dans les réglementations juridiques peuvent avoir un impact sur les pratiques d'achat et d'approvisionnement de Biopharm, notamment en ce qui concerne la gestion des risques liés à la conformité.

Comment peut-on utiliser les techniques de la data science et de l'intelligence artificielle pour améliorer le processus d'achat et d'approvisionnement des pièces de rechange indispensables à la production ? Plus précisément, comment peut-on prévoir le nombre de pièces nécessaires pour les mois et les années à venir pour chaque articles spécifiques ?

En développant une interface de prédiction de la quantité d'achat de chaque pièce de rechange dans une période donnée, basée sur l'analyse des données historiques et en identifiant les principaux facteurs qui influencent les besoins en pièces de rechange, tout en intégrant les prévisions d'achat avec les processus d'achat existants de l'entreprise et en identifiant les fournisseurs les plus fiables.

4.2. Diagnostic Interne

Le département des achats et des approvisionnements occupe une position cruciale au sein de l'entreprise en raison de son rôle dans :

- La satisfaction des besoins des divers services de l'entreprise (Supply Chain, Maintenance, Production, LCQ, Développement, Métrologie) ;
- La livraison de produits conformes dans les délais impartis à un coût optimisé, tout en respectant le budget approuvé. Ce département constitue l'amont de la chaîne d'approvisionnement et sa performance influence directement les autres processus.

Au sein de Biopharm SPA, ce département est responsable de l'achat et de l'approvisionnement des matières premières, des articles d'emballage, des

équipements , **des pièces de rechange**, des produits chimiques, ainsi que des divers consommables et services nécessaires à l'entreprise.

Le processus d'achat (y compris la prospection) est généralement similaire pour les différents types de produits, bien que certains détails techniques et réglementaires puissent varier en fonction de la nature de l'achat. Nous distinguons ainsi :

Les matières premières et les articles d'emballage : La demande provient du service de planification et concerne les produits de routine.

Ces demandes sont généralement régulières, mais parfois difficiles à anticiper en raison de la volatilité du marché.

Les équipements et **les pièces de rechange** : Les demandes sont généralement régulières, en particulier pour les pièces de rechange et les consommables, à l'exception des investissements.

Les achats LCQ : Ils passent systématiquement par la procédure de prospection, qui vise à rechercher de nouveaux fournisseurs et à obtenir les meilleures offres en termes de coûts et de délais.

Nous nous intéressons aux pièces de rechange. Comment sont décidées les quantités de matières premières à acheter ? Suite à des entretiens menés auprès de différents départements, voici comment les décisions sont prises :

Les départements établissent un **seuil d'approvisionnement et un seuil maximal** pour chaque article (pièce de rechange).

Ces seuils sont déterminés en tenant compte des trois éléments suivants :

a. Consommation des périodes précédentes, ajustée légèrement à la hausse ou à la baisse

Les départements étudient les données historiques de consommation afin d'identifier les tendances et variations saisonnières. Au lieu de se baser uniquement sur des calculs empiriques, ils s'appuient principalement sur la consommation annuelle, soit sur une période d'un an.

Ils ajustent ensuite les seuils d'approvisionnement en fonction de ces tendances, en ajoutant ou en retranchant une certaine quantité pour tenir compte des variations imprévues, telles que des ruptures de stock ou des changements dans la demande.

b. Guide manuel de la machine

Les départements consultent les guides manuels des machines pour connaître les spécifications techniques des pièces de rechange nécessaires, ainsi que les intervalles de maintenance recommandés.

Ils tiennent compte de ces informations pour estimer les besoins en pièces de rechange, en s'assurant que les seuils d'approvisionnement permettent de maintenir les machines en bon état de fonctionnement.

c. Politique algérienne en matière de dédouanement

Les départements prennent en compte les régulations et les procédures douanières algériennes, qui peuvent influencer les délais et les coûts d'approvisionnement des pièces de rechange importées.

Ils tiennent compte des éventuelles taxes, droits de douane et autres frais associés à l'importation de ces pièces, ainsi que des délais de dédouanement, pour s'assurer que les seuils d'approvisionnement sont suffisamment élevés pour éviter les ruptures de stock et les retards de production.

Ces informations, collectées lors des entretiens, décrivent le processus de décision actuel pour déterminer les quantités de matières premières à acheter, en particulier pour les pièces de rechange.

4.3. Synthèse du diagnostic

Nous pouvons résumer le diagnostic dans ce qui suit :

Dysfonctionnements

- **Prévisions inexactes** : Les prévisions actuelles des besoins en pièces de rechange sont basées sur l'expérience humaine et l'intuition, ce qui peut entraîner des erreurs et des inefficacités.
- **Manque de coordination** : Les différents départements peuvent avoir du mal à partager les informations pertinentes et à collaborer efficacement dans la prise de décision concernant les achats de pièces de rechange.
- **Insuffisance d'analyse de données** : Les données historiques et les informations issues des guides manuels des machines et de la politique algérienne en matière de dédouanement ne sont pas pleinement exploitées pour optimiser le processus d'achat.

Obstacles

Chapitre 1. Etat des lieux

- Résistance au changement : L'introduction de nouvelles méthodes et technologies basées sur la data science et l'IA pourrait rencontrer une résistance de la part des employés habitués aux approches traditionnelles.
- Compétences en data science : Les employés de l'entreprise peuvent manquer de compétences en data science et en intelligence artificielle, ce qui rend difficile l'adoption et la mise en œuvre de solutions basées sur ces technologies.
- Complexité des données : La collecte, la gestion et l'analyse de grandes quantités de données pertinentes pour les achats de pièces de rechange peuvent s'avérer complexes et nécessiter des ressources supplémentaires en termes de temps, de personnel et d'investissements.
- Intégration des systèmes existants : L'intégration de nouvelles solutions de data science et d'IA avec les systèmes et processus existants au sein de l'entreprise peut présenter des défis techniques et organisationnels.
- Confidentialité et sécurité des données : La gestion et l'analyse des données sensibles de l'entreprise, telles que les informations sur les fournisseurs et les transactions, peuvent soulever des préoccupations concernant la confidentialité et la sécurité des données.

Cela nous permet de formuler nos **objectifs**

1. Développer une interface de prédiction de la quantité d'achat de chaque pièce de rechange dans une période donnée, basée sur l'analyse des données historiques.
2. Concevoir des modèles de prédiction précis pour chaque pièce de rechange en utilisant des techniques de data science appropriée.
3. Identifier les principaux facteurs qui influencent les besoins en pièces de rechange, tels que la demande, les taux d'utilisation et les cycles de maintenance des équipements.
4. Intégrer les prévisions d'achat de pièces de rechange avec les processus d'achat existants de l'entreprise.
5. Identifier les fournisseurs les plus fiables, les délais de livraison les plus courts et les coûts les plus bas pour chaque pièce de rechange, en utilisant des techniques d'analyse de données.
6. Évaluer l'impact de l'interface sur la précision des prévisions d'achat de pièces de rechange et les performances globales de l'entreprise.
7. Fournir des recommandations pour améliorer l'efficacité et la rentabilité des achats de pièces de rechange basées sur les résultats de l'analyse des données.

8. Développer des compétences en data science et en intelligence artificielle chez les employés de l'entreprise pour permettre une adoption et une utilisation efficaces de l'interface.
9. Garantir la confidentialité et la sécurité des données sensibles de l'entreprise à travers la mise en place de mesures de protection appropriées.

Énoncé de la problématique

L'achat de pièces de rechange est un processus qui sert à garantir le fonctionnement efficace des machines et des équipements de l'entreprise. Cependant, les prévisions actuelles des besoins en pièces de rechange sont souvent basées sur l'expérience humaine et l'intuition, ce qui peut entraîner des erreurs et des inefficacités. De plus, le manque de coordination entre les différents départements de l'entreprise et l'insuffisance d'analyse de données empêchent une optimisation efficace du processus d'achat.

Dans ce contexte, l'objectif de ce travail est de développer une interface de prédiction de la quantité d'achat de chaque pièce de rechange dans une période donnée, basée sur l'analyse des données historiques. Cette interface permettra à l'entreprise de prendre des décisions d'achat plus éclairées et basées sur des données prévisionnelles issues de méthodes scientifiques et techniques de prédiction, plutôt que sur l'expérience humaine et l'intuition. Les modèles de prédiction précis pour chaque pièce de rechange seront développés en utilisant des techniques de data science appropriée, en identifiant les principaux facteurs qui influencent les besoins en pièces de rechange. Les prévisions d'achat de pièces de rechange seront ensuite intégrées avec les processus d'achat existants de l'entreprise, tout en identifiant les fournisseurs les plus fiables, les délais de livraison les plus courts et les coûts les plus bas pour chaque pièce de rechange, en utilisant des techniques d'analyse de données.

L'énoncé complet de la problématique sera donc

Comment peut-on utiliser les techniques de la data science et de l'intelligence artificielle pour améliorer le processus d'achat et d'approvisionnement des pièces de rechange indispensables à la production ? Plus précisément, comment peut-on prévoir le nombre de pièces nécessaires pour les mois et les années à venir pour chaque articles spécifique ?

En développant une interface de prédiction de la quantité d'achat de chaque pièce de rechange dans une période donnée, basée sur l'analyse des données historiques et en identifiant les principaux facteurs qui influencent les besoins en pièces de rechange, tout en intégrant les prévisions d'achat avec les processus

d'achat existants de l'entreprise et en identifiant les fournisseurs les plus fiables.

Conclusion

Dans ce chapitre d'état des lieux, une analyse exhaustive de l'entreprise, son domaine d'activité et ses défis a été menée. Un examen externe et interne a permis d'évaluer les éléments macro-environnementaux, les enjeux, les forces, les faiblesses, les obstacles, et les dysfonctionnements à adresser grâce à la solution suggérée.

Ce chapitre a éclairé le contexte, la problématique et les objectifs. Il a mis en évidence les problèmes que l'entreprise rencontre et que la data science peut résoudre pour optimiser les achats de pièces de rechange.

Les chapitres suivants détailleront les concepts, outils et élaboration de la proposition de résolution."

Chapitre 2

État de l'art

Chapitre 2 Etat de l'art

Introduction

La prévision et la prédiction jouent un rôle essentiel dans la gestion des données pour anticiper les tendances et prendre des décisions éclairées. Ce document explore les concepts fondamentaux de la prévision d'achat en utilisant des méthodes avancées telles que l'intelligence artificielle, le machine learning et les séries temporelles. Nous examinerons les différences entre la prévision et la prédiction, l'importance des bases de données, ainsi que l'utilisation des séries temporelles et du machine learning, y compris le deep learning. Nous aborderons également l'évaluation des modèles de prédictions pour améliorer leur précision. Ce document vise à fournir une compréhension approfondie des concepts clés nécessaires pour développer des modèles de prédictions efficaces dans le domaine de l'achat.

1. Termes et définitions

1.1 Prévision

La prévision est un processus qui utilise les données historiques pour prédire les événements futurs. Elle est couramment utilisée dans de nombreux domaines, tels que la météorologie, l'économie, la finance et la planification des opérations. Les méthodes de prévision peuvent varier de simples approches basées sur des tendances historiques à des modèles statistiques complexes. [1], [2]

1.2 Prédiction

La prédiction est le processus d'utilisation d'un modèle ou d'un algorithme pour estimer le résultat pour une entrée donnée. En apprentissage automatique, la prédiction est souvent associée à des résultats probabilistes, où le modèle prédit la probabilité de chaque résultat possible pour une entrée donnée. [3]

1.3 Prévision vs Prédiction

Bien que ces termes soient souvent utilisés de manière interchangeable dans de nombreux domaines, en statistique, une prévision est une prédiction des valeurs futures d'une série temporelle, tandis qu'une prédiction est souvent associée à des modèles d'apprentissage automatique et peut concerner tout type d'événement. La prévision est généralement utilisée dans le contexte de l'analyse

de séries temporelles pour prédire des événements futurs sur la base de données Prévision d'achat historiques. [1]

La prévision d'achat fait référence à la prédiction du comportement d'achat futur des clients. Elle est souvent utilisée dans les domaines de la vente et du marketing pour planifier les stratégies d'inventaire et de vente. Ces prévisions peuvent être basées sur des données historiques d'achat, des tendances du marché, des facteurs saisonniers et d'autres variables pertinentes.[4]

1.4 Prévision d'achat des pièces de rechange dans le domaine de la pharma

La prévision d'achat des pièces de rechange dans le domaine de l'industrie pharmaceutique est un processus critique pour maintenir la continuité des opérations. Les équipements utilisés dans la production pharmaceutique sont généralement très spécialisés et coûteux. De ce fait, avoir un stock approprié de pièces de rechange est essentiel pour minimiser les temps d'arrêt en cas de défaillance de l'équipement.

L'objectif de la prévision d'achat des pièces de rechange est d'assurer que l'entreprise dispose toujours des pièces nécessaires pour maintenir ses opérations en cours, tout en minimisant les coûts associés au stockage excessif de pièces. L'utilisation de modèles mathématiques, de logiciels de prévision et d'algorithmes d'intelligence artificielle peut grandement aider dans ce processus.

