



**HAL**  
open science

# Prévision de la demande par apprentissage automatique pour le dimensionnement et la gestion de stocks des médicaments dans les pharmacies hospitalières

Denis Koala

► **To cite this version:**

Denis Koala. Prévision de la demande par apprentissage automatique pour le dimensionnement et la gestion de stocks des médicaments dans les pharmacies hospitalières. Base de données [cs.DB]. Université Grenoble Alpes [2020-..], 2023. Français. NNT: 2023GRALI084 . tel-04861139

**HAL Id: tel-04861139**

**<https://theses.hal.science/tel-04861139v1>**

Submitted on 2 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES**

École doctorale : I-MEP2 - Ingénierie - Matériaux, Mécanique, Environnement, Energétique, Procédés, Production

Spécialité : GI - Génie Industriel : conception et production

Unité de recherche : Laboratoire des Sciences pour la Conception, l'Optimisation et la Production de Grenoble

**Prévision de la demande par apprentissage automatique pour le dimensionnement et la gestion de stocks des médicaments dans les pharmacies hospitalières**

**Machine learning based demand forecasting for inventory sizing and management of medicines in hospital pharmacies**

Présentée par :

**Denis KOALA**

Direction de thèse :

**Gülğün ALPAN-GAUJAL**

PROFESSEURE DES UNIVERSITES, GRENOBLE INP

Directrice de thèse

**Zakaria YAHOUNI**

MAITRE DE CONFERENCES, Université Grenoble Alpes

Co-encadrant de thèse

Rapporteurs :

**SEBASTIEN THOMASSEY**

PROFESSEUR DES UNIVERSITES, ENS ARTS INDUST TEXTIL ROUBAIX

**LATEFA GHOMRI**

FULL PROFESSOR, UNIVERSITE ABOU BEKR BELKAID TLEMCCEN

Thèse soutenue publiquement le **14 novembre 2023**, devant le jury composé de :

**GULGUN ALPAN-GAUJAL**

PROFESSEURE DES UNIVERSITES, GRENOBLE INP

Directrice de thèse

**EVREN SAHIN**

PROFESSEURE DES UNIVERSITES, ECOLE CENTRALE SUPELEC

Examinatrice

**VAN-DAT CUNG**

PROFESSEUR DES UNIVERSITES, GRENOBLE INP

Président

**SEBASTIEN THOMASSEY**

PROFESSEUR DES UNIVERSITES, ENS ARTS INDUST TEXTIL ROUBAIX

Rapporteur

**LATEFA GHOMRI**

FULL PROFESSOR, UNIVERSITE ABOU BEKR BELKAID TLEMCCEN

Rapporteuse

Invités :

**DJAMAL SI MOHAND**

DOCTEUR EN SCIENCES, KLS-GROUP



# Remerciements

Cette thèse touche à sa fin, et est le fruit de nombreux efforts consentis par nombreuses personnes tout au long de mon parcours. C'est pourquoi, je tiens à exprimer ma gratitude à tous ceux d'une manière ou d'une autre, ont contribué à la réalisation de ce travail.

Je voudrais tout d'abord exprimer mes plus profonds remerciements à ma directrice de thèse, Professeure Gülgün ALPAN sans qui ce travail n'aurait pas vu le jour. Merci pour votre disponibilité, votre rigueur scientifique, vos remarques judicieuses et tous les travaux réalisés ensemble. Merci également pour la chaleur humaine que vous dégagez, auprès de vous chère « Maître », j'ai beaucoup appris.

Je tiens ensuite à remercier Docteur Zakaria YAHOUNI de m'avoir donné la chance de préparer cette thèse. Ses remarques pertinentes m'ont aidé à cadrer cette thèse. J'ai énormément profité de son abnégation au travail, désolé de vous avoir volé certains de vos week-ends et vos nuits.

Merci également au Professeur Yannick FREIN dont j'ai bénéficié des conseils judicieux avant son départ à la retraite.

Mes remerciements vont à l'endroit de Docteur Djamel SI-MOHAND, Directeur R&D au sein de KLS-GROUP, pour la qualité de l'encadrement et du soutien qu'il m'a accordée tout le long de cette étude. Grâce à lui, je dispose d'un meilleur recul par rapport à la recherche que j'ai effectué.

Je remercie Gilbert GARCIA, Président directeur général de KLS-GROUP, et Grégoire GARCIA, directeur général KLS-GROUP pour m'avoir accueilli dans votre société et de m'avoir ainsi permis d'effectuer ce projet de thèse. Et merci également pour la confiance et l'opportunité que vous m'avez données de continuer l'aventure avec vous.

C'est le lieu pour moi de remercier le personnel de l'entreprise KLS-GROUP, en particulier ceux de KLS-LOGISTIC, et de KLS-MEDICAL. J'évite de citer des noms par crainte d'en oublier, j'ai pratiquement intéressé, souvent agacé certains avec mes questions et mes recherches d'information. Merci pour vos sympathies, aides indispensables et les activités réalisées. Permettez-moi tout de même de citer Christelle REVERDY pour cette joie de vivre qu'elle m'a apportée.

Je remercie aussi les membres du comité de suivi de cette thèse, Professeur Vincent AUGUSTO et Professeur Bernard PENZ pour leurs remarques et encouragements.

Qu'il me soit également permis de remercier Professeur Van-Dat CUNG pour avoir accepté de présider le jury de soutenance, aux Professeur Sébastien THOMASSEY et Professeure Latéfa GHOMRI de m'avoir fait l'honneur d'accepter d'être rapporteurs de cette thèse et au Professeure Evren SAHIN pour avoir accepté d'être également membre de mon jury de soutenance.

J'ai connu trois directeurs de laboratoire G-SCOP, les Professeurs François VILLENEUVE, Bernard PENZ et Peggy ZWOLINSKI merci pour l'accueil au sein de votre structure. Mes

---

remerciements vont à l'endroit de toutes les équipes du laboratoire et tout particulièrement DOME2S dont j'étais membre. Les nombreuses discussions que j'ai pu avoir avec chacun m'ont beaucoup apporté.

Merci à toute l'administration, au personnel enseignant et chercheur, aux équipes techniques de GSCOP, de Grenoble INP-génie industriel en général, à l'ADOC, et tous les étudiants et doctorants que j'ai pu rencontrer. J'évite également de citer des noms de peur d'en oublier.

Je remercie mes colocos, les membres de l'association des étudiants Burkinabè de l'Isère (AESBI), mes amis de Grenoble, de Saint-Etienne, Lyon, Strasbourg et du Burkina Faso, tout particulièrement à Dr Mohamed SANA et à Dr Émile OUEDRAOGO pour les discussions riches autour de mes travaux.

Enfin, je rends hommage à ceux sans lesquels ce travail n'aurait pas pu être fait : à toute ma grande famille, en particulier mes sœurs et à ma mère. Désolé de n'avoir pas été présent surtout dans les moments très difficiles, Paix à l'âme de mon bien-aimé père et papa ! Ce travail vous est dédié.