1.5 Base de données

Une base de données est une collection organisée de données stockées et accessibles électroniquement. Les bases de données sont conçues pour offrir un moyen efficace de stocker, récupérer et gérer des informations. Elles peuvent stocker divers types de données, y compris du texte, des images, des données audio et vidéo, et des données structurées telles que les listes et les tableaux. Les bases de données sont essentielles pour de nombreuses applications, allant des sites web et des applications mobiles aux systèmes d'information d'entreprise et aux centres de recherche scientifique.[5]

1.6 Série temporelle

Une série temporelle est une séquence de points de données numériques en ordre successif. Dans une série temporelle, le temps est souvent la variable indépendante et l'objectif est généralement de faire une prévision pour l'avenir. Les séries temporelles sont couramment utilisées dans de nombreux domaines, y compris l'économie, la finance, la météorologie, et l'ingénierie, pour analyser les tendances temporelles, les cycles saisonniers, et les effets de certaines interventions ou événements.[6]

1.7 Intelligence Artificielle (IA)

L'IA fait référence à la simulation des processus d'intelligence humaine par des machines, en particulier des systèmes informatiques. Ces processus comprennent l'apprentissage (la capacité d'acquérir et de comprendre les informations), le raisonnement (l'utilisation des règles pour atteindre des conclusions approximatives ou définitives), la résolution de problèmes, la perception (la reconnaissance et l'interprétation des stimuli sensoriels) et la compréhension du langage (la compréhension et la génération du langage naturel). L'objectif de l'IA est de créer des systèmes informatiques qui sont capables de réaliser des tâches qui nécessiteraient normalement l'intelligence humaine.[7]

1.8 Machine Learning (ML)

Le ML est un sous-domaine de l'IA qui se concentre sur la conception de systèmes qui peuvent apprendre à partir de données, plutôt que d'être explicitement programmés. Le ML utilise des algorithmes pour analyser les données, apprendre à partir de ces données, et faire des déterminations ou des prédictions sur le monde. Par exemple, un algorithme d'apprentissage automatique pourrait être formé sur un ensemble de données d'e-mails pour apprendre à distinguer les spams des e-mails légitimes.[8]

2. Concepts fondamentaux sur les séries temporelles

2.1 Types de séries temporelles

2.1.1 Séries temporelles univariées

Une série temporelle univariée est une séquence d'observations de données recueillies à des intervalles de temps successifs. Chaque observation dans la séquence est une seule variable, d'où le terme "univariée". Les modèles de séries temporelles univariées tentent de comprendre et de prédire les futures observations en fonction des observations passées.[9]

2.1.2 Séries temporelles multivariées

Une série temporelle multivariée est une généralisation de la série temporelle univariée. Au lieu d'une seule observation à chaque intervalle de temps, on a plusieurs observations, ce qui peut permettre de comprendre les relations entre

différentes variables au fil du temps.[10]

2.1.2 Composantes de séries temporelles

Les séries temporelles peuvent souvent être décomposées en plusieurs composantes principales :

- **Tendance** Il s'agit de la direction générale dans laquelle se déplace une série temporelle sur une longue période. Par exemple, la population d'un pays peut avoir une tendance à la hausse sur plusieurs décennies.
- **Saisonnalité** Il s'agit de variations régulières et prévisibles qui se produisent sur une période spécifique, comme une année, un mois ou une semaine. Par exemple, les ventes de glaces peuvent augmenter pendant les mois d'été et diminuer pendant l'hiver.
- **Cycles** Il s'agit de fluctuations qui ne sont pas de nature saisonnière et dont la durée n'est pas fixe. Par exemple, les cycles économiques qui se produisent sur plusieurs années.
- **Bruit ou irrégularité** Il s'agit des variations aléatoires qui ne peuvent pas être expliquées par la tendance, la saisonnalité ou les cycles.

TIME SERIES DATA COMPONENTS

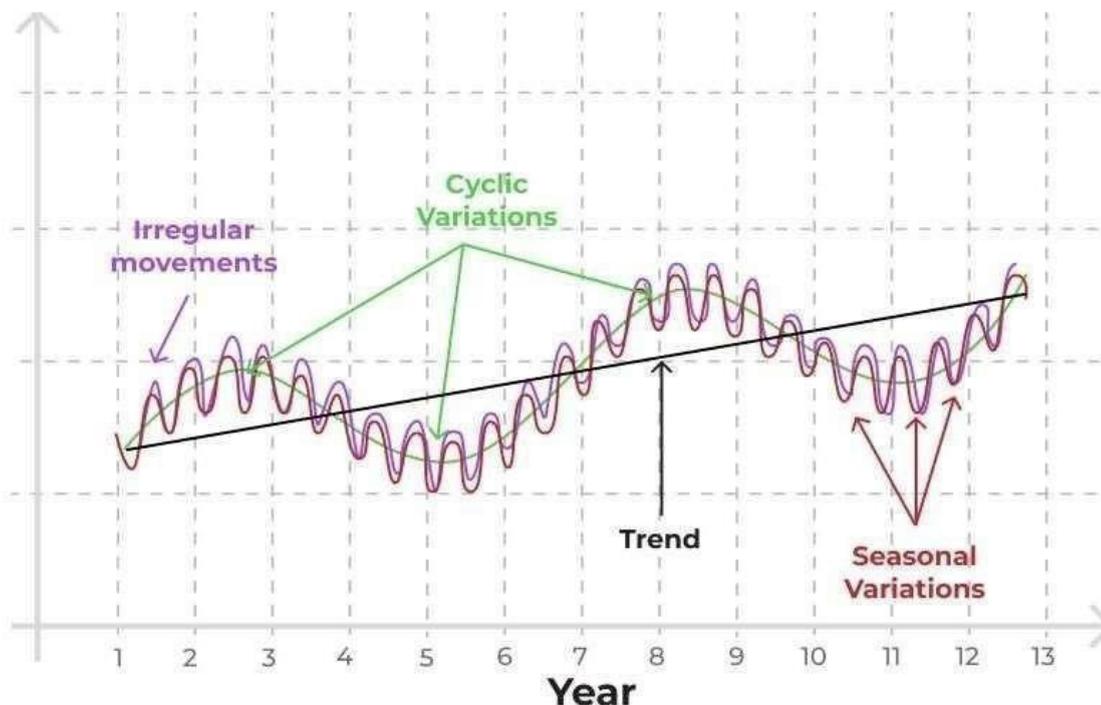


Figure 5 Les composants d'une série temporelle [3]

2.2 Stationnarité

C'est un concept fondamental dans les processus stochastiques où tous les moments, les distributions conjointes et les caractéristiques du processus restent inchangés dans le temps, ce qui garantit que les propriétés statistiques telles que la moyenne, la variance et l'autocovariance restent constantes, permettant une analyse et une modélisation fiables et cohérentes du processus sur toute sa durée.

2.3 Retard (Lag)

Dans le contexte des séries temporelles, un retard (ou lag) fait référence à l'écoulement du temps entre deux points de données. Par exemple, un retard de 1 signifie que l'observation à un moment donné est comparée à l'observation au moment précédent. Les retards sont souvent utilisés dans les modèles de séries temporelles pour capturer les dépendances temporelles.[12]

2.4 Autocorrélation

L'autocorrélation est une mesure statistique qui décrit le degré de corrélation entre les valeurs d'une série temporelle et les valeurs précédentes de la même série. Une autocorrélation positive signifie que les valeurs plus élevées d'une série sont généralement suivies par des valeurs plus élevées, et les valeurs plus basses par des valeurs plus basses. Une autocorrélation négative signifie le contraire.[13]

2.5 Transformations et différenciations :

Les transformations et les différenciations sont des techniques utilisées pour rendre une série temporelle plus facile à modéliser. Les transformations peuvent inclure des choses comme la prise du logarithme de la série, tandis que la différenciation implique de soustraire la valeur précédente de chaque valeur, ce qui peut aider à rendre la série plus stationnaire.

2.6 Transformation Logarithmique

La transformation logarithmique est une technique utilisée en statistiques et en science des données pour rendre une distribution fortement asymétrique plus symétrique, et donc plus conforme à la loi normale, ce qui est une hypothèse nécessaire pour de nombreuses techniques statistiques. La transformation logarithmique est particulièrement utile lorsque les données suivent une loi de puissance, comme c'est souvent le cas dans les sciences de l'information. En pratique, cela implique de prendre le logarithme de chaque observation dans le jeu de données. Cela peut aider à réduire l'effet des valeurs aberrantes et à rendre

les modèles statistiques plus robustes.[14]

2.7 Transformation de Box-Cox

La transformation de Box-Cox est une transformation de données statistiques qui vise à améliorer la normalité, la linéarité et l'homoscédasticité (variance constante) d'une variable. Elle est particulièrement utile pour les modèles de régression où ces conditions sont souvent nécessaires. La transformation de Box- Cox est définie comme une transformation paramétrique de données non négatives à l'aide d'un paramètre de puissance. Si les données originales suivent une distribution log-normale, alors les données transformées suivront une distribution normale. [16]

2.8 Normalisation

La normalisation est une technique de mise à l'échelle ou de mappage, ou une étape de prétraitement. Elle permet de trouver une nouvelle plage à partir d'une plage existante. Elle peut être utile pour la prédiction ou la prévision, car elle permet de réduire les grandes variations de prédiction et de prévision.[17]

Il existe plusieurs techniques de normalisation, notamment la normalisation Min-Max, la normalisation Z-score et la mise à l'échelle décimale. La normalisation Min- Max est une technique qui fournit une transformation linéaire sur la plage originale des données.

Elle maintient la relation entre les données originales et peut spécifiquement adapter les données dans une limite prédéfinie avec une limite prédéfinie. La normalisation Z- score donne les valeurs normalisées ou la plage de données à partir des données non structurées originales en utilisant des concepts comme la moyenne et l'écart type. La mise à l'échelle décimale fournit la plage entre -1 et 1. Ces techniques de normalisation permettent de transformer les données pour qu'elles se situent dans une plage spécifique, ce qui peut être utile pour de nombreux algorithmes d'apprentissage automatique.

2.9 Considérations pour l'Interprétation

Il est important de noter que si ces transformations facilitent l'analyse des séries temporelles, elles modifient également les propriétés de la série, et ces modifications doivent être prises en compte lors de l'interprétation des résultats. Par exemple, une prévision basée sur une série temporelle qui a subi une transformation logarithmique devra être retransformée (à l'aide d'une fonction exponentielle) pour fournir une prévision sur l'échelle originale.

Enfin, tout comme avec d'autres formes de transformation de données, il est crucial de se rappeler que les résultats obtenus à partir de données normalisées doivent être ramenés à leur échelle originale pour une interprétation précise.

2.10 Différenciation des séries temporelles

La différenciation est une technique couramment utilisée dans l'analyse des séries temporelles pour rendre une série temporelle stationnaire (c'est-à-dire rendre sa moyenne et sa variance constantes dans le temps). En pratique, cela implique généralement de prendre la différence entre les observations consécutives. La différenciation peut aider à stabiliser la moyenne d'une série temporelle en supprimant les changements de niveau et en réduisant la tendance. Une série temporelle qui a été rendue stationnaire par la différenciation peut être plus facile à modéliser et à prévoir.[15]

3. Concepts fondamentaux sur le machine learning

3.1 Type de machine learning

3.1.1 Apprentissage supervisé (Supervised Learning)

L'apprentissage supervisé est une méthode d'apprentissage machine qui nécessite un grand nombre de données étiquetées pour l'entraînement. Ces données étiquetées sont souvent coûteuses et prennent du temps à produire. Dans ce contexte, l'apprentissage auto-supervisé est apparu comme un paradigme prometteur pour surmonter le goulot d'étranglement de l'annotation en générant automatiquement des exemples d'entraînement bruyants à partir de données non étiquetées. En particulier, l'auto-supervision spécifique à une tâche convertit les connaissances préalables en modèles d'auto-supervision pour la génération d'étiquettes, comme dans la supervision à distance, la programmation de données, et l'inférence conjointe. [18]

3.1.2 Apprentissage non supervisé (Unsupervised Learning)

L'apprentissage non supervisé est une méthode d'apprentissage machine où les exemples d'entraînement ne sont pas étiquetés. L'apprentissage contrasté a montré des performances exceptionnelles à la fois en apprentissage supervisé et non supervisé. Il apprend de bonnes représentations d'observations de haute dimension à partir d'une grande quantité de données non étiquetées, en rapprochant une ancre et ses vues augmentées dans l'espace d'incorporation.[19]

Apprentissage par renforcement (Reinforcement learning)

L'apprentissage par renforcement est un processus d'apprentissage où un agent (ou plusieurs agents) interagit avec son environnement pour atteindre un comportement optimal. Dans le cas de l'apprentissage par renforcement à agent unique, un seul agent intelligent est impliqué dans le processus d'apprentissage. Il apprend en interagissant avec son environnement afin d'atteindre l'objectif d'un comportement individuel optimal.[20]

3.1.3 Apprentissage semi-supervisé (Semi-Supervised Learning)

L'apprentissage semi-supervisé est une méthode d'apprentissage machine qui se situe entre l'apprentissage supervisé et non supervisé. Dans l'apprentissage semi-supervisé, l'ensemble d'apprentissage est composé d'un petit nombre d'échantillons d'apprentissage étiquetés et d'un grand nombre d'échantillons non étiquetés. L'objectif est d'utiliser les informations fournies par les échantillons non étiquetés pour améliorer la performance de l'apprentissage à partir des échantillons étiquetés.[21]

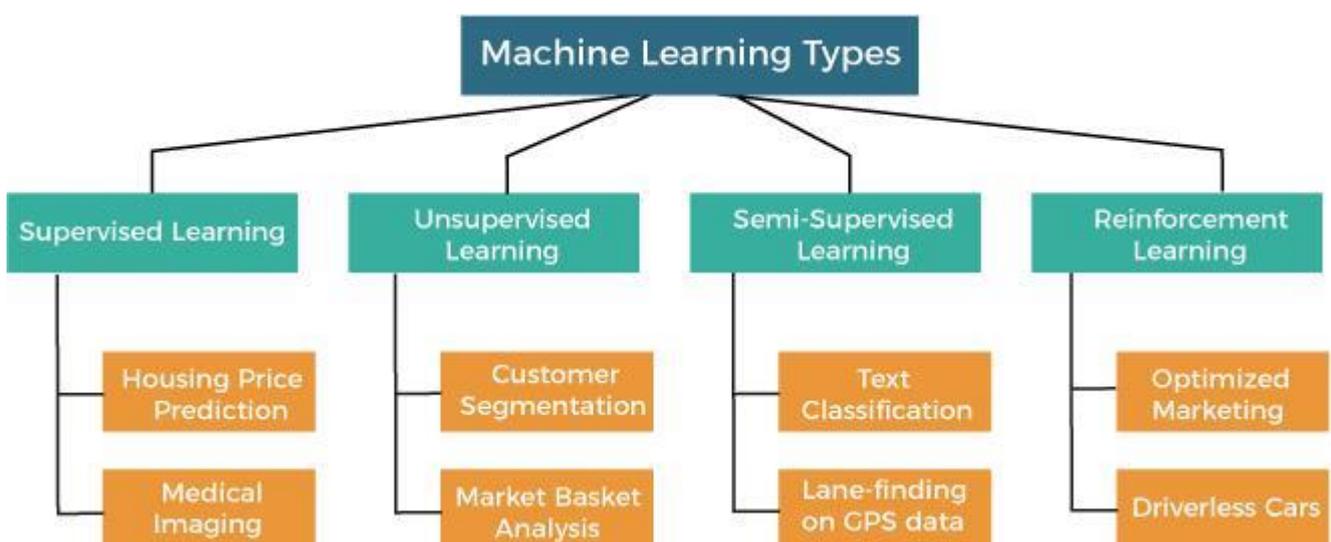


Figure 6 Types d'apprentissage automatique [4]

3.2 Concepts basiques du machine learning

3.2.1 Données (data)

Les informations brutes qui sont utilisées pour entraîner un modèle d'apprentissage automatique. Elles peuvent être structurées (par exemple, dans un format tabulaire avec des lignes et des colonnes) ou non structurées (par exemple, du texte, des images, du son).

3.2.2 Caractéristiques (features)

Aussi appelées variables ou attributs, les caractéristiques sont les caractéristiques ou propriétés mesurables des données qui sont utilisées comme entrées pour un modèle d'apprentissage automatique.

3.2.3 Etiquettes (labels)

En apprentissage supervisé, les étiquettes sont les sorties connues ou les variables cibles associées aux données d'entrée. Le modèle apprend à partir des paires d'entrée- sortie pendant l'entraînement pour faire des prédictions sur des données non vues.

3.2.4 Données d'entraînement (training data)

Le jeu de données étiqueté utilisé pour entraîner un modèle d'apprentissage automatique. Il est composé de données d'entrée avec les étiquettes correspondantes ou les valeurs cibles.

3.2.5 Données de test (test data)

Des données non vues utilisées pour évaluer les performances d'un modèle d'apprentissage automatique entraîné. Les prédictions du modèle sur les données de test sont comparées aux étiquettes réelles pour évaluer sa précision et sa capacité de généralisation.

3.2.6 Modèle (model)

Une représentation mathématique ou informatique des relations entre les caractéristiques d'entrée et les étiquettes de sortie. Il est entraîné sur les données d'entraînement pour faire des prédictions sur de nouvelles données non vues.

3.2.7 Entraînement (training)

Le processus d'ajustement d'un modèle d'apprentissage automatique aux données d'entraînement en modifiant ses paramètres internes ou ses poids pour minimiser la différence entre les étiquettes prédites et réelles.

3.2.8 Prédiction (prediction)

La sortie ou l'inférence faite par un modèle d'apprentissage automatique entraîné lorsqu'il est fourni avec de nouvelles données non vues. Le modèle généralise à partir des données d'entraînement pour faire des prédictions sur des instances similaires, mais auparavant non vues.

3.2.9 Fonction de perte (loss function)

Une fonction de perte est une mesure de la performance d'un modèle de prédiction.

Dans l'apprentissage supervisé, la fonction de perte est utilisée pour estimer l'écart entre les prédictions du modèle et les valeurs réelles. Les fonctions de perte peuvent être différentes en fonction du type de problème d'apprentissage (par exemple, classification, régression, etc.) et de la nature des données. Par exemple, la fonction de perte d'entropie croisée est couramment utilisée pour les problèmes de classification, tandis que la fonction de perte quadratique est couramment utilisée pour les problèmes de régression.[22]

3.2.10 Algorithme d'optimisation (optimization algorithm)

Les algorithmes d'optimisation sont des outils essentiels pour résoudre de nombreux problèmes en apprentissage automatique. Ces algorithmes visent à trouver la meilleure solution possible (optimum) pour un problème donné. Dans le contexte de l'apprentissage automatique, les algorithmes d'optimisation sont souvent utilisés pour minimiser une fonction de perte, c'est-à-dire pour trouver les paramètres du modèle qui donnent la meilleure performance possible sur les données d'apprentissage. [23]

3.2.11 Métriques d'évaluation (evaluation metrics)

Des mesures utilisées pour évaluer les performances d'un modèle d'apprentissage automatique. Les exemples incluent l'exactitude, la précision, le rappel, le score F1 et l'erreur quadratique moyenne (MSE).

3.2.12 Généralisation (generalization)

La généralisation est un concept clé en apprentissage automatique. Il se réfère à la capacité d'un modèle d'apprentissage automatique à bien fonctionner sur de nouvelles données inédites qui n'ont pas été utilisées pendant la phase d'entraînement. En d'autres termes, un modèle qui généralise bien est capable de faire des prédictions précises non seulement sur les données d'entraînement, mais aussi sur les données qu'il n'a jamais vues auparavant.[24]

3.3 Concepts avancés du machine learning

3.3.1 Trade-off bias-variance

Le trade-off bias-variance est un concept fondamental en apprentissage automatique qui fait référence à la relation entre le biais et la variance des prédictions d'un modèle. Le biais mesure l'erreur systématique d'un modèle, c'est-à-dire à quel point les prédictions du modèle sont éloignées des valeurs réelles en moyenne. La variance, quant à elle, mesure la variabilité des prédictions du modèle pour différentes instances d'entraînement.[25]

Le trade-off entre le biais et la variance se résume au fait que les modèles avec un faible biais ont tendance à avoir une haute variance, et vice versa. Un modèle avec un biais élevé est souvent trop simplifié et ne capture pas pleinement la complexité des données, ce qui conduit à des prédictions inexactes. En revanche, un modèle avec une variance élevée est trop sensible aux variations des données d'entraînement et peut être sur ajusté, ce qui entraîne une mauvaise performance sur de nouvelles données.

3.3.2 Techniques de régularisation

Les techniques de régularisation sont des méthodes utilisées pour prévenir le surajustement (overfitting) dans un modèle. Le surajustement se produit lorsque le modèle devient trop complexe et capture le bruit ou des motifs non pertinents dans les données d'entraînement, ce qui entraîne de mauvaises performances sur de nouvelles données. La régularisation ajoute un terme de pénalité à la fonction d'objectif du modèle, encourageant des solutions plus simples et plus généralisables.

Les techniques de régularisation courantes comprennent la régression Ridge (régularisation L2), la régression Lasso (régularisation L1) et la régression Elastic Net (combinaison des régularisations L1 et L2). La régression Ridge réduit les coefficients du modèle, réduisant ainsi sa complexité. La régression Lasso

favorise la parcimonie en encourageant certains coefficients à devenir exactement nuls, ce qui permet une sélection de caractéristiques. La régression Elastique Net combine les deux types de régularisation.

3.3.3 Techniques de sélection e caractéristiques (features selection techniques)

Les techniques de sélection de caractéristiques visent à identifier le sous-ensemble le plus pertinent de caractéristiques à partir d'un ensemble plus large. Cela permet d'améliorer les performances du modèle, de réduire le surajustement et d'améliorer l'interprétabilité.

Les méthodes de filtrage évaluent la pertinence des caractéristiques indépendamment du modèle en utilisant des mesures statistiques telles que la corrélation ou l'information mutuelle. Les méthodes d'enveloppement évaluent les sous-ensembles de caractéristiques en entraînant et testant des modèles avec différentes combinaisons. Les méthodes intégrées intègrent la sélection de caractéristiques dans le processus d'entraînement du modèle.

3.3.4 Apprentissage profond (Deep Learning)

Le Deep Learning est une branche de l'apprentissage automatique qui se concentre sur l'entraînement de réseaux neuronaux avec plusieurs couches pour apprendre des représentations des données. Les réseaux neuronaux profonds apprennent automatiquement des caractéristiques hiérarchiques, ce qui leur permet de capturer des motifs et des dépendances complexes.

Les architectures de deep learning comprennent les réseaux neuronaux convolutifs (CNN) pour l'analyse d'images et les réseaux neuronaux récurrents (RNN) pour les données séquentielles. Ces réseaux apprennent des représentations hiérarchiques en extrayant des caractéristiques de plus haut niveau à partir des données brutes, ce qui permet d'obtenir des performances de pointe dans divers domaines.

4. Modèles de prédiction

4.1 Modèles de prédiction pour les séries temporelles

4.1.1 Modèle autorégressif (AR)

Le modèle autorégressif suppose que la valeur future d'une série temporelle

Chapitre 2. Etat de l'art

dépend linéairement de ses valeurs passées. La prédiction est basée sur une régression linéaire des observations précédentes. La formule générale d'un modèle AR d'ordre p est la suivante :

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t$$

- X_t : La valeur de la série temporelle à l'instant t .
- c : Une constante.
- φ_i : Le coefficient associé à la valeur de la série temporelle à l'instant $t-i$.
- X_{t-i} : La valeur de la série temporelle à l'instant $t-i$.
- ε_t : Un terme d'erreur aléatoire.

4.1.2 Modèle à moyenne mobile (MA)

Le modèle à moyenne mobile suppose que la valeur future d'une série temporelle est une combinaison linéaire des erreurs de prédiction passées. La formule générale d'un modèle MA d'ordre q est la suivante :

$$X_t = \mu + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

- X_t : La valeur de la série temporelle à l'instant t .
- μ : La moyenne de la série temporelle.
- θ_i : Le coefficient associé à l'erreur de prédiction à l'instant $t-i$.
- ε_{t-i} : L'erreur de prédiction à l'instant $t-i$.
- ε_t : Un terme d'erreur aléatoire.

4.1.3 Modèle autorégressif à moyenne mobile (ARMA)

Le modèle ARMA combine les caractéristiques des modèles AR et MA. Il utilise à la fois les valeurs passées et les erreurs de prédiction passées pour prédire la valeur future. La formule générale d'un modèle ARMA est la suivante :

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

- X_t : La valeur de la série temporelle à l'instant t .
- c : Une constante.
- φ_i : Le coefficient associé à la valeur de la série temporelle à l'instant $t-i$.
- X_{t-i} : La valeur de la série temporelle à l'instant $t-i$.
- ε_t : Un terme d'erreur aléatoire.

4.1.4 Modèle autorégressif intégré à la moyenne mobile (ARIMA)

Le modèle ARIMA est une extension du modèle ARMA qui prend également en compte les différences entre les observations successives. Il est utile pour les séries temporelles non stationnaires, c'est-à-dire les séries dont les propriétés statistiques varient avec le temps. La formule générale d'un modèle ARIMA est la suivante :

- $$\nabla^d X_t = c + \sum_{i=1}^p \phi_i \nabla^d X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$
- $\nabla^d X_t$: La série temporelle différenciée d'ordre d à l'instant t.
 - ϕ_i : Le coefficient associé à la série temporelle différenciée à l'instant t-i.
 - θ_i : Le coefficient associé à l'erreur de prédiction à l'instant t-i
 - ε_{t-i} : L'erreur de prédiction à l'instant t-i.
 - ε_t : Un terme d'erreur aléatoire.

4.2 Modèles de prédiction pour le machine learning

4.2.1 Régression linéaire

La régression linéaire est l'une des méthodes les plus simples et les plus anciennes en statistiques et en apprentissage automatique. Elle est utilisée lorsque la variable cible (dans ce cas, l'achat) a une relation linéaire avec les variables prédictives. Dans la régression linéaire, le but est de trouver la meilleure ligne (ou hyperplan dans un espace multidimensionnel) qui minimise la somme des carrés des résidus, c'est-à-dire la distance entre les points de données réels et les points prédits par le modèle. La régression linéaire est facile à comprendre et à interpréter, et elle peut servir de base pour des méthodes plus avancées.

En régression linéaire, les entrées font référence aux variables indépendantes utilisées pour prédire une variable de sortie continue. Les entrées peuvent représenter des caractéristiques, des mesures ou des observations qui sont supposées avoir une relation linéaire avec la variable de sortie. La variable de sortie, quant à elle, est la valeur que l'on cherche à prédire à l'aide du modèle de régression linéaire. L'objectif est d'établir la meilleure relation linéaire entre les entrées et la sortie afin de pouvoir faire des prédictions précises pour de nouvelles observations basées sur les entrées fournies.

4.2.2 Régression par forêts aléatoires

La régression par forêts aléatoires est une méthode qui utilise un ensemble d'arbres de décision pour effectuer une prédiction. Chaque arbre de décision est formé sur un sous-ensemble des données d'entraînement choisies au hasard, et la prédiction finale est la moyenne des données d'entraînement choisies au hasard, et la prédiction finale est la moyenne des prédictions de tous les arbres. Cette méthode est utile pour éviter le surapprentissage et pour gérer les relations non linéaires. De plus, elle peut gérer les valeurs manquantes et les variables catégorielles sans avoir besoin de prétraitement. La régression par forêts aléatoires peut être un peu plus difficile à interpréter qu'un simple arbre de décision, mais elle offre généralement de meilleures performances.

Les entrées font référence aux variables indépendantes utilisées pour prédire une variable de sortie continue. Ces entrées peuvent être des caractéristiques, des mesures ou des variables observées qui servent de base à la construction de l'ensemble d'arbres aléatoires. La variable de sortie, quant à elle, est la valeur continue que l'on souhaite prédire à l'aide du modèle de régression par forêts aléatoires. L'objectif de ce modèle est de combiner les prédictions de multiples arbres de décision pour obtenir une prédiction plus robuste et précise. En utilisant les entrées fournies, le modèle de régression par forêts aléatoires est capable de prédire la variable de sortie pour de nouvelles observations en exploitant les relations complexes et non linéaires présentes dans les données.

4.2.3 Régression par Vecteurs de Support

La Régression par Vecteurs de Support (Support Vector Regression, SVR) est une version de la Machine à Vecteurs de Support (Support Vector Machine, SVM) qui est utilisée pour la régression. Elle utilise les mêmes principes que la SVM pour la classification, mais avec quelques modifications pour permettre des prédictions continues au lieu de classifications discrètes.

Dans le cas de la régression par vecteurs de support (SVM pour Support Vector Regression), les entrées se réfèrent aux variables indépendantes utilisées pour prédire une variable de sortie continue. Ces entrées peuvent être des caractéristiques, des mesures ou des observations qui sont utilisées pour construire le modèle de régression par vecteurs de support. La variable de sortie, quant à elle, est la valeur continue que l'on souhaite prédire à l'aide du modèle. L'objectif de la régression par vecteurs de support est de trouver une fonction d'ajustement optimale qui minimise l'erreur de prédiction en maximisant la marge entre les points de données et la droite (ou l'hyperplan) de régression. En utilisant les entrées fournies, le modèle de régression par vecteurs de support est

en mesure d'estimer avec précision la variable de sortie pour de nouvelles observations en se basant sur la position relative des points de données par rapport à l'hyperplan de régression.