---

## Résumé

Afin d'améliorer la prise en charge des patients et réduire les coûts logistiques hospitaliers, le dimensionnement et la gestion des stocks de produits pharmaceutiques se posent aujourd'hui encore comme des défis majeurs pour les établissements de santé. Les surstockages, les pénuries et la mobilisation de ressources sont des défis de la gestion des stocks au quotidien. Le dimensionnement de stock, quant à lui, concerne plus particulièrement l'implantation d'un nouvel entrepôt de pharmacie hospitalière ou la création d'une unité médicale. Ce dernier est complexe par manque de données historiques pour prédire la consommation des produits pharmaceutiques ; une étape cruciale pour le bon dimensionnement des stocks. En pratique, la difficulté de prévoir la demande conduit les gestionnaires à corriger les niveaux de stocks par essais-et-erreurs, entraînant plusieurs mois d'actions correctives, et une mauvaise gestion de stocks. Dans ce contexte, l'importance de l'estimation de la consommation des produits pharmaceutiques résonne profondément à la fois pour la gestion mais, aussi le dimensionnement de stocks pour le bon fonctionnement des établissements de santé. Ces défis sont exacerbés par le manque de données fiables et les complexités inhérentes à la prévision de la demande, notamment les fortes volatilités de demande, les interventions médicales imprévisibles, le risque d'expiration, la diversité des produits, etc. Face à cette réalité, la mise en place d'outils d'aide à la décision pour la planification et la prévision des besoins en produits de santé se révèle d'une importance critique.

L'objectif de cette thèse est de proposer des méthodes et des outils de prévision de la demande de produits pharmaceutiques pour les hôpitaux dépourvus d'historique de consommation. L'idée fondamentale réside dans l'utilisation de données provenant d'autres hôpitaux pour prédire les besoins d'un nouvel établissement de santé qui ne dispose pas de données historiques. Pour cela, on cherche à caractériser les hôpitaux par les facteurs qui influent sur les tendances de consommation, tels que le nombre de lits, le nombre de séjours d'hospitalisation, lieu de l'hôpital, etc. Ceci permet de faire de rapprochement entre les hôpitaux pour se servir des données historiques de consommation existantes. Les méthodes de prévision s'appuient sur des techniques d'apprentissage automatique et sont validées grâce à des données réelles provenant de plusieurs hôpitaux en France.

Pour mener à bien cette thèse, tout d'abord une analyse des facteurs qui influent sur la consommation de produits pharmaceutiques dans le contexte français est réalisée. Cette analyse englobe une revue bibliographique, une étude de terrain et une analyse statistique des corrélations entre ces facteurs. Par la suite, une étude comparative entre les méthodes de prévision quantitatives classiques et celles basées sur l'apprentissage automatique est entreprise, en tenant compte des facteurs de consommation. Dans la continuité de ces travaux, une exploration de l'approche causale de la prévision de la demande est menée pour atteindre notre objectif principal. Les résultats numériques montrent que les outils d'apprentissage automatique sont les meilleurs et atteignent plus de 90% de précision en termes de prédiction. Aussi, certains algorithmes se montrent efficaces pour la prédiction de consommation de certains produits pharmaceutiques dans de nouvelles structures. L'ensemble de ces contributions permet de formuler des recommandations précieuses pour les décideurs du secteur hospitalier, en lien avec la problématique traitée.

**Les mots - clés :** Logistique hospitalière, Dimensionnement et gestion des stocks, Prévision de la demande de médicament, Fouille de données, Apprentissage automatique.

# Abstract

To enhance patient care and reduce hospital logistical costs, the sizing and management of pharmaceutical product stocks continue to present significant challenges for healthcare facilities nowadays. Issues like overstocking, shortages, and resource mobilization are daily stock management challenges. Stock sizing, on the other hand, is particularly complex when there is a lack of historical data to predict pharmaceutical product consumption, a crucial step for proper stock sizing. In practice, this leads managers to adjust stock levels through trial and error, resulting in several months of corrective actions and poor stock management.

In this context, the importance of estimating pharmaceutical product consumption resonates deeply, both for management and stock sizing, to ensure the smooth functioning of healthcare establishments. These challenges are exacerbated by the lack of reliable data and the inherent complexities of demand forecasting, including high demand volatility, unpredictable medical interventions, expiration risks, product diversity, and more. Faced with this reality, the implementation of decision-support tools for healthcare product planning and demand forecasting is critically essential.

The objective of this thesis is to propose methods and tools for forecasting the demand for pharmaceutical products for hospitals lacking consumption history. The fundamental idea is to utilize historical consumption data of pharmaceutical products from multiple hospitals to predict and manage the needs of a new healthcare facility that lacks historical data. To achieve this, we aim to characterize hospitals based on factors influencing consumption trends, such as the number of beds, the number of hospital visits, the location of the hospital, and more. This allows for cross-referencing between hospitals to make use of data from existing healthcare structures. Forecasting methods rely on machine learning techniques and are validated using real data from several hospitals in France.

To successfully complete this thesis, we first conduct an analysis of factors influencing pharmaceutical product consumption in the French context. This analysis includes a literature review, field studies, and a statistical analysis of correlations between these factors. Subsequently, a comparative study between traditional quantitative forecasting methods and those based on machine learning is undertaken, taking consumption factors into account. In continuation of this work, a causal approach to demand forecasting is explored to achieve our main objective. Numerical results demonstrate that machine learning tools perform the best and achieve over 90% accuracy in prediction. Additionally, certain algorithms effectively predict the consumption of specific pharmaceutical products in new healthcare facilities. All of the contributions provide valuable recommendations for decision-makers in the hospital sector, related to the addressed issue.

**Keywords :** Hospital logistics, Inventory sizing and management, Medicine demand forecasting, Data mining, Machine learning.

# Table des matières

Table des matières	v
Table des figures	ix
Liste des tableaux	xi
Liste des Abréviations	xiii
Introduction générale	1
<b>I Contexte et problématique de la thèse</b>	<b>5</b>
<b>1 Logistique hospitalière et produits pharmaceutiques</b>	<b>6</b>
1.1 La logistique hospitalière	6
1.1.1 Définitions de la logistique hospitalière	8
1.1.2 Spécificités et importance des activités de logistique hospitalière	10
1.2 Les produits de santé	14
1.2.1 Définitions et types de produits de santé	14
1.2.1.1 Les produits pharmaceutiques	15
1.2.1.2 Les dispositifs médicaux	16
1.2.2 Régulations et sérialisations des produits de santé	17
1.2.2.1 Régulations des produits de santé en France	17
1.2.2.2 Sérialisation des médicaments	17
1.3 Dimensionnement et gestion de stocks des produits de santé	23
1.3.1 Dimensionnement et gestion des stocks	23
1.3.2 Méthodes et techniques du dimensionnement et gestion des stocks	25
1.4 Conclusion	26
<b>2 Préviation de la demande des produits pharmaceutiques</b>	<b>27</b>
2.1 Méthodes classiques pour l'estimation des besoins en produits de santé	28
2.1.1 Méthode de consommation	29
2.1.2 Méthode de morbidité	30
2.1.3 Méthodes de consommation par procuration	30
2.1.4 Projection du besoin budgétaire au niveau du service	31

2.1.5	Méthode de la demande des hôpitaux . . . . .	31
2.1.6	Méthode basée sur la population . . . . .	32
2.1.7	Comparaison des méthodes classiques de quantification . . . . .	32
2.2	Méthodes statistiques de prévision de la demande . . . . .	36
2.2.1	Méthodes quantitatives de prévision de la demande . . . . .	38
2.2.1.1	Méthodes de séries temporelles . . . . .	38
2.2.1.2	Méthodes causales . . . . .	38
2.2.2	Machine Learning pour la prévision de la demande . . . . .	39
2.2.2.1	Le Machine Learning dans le domaine de la santé . . . . .	41
2.2.2.2	Machine Learning pour la prévision de demande des produits de santé . . . . .	43
2.3	Conclusion . . . . .	47
<b>II Contributions de la thèse</b>		<b>48</b>
<b>3 Description, structuration et pré-traitement des données</b>		<b>49</b>
3.1	Descriptions des sources de données . . . . .	50
3.1.1	Bases de données internes . . . . .	50
3.1.1.1	Base de données des articles . . . . .	50
3.1.1.2	Unité fonctionnelle et base de données des unités fonctionnelles a) Qu'est ce qu'une unité fonctionnelle dans l'hôpital? . . . . .	51
	b) Description des bases de données des unités fonctionnelles . . . . .	55
3.1.1.3	Bases de données historiques . . . . .	55
3.1.2	Bases de données externes . . . . .	58
3.1.2.1	Données open sources : . . . . .	58
3.1.2.2	Base de données nécessitant une autorisation ou payantes . . . . .	62
3.1.3	Étude de terrain . . . . .	62
3.2	Organisation et structuration des données . . . . .	64
3.2.1	Pré-traitements des données . . . . .	64
3.2.1.1	Pré-traitements des données des produits . . . . .	65
	a) Démarche employée . . . . .	65
	b) Résultats du traitement . . . . .	68
3.2.1.2	Pré-traitements des données des unités fonctionnelles . . . . .	70
	a) Démarche employée . . . . .	70
	b) Résultats du traitement . . . . .	70
3.2.2	Premier croisement des données . . . . .	73
3.2.3	Second croisement des données . . . . .	74
3.3	Conclusion . . . . .	78
<b>4 Étude et analyse des facteurs impactant la consommation des produits pharmaceutiques</b>		<b>79</b>
4.1	Introduction . . . . .	79

4.2	Méthodologie . . . . .	80
4.3	Les Facteurs impactant la consommation de médicaments . . . . .	82
4.3.1	Identifications des facteurs selon la littérature . . . . .	82
4.3.2	Choix des facteurs dans le cadre de notre étude . . . . .	86
4.3.2.1	Critère de non conformité de contexte français . . . . .	87
4.3.2.2	Critère lié à la disponibilité de données . . . . .	87
4.3.2.3	Avis des experts sur le choix des facteurs . . . . .	90
4.4	Description des données et méthodes d'analyse de corrélation . . . . .	91
4.4.1	Description des données . . . . .	91
4.4.2	Méthodes d'analyse de corrélation des variables . . . . .	92
4.5	Résultats et analyses . . . . .	95
4.5.1	Corrélations pour tous les médicaments confondus . . . . .	95
4.5.2	Corrélations par médicament . . . . .	99
4.5.3	Etude complémentaire sur la variable "MOIS" . . . . .	102
4.6	Conclusion . . . . .	104
<b>5</b>	<b>Étude comparative des modèles de machine learning vs méthodes classiques</b>	<b>105</b>
5.1	Introduction . . . . .	105
5.2	Méthodologie et démarche . . . . .	109
5.2.1	Extraction et collecte des données . . . . .	109
5.2.2	Exploration et pré-traitement des données . . . . .	110
5.2.3	Sélection des données . . . . .	111
5.2.4	Application des modèles . . . . .	112
5.2.4.1	Méthode de consommation . . . . .	112
5.2.4.2	Modèles d'apprentissage automatique . . . . .	113
5.2.5	Évaluation et comparaison . . . . .	116
5.3	Résultats expérimentaux et analyses . . . . .	117
5.3.1	Évaluation des performances des techniques d'apprentissage automatique sélectionnées . . . . .	118
5.3.2	Comparaison des modèles de Random Forest, du DNN et de la méthode de consommation . . . . .	120
5.4	Conclusion et perspectives . . . . .	122
<b>6</b>	<b>Dimensionnement des besoins pour un nouvel hôpital</b>	<b>124</b>
6.1	Introduction . . . . .	124
6.2	Méthodes et Matériels . . . . .	125
6.2.1	Description, pré-traitement et séparation des données . . . . .	125
6.2.2	Modèles d'apprentissage automatique de base et métriques de performance . . . . .	128
6.2.3	Modèle hybride . . . . .	128
6.3	Résultats expérimentaux et analyses . . . . .	130
6.3.1	Conditions expérimentales . . . . .	130

---

6.3.2	Évaluations de la performance des prévisions après l'entraînement . . .	132
6.3.2.1	Analyse des performances selon les médicaments . . . . .	135
6.3.2.2	Analyse des performances des modèles . . . . .	137
6.3.2.3	Résumé de cette première analyse . . . . .	137
6.3.3	Prévision pour d'autres hôpitaux . . . . .	138
6.3.3.1	Analyse des performance par médicament . . . . .	142
6.3.3.2	Analyse des performances selon les modèles . . . . .	145
6.4	Conclusions . . . . .	149
	<b>Conclusion générale et perspectives</b>	<b>150</b>
	<b>Bibliographie</b>	<b>155</b>
	<b>III Les annexes</b>	<b>167</b>
	<b>A Annexe du chapitre 3</b>	<b>168</b>
	<b>B Étude de Terrain : Questionnaire</b>	<b>171</b>
	<b>C Annexe du chapitre 5</b>	<b>180</b>
C.1	Les performances des modèles RF et DNN avec l'ensemble de données de test	180
C.2	Comparaison des modèles de Random Forest, du DNN et de la méthode de consommation : Exemple avec le jeu de donnée de l'hôpital H3 . . . . .	181
	<b>D Annexe du chapitre 6</b>	<b>183</b>
D.1	Résultats suite aux tests des modèles construites . . . . .	183
D.2	Résultats suite aux validations des modèles dans les deux hôpitaux (H4 & H5)	193
D.2.1	Pour l'hôpital H4 . . . . .	193
D.2.2	Pour l'hôpital H5 . . . . .	195

# Table des figures

1.1	Ensemble des activités de la logistique hospitalière (Source : Chow and Heaver (1994) adaptées par Beaulieu et al.[15] . . . . .	9
1.2	Description de la chaîne logistique pharmaceutique (Source : Di Martinelly [85])	10
1.3	Comparaison entre la logistique manufacturière et la logistique hospitalière [49]	11
1.4	Les proportions des coûts de la logistiques hospitalière (Source : [72, 49]) . .	13
1.5	Exemple sérialisation de médicament (source FMVO) . . . . .	18
1.6	Exemple Code UCD (source : Club Inter Pharmaceutique [117]) . . . . .	20
1.7	Exemple de codification CIP (source : Club Inter Pharmaceutique [117]) . .	22
2.1	Méthodes quantitatives de la prévision de la demande . . . . .	37
2.2	Tendance des publications sur le ML dans le milieu sanitaire . . . . .	42
2.3	Visualisation du réseau de mots-clés sur le ML dans la santé . . . . .	43
3.1	Organisation et hiérarchisation des Hôpitaux . . . . .	53
3.2	Process global d’organisation des données . . . . .	65
3.3	Démarche employée pour le pré-traitement des données de produits de santé	67
3.4	Processus de pré-traitement des données des unités fonctionnelles . . . . .	71
4.1	Résumé de la méthodologie employée . . . . .	81
4.2	Exemples de tendances de consommation de 5 médicaments à l’Hôpital H3 (De janvier 2017 à janvier 2019) . . . . .	98
4.3	Exemples de tendances de consommation de 5 médicaments à l’Hôpital H6 (De janvier 2017 à janvier 2019) . . . . .	99
5.1	Approche proposée . . . . .	109
5.2	Structure basique de notre DNN . . . . .	115
5.3	Prédiction pour le médicament 3 à l’hôpital H3 . . . . .	121
5.4	Prédiction pour le médicament 6 à l’hôpital de H3 . . . . .	122
5.5	Prédiction pour le médicament 9 à l’hôpital de H3 . . . . .	122
6.1	Diagramme de flux de la méthodologie proposée. . . . .	126
6.2	Structure du modèle hybride proposé . . . . .	129
6.3	Performances des modèles des médicaments en commun (score $R^2$ ) . . . . .	133
6.4	Performances des modèles des médicaments en commun (score MAPE) . . .	134