4.2.4 Amplification Extrême du Gradient (XGBOOST)

XGBoost, ou Extreme Gradient Boosting, est un algorithme d'apprentissage automatique basé sur des arbres de décision. Il est conçu pour être hautement efficace, flexible et portable. XGBoost utilise le gradient boosting (optimisation par descente de gradient) pour minimiser l'erreur de prédiction en ajoutant de manière itérative de nouveaux modèles. [26]

Dans le cas de l'Amplification Extrême du Gradient (XGBoost), les entrées désignent les variables indépendantes utilisées pour prédire une variable de sortie continue. Ces entrées peuvent être des caractéristiques, des mesures ou des observations qui sont utilisées pour construire le modèle XGBoost. La variable de sortie, quant à elle, est la valeur continue que l'on souhaite prédire à l'aide du modèle. L'objectif de XGBoost est de créer un ensemble de modèles d'arbres de décision, appelé "boosting", où chaque modèle apprend à corriger les erreurs du modèle précédent. En utilisant les entrées fournies, le modèle XGBoost est capable de prédire avec précision la variable de sortie pour de nouvelles observations en utilisant la combinaison des prédictions de tous les modèles d'arbres dans l'ensemble. Le processus d'amplification extrême du gradient permet d'obtenir des performances de prédiction supérieures grâce à des techniques sophistiquées telles que la régularisation, la gestion des gradients et l'utilisation de valeurs manquantes.

4.2.5 Modèles basés sur les réseaux de neurones

Les réseaux de neurones, tels que les réseaux de neurones récurrents (RNN) et les réseaux de neurones à mémoire à court terme (LSTM), sont capables de capturer les motifs complexes et les dépendances à long terme dans les séries temporelles. Les formules spécifiques dépendent de l'architecture du réseau de neurones utilisé.

En ce qui concerne les modèles basés sur les réseaux de neurones, les entrées font référence aux variables indépendantes utilisées pour prédire une variable de sortie continue. Ces entrées peuvent être des caractéristiques, des mesures ou des observations qui servent de données d'entrée au réseau de neurones. La variable de sortie, quant à elle, est la valeur continue que l'on cherche à prédire à l'aide du modèle basé sur les réseaux de neurones. L'objectif de ces modèles est d'apprendre des relations complexes et non linéaires entre les entrées et la sortie

en utilisant des architectures de réseau de neurones avec des couches intermédiaires. Les modèles basés sur les réseaux de neurones utilisent des fonctions d'activation, des poids et des biais pour calculer des prédictions en passant les données à travers le réseau de neurones. En utilisant les entrées fournies, le modèle basé sur les réseaux de neurones est capable de prédire avec précision la variable de sortie pour de nouvelles observations en exploitant les capacités de modélisation puissantes et flexibles des réseaux de neurones.

5. Evaluation des modèles

5.1 Modèles de régression

Les mesures d'évaluation les plus courantes pour les modèles de régression sont :

5.1.1 Erreur quadratique moyenne (MSE)

C'est la moyenne des carrés des erreurs. L'erreur est la différence entre la valeur prédite et la valeur réelle.

Tapez une équation ici.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

- Y_i : La valeur réel a l'instant i
- \hat{Y}_i : La valeur prédite a l'instant i

5.1.2 Erreur absolue moyenne (MAE)

C'est la moyenne des valeurs absolues des erreurs.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

- Y_i : La valeur réel a l'instant i
- \hat{Y}_i : La valeur prédite a l'instant i

5.1.3 Coefficient de détermination (R^2)

Il indique la proportion de la variance de la variable dépendante qui est prévisible à partir des variables indépendantes. Un R^2 de 1 indique que les

prédictions correspondent parfaitement aux valeurs réelles.

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

- Y_i : La valeur réel a l'instant i
- \hat{Y}_i : La valeur prédite a l'instant i

5.2 Modèles de séries temporelles

Les mesures d'évaluation courantes pour les modèles de séries temporelles sont similaires à celles des modèles de régression. Cependant, il y a quelques considérations supplémentaires à prendre en compte :

5.2.1 Erreur quadratique moyenne (RMSE)

C'est une mesure couramment utilisée dans les séries temporelles qui donne plus de poids aux grandes erreurs.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$$

- Y_i : La valeur réel a l'instant i
- \hat{Y}_i : La valeur prédite a l'instant i

5.2.2 Erreur absolue en pourcentage (MAPE)

C'est la moyenne des erreurs absolues en pourcentage. Elle peut être utile pour comparer les erreurs.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100$$

- Y_i : La valeur réel a l'instant i
- \hat{Y}_i : La valeur prédite a l'instant i

6. CRISP-DM

La méthodologie CRISP-DM, pour Cross Industry Standard Process for Data Mining, est une approche structurée pour résoudre les problèmes dans le domaine de l'analyse, de l'extraction et des sciences des données. Créée par un consortium formé des entreprises NCR, SPSS et Daimler-Benz, elle est largement reconnue et utilisée dans les projets d'analyse de données et de data science. Cette méthode a été conçue pour être itérative et agile, chaque itération apportant une connaissance supplémentaire qui permet d'améliorer les suivantes.

La méthode CRISP-DM se décompose en six phases principales :

6.1 Compréhension du problème métier

Cette phase consiste à comprendre en profondeur les éléments métiers et problématiques que la Data Science vise à résoudre ou à améliorer. Compréhension des données

Cette phase vise à déterminer précisément les données à analyser, à identifier la qualité des données disponibles et à faire le lien entre les données et leur signification d'un point de vue métier.

6.3 Préparation des données

Cette phase, aussi appelée "construction du Data Hub", regroupe les activités liées à la construction de l'ensemble précis des données à analyser à partir des données brutes. Cela inclut le classement des données en fonction de critères choisis, le nettoyage des données, et surtout leur recodage pour les rendre compatibles avec les algorithmes qui seront utilisés.

6.4 Modélisation des données

Cette phase de Data Science proprement dite comprend le choix, le paramétrage et le test de différents algorithmes ainsi que leur enchaînement, qui constitue un modèle.

6.5 Évaluation

Cette phase vise à vérifier les modèles ou les connaissances obtenues afin de

s'assurer qu'ils répondent aux objectifs formulés au début du processus.

6.6 Déploiement

Il s'agit de l'étape finale du processus. Elle consiste en une mise en production pour les utilisateurs finaux des modèles obtenus.

Voici la figure suivante qui illustre cette méthodologie.

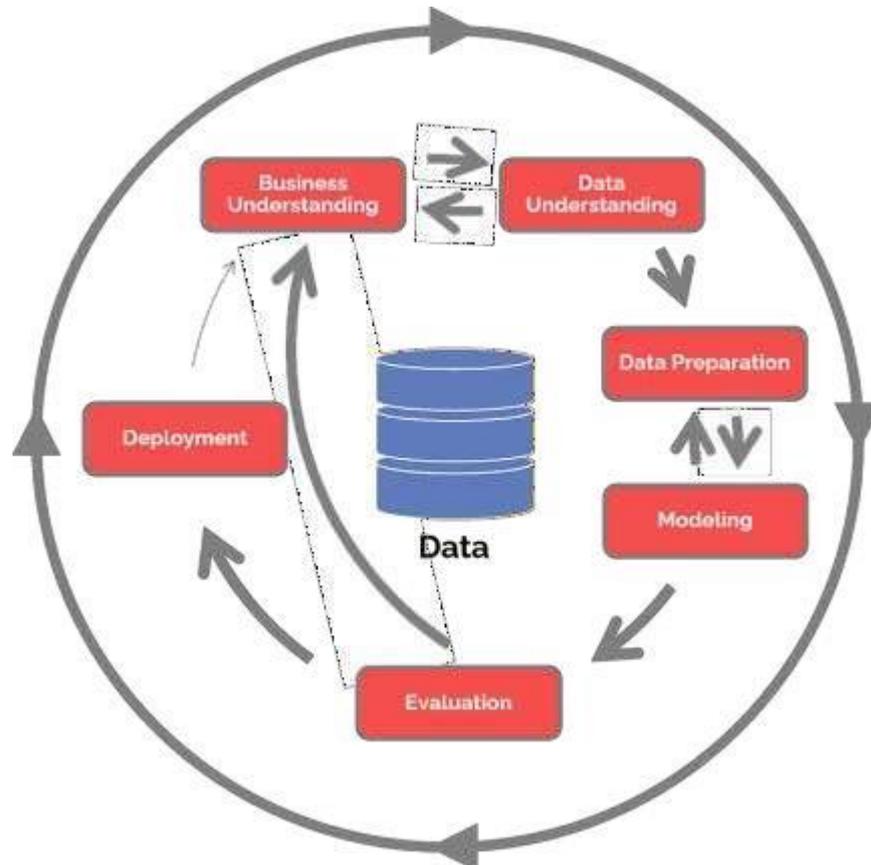


Figure 7 Les étapes de la méthodologie CRISP-DM [5]

Conclusion

Dans ce deuxième chapitre, nous avons exploré en profondeur l'état de l'art des séries temporelles et du machine learning. Nous avons d'abord défini les termes clés et introduit les concepts fondamentaux des séries temporelles, incluant différents types de séries temporelles, leurs composantes, le concept de stationnarité, le retard, l'autocorrélation, et diverses transformations et différenciations.

Ensuite, nous nous sommes penchés sur les concepts essentiels du machine

Chapitre 2. Etat de l'art

learning, y compris ses différents types, les concepts basiques et avancés de cette discipline. L'objectif était de construire une compréhension solide des principes clés qui sous-tendent les techniques de machine learning. Dans la troisième partie du chapitre, nous avons examiné les modèles de prédiction, en particulier pour les séries temporelles et le machine learning. Nous avons présenté les différences et les particularités des modèles de prédiction dans ces deux domaines.

Enfin, nous avons discuté de l'évaluation des modèles, en couvrant les modèles de régression et ceux des séries temporelles. Cet examen vise à aider à comprendre comment déterminer l'efficacité d'un modèle dans la prédiction des données futures.

En somme, ce chapitre a fourni une base solide pour comprendre les aspects théoriques des séries temporelles et du machine learning, qui seront essentiels pour les chapitres suivants où nous appliquerons ces concepts à des problèmes pratiques.

Chapitre 3
Apports
Et
Solutions

Chapitre 3 Apports et solutions

Introduction

Dans ce chapitre intitulé "Solution", nous allons détailler l'approche que nous avons adoptée pour résoudre le problème de la prévision d'achat de pièces de rechange chez Biopharm. Cette partie est l'aboutissement de tout le travail que nous avons accompli jusqu'à présent et représente le cœur de notre projet.

Tout d'abord, nous décrivons la méthodologie CRISP-DM (Cross-Industry Standard Process for Data Mining) que nous avons utilisée, une approche reconnue dans le domaine du data mining. Nous examinerons les étapes que nous avons suivies et comment nous les avons adaptées à notre contexte spécifique.

Ensuite, nous aborderons l'analyse des séries temporelles et le forecasting que nous avons réalisés. Cette analyse a été essentielle pour comprendre les tendances et les modèles de l'achat de pièces de rechange, et nous montrerons comment nous l'avons utilisée pour faire des prévisions précises.

Nous présenterons également l'interface que nous avons développée pour faciliter l'utilisation de nos prévisions. Cette interface permet à l'utilisateur de choisir une pièce de rechange et de déterminer le nombre de mois pour lequel une prévision est nécessaire, fournissant ensuite les résultats de différents modèles.

Nous espérons que ce chapitre fournira une compréhension claire de la manière dont nous avons abordé et résolu le problème de la prévision d'achat de pièces de rechange chez Biopharm.

1. Compréhension du problème métier

Dans le secteur pharmaceutique, la gestion efficace de la chaîne d'approvisionnement est un élément crucial pour garantir la disponibilité de médicaments de qualité et pour soutenir les opérations de l'entreprise. Cela est particulièrement vrai pour une entreprise comme Biopharm, qui opère dans un environnement où la fiabilité, l'efficacité et la conformité aux réglementations sont essentielles. L'un des aspects importants de la chaîne d'approvisionnement concerne l'achat des pièces de rechange (PDR) qui sont nécessaires pour maintenir l'équipement en bon état de fonctionnement.

La première étape dans l'élaboration d'un modèle de prévision d'achat de pièces de rechange est de bien comprendre les objectifs et les contraintes spécifiques du métier. Cela signifie de saisir non seulement les aspects techniques de la gestion des pièces de rechange mais aussi les enjeux stratégiques, opérationnels

et financiers qui y sont associés.

Comprendre le métier est essentiel car cela permet de :

- Identifier clairement le problème que nous cherchons à résoudre.
- Assurer que la solution développée est alignée avec les objectifs de l'entreprise.
- Établir une communication efficace avec les parties prenantes pour obtenir le soutien nécessaire.
- Définir des critères de succès pertinents et mesurables.

Ce processus de compréhension du métier est d'autant plus important dans le contexte de Biopharm, où l'achat de PDR a été traditionnellement guidé par l'expérience des utilisateurs plutôt que par des méthodes analytiques. Il existe un potentiel significatif d'amélioration grâce à l'application de techniques de science des données, mais pour que ces efforts soient fructueux, ils doivent être fermement ancrés dans une compréhension approfondie de l'environnement opérationnel de Biopharm et des défis auxquels l'entreprise est confrontée auxquels sont déjà définis au préalable dans les sections précédentes.

2. Compréhension des données

2.1 Données fournis

Les données analysées proviennent du système ERP (Enterprise Resource Planning) de Biopharm, qui joue un rôle central dans la coordination des ressources, de l'information et des processus au sein de l'entreprise.

Dans ce contexte, nous nous concentrerons sur trois tables de données spécifiques extraites du système ERP, qui contiennent des informations pertinentes sur les niveaux de stocks, les mouvements de stocks et les seuils d'achat. Ces tables sont intitulées "STOCK 2023", "STOJOU" et "Seuils".

Table STOCK 2023

La table "STOCK2023" contient des données relatives au stock d'articles en 2023. Pour cette analyse, deux colonnes spécifiques ont été extraites :

- **ITMREF_o** : Il s'agit du code de référence de l'article. Ce code est unique pour chaque article et sert à l'identifier de manière univoque.
- **QTYSTU_o** : Cette colonne indique la quantité d'articles en stock, exprimée

en unités. Elle permet d'avoir une idée de la quantité d'articles disponibles à un Moment donné.

Table STOJOU

La table "STOJOU" contient des données journalières concernant les mouvements de stock. Trois colonnes ont été extraites pour cette analyse:

- **IPTDAT_o** : Cette colonne stocke la date de l'entrée en stock. Elle permet de suivre la chronologie des mouvements de stock.
- **ITMREF_o** : Tout comme dans la table "STOCK2023", cette colonne contient le code de référence de l'article, qui est unique pour chaque article.
- **QTYSTU_o** : Cette colonne indique également la quantité d'articles en unités, mais contrairement à la table "STOCK2023", elle donne une image plus dynamique de la façon dont les stocks évoluent au jour le jour

Table Seuils

La table "**Seuils**" contient des informations sur les seuils de stock minimum et maximum qui sont utilisés comme critères pour la prise de décisions concernant la quantité d'articles à acheter lors d'une commande. Les colonnes pertinentes sont :

- **Seuil Minimum** : Cette colonne indique la quantité minimale d'un article qui peut être achetée lors d'une commande. Cela signifie que si le seuil minimum est de 1, par exemple, le nombre d'articles commandés ne doit pas être inférieur à ce seuil.
- **Seuil Maximum** : Cette colonne indique la quantité maximale d'un article qui peut être achetée lors d'une commande. Par exemple, si le seuil maximum est de 5, cela signifie que le nombre d'articles commandés ne doit pas dépasser ce seuil.