6.5	Visualisation des performances des modèles pour un médicament dans le groupe 1 (Ex. médicament 2 : H4 vs H5) . . . . .	147
6.6	Visualisation des performances des modèles pour un médicament dans le groupe 2 (Ex. médicament 1 : H4 vs H5) . . . . .	147
6.7	Visualisation des performances des modèles pour un médicament dans le groupe 3 (Ex. médicament 27 : H4 vs H5) . . . . .	148
6.8	Visualisation des performances des modèles pour un médicament dans le groupe 5 (Ex. médicament 44 : H4 vs H5) . . . . .	148

# Liste des tableaux

1.1	Extrait de codes UCD en lien avec le code ATC N02BE01 . . . . .	21
1.2	Exemples entres les trois codes CIP, UCD et ATC (Source PHAST) . . . . .	22
1.3	Méthodes de gestion des stocks en fonction des questions posées . . . . .	25
2.1	Avantages et limites des méthodes classiques de quantification . . . . .	36
2.2	Présentation succincte de quelque modèles de machine learning . . . . .	47
3.1	Table des caractéristiques des produits de santé de la base . . . . .	52
3.2	Répartition des informations par hôpital . . . . .	56
3.3	Synthèse des caractéristiques de la base de données . . . . .	57
3.4	Répartition des données de consommation par hôpital dans la base . . . . .	58
3.5	Répartition des codes de traçabilité par hôpital . . . . .	66
3.6	Répartition des codes de médicaments par hôpital après traitements . . . . .	69
3.7	Esquisse de la table de référence des médicaments traçables dans les 6 hôpitaux	69
3.8	Répartition des destinations de consommations par hôpital . . . . .	72
3.9	Répartition des données de consommation après croisement avec les autres bases de données pré-traitées . . . . .	73
3.10	Visualisation des médicaments en commun consommées dans les hôpitaux . .	75
3.11	Un aperçu de la table de données finale . . . . .	77
4.1	Facteurs affectant la consommation de médicaments dans les hôpitaux . . . . .	85
4.2	Les facteurs sélectionnés . . . . .	88
4.3	Description des colonnes du jeu de données . . . . .	92
4.4	Résultats du test de Kolmogorov-Smirnov pour la normalité des données (mé- dicament sans distinction) . . . . .	96
4.5	Corrélations de Pearson . . . . .	96
4.6	Corrélations de Kendall . . . . .	97
4.7	Corrélations de Spearman . . . . .	97
4.8	Correlations de Kendall par médicament . . . . .	100
4.9	Correlations de Spearman par médicament . . . . .	101
4.10	Test statistique pour la corrélation entre la consommation et le "MOIS" . . .	103
5.1	Résumé des gaps de la littérature et nouveauté de ce travail. . . . .	108
5.2	Les caractéristiques principales l'ensemble de données . . . . .	110
5.3	Les médicaments sélectionnés . . . . .	111

5.4	Comparaison des performances des modèles suite à la validation croisée . . . .	119
5.5	Présentation des résultats par médicament et par modèle sur l'ensemble de données de validation . . . . .	120
6.1	Répartition du nombre de lignes d'informations par médicament par hôpital et total . . . . .	127
6.2	Hyperparamètres utilisés pour chaque modèle . . . . .	131
6.3	Modèle hybride 2 par médicament . . . . .	131
6.4	Résumé des performances des modèles selon la métrique $R^2$ (avec les données tests) . . . . .	132
6.5	Résumé des performances des modèles selon la métrique MAPE (avec les données tests) . . . . .	133
6.6	Résumé de la performance des modèles selon la métrique $R^2$ . . . . .	135
6.7	Répartition des groupes en fonction du score MAPE . . . . .	136
6.8	Répartition des médicaments en fonction des catégories de score $R^2$ . . . . .	138
6.9	Répartition des médicaments en fonction des catégories de score MAPE . . . .	139
6.10	Performances des modèles sur les 18 médicaments dans l'hôpital H4 selon la métrique $R^2$ & MAPE . . . . .	140
6.11	Performances des modèles sur les 18 médicaments dans l'hôpital H5 selon la métrique $R^2$ & MAPE . . . . .	141
6.12	Récapitulatif performances des modèles dans les hôpitaux H4 et H5 selon les seuils de $R^2$ & MAPE fixés . . . . .	145
A.1	Description des variables de la base de données structurée et exploitable . . . .	170
D.1	Scores selon $R^2$ (Données test) . . . . .	184
D.2	Scores selon RMSE (donnée test) . . . . .	185
D.3	Scores selon MAE (donnée test) . . . . .	186
D.4	Scores selon MAPE (donnée test) . . . . .	187
D.5	Performance des modèles par médicament selon la métrique $R^2$ . . . . .	191
D.6	Performance des modèles par médicament selon la métrique MAPE . . . . .	193
D.7	Performance des modèles par médicament selon la métrique $R^2$ & MAPE dans H4 . . . . .	194
D.8	Performance des modèles par médicament selon la métrique $R^2$ & MAPE dans H5 . . . . .	196

# Liste des Abréviations

ACP	Analyse en Composantes Principales
AFSSAPS	Agence Française de Sécurité Sanitaire du médicament et des Produits de Santé
AISS	Association Internationale de la Sécurité Sociale
AMM	Autorisation de Mise sur Marché
ANOVA	Analysis of Variance
ANN	Artificial Neural Network
ANSM	Agence Nationale de Sécurité du Médicament et des produits de santé
ARIMA	AutoRegressive Integrated Moving Average
ARMA	AutoRegressive Moving Average
ARV	Antirétroviral
ASLOG	Association française pour la LOGistique
ATC	Anatomical Therapeutic Chemical
ATIH	Agence technique de l'informatique sur l'hospitalisation
CH	Centre Hospitalier
CHR	Centre Hospitalier Régional
CHS	Centres Hospitaliers Spécialisés
CHU	Centre Hospitalier Universitaire
CHRU	Centre Hospitalier Régional Universitaire
CIFRE	Conventions Industrielles de Formation par la Recherche
CIO	Codification Inter-Opérable
CIP	Code Identifiant de Présentation
CLCC	Centres de Lutte Contre le Cancer
CNAMTS	Caisse Nationale d'Assurance Maladie des Travailleurs Salariés
CNIL	Commission Nationale de l'Informatique et des Libertés
DNN	Deep Neural Network
DREES	Direction de la Recherche, des Études, de l'Évaluation et des Statistiques

DSES	Direction de la Stratégie, des Études et des Statistiques
EG	Entité Géographique
EJ	Entité Juridique
EMVS	European Medicines Verification System
EOQ	Economic Order Quantity
FIFO	First In First Out
FILO	First In Last Out
FINESS	Fichier National des Établissements Sanitaires et Sociaux
FMVO	France Medicines Verification Organisation
GRNN	General Regression Neural Network
GRU	Gated Recurrent Unit
HAD	Hospitalisation à domicile
INSEE	Institut National de la Statistique et des Études Économiques
LPP	Liste des Produits et Prestations
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MAR	Multivariate Autoregressive
MCO	Médecine, Chirurgie, Obstétrique
MSH	Management Science Health
ML	Machine Learning
MLP	Multilayer Perceptron
MLR	Multiple Linear Regression
MRP	Material Requirements Planning
MSE	Mean Square Error
NMVS	National Medicines Verification System
OCDE	Organisation de Coopération et de Développement Économiques
OMS	Organisation Mondiale de la Santé
PMSI	Programme de Médicalisation des Systèmes d'Information