Utilisation des seuils pour la prise de décisions d'achat :

C'est ces min et max qui sont utilisés actuellement pour l'achat de pièces de rechange pour le prochain achat.

Les seuils minimum et maximum sont utilisés pour déterminer la plage de quantités d'articles qui peuvent être achetés lors d'une commande. Par exemple, si le seuil minimum est de 1 et le seuil maximum est de 5, cela signifie que lors de la passation d'une commande, la quantité d'articles à acheter doit être comprise entre 1 et 5 inclus. Ces seuils aident à contrôler les dépenses en évitant des achats excessifs tout en garantissant qu'une quantité minimale adéquate est achetée.

Cela devrait mieux refléter la façon dont les seuils sont utilisés pour guider la prise de décisions en ce qui concerne la quantité d'articles à acheter lors de chaque commande.

2.2 La Sélection des données nécessaires

Après avoir effectué des joints sur ces tables sur seulement les données nécessaires (tout est inclus dans le notebook), nous avons obtenu les données préliminaires suivantes, avec les colonnes : ID article, stock actuel, date, quantité mouvementé (combien est entré ou sortie d'article avec un certain "ID article" dans une date donnée), minimum théorique et maximum théorique.

Voici une figure illustrant les données initiales :



	ID article	stock_actuel	date	quantité mvt	min th	max th
43534	GLGE2087	2	01/01/19	-1	1	2
11692	SDELC455	1	01/01/19	-3	2	5
32174	SASC1008	29	01/01/19	-1	8	16
22846	SDELC390	45	01/01/19	100	25	100
37620	CDBL4020	2	01/01/19	-1	1	2
...
55174	SPRE1023	16	31/12/22	-1	15	40
24948	GLGE2426	6	31/12/22	3	0	0
43258	CDEN381	2	31/12/22	1	1	2
12887	SDELC346	22	31/12/22	-4	4	12
27795	SASC2084	2	31/12/22	-2	0	1

Figure 8 Récapitulatif des données

3. Préparation des données

3.1 Valeurs nulles

Avant d'appliquer un algorithme d'apprentissage automatique sur un ensemble de données, il est crucial d'identifier les valeurs manquantes. Sinon, cela peut entraîner des prédictions erronées. Il existe différentes méthodes pour traiter les valeurs manquantes ou nulles d'une variable. On peut les remplacer par une valeur fixe, comme la moyenne (ce qu'on appelle l'imputation de données), ou bien supprimer les observations contenant ces valeurs manquantes.

Dans notre cas, nous avons la chance de ne pas avoir de valeurs manquantes dans notre jeu de données. Cela signifie que toutes les observations sont complètes et ne comportent aucun élément manquant. Cette situation est extrêmement favorable, car elle nous dispense de devoir effectuer des opérations de traitement supplémentaires pour gérer ces valeurs manquantes. Nous pouvons donc directement procéder à l'application de notre algorithme de machine learning sur l'ensemble des données disponibles. Cette absence de valeurs manquantes simplifie grandement le processus d'analyse et de prédiction, nous permettant de tirer pleinement parti des informations contenues dans notre jeu de données sans aucun biais ou perte potentielle. Ainsi, nous pouvons avoir une confiance accrue dans les résultats obtenus à partir de notre algorithme de machine learning, sachant que toutes les observations sont prises en compte de manière exhaustive.

```
def resume_table(data):
    #print(f'Data Shape: {data.shape}')
    summary = (pd
                .DataFrame(data.dtypes, columns=['dtypes'])
                .reset_index()
                .rename(columns={'index': 'name'}))
    summary.loc[:, 'missing'] = data.isna().sum().values
    summary.loc[:, '%missing'] = np.round(summary.missing / data.shape[0] * 100, 2)
    summary.loc[:, 'uniques'] = data.nunique().values
    return summary, data.shape
```

Figure 9 Résumé des caractéristiques des données (Fonction `resume_table`)

```

✓ [9] 1 resume_table(final_df)
0s
(
      name  dtypes  missing  %missing  uniques
0  ID article  object        0        0.0    4628
1 stock_actuel  int64        0        0.0     86
2      date  object        0        0.0   1371
3 quantité mvt  int64        0        0.0    283
4      min th  int64        0        0.0     26
5      max th  int64        0        0.0    41,

```

Figure 10 Résultats de la fonction (*resume_table*) sur notre jeu de données

3.2 Séparation des données

Nous avons effectué une séparation minutieuse des données en deux tables distinctes afin de mieux traiter les informations constantes et dynamiques. Dans la première table, appelée "Table 1", nous avons regroupé les valeurs constantes des articles, telles que l'ID de l'article, le stock actuel, le seuil minimal (*min_th*) et le seuil maximal (*max_th*). Ces valeurs restent constantes dans le temps et représentent les caractéristiques fondamentales des articles, indépendamment des fluctuations de stock.

La deuxième table, nommée Table 2", est conçue pour capturer les séries chronologiques (time series) des mouvements de stock. Elle comprend des informations dynamiques telles que l'ID de l'article, la date et la quantité en stock à une date donnée. Grâce à cette table, nous pouvons analyser en détail l'évolution des stocks au fil du temps, en mettant l'accent sur les tendances, les fluctuations saisonnières et les schémas récurrents.

La distinction claire entre la table statique (Table 1) et la table dynamique (Table 2) permet d'approfondir l'analyse des données. La première fournit une vue instantanée des caractéristiques fixes des articles, telles que le stock actuel et les seuils minimaux et maximaux recommandés. Cela aide à évaluer rapidement la situation actuelle de chaque article par rapport à ses paramètres préétablis.

En revanche, la table dynamique (Table 2) offre une vision temporelle, nous permettant de suivre les changements de stock au fil du temps et d'observer des tendances et des comportements spécifiques. Grâce à cette approche, nous pouvons détecter des schémas saisonniers, identifier les moments où les niveaux

de stock sont les plus bas ou les plus élevés, et prendre des décisions informées en matière de réapprovisionnement, de gestion des stocks et de prévision des demandes futures.

En résumé, cette séparation précise en deux tables distinctes nous permet d'analyser à la fois les aspects constants et dynamiques de la gestion des stocks. Cela nous donne une meilleure compréhension des caractéristiques fondamentales des articles, tout en nous permettant d'explorer les variations de stock au fil du temps pour une prise de décision éclairée et une optimisation de la gestion des stocks.

Voici ci-joint les 2 tables en figures (exemple de l'art avec ID "CDBL4005")

Données statiques

	ID article	date	quantity_in_stock
0	CDBL4005	2018-01-31	2
1	CDBL4005	2018-02-28	0
2	CDBL4005	2018-03-31	0
3	CDBL4005	2018-04-30	0
4	CDBL4005	2018-05-31	0

Figure 11 Nos données statiques

Données dynamiques

	ID article	stock_actu	min_th	max_th
0	CDBL4005	2	0	2
1	CDBL4007	2	0	0
2	CDBL4008	7	4	8
3	CDBL4009	11	1	2
4	CDBL4010	3	0	0

Figure 12 Nos données dynamiques

3.3 Éliminer certaines valeurs

L'utilisation d'une méthode de prévision, telle qu'un modèle de prévision, sur une série temporelle qui ne possède pas suffisamment de données peut être problématique. En effet, le manque de données peut limiter la capacité du modèle à identifier des motifs et des tendances significatifs dans la série. Dans cette introduction, nous examinerons pourquoi il est difficile d'utiliser des prévisions sur une série temporelle avec un nombre limité de points de données, et les conséquences potentielles de cette limitation.

Dans notre situation particulière, nous faisons face à un défi concernant plusieurs identifiants d'articles qui ne disposent pas d'un nombre suffisant de données. Par conséquent, afin de garantir des résultats de prévision robustes et fiables, nous avons pris la décision d'éliminer les articles qui présentent moins de quatre points de données dans leur série temporelle. Cette approche nous permet de travailler avec des ensembles de données plus complets et de maximiser notre capacité à identifier des schémas et des modèles significatifs pour effectuer des prévisions précises. En éliminant ces articles à faible disponibilité de données, nous nous assurons de disposer d'un ensemble de données adéquat pour mener des analyses et des prévisions plus solides.

3.4 Catégorisation de la demande

La précision des prévisions de la demande d'un produit dépend grandement de la prévisibilité de ce produit. La catégorisation de la demande des produits permet de les classer en fonction de leur degré de prévisibilité en utilisant certains

critères. Cela permet aux décideurs de choisir la meilleure méthode de prévision tout en étant conscients de la fiabilité des résultats.

Pour catégoriser la demande des différents produits chez Biopharm, nous avons calculé les deux critères suivants :

1. **Interval moyen de demande (ADI)** Mesure la régularité de la demande dans le temps en calculant l'intervalle moyen entre deux demandes.
2. **Carré du coefficient de variation (CV²)** Mesure la variation des quantités.

En se basant sur ces deux mesures, la littérature classe les profils de demande en quatre catégories distinctes :

1. **Demande régulière (Smooth) (ADI < 1,32 et CV² < 0,49)** La demande est très régulière dans le temps et en quantité. Il est donc facile à prévoir et il y a peu d'erreurs de prévision.
2. **Demande intermittente (Intermittent) (ADI ≥ 1,32 et CV² < 0,49)** L'historique de la demande montre peu de variation dans les quantités demandées, mais une forte variation dans les intervalles entre deux demandes. Bien qu'il existe des méthodes de prévision spécifiques pour ce type de demande intermittente, la marge d'erreur de prévision est considérablement plus élevée.
3. **Demande erratique (Erratic) (ADI < 1,32 et CV² ≥ 0,49)** La demande présente des occurrences régulières dans le temps avec des variations importantes des quantités. La précision des prévisions reste fragile.
4. **Demande instable (Lumpy) (ADI ≥ 1,32 et CV² ≥ 0,49)** La demande se caractérise par une grande variation tant au niveau des quantités que des intervalles temporels. Il est en fait impossible de produire des prévisions fiables, quelles que soient les méthodes de prévision utilisées. Ce type particulier de modèle de demande est imprévisible.

Voici une figure qui illustre cela :

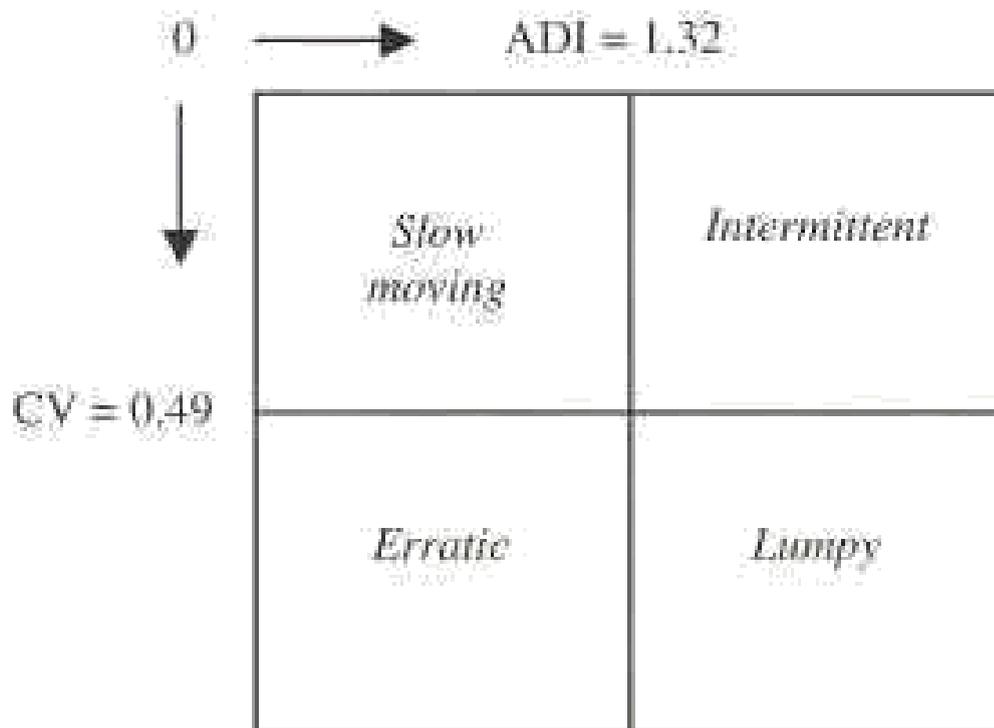


Figure 13 Les 4 types de demandes

Nous avons créé un dataframe contenant ces informations, et nous avons constaté que la quasi-totalité des demandes étaient instables (lumpy demand) (voire la totalité, car une seule demande n'est pas irrégulière).

La figure ci-dessous illustre cela

```

1 demand_type_df.head(5)

```

	article_id	demand_type
0	CDBL4005	Lumpy Demand
1	CDBL4007	Lumpy Demand
2	CDBL4008	Lumpy Demand
3	CDBL4009	Lumpy Demand
4	CDBL4010	Lumpy Demand

```

1 demand_type_df['demand_type'].value_counts()

```

```

Lumpy Demand      4002
Smooth Demand      1
Name: demand_type, dtype: int64

```

Figure 14 Répartition des types de demandes sur notre jeu de données

3.5 Visualisation temporelle des séries chronologiques pour chaque article

En traçant les séries chronologiques, nous pourrions observer les variations et les évolutions au fil du temps, ce qui nous permettra d'identifier les modèles récurrents, les tendances significatives et les éventuels événements marquants. Cette approche nous offre une perspective dynamique et détaillée sur les données, nous permettant de mieux comprendre les phénomènes sous-jacents et d'extraire des informations précieuses.

Cette visualisation temporelle des séries chronologiques pour chaque article nous permettra de mettre en évidence les relations, les corrélations ou les divergences entre les différentes variables étudiées. Nous pourrions ainsi explorer les interactions complexes entre les articles et identifier d'éventuelles influences mutuelles ou dépendances temporelles.