PSY	Psychiatrie
ReLU	Rectified Linear Unit
RF	Random Forest
RNN	Recurrent Neural Network
RMSE	Root Mean Square Error
SI	Système d'Information
SAE	Statistique annuelle des établissements de santé
SLD	Soins de Longue Durée
SNDS	Système National des Données de Santé
SNIIRAM	Système National d'Information Inter-régimes de l'Assurance Maladie
SSR	Soins de Suite et de Réadaptation
SVR	Support Vector Regression
UM	Unités Médical
UCD	Unité Commune de Dispensation
UF	Unité Fonctionnelle
USLD	Unités de Soins de Longue Durée
VAR	Vector Autoregression
VARMA	Vector Autoregressive Moving Average
VMI	Vendor Managed Inventory
WMS	Warehouse Management System

# Introduction générale

## Contexte

Comme dans la plupart des pays développés, le secteur de la santé en France est en constante transformation pour s'adapter à un environnement complexe et changeant, avec des budgets de plus en plus serrés. Selon l'Organisation de Coopération et de Développement Économiques (OCDE), les dépenses totales de santé en France ont augmenté rapidement au cours des dernières décennies. De 2000 à 2019, ces dépenses ont presque doublé, passant de 2 687,5 dollars par habitant à 5 167,8 dollars par habitant. Avec la situation de la Covid-19 en 2020, elles ont atteint 5 468,4 dollars par habitant [108]. Environ 20% de ces dépenses sont attribuées aux produits de santé (principalement les médicaments) [108, 107, 106]. Comme le soulignent Volland et al. [160] et Nachtmann & Edward A. Pohl [101], d'autres dépenses concernent les coûts logistiques qui représentent le deuxième plus gros poste de dépenses après le personnel, soit environ 30% à 46% du coût total des dépenses de santé. Dans ce contexte, la planification de la demande, le dimensionnement et la gestion optimale des stocks de produits de santé sont cruciaux, non seulement d'un point de vue de réduction des coûts, mais aussi pour de meilleure prise en charge de soins des patients. Plusieurs séries de programmes d'investissement dans le but d'optimiser la gestion au sein des centres hospitaliers ont été lancées, tout en garantissant la sécurité et la qualité des soins prodigués aux patients. Le plan Hôpital 2007-2012 en France en est un exemple [131]. Cependant, celui-ci a montré ses limites, comme le souligne Bernardini-Perinciolo et al. [17].

La logistique hospitalière est identifiée au cours des vingt dernières années comme un levier clé pour maîtriser les coûts de santé et les situations désagréables dans les services hospitaliers. Des recherches intéressantes ont été menées et des méthodes ont été mises en œuvre pour optimiser et remédier à certaines lacunes, notamment les problèmes de gestion des stocks (pénuries chroniques et généralisées, surplus ou surstockage, ...), les problèmes de chaîne d'approvisionnement, etc. Cependant, les problèmes liés à la gestion et dimensionnement de stocks demeurent. Les pénuries de médicaments sont devenues de plus en plus fréquentes ces dernières années et mettent à rude épreuve la logistique hospitalière. En France, par exemple, l'ANSM (Agence nationale de sécurité du médicament et des produits de santé) signale de plus en plus de ruptures de stock et de tensions d'approvisionnement pour plusieurs médicaments chaque année. France Assos Santé, une organisation de référence pour représenter les patients et les usagers du système de santé et défendre leurs intérêts en France, à travers un sondage révélait, en 2018 bien avant la situation de la Covid-19, que 25 % des Français ont été confrontés à une pénurie de médicaments [135]. Les raisons de ces

---

pénuries, comme le soulignent AlRuthia et al. [3], Gu et al. [43] et Shukar et al. [144], incluent l'indisponibilité des matières premières, les ralentissements de production ou les difficultés de fabrication, les fusions et acquisitions de fabricants de médicaments, les rappels volontaires, les problèmes réglementaires, la demande imprévue, les catastrophes naturelles, les perturbations du travail, etc. Même si les facteurs perturbateurs ne sont pas nécessairement liés aux opérations hospitalières, les hôpitaux doivent disposer d'outils de gestion et dimensionnement de stocks efficaces en vue d'appréhender de façon optimale les besoins. La manière dont les produits de santé sont estimés ou quantifiés a des implications sérieuses pour le fonctionnement des établissements de santé. Certaines recherches estiment qu'une gestion efficace permet d'éliminer environ la moitié des coûts liés à la logistique dans les hôpitaux [120, 49, 72].

De plus, selon Vila-Parrish & Ivy [159] et Volland et al. [160], la gestion des stocks des produits de santé, notamment les médicaments, impose des exigences élevées et diffère de celle d'autres catégories de produits en raison de leurs caractéristiques spécifiques, telles que les dates d'expiration, les exigences de stockage et la nécessité d'une gestion prudente pour garantir la sécurité des patients. De nombreuses personnes dépendent des médicaments, telles que les personnes âgées et les personnes atteintes de comorbidités. Une gestion inappropriée peut entraîner la mort de patients. La logistique dans le secteur de la santé est un scénario de vie ou de mort, car si le bon produit pharmaceutique ou l'outil médical approprié n'est pas disponible là où il est nécessaire, les patients en subissent les conséquences. De plus, le secteur pharmaceutique se caractérise par une volatilité extrême de la demande et des caractéristiques diverses, ce qui contribue à sa complexité. Il est donc crucial de mettre en place des stratégies efficaces pour optimiser les stocks de médicaments dans les hôpitaux, notamment des techniques de prévision de la demande et de contrôle des stocks. Par conséquent, la mise en place d'outils d'aide à la décision dans l'optique d'une planification et de la prévision des besoins en produits de santé se révèle d'une grande importance. D'où notre première motivation. Cette thèse s'intéresse donc au dimensionnement et à la gestion du stock, en particulier la prévision de la demande des produits de santé dans les centres hospitaliers.

## Problématique et objectif de recherche

Cette thèse est ancrée dans la pratique grâce à une synergie avec le secteur industriel, en cohérence avec le contexte d'un contrat CIFRE réalisé en partenariat avec l'entreprise KLS Group<sup>1</sup>, en particulier sa filiale KLS Médical, spécialisée dans le développement et l'intégration d'outils d'optimisation de la logistique hospitalière et de gestion des stocks comme le WMS (Warehouse Management System). Le WMS est un des leviers essentiels permettant l'aboutissement de cette gestion effectuée de manière efficace et automatisée dans le cadre des activités d'entrepôt. Cette catégorie de logiciel vise à gérer les opérations d'un entrepôt dans le but de les optimiser. L'usage du WMS est d'autant plus nécessaire lorsque les hôpitaux choisissent de passer le cap de l'automatisation. L'entreprise partenaire, suite aux intégrations de leurs technologies dans plus de 50 centres hospitaliers en Europe, avec une

---

1. Lien accès KLS-GROUP : <https://www.kls-group.fr/>

---

expérience de plus de 37 ans, a constaté que la mise en place d'un nouvel entrepôt pharmaceutique pour un hôpital ou une unité médicale peut nécessiter de 3 à 12 mois d'actions correctives jusqu'à ce que le dimensionnement adéquat et la gestion des stocks soient établis. Cela est essentiellement dû à des estimations inadéquates de la consommation des produits de santé dans les unités et services des hôpitaux, en particulier de nouveaux clients (nouveaux établissements de santé). Une conséquence immédiate est la perte de ressources, de temps, un risque accru de ruptures de stock, ou de coûts élevés. Afin d'assurer un dimensionnement le plus précis possible et de garantir la performance de la chaîne d'approvisionnement pharmaceutique au sein d'une structure hospitalière, il serait nécessaire de mettre en œuvre des outils d'aide à la décision à la fois rapides et efficaces tout en tenant compte des facteurs tels que la taille de ces structures, leur emplacement, etc. De plus, la disponibilité croissante de grandes quantités de données (*Big data*) dans le domaine médical, peut également aider dans ce processus.