Time Series - ALL ID Articles

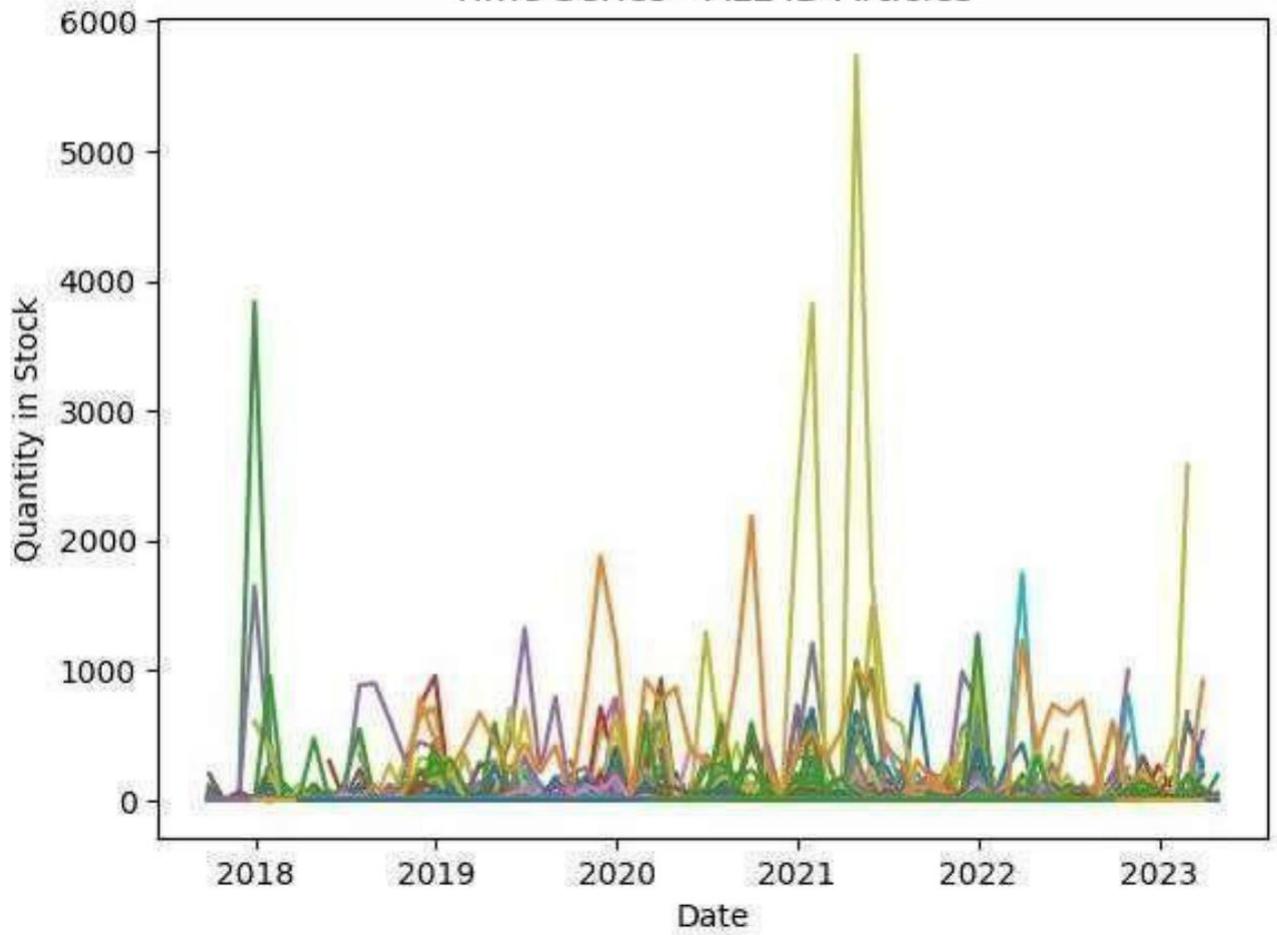


Figure 15 Visualisation des séries temporelles de notre jeu de données pour chaque article

Après avoir examiné la figure, il est clair que chaque article présente un ensemble unique de séries chronologiques, reflétant sa propre dynamique temporelle. Les couleurs distinctes pour chaque article nous aident à différencier les variations et les tendances de chacun. On remarque des modèles récurrents, des similitudes et des différences notables qui témoignent de la diversité des comportements au fil du temps.

Les variations colorées au sein de chaque série chronologique reflètent également la manière dont différentes variables peuvent interagir ou influencer un article particulier. Par exemple, des pics de couleur peuvent indiquer des événements clés qui ont eu un impact significatif sur l'article. En outre, nous pourrions également détecter des modèles de synchronicité ou de correspondance entre les séries chronologiques de différents articles, indiquant des relations ou des dépendances temporelles entre eux.

Il convient maintenant de passer à une analyse plus détaillée de ces séries chronologiques, en se concentrant sur l'identification de ces interactions complexes et en essayant de comprendre les facteurs qui peuvent les expliquer.

3.6 Complétion des données

Notre objectif est de prédire les 24 prochains mois dans notre étude. Cependant, nous avons remarqué dans les graphiques que certaines séries chronologiques pour certains articles ne vont pas jusqu'à la date de juin. Par conséquent, nous devons effectuer une imputation de données jusqu'à la date actuelle (juin 2023).

Pour cela, nous avons créé un dataframe (DF) contenant, pour chaque article, la dernière date disponible avant juin 2023.

Voici un exemple des 5 premiers articles :



```
1 last_dates.head(5)
```

	ID article	last_date
0	CDBL4005	2022-03-31
1	CDBL4007	2021-12-31
2	CDBL4008	2022-05-31
3	CDBL4009	2022-05-31
4	CDBL4010	2021-12-31

Figure 16 Dernières dates pour chaque article

En d'autres termes, pour prendre un exemple avec la PDR "**CDBL4005**", les données disponibles remontent jusqu'au 31 mars 2022. Par conséquent, il est nécessaire de compléter les données jusqu'au 30 juin 2023.

Pour compléter les données manquantes dans les séries temporelles des stocks de produits, on utilise la méthode de la moyenne récurrente. Elle calcule la moyenne des quantités en stock jusqu'à une date donnée, puis utilise cette moyenne comme estimation pour les dates futures. Ainsi, les données manquantes sont remplacées de manière cohérente en tenant compte de l'évolution des quantités au fil du temps.

4. Modélisation des données

Il existe un grand nombre de méthodes et modèles de prévision et comme il n'y a pas un standard qui permet d'attribuer un modèle spécifique pour un problème donné, nous avons opté pour une présélection d'un ensemble de méthodes candidates que nous avons soigneusement choisies en fonction des résultats de multiples études présentes dans la littérature et qui ont été appliquées sûrement des données séquentielles similaires à nos données.

Dans cette étude, nous avons exploré plusieurs modèles de prévision, couvrant une variété de techniques, allant des méthodes statistiques traditionnelles aux modèles d'apprentissage automatique et d'apprentissage en profondeur. Parmi les modèles statistiques, nous avons utilisé les méthodes de CROSTON et Holt- Winters pour analyser les séries temporelles et prévoir les tendances futures. Dans le domaine de l'apprentissage automatique, nous avons exploité les avantages de XGBoost et RandomForest, des algorithmes populaires basés sur les arbres de décision, pour capturer les relations complexes entre les caractéristiques et les variables cibles. Enfin, nous avons également utilisé un modèle d'apprentissage en profondeur, à savoir LSTM, pour modéliser les séquences et prendre en compte les dépendances à long terme. Chacun de ces modèles possède des hyperparamètres spécifiques qui peuvent être ajustés pour améliorer les performances prédictives. Dans le tableau ci-dessous, nous récapitulons les différents modèles et leurs hyperparamètres respectifs, offrant ainsi une vue d'ensemble des techniques utilisées dans notre étude.

Tableau 2 Récapitulatif des modèles utilisés

Modèle	Type	Hyperparamètres
CROSTON	Statistique	-None
Holt-Winters	Statistique	-None
XGBoost	Machine Learning	- max_depth- learning_rate- n_estimators - subsample

Modèle	Type	Hyperparamètres
RandomForest	Machine Learning	- n_estimators- max_depth- min_samples_split - max_features
LSTM	Deep Learning	- hidden_size- num_layers- learning_rate- dropout

5. Évaluation et résultats

Dans cette section, nous évaluons et analysons les performances de différents modèles de prédiction appliqués à notre jeu de données. Les modèles considérés dans cette analyse comprennent LSTM, Holt-Winters, Croston, SVR (Support Vector Regression), RF (Random Forest) et XGBoost. Nous utilisons l'erreur quadratique moyenne (RMSE) comme métrique d'évaluation, qui est couramment utilisée pour mesurer l'exactitude des valeurs prédites par rapport aux valeurs réelles.

5.1 Méthodologie

Pour chaque modèle, nous avons formé les algorithmes en utilisant des ensembles de données d'entraînement, puis validé les performances sur des ensembles de données de test. L'erreur quadratique moyenne (RMSE) a été calculée pour chaque modèle.

5.2 Résultats

Les figures qui suivent contiennent une série de résultats provenant de nos prédictions, ainsi qu'une analyse détaillée des erreurs qui ont été relevées. Ces résultats sont présentés dans le but de fournir une vision complète et transparente de notre processus de prédiction, mettant en évidence à la fois les succès et les limitations de notre modèle. En examinant attentivement ces figures, vous pourrez vous faire une idée précise des performances de notre système et des domaines dans lesquels des améliorations peuvent être apportées.

Ainsi que Les modèles sont classés en ordre croissant en fonction de leur RMSE, c'est-à-dire que le modèle avec la plus faible RMSE est classé en premier, tandis que celui avec la plus élevée est classé en dernier. Voici les rangs obtenus :

1. LSTM
2. Holt-Winters
3. Croston
4. SVR (Support Vector Regression)
5. RF (Random Forest)
6. XGBoost

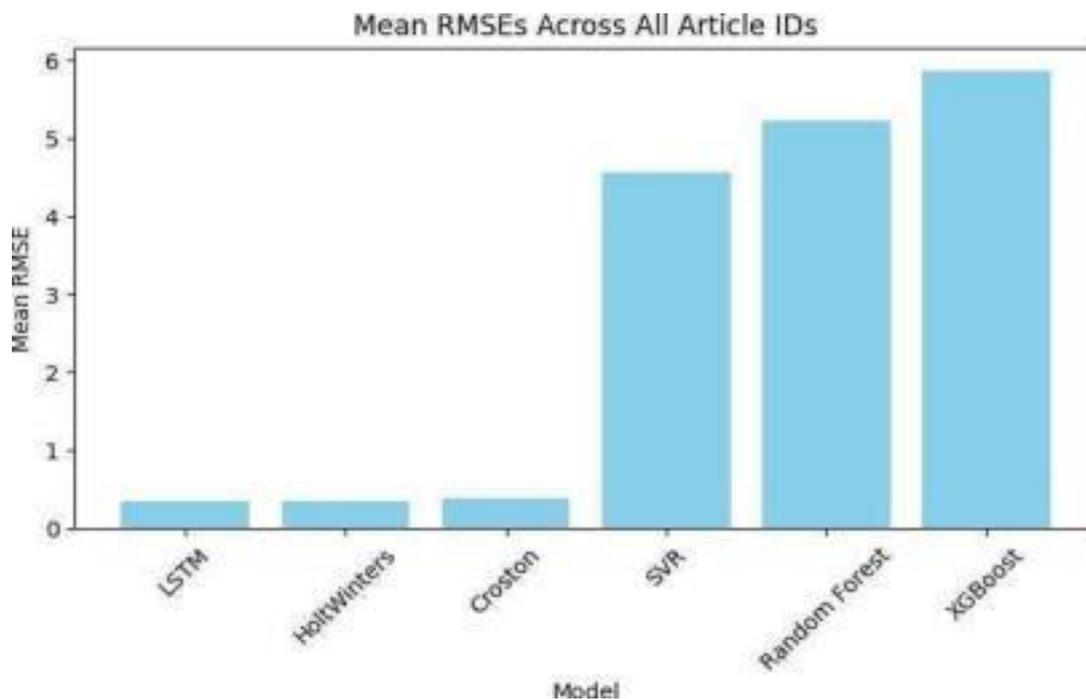


Figure 19 Résultats des modèles sur notre jeu de données

5.3 Analyse des Résultats

- LSTM (Long Short-Term Memory)

Le modèle LSTM s'est avéré être le plus performant parmi tous les modèles testés. Cela peut être attribué à la capacité des LSTM de capturer des dépendances à long terme et des motifs complexes dans des séries temporelles, ce qui est particulièrement utile pour des données de demande de type lumpy.

- Holt-Winters

Ce modèle a obtenu la deuxième meilleure performance. Holt-Winters est une technique de lissage exponentiel qui est efficace pour modéliser des séries temporelles avec une tendance et une saisonnalité, ce qui suggère que ces composants peuvent être présents dans les données.

- **Croston**

Le modèle de Croston est spécifiquement conçu pour la prévision de séries temporelles intermittentes, il se concentre donc sur la prédiction des niveaux de demande non nuls. Il est souvent utilisé pour les demandes de type lumpy et a obtenu de bons résultats dans notre cas, mais il n'a pas surpassé LSTM ou Holt-Winters.

- **SVR (Support Vector Regression)**

SVR est une forme de SVM utilisée pour la régression. Sa performance modérée suggère qu'il pourrait ne pas être le meilleur choix pour ce type de données, mais il est tout de même capable de capturer certains motifs.

- **RF (Random Forest)**

Random Forest est un algorithme d'apprentissage ensembliste. Il semble moins efficace dans ce contexte, ce qui pourrait être dû à la nature des données de demande de type lumpy.

- **XGBoost**

En dépit d'un algorithme puissant, XGBoost a eu les performances les moins bonnes. Cela pourrait suggérer que les données ne sont pas bien adaptées à des modèles de boosting d'arbres de décision.

En se basant sur l'erreur quadratique moyenne, le modèle LSTM a montré les meilleures performances pour la prédiction de données de demande de type lumpy, suivi de Holt-Winters et Croston. Il est important de considérer le contexte et la nature des données lors du choix d'un modèle de prédiction. Dans ce cas, des modèles capables de capturer des dépendances à long terme et des motifs complexes dans les séries temporelles se sont avérés plus efficaces.

6. Déploiement

6.1. Exploration de l'Interface Utilisateur : Présentation du GUI (Graphical User Interface) développé

L'interface **SPB Decision Helper (Spare Parts Buyer Decision Helper)** a été nommée ainsi pour refléter son rôle dans l'aide à la prise de décision pour les acheteurs de pièces de rechange. Cette interface a été conçue pour faciliter le processus de sélection en fournissant des informations claires et pertinentes.