Partant de ces faits, l'idée principale de cette thèse est de se servir des expériences (données historiques) de consommation de produits de santé de plusieurs hôpitaux pour prédire en amont les besoins d'un nouvel hôpital. Cette prévision pourrait aider à dimensionner et gérer correctement les stocks de produits de santé dans le nouvel établissement ou dans un hôpital qui n'a pas de données de consommation. Ce problème peut être expliqué par la formulation suivante :

Soient  $H_1, H_2, \dots, H_m$ ,  $m$  hôpitaux connus avec  $n$  caractéristiques  $X_1, X_2, \dots, X_n$  (également appelées facteurs tels que la taille de l'hôpital, le nombre de personnels médicaux, etc.) et  $H_0$  un nouvel hôpital ou un autre hôpital dont les besoins sont à estimer. L'estimation des besoins en produits  $i$  pour l'hôpital  $H_0$  est formulée comme suit :

$$Y_i^{(H_0)} = f_i^{(H_1, H_2, \dots, H_m)}(X_1, X_2, \dots, X_n) \quad (1)$$

où la fonction  $f_i^{(H_1, H_2, \dots, H_m)}$  est à déterminer.

Ainsi, l'objectif de la thèse est de proposer des outils et méthodes de prévision de demande de produits de santé dans un hôpital sans historique de consommation, en s'appuyant sur les données d'autres hôpitaux et en prenant en compte les facteurs impactant les consommations. A l'ère de la digitalisation et production massive de données dans les hôpitaux, ces méthodes d'aide à la décision reposeront sur les techniques Machine Learning, en utilisant les données fournies par l'entreprise partenaire KLS-GROUP ainsi que d'autres données externes aux hôpitaux. Ces outils, nous aideront à déterminer la fonction  $f_i^{(H_1, H_2, \dots, H_m)}$  énoncée dans l'équation (1). Cette approche facilitera le dimensionnement et la gestion appropriés des stocks de produits de santé lors la mise en place d'une nouvelle structure sanitaire.

Pour relever les défis liés à cet objectif principal, plusieurs questions de recherche se posent. Il s'agit dans un premier temps de l'identification des facteurs significatifs impactant la consommation des produits de santé. Cette étape nécessite une analyse de la littérature existante ainsi qu'une étude sur le terrain pour recueillir ces informations. De plus, une fois que les corrélations entre ces facteurs et les quantités de consommations sont établies, il est impératif de passer en revue les méthodes quantitatives existantes de prédiction de la demande et de proposer une étude comparative dans le cadre de la gestion de stocks. Le but est

---

de développer des modèles de prévision de la demande qui intègrent tous les facteurs identifiés précédemment. Ces modèles seront construits en utilisant les données pharmaceutiques provenant de plusieurs hôpitaux.

Il s'agit en dernière étape d'appliquer et d'analyser les performances des outils de Machine Learning relativement à notre objectif principal.

Une tâche qui va aider dans le dimensionnement d'un stock pharmaceutique pour un nouvel entrepôt hospitalier. L'ensemble de ce travail nécessite une analyse et un traitement des données, qui représentent généralement 80 % du temps consacré à une tâche d'un *data scientist*.

## La structure du manuscrit

Cette thèse est organisée en deux parties.

1. La partie I de ce manuscrit est consacrée aux notions concernant la logistique hospitalière en particulier, le dimensionnement et gestion des stocks des produits de santé. Cette partie est subdivisée en deux chapitres.
  - Le chapitre 1 est relatif à la littérature et au positionnement de nos travaux par rapport à la logistique hospitalière. Ce chapitre revient également sur les définitions et types de produits de santé.
  - Dans le deuxième chapitre, nous revenons sur les méthodes de prévision de la demande rencontrées dans la littérature pour les produits pharmaceutiques.
2. La partie II de ce manuscrit est consacrée à l'ensemble des travaux contributifs de cette thèse. Cette partie est divisée comme suit :
  - Dans le chapitre 3, nous fournissons les descriptions de bases de données utilisées dans le cadre de nos travaux. Ce chapitre est également consacré aux prétraitements et à la structuration des données pour aider à répondre à nos objectifs.
  - Le chapitre 4 se focalise sur l'étude et analyse des facteurs impactant la consommation des produits de santé dans le contexte français.
  - Le chapitre 5 est dédié à l'utilisation de méthodes de machine learning pour la prévision de demande de produits de santé issue de plusieurs hôpitaux tout en tenant compte des facteurs identifiés dans le chapitre précédent. Une comparaison des méthodes de machine learning et méthodes classiques de quantification est également présentée dans ce chapitre.
  - Dans le chapitre 6, nous nous consacrons à l'étude du dimensionnement de produits de santé pour un nouvel hôpital en nous basant sur les expériences de consommation d'autres hôpitaux. Cette étude expérimentale se base sur l'approche causale reposant sur des méthodes de machine learning.

Enfin, nous clôturons cette thèse par une synthèse des principaux résultats apports de nos travaux et suggérons des pistes d'améliorations ainsi que des perspectives de recherche.

# Première partie

## Contexte et problématique de la thèse

# Chapitre 1

## Logistique hospitalière et produits pharmaceutiques

### Sommaire

---

<b>1.1 La logistique hospitalière</b> . . . . .	<b>6</b>
1.1.1 Définitions de la logistique hospitalière . . . . .	8
1.1.2 Spécificités et importance des activités de logistique hospitalière . . . . .	10
<b>1.2 Les produits de santé</b> . . . . .	<b>14</b>
1.2.1 Définitions et types de produits de santé . . . . .	14
1.2.2 Régulations et sérialisations des produits de santé . . . . .	17
<b>1.3 Dimensionnement et gestion de stocks des produits de santé</b> . . . . .	<b>23</b>
1.3.1 Dimensionnement et gestion des stocks . . . . .	23
1.3.2 Méthodes et techniques du dimensionnement et gestion des stocks . . . . .	25
<b>1.4 Conclusion</b> . . . . .	<b>26</b>

---

Dans ce chapitre, nous commençons par une présentation générale de la logistique hospitalière. Cela passe par les définitions de la logistique dans le domaine de la santé, tout en soulignant les particularités et l'importance de ses activités. En outre, nous présentons les produits de santé, y compris les types de produits et leur traçabilité. Par la suite, nous revenons sur les notions de dimensionnement et gestions de stocks des produits pharmaceutiques, ainsi que les méthodes relatives. L'ensemble des notions décrites dans ce premier chapitre nous permettra de mieux appréhender le contexte et la suite de nos travaux.

### 1.1 La logistique hospitalière

Bien que l'on parle le plus souvent de « la logistique », dans son ensemble, différents types de logistique doivent être distingués en fonction du stade auquel se trouvent les produits, en fonction de la stratégie et des flux ainsi que du domaine d'activité. De la source d'origine à l'exécution de la commande, différents types de logistique et aspects de la logistique se combinent pour rendre possible l'exécution de la commande.

La logistique est d'une importance cruciale pour toutes activités dans le sens où elle permet d'optimiser leur efficacité opérationnelle, de réduire les coûts, d'améliorer la satisfaction client et de maintenir leur compétitivité. Cependant les objectifs peuvent varier selon les secteurs d'activité et champs d'application. Mais en général ils comprennent :

- Gestion efficace des stocks : la logistique vise à garantir que les stocks soient maintenus au niveau optimal, ni trop bas ni trop élevé, afin d'éviter les coûts excessifs liés au stockage et à l'entreposage.
- Optimisation des processus : la logistique permet d'identifier les processus inefficaces et de les améliorer, de sorte que les flux de produits et de matériaux puissent être gérés plus efficacement.
- Réduction des coûts : en améliorant l'efficacité des processus, la logistique permet aux entreprises de réduire les coûts liés à l'approvisionnement, au stockage, à la maintenance, au transport et à la distribution.
- Amélioration de la satisfaction client : en veillant à ce que les produits soient livrés à temps, en bon état et au bon endroit, la logistique contribue à améliorer la satisfaction client.
- Renforcement de la compétitivité : en réduisant les coûts et en améliorant la qualité de service, la logistique permet aux entreprises de maintenir leur compétitivité.

Si nous restons dans les milieux hors militaire, nous pouvons énumérer nombreux secteurs d'activité et champs d'application de la logistique parmi lesquels nous pouvons citer :

- Industrie manufacturière : la logistique est indispensable pour l'approvisionnement des matières premières, la gestion des stocks, la production, la gestion des entrepôts et la distribution des produits finis.
- Secteur de la santé : la logistique hospitalière est un domaine important pour la gestion des approvisionnements en produit de santé, en équipements médicaux et en fournitures, ainsi que pour la gestion des déchets médicaux.
- Services publics : la logistique est utilisée pour la gestion des approvisionnements, la maintenance des infrastructures, la gestion des déchets, l'acheminement des produits alimentaires, et des fournitures aux institutions publiques.
- Commerce de détail : la logistique joue un rôle crucial pour la gestion des stocks, la préparation des commandes et la livraison des produits aux clients.
- Transport et logistique : la logistique est au cœur de ce secteur, qui englobe notamment le transport de marchandises, la gestion des entrepôts et des centres de distribution, la planification des itinéraires, la gestion des flottes de véhicules, la préparation des commandes et la gestion des retours.
- etc.

Dans cette thèse, nous nous focalisons sur la logistique relative au domaine de la santé : la logistique hospitalière. Dans les sections qui suivent nous décrivons les notions en lien avec cette catégorie de logistique tout en spécifiant les points qui nous intéressent.

### 1.1.1 Définitions de la logistique hospitalière

Tout comme la logistique en générale, la logistique hospitalière compte plusieurs définitions. Les travaux de Beaulieu et al. [83], Hassan Taher [49], de Narayana et al.[102], Volland et al.,[160] ainsi que Jawab et al. [61] pour ne citer que ces derniers, nous permettent de réaliser l'ampleur et l'importance de la logistique dans le domaine de la santé. Dans la littérature les mots clés qui facilitent la recherche relative à la logistique hospitalière en anglais sont en général : "Hospital Logistics", "Healthcare logistics", "medical logistics" et "health logistics". Sans être exhaustif nous proposons quelques unes des définitions rencontrées dans la littérature.

Selon Landry et Beaulieu [72] définissent la logistique hospitalière comme « *un ensemble d'activités de conception, de planification et d'exécution qui permettent l'achat, la gestion des stocks et le réapprovisionnement de biens et services entourant la prestation de services médicaux aux patients* ».

Maria Costin [29] rappelle que la commission hospitalière de l'Association Française pour la Logistique (ASLOG) définit la logistique hospitalière comme « *la gestion des flux de patients, produits, matières, des services et informations qui s'y rapportent, depuis le fournisseur jusqu'au bénéficiaire final, à un niveau défini de performance au service de la qualité et de la sécurité des soins prodigués aux patients* ». L'Association nationale des cadres et experts techniques hospitaliers, H360<sup>1</sup>, explique en d'autres termes que « *la logistique hospitalière est l'étude et l'optimisation de deux types de flux physiques hospitaliers : les flux de personnes (qui comprennent les mouvements des patients depuis leur entrée jusqu'à leur sortie de l'établissement, mais aussi ceux du personnel et des visiteurs), et les flux de matières* »[103]. Cette définition rejoint celle fournie par Nathalie Sampieri-Teissier qui déclare que : « *la logistique hospitalière peut se définir comme la technologie de la maîtrise des flux physiques (hôteliers, de médicaments, d'autres matériels médicaux et de personnels et de patients) par les flux d'informations (médicaux et administratifs)* » [133].

De ces définitions découle une démarche globale de la logistique hospitalière, qui vise à optimiser la gestion des flux physiques (personnes, matériaux, produits) et des flux d'informations (médicaux et administratifs) tout au long de la chaîne d'approvisionnement, depuis les fournisseurs jusqu'aux patients. Cette démarche comprend la planification et la conception de processus logistiques efficaces, la gestion des stocks, le suivi des trajectoires des patients, la coordination entre les différents acteurs de la chaîne, ainsi que la recherche de performances optimales en termes de qualité et de sécurité des soins prodigués aux patients. De manière générale, la logistique hospitalière vise à améliorer la qualité des soins et à optimiser les coûts en garantissant une gestion efficace des ressources et une réduction des temps d'attente.

La logistique hospitalière regroupe un ensemble d'activités que Beaulieu et al.[15] ont pu schématiser afin de voir les différentes ramifications de la logistique hospitalière (voir la figure 1.1).

Dans Fouad Jawab et al. [61], les auteurs donnent une vue d'ensemble des activités de logistique hospitalière. Ils fournissent une liste de 27 activités incluses dans le domaine de

---

1. H360 : lien pour plus de détail sur l'association <http://www.h360.fr/>



FIGURE 1.1 – Ensemble des activités de la logistique hospitalière (Source : Chow and Heaver (1994) adaptées par Beaulieu et al.[15])

la logistique hospitalière : Achat, Approvisionnement, Gestion des stocks, Distribution, Réception, Ravitaillement, Gestion des systèmes d'information, Transport, Pharmacie, Flux de patients, Service de restauration, Blanchisserie, Nettoyage/Hygiène, Stérilisation, Impression, Télécommunication/Télémédecine, Service de courrier, Planification/programmation, Conception, Prévision, Service de réception, Gestion des déchets, Logistique de soins à domicile, Maintenance, Sécurité et sûreté, Construction, Entretien des espaces verts. Cependant cette liste n'est pas exhaustive, il existe des activités comme la gestion des déchets médicaux et dangereux, la gestion des équipements biomédicaux, la gestion des archives médicales, etc. Costin appuie cette assertion en affirmant que le domaine de la logistique hospitalière est différent d'un établissement de santé à l'autre, en fonction de sa taille, de ses capacités, de ses activités et de sa culture interne [29].

Il existe également plusieurs sous-ensemble de la logistique hospitalière dont la logistique pharmaceutique qui nous intéresse plus particulièrement. La logistique pharmaceutique est un sous-ensemble de la logistique hospitalière qui englobe plusieurs aspects tels que le transport, le stockage, la manutention et la distribution de produits de santé. L'objectif de la logistique pharmaceutique est de garantir la disponibilité et l'accessibilité des produits de santé tout en assurant leur qualité, leur sécurité et leur traçabilité tout au long du processus de leur fabrication jusqu'à leur distribution et leur mise à disposition aux patients. La logistique pharmaceutique est donc essentielle pour assurer la gestion efficace et sécurisée de la chaîne

d'approvisionnement des produits de santé, y compris le dimensionnement et la gestion des stocks de médicaments, tout en respectant les réglementations en vigueur. Elle est également un élément clé de la logistique hospitalière qui assure le support aux activités de soins apportés aux malades.