Pour commencer, les données ont été divisées en deux catégories distinctes : les données statiques et les données dynamiques. Les données statiques incluent le stock actuel, le seuil minimal et le seuil maximal.

Chapitre 03 : Apports et solutions

Le stock actuel fait référence à la quantité actuelle de chaque pièce de rechange, identifiée par un ID spécifique. Le seuil minimal indique la quantité minimale souhaitée pour chaque pièce, afin de ne pas tomber en dessous de ce seuil. De même, le seuil maximal indique la quantité maximale souhaitée pour éviter les excès.

Ces informations s'afficheront automatiquement dès que l'utilisateur choisit l'ID de l'article en question. Il est ainsi possible de visualiser rapidement l'état du stock et de prendre des décisions éclairées.

En ce qui concerne les données dynamiques, nous avons développé des modèles basés sur des séries chronologiques pour chaque pièce (ID de l'article). Ces modèles permettent de prédire les quantités nécessaires sur différentes périodes, allant de 1 à 24 mois.

Une fois que l'utilisateur a sélectionné l'article, il doit ensuite choisir la période souhaitée en mois. Les modèles génèrent ensuite des prédictions basées sur cette période spécifique.

Le meilleur modèle est ensuite sélectionné en fonction de la prédiction de la quantité à acheter. Cette sélection est basée sur des critères définis pour garantir l'exactitude et la fiabilité des prédictions.

Enfin, l'interface fournit également des suggestions de fournisseurs, indiquant où acheter la pièce de rechange recommandée avec la quantité déterminée précédemment. Cette fonctionnalité permet d'optimiser le processus d'achat en proposant des options appropriées.

En résumé, l'interface SPB Decision Helper offre une approche structurée et logique pour aider les acheteurs de pièces de rechange. Elle fournit des informations sur l'état actuel du stock, des prédictions basées sur des modèles de séries chronologiques et des suggestions de fournisseurs, permettant ainsi de prendre des décisions d'achat éclairées et efficaces.

Voici ci-dessous les figures illustrant toutes ces informations de manière claire et visuelle.

Chapitre 03 : Apports et solutions



Figure 20 Page 1 de l'interface (SBP decision helper)

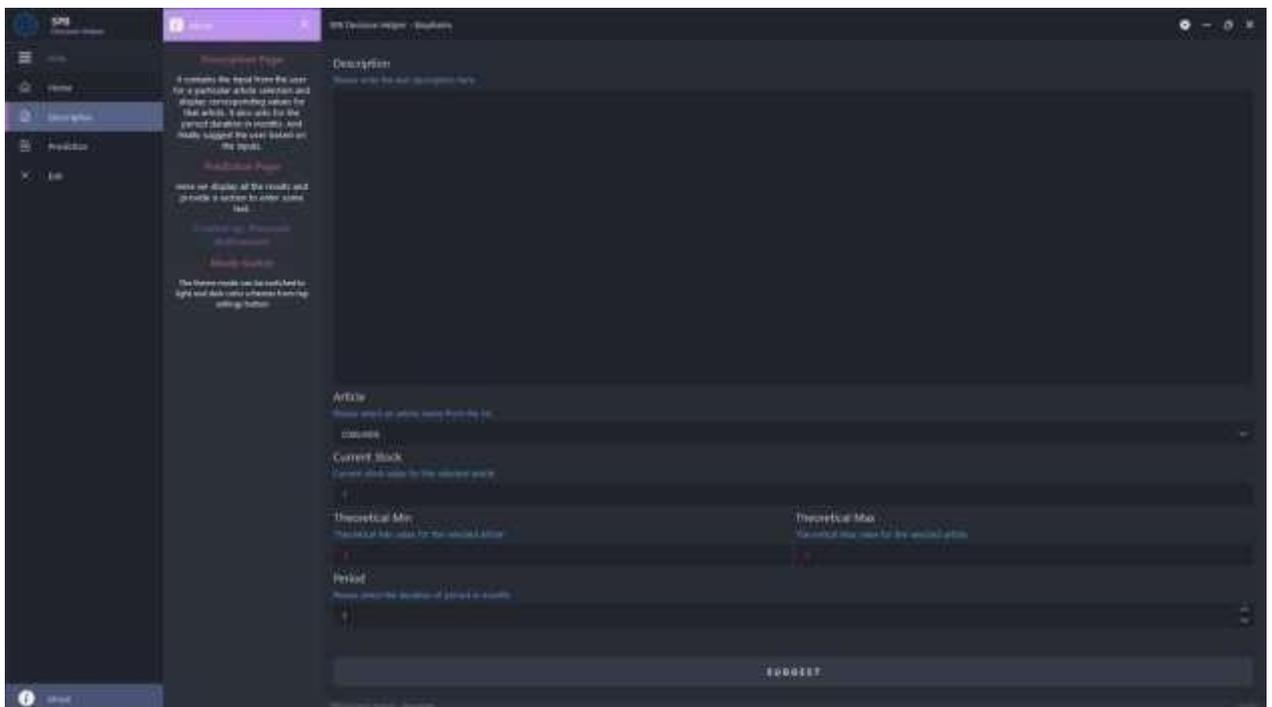


Figure 21 Page 2 de l'interface (SBP decision helper)

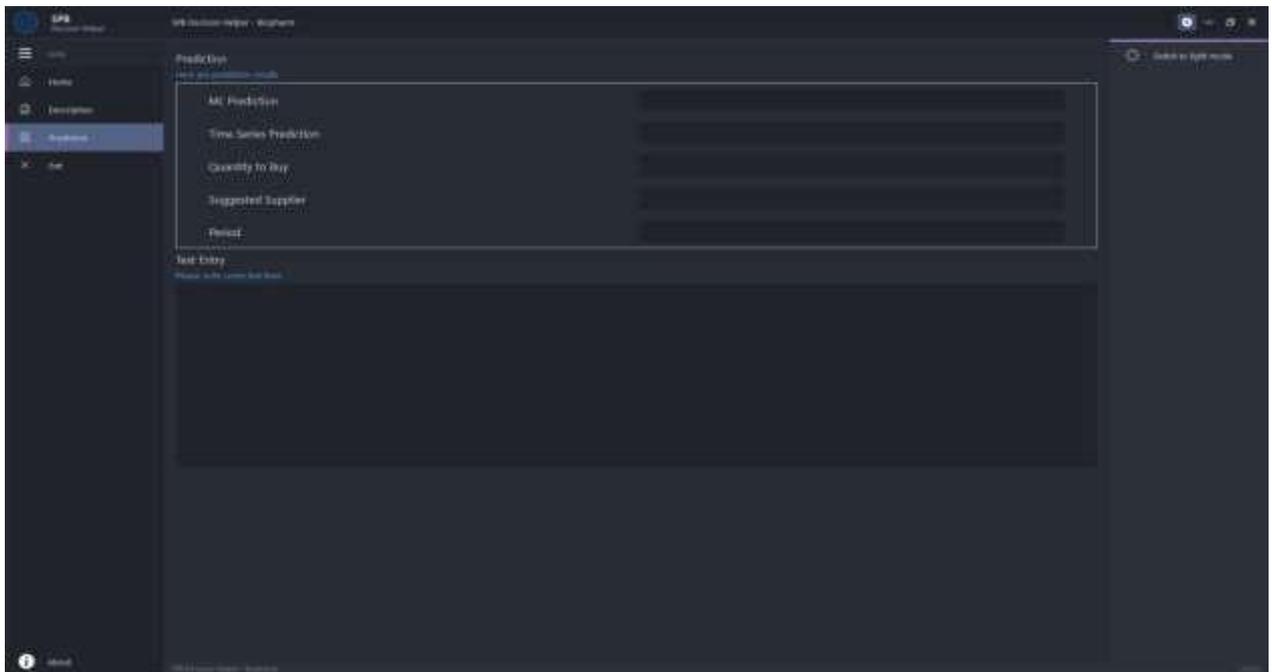


Figure 22 Page 3 de l' interface (SBP decision helper)

6.2. Intégration Harmonieuse des Modèles de Prédiction au sein de l'Interface

Les modèles développés comprennent une approche statistique (Croston-Holtwinters), des algorithmes d'apprentissage automatique (XGBoost et Random Forest), ainsi qu'un modèle de réseau de neurones récurrents (LSTM) basé sur l'apprentissage profond.

Pour intégrer ces modèles à l'interface utilisateur, on peut suivre les approches suivantes : Chargement des modèles

On peut sauvegarder les modèles entraînés dans des fichiers spécifiques, tels que des fichiers pickle ou des fichiers de poids selon le modèle. Dans l'interface utilisateur, on peut ajouter une fonctionnalité permettant de charger ces modèles pré-entraînés à partir de leurs fichiers respectifs. Une fois chargés, on peut utiliser ces modèles pour effectuer des prédictions en fonction des entrées fournies par l'utilisateur.

1. Intégration du code des modèles

On peut copier et coller le code des modèles directement dans l'application GUI. Afin d'assurer une meilleure encapsulation et une plus grande modularité,

on peut créer des fonctions dédiées encapsulant la logique des modèles pour la prédiction. Ces fonctions peuvent être appelées depuis l'interface utilisateur lorsque l'utilisateur souhaite obtenir des prédictions.

1. Utilisation de modules externes

Pour faciliter le processus de prédiction, on peut utiliser des bibliothèques Python spécialisées dans la prédiction des séries temporelles. Par exemple, on peut intégrer la bibliothèque statsmodels pour les modèles statistiques, ainsi que des bibliothèques spécifiques pour les modèles d'apprentissage automatique et de deep learning. Ces bibliothèques peuvent être intégrées dans l'application GUI, permettant ainsi d'utiliser leurs fonctions et méthodes pour effectuer les prédictions.

2. Utilisation de services web

On peut également explorer l'option de déployer les modèles de prévision en tant que services web indépendants. Pour cela, on peut utiliser des frameworks tels que Flask ou Django pour créer des API pour chaque modèle. Ces services web permettent d'envoyer les données nécessaires et de récupérer les prédictions correspondantes pour les afficher dans l'interface utilisateur.

3. Utilisation de bibliothèques de machine learning intégrées à PySide 6

PySide 6 offre une intégration avec certaines bibliothèques de machine learning, comme scikit-learn. On peut utiliser ces bibliothèques pour entraîner et charger les modèles de prévision. En intégrant directement les modèles entraînés dans l'interface utilisateur en utilisant les fonctionnalités fournies par PySide 6 pour l'apprentissage automatique, on peut offrir une expérience fluide aux utilisateurs.

4. Utilisation des résultats des modèles comme données

Une fois les prédictions effectuées par les modèles, on peut utiliser ces résultats comme données pour d'autres fonctionnalités de l'interface utilisateur. Par exemple, on peut les afficher dans des graphiques ou les comparer avec d'autres données pertinentes pour faciliter l'analyse et la prise de décision.

Dans notre projet, nous avons utilisé la sixième option, qui consiste à utiliser les résultats des modèles de prévision comme données pour notre interface utilisateur. Étant donné que chaque article possède sa propre série temporelle, nous avons pu mettre en place un mécanisme permettant de sélectionner l'identifiant de l'article et la période de prédiction souhaitée. Cela nous a permis d'obtenir des données exhaustives de toutes les prédictions effectuées.

Lorsque l'utilisateur sélectionne un article spécifique et indique la période de prédiction souhaitée, notre interface utilisateur utilise les modèles de prévision

des séries temporelles correspondantes pour générer les prédictions associées. Ces

prédictions sont ensuite traitées comme des données et peuvent être utilisées de différentes manières.

En résumé, en utilisant l'option "**Utilisation des résultats des modèles comme données**", nous avons pu fournir à notre interface utilisateur des données exhaustives de toutes les prédictions effectuées pour chaque article, en utilisant l'identifiant de l'article et la période de prédiction comme paramètres. Cela permet aux utilisateurs d'analyser les prévisions, de les comparer avec d'autres données pertinentes et de prendre des décisions basées sur ces informations.

7. Perspectives et améliorations

En plus des aspects mentionnés précédemment, il y a plusieurs perspectives d'amélioration pour votre interface GUI de prévision d'achat de pièces de rechange. Voici quelques idées pour enrichir votre application :

7.1. Ajout des fournisseurs

Vous pouvez envisager d'intégrer des informations sur les fournisseurs dans votre interface. Cela permettrait aux utilisateurs de sélectionner un fournisseur spécifique pour chaque article et de prendre en compte les délais de livraison dans leurs prévisions. En tenant compte des fournisseurs, vous pouvez aider les utilisateurs à déterminer le moment optimal pour passer une commande afin d'éviter les ruptures de stock.

7.2. Prédiction des niveaux de stock

En plus de prédire le nombre de pièces de rechange à acheter, vous pouvez également étendre votre modèle pour prédire les niveaux de stock futurs. Cela permettrait de visualiser graphiquement l'évolution des niveaux de stock dans le temps et d'identifier les périodes où des ajustements sont nécessaires pour éviter les pénuries ou les excès de stock.

7.3. Analyse des ruptures de stock

Vous pouvez inclure une fonctionnalité d'analyse des ruptures de stock dans votre interface. Cela pourrait aider les utilisateurs à identifier les raisons des ruptures de stock passées, telles que des variations soudaines de la demande, des problèmes d'approvisionnement ou des retards de livraison. L'analyse des ruptures de stock peut fournir des informations précieuses pour prendre des mesures préventives et optimiser la gestion des stocks.

7.4. Déclenchement automatique des commandes

Une amélioration supplémentaire serait d'incorporer une fonctionnalité de déclenchement automatique des commandes. Sur la base des prévisions de demande et des niveaux de stock actuels, vous pourriez mettre en place un système qui déclenche automatiquement les commandes de réapprovisionnement lorsque les seuils de réapprovisionnement sont atteints. Cela réduirait la charge de travail manuelle et aiderait à maintenir les niveaux de stock optimaux.

En intégrant ces améliorations, votre interface GUI de prévision d'achat de pièces de rechange deviendrait plus complète et offrirait une gamme plus large d'outils et de fonctionnalités pour aider les utilisateurs à gérer efficacement leurs stocks et leurs approvisionnements.

Conclusion

Dans le chapitre 3, "Apports et solutions", l'accent a été mis sur l'utilisation de la méthodologie CRISP-DM pour aborder le projet de prévision d'achat de pièces de rechange en utilisant la science des données et l'intelligence artificielle. Après avoir examiné la compréhension du problème, la préparation des données, et développé des modèles statistiques, de machine learning et de deep learning, le chapitre a traité de la phase de déploiement, comprenant la mise en place d'une interface utilisateur et l'intégration des modèles de prévision au sein de cette interface.

En outre, le chapitre s'est terminé par une discussion sur les perspectives et les améliorations futures, abordant des aspects tels que l'ajout des fournisseurs, la prédiction des niveaux de stock, l'analyse des ruptures de stock, et le déclenchement automatique des commandes. Ces perspectives soulignent les avenues possibles pour améliorer davantage le système de prévision et l'adapter aux besoins changeants et aux défis futurs dans le domaine de la gestion des pièces de rechange.

Conclusion Générale

La gestion efficace de l'achat de pièces de rechange est cruciale dans tous les secteurs, y compris dans l'industrie pharmaceutique. Ce mémoire a abordé le défi de la prévision d'achat de pièces de rechange en utilisant la science des données et l'intelligence artificielle (DSIA) au sein de l'entreprise BIOPHARM INDUSTRY en Algérie.

Le Chapitre 1 a établi le contexte en présentant l'industrie pharmaceutique, aussi bien au niveau mondial qu'en Algérie, et en décrivant en détail l'entreprise BIOPHARM INDUSTRY. Ce chapitre a également défini le périmètre du projet, en identifiant les départements concernés, et a réalisé un diagnostic externe et interne afin de cerner la problématique de la gestion des pièces de rechange au sein de l'entreprise. Le Chapitre 2, l'état de l'art, a fourni une base théorique en expliquant des termes clés et des concepts essentiels liés aux séries temporelles, au machine learning et à l'évaluation des modèles. Cela a permis d'établir un cadre de référence solide pour comprendre et appliquer les méthodologies et les techniques nécessaires dans le cadre de la prévision d'achat.

Le Chapitre 3, "Apports et solutions", est le cœur de ce mémoire. En utilisant la méthodologie CRISP-DM, il a guidé le lecteur à travers les étapes clés du processus de développement de modèles de prévision. Après une compréhension approfondie du problème métier, une attention particulière a été accordée à la préparation des données. Trois types de modèles de prévision - statistiques, de machine learning et de deep learning - ont été développés et évalués. Le chapitre a également traité du déploiement des modèles à travers une interface utilisateur bien conçue, rendant les prévisions accessibles aux parties prenantes. Finalement, il a abordé les perspectives et améliorations futures, notamment en matière d'intégration des fournisseurs, de prédiction des niveaux de stock, d'analyse des ruptures de stock et de déclenchement automatique des commandes.

En somme, ce mémoire démontre l'importance et la faisabilité d'utiliser des approches avancées en science des données et en intelligence artificielle pour optimiser le processus d'achat de pièces de rechange. En adaptant ces techniques au contexte spécifique de l'industrie pharmaceutique et de l'entreprise BIOPHARM INDUSTRY, il est possible d'améliorer l'efficacité opérationnelle, de réduire les coûts et d'augmenter la satisfaction des parties prenantes. L'adoption et l'intégration de telles solutions, ainsi que l'exploration de pistes d'amélioration future, marquent un pas en avant dans l'adoption de l'innovation pour répondre aux défis opérationnels de l'industrie pharmaceutique.

Bibliographie

Bibliographie

- [1] S. Makridakis, E. Spiliotis, et V. Assimakopoulos, « Statistical and Machine Learning forecasting methods: Concerns and ways forward », *PLOS ONE*, vol. 13, n° 3, p. e0194889, mars2018, doi: 10.1371/journal.pone.0194889.
- [2] L. H. Meyer, « Issues Concerning the Definition, Measurement and Forecasting of Productive Capacity ». <https://research.stlouisfed.org/publications/review/1977/12/01/issues-concerning-the-definition-measurement-and-forecasting-of-productive-capacity> (consulté le 28juin 2023).
- [3] G. Shmueli, « To Explain or to Predict? », *Stat. Sci.*, vol. 25, n° 3, p. 289-310, août 2010, doi:10.1214/10-STS330.
- [4] M.-C. Chen, A.-L. Chiu, et H.-H. Chang, « Mining changes in customer behavior in retailmarketing », *Expert Syst. Appl.*, vol. 28, n° 4, p. 773-781, mai 2005, doi: 10.1016/j.eswa.2004.12.033.
- [5] E. F. Codd, « A relational model of data for large shared data banks », *Commun. ACM*, vol.13, n° 6, p. 377-387, 1970.
- [6] G. E. Box, G. M. Jenkins, G. C. Reinsel, et G. M. Ljung, *Time series analysis: forecastingand control*. John Wiley & Sons, 2015.
- [7] S. Moein, « Definition of Artificial Neural Network », in *Artificial Intelligence: Concepts,Methodologies, Tools, and Applications*, IGI Global, 2017, p. 1-11.
- [8] G. Rebala, A. Ravi, S. Churiwala, G. Rebala, A. Ravi, et S. Churiwala, « Machine learningdefinition and basics », *Introd. Mach. Learn.*, p. 1-17, 2019.
- [9] J. S. Racine, « Univariate Linear Time Series Models », in *Reproducible EconometricsUsing R*, J. S. Racine, Éd., Oxford University Press, 2019, p. 0. doi: 10.1093/oso/9780190900663.003.0003.
- [10] X. Liu, S. Swift, A. Tucker, G. Cheng, et G. Loizou, « Modelling multivariate time series », *Intell. Data Anal. Med. Pharmacol.*, p. 67-78, 1999.
- [11] J. D. Hamilton, *Time Series Analysis*. Princeton University Press, 1994. doi:10.2307/j.ctv14jx6sm.
- [12] N. Uras, L. Marchesi, M. Marchesi, et R. Tonelli, « Forecasting Bitcoin closing price seriesusing linear regression and neural networks models », *PeerJ Comput. Sci.*, vol. 6, p. e279, 2020.
- [13] L. Leydesdorff et S. Bensman, « Classification and Powerlaws:

The Logarithmic Transformation »

- [14] D. Mao, Y. Wang, et Q. Wu, « A new approach for physiological time series ». arXiv, 23 avril 2015. Consulté le: 28 juin 2023. [En ligne]. Disponible sur: <http://arxiv.org/abs/1504.06274>.
G. K. Patro et K. K. Sahu, « Normalization: A Preprocessing Stage ». arXiv, 19 mars 2015. doi: 10.48550/arXiv.1503.06462.
- [15] G. Sparling, « The Xi-transform for conformally flat space-time ». arXiv, 30 novembre 2006. Consulté le: 28 juin 2023. [En ligne]. Disponible sur: <http://arxiv.org/abs/gr-qc/0612006>
- [16] H. Lang et H. Poon, « Self-supervised self-supervision by combining deep learning and probabilistic logic ». arXiv, 22 décembre 2020. doi: 10.48550/arXiv.2012.12474.
- [17] V. K. Garg et A. T. Kalai, « Meta-Unsupervised-Learning: A supervised approach to unsupervised learning ». arXiv, 3 janvier 2017. doi: 10.48550/arXiv.1612.09030.
- [18] K. Wu et X.-J. Zeng, « Group-Agent Reinforcement Learning ». arXiv, 30 avril 2022. doi:10.48550/arXiv.2202.05135.
- [19] M. Schlesinger et E. Vodolazskiy, « Minimax deviation strategies for machine learning and recognition with short learning samples ». arXiv, 16 juillet 2017. doi: 10.48550/arXiv.1707.04849.
- [20] J. Grabocka, R. Scholz, et L. Schmidt-Thieme, « Learning Surrogate Losses ». arXiv, 24 mai 2019. doi: 10.48550/arXiv.1905.10108.
- [21] « An aggregate and iterative disaggregate algorithm with proven optimality in machine learning | SpringerLink ». <https://link.springer.com/article/10.1007/s10994-016-5562-z> (consulté le 28 juin 2023).
- [22] R. J. Chase, D. R. Harrison, A. Burke, G. M. Lackmann, et A. McGovern, « A Machine Learning Tutorial for Operational Meteorology, Part I: Traditional Machine Learning », *WeatherForecast.*, vol. 37, n° 8, p. 1509-1529, août 2022, doi: 10.1175/WAF-D-22-0070.1.
- [23] « Understanding the Bias-Variance Tradeoff ». <https://scott.fortmann-roe.com/docs/BiasVariance.html> (consulté le 28 juin 2023).
- [24] « Machine Learning for Skin Permeability Prediction: Random Forest and XGBoost Regression by Kevin Ita, Joyce Prinze :: SSRN ». https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4415496 (consulté le 28 juin 2023).

Webographie

Webographie

[1] Euronews. (2022, Septembre 19). Algérie : Une industrie pharmaceutique en plein développement. Consulté le 28 juin 2023, de <https://fr.euronews.com/2022/09/19/algerie-une-industrie-pharmaceutique-en-plein-developpement>.

[2] Xerfi. (2023). Industrie pharmaceutique: hausse de la production en 2023. Consulté le 28 juin 2023, de https://www.xerfi.com/blog/Industrie-pharmaceutique-hausse-de-la-production-en-2023_344.

[3] Analytics Vidhya. (2021, Juillet). Time Series Analysis: A Beginner- friendly Guide. Consulté le 28 juin 2023, de <https://www.analyticsvidhya.com/blog/2021/07/time-series-analysis-a-beginner-friendly-guide/>.

[4] JavaTpoint. (s.d.). Types of Machine Learning. Consulté le 28 juin 2023, de <https://www.javatpoint.com/types-of-machine-learning>.

[5] Wikipédia. (s.d.). Cross Industry Standard Process for Data Mining. Consulté le 28 juin 2023, de https://fr.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining.

Annexes

Annexe A

Présentation des outils

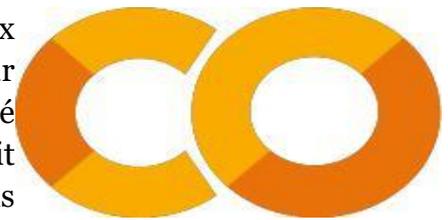
Python

Python est un langage de programmation de haut niveau qui est célèbre pour sa syntaxe propre et lisible. Très polyvalent, Python est couramment utilisé dans diverses applications, y compris le développement web, l'analyse de données, l'intelligence artificielle, et l'automatisation. Il dispose également d'une vaste bibliothèque standard et d'un écosystème de paquets tiers, qui étendent ses capacités.



Google colab

Google Colab, ou Collaboratory, est un service en nuage qui offre un environnement de notebooks Jupyter. Il permet aux utilisateurs d'exécuter des codes Python dans un navigateur web sans configuration locale. Il est particulièrement apprécié des scientifiques de données et chercheurs, car il fournit gratuitement un accès limité à des accélérateurs matériels tels que des GPU et des TPU, facilitant l'exécution de tâches de calcul intensif.



Sage X3

Sage X3, aussi connu sous le nom de Sage Business Cloud X3, est un système de planification des ressources d'entreprise (ERP) destiné aux entreprises de taille moyenne et grande. Il combine diverses fonctionnalités, telles que la gestion financière, les achats, la gestion des stocks et des ventes, dans un système intégré. Sage X3 est reconnu pour sa flexibilité, permettant un déploiement sur site ou en cloud, et peut être personnalisé pour répondre aux exigences spécifiques de différentes industries.



Bibliothèques utilisées

Matplotlib

Matplotlib est une bibliothèque populaire de visualisation de données en Python. Elle permet de créer une grande variété de graphiques tels que des diagrammes en barres, des histogrammes, des diagrammes circulaires, des graphiques en nuage de points, etc. Matplotlib offre une flexibilité considérable pour personnaliser les graphiques et les rendre esthétiquement plaisants.



Scikit-learn (sklearn)

Scikit-learn est une bibliothèque d'apprentissage automatique (machine learning) en Python. Elle propose un ensemble d'outils et d'algorithmes pour effectuer des tâches courantes en apprentissage automatique, telles que la classification, la régression, le regroupement, la réduction de dimensionnalité, la sélection de modèle, etc. Scikit-learn est appréciée pour sa simplicité d'utilisation et sa documentation détaillée.



Pandas

Pandas est une bibliothèque Python pour manipuler et analyser des données tabulaires. Elle offre des fonctionnalités puissantes pour le nettoyage, la fusion, l'indexation, le filtrage et l'agrégation de données.



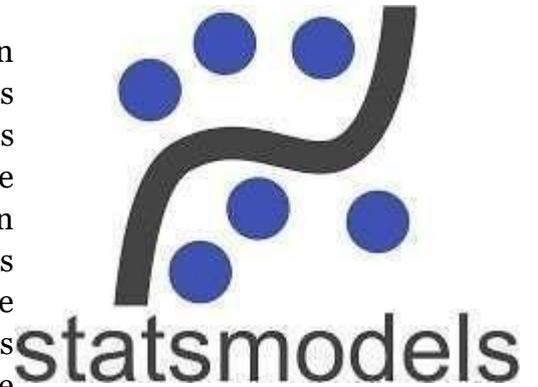
NumPy

NumPy est une bibliothèque fondamentale pour le calcul scientifique en Python. Elle fournit des structures de données de tableau multidimensionnel, des fonctions mathématiques avancées, des outils pour l'intégration de code C/C++ et Fortran, ainsi que des fonctionnalités pour travailler avec des tableaux de données.



Statsmodels

Statsmodels est une bibliothèque statistique en Python. Elle propose des outils pour effectuer des analyses statistiques et des modélisations économétriques. Statsmodels offre une vaste gamme de modèles statistiques, tels que la régression linéaire, la régression logistique, les séries temporelles, les modèles ARIMA, les modèles de survie, etc. Elle fournit également des fonctionnalités pour effectuer des tests statistiques, des diagnostics de modèle et des visualisations associées.



PySide6

PySide 6 est une bibliothèque Python basée sur Qt, permettant de créer facilement des interfaces utilisateur multi plateformes attrayantes et interactives. Elle offre une large gamme de fonctionnalités et une compatibilité avec les principaux systèmes d'exploitation tels que Windows, macOS et Linux.



Annexe B

Lien utile

B.01 Voici le lien qui contient tout le projet (et j'accepte)

<https://drive.google.com/drive/folders/16aGRdWP6SoMzuNrvYH Zf 9qBqCG DKyHo8?usp=sharing>

Voici son contenu :

- Le code de l'interface de prédiction
- Les données brutes de l'ERP de biopharme après extraction
- Le notebook qui contient tout le code de mon étude
- Les résultats de l'étudew