Di Martinelly et al. dans un schéma citant Beretz, illustre la description de la chaîne logistique pharmaceutique [85] (voir figure 1.2). Pour eux, la chaîne logistique pharmaceutique doit permettre de mettre à disposition des patients le plus efficacement possible les produits pharmaceutiques qui leur seront administrés, dans des conditions garantissant sécurité et traçabilité tout en respectant les nombreuses réglementations entourant les produits pharmaceutiques et leur dispensation. Le schéma de Beretz permet de mettre en évidence certaines spécificités. Outre les flux mentionnés, il est important de prendre en compte le flux de retours des produits en particulier les médicaments. Ce processus concerne les produits non administrés, défectueux, endommagés ou périmés qui sont renvoyés des unités de soins vers la pharmacie centrale. Une gestion particulière est nécessaire pour décider de leur réintégration dans le stock, de leur élimination ou de leur retour aux fournisseurs en vue de leur destruction.

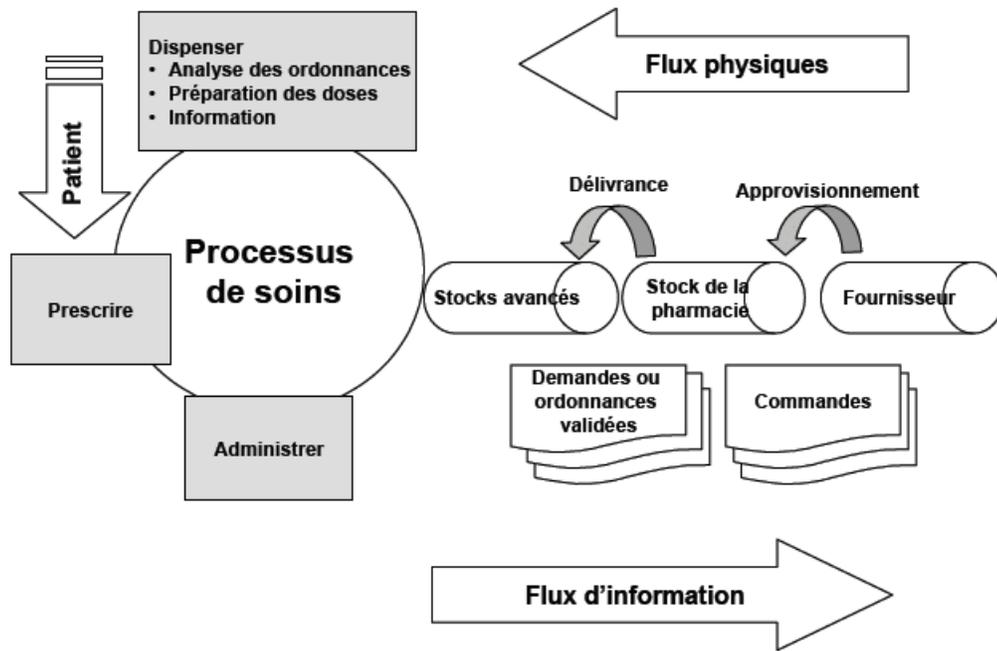


FIGURE 1.2 – Description de la chaîne logistique pharmaceutique (Source : Di Martinelly [85])

### 1.1.2 Spécificités et importance des activités de logistique hospitalière

Dans sa thèse, Taher Hassan [49] nous aide à établir une comparaison entre la logistique manufacturière et la logistique hospitalière comme indiqué dans la table de synthèse 1.3. La

comparaison a été effectuée sur la base de cinq critères : le cycle de vie du produit, les activités de base, les activités de soutien, les activités de transformation et les flux spécifiques. Il en ressort que, de manière similaire à l'industrie, les activités logistiques concernent les services de soutien (alimentation, linge, gestion des magasins) ainsi que les services de base (prestation de soins pour les hôpitaux) du centre hospitalier. Bien que la logistique hospitalière s'inspire de la logistique manufacturière, où les exigences clés sont d'ordre financier, elle met l'accent sur la sécurité, l'efficacité et la qualité des soins pour les patients. Cependant, étant donné les chiffres des différents rapports de l'OCDE [107, 108], qui indiquent des dépenses de plus en plus restreintes, l'optimisation des coûts liés à l'approvisionnement, au stockage et à la distribution des produits de santé doit être considérée comme l'une des principales exigences de la logistique hospitalière.

Définition		La logistique manufacturière	La logistique hospitalière
<b>Critères</b>			
<b>Cycle de vie de produit</b>	Etude de marché	✓	
	Conception	✓	
	Approvisionnement	✓	✓
	Production	✓	✓
	Distribution	✓	✓
	Exploitation	✓	
	Recyclage	✓	
<b>Les activités de base</b>	Conception de produit	✓	
	Achat	✓	✓
	Fabrication	✓	✓
	Emballage	✓	Conditionnement
	Transport	✓	✓
	Manutention	✓	✓
	Stockage	✓	✓
	Vente	✓	
Service après vente	✓		
<b>Les activités de soutien</b>	Etude de marché	✓	
	Informatique	✓	✓
	Planification	✓	✓
	Contrôle	✓	✓
	Finance	✓	✓
	Personnel	✓	✓
	Engineering	✓	✓
Maintenance	✓	✓	
<b>Les activités de transformation</b>	Analyse médicale		✓
	Cuisines	✓	✓
	Stérilisation		✓
	Déchets	✓	✓
	Méd. DMS, O. Pansements		✓
	Flux de patients		✓
<b>Les flux spécifiques</b>	Médicament dérivé de sang		✓
	Stupéfiant		✓
	Lingerie		✓

FIGURE 1.3 – Comparaison entre la logistique manufacturière et la logistique hospitalière [49]

La logistique hospitalière et la logistique manufacturière diffèrent principalement par la nature des produits manipulés et les conséquences des échecs logistiques. En effet, dans la logistique hospitalière, les produits sont souvent essentiels pour les soins des patients et peuvent avoir des conséquences critiques pour leur santé, voire entraîner des situations de mort de

patient. Il est donc crucial que l'équipement médical, les fournitures médicales et les médicaments soient livrés au bon moment pour garantir des résultats positifs pour les patients. En revanche, la logistique manufacturière concerne généralement des articles moins critiques tels que des matières premières, des produits finis ou de l'équipement qui peuvent subir des retards de production ou des interruptions de la chaîne d'approvisionnement. Toutefois, une logistique défaillante dans le secteur industriel, par exemple des retards à répétition, peut engendrer "*la mort*" d'une entreprise.

Par ailleurs, la logistique hospitalière requiert généralement une manipulation et un transport spécialisés en raison de la sensibilité des produits de santé. Par exemple, certains médicaments et fournitures médicales nécessitent des environnements à température contrôlée, tandis que d'autres doivent être manipulés avec précaution en raison de leur fragilité ou des risques de contamination. En revanche, la logistique manufacturière peut avoir des exigences différentes, telles que des machines lourdes ou un équipement surdimensionné qui nécessitent un transport spécialisé.

De plus, la réglementation de la logistique de la santé est généralement plus complexe que celle de la logistique manufacturière en raison de l'existence de réglementations et de directives strictes pour la manipulation et le transport de produits médicaux afin d'assurer la sécurité des patients et la conformité réglementaire. Il faut ajouter deux spécificités : d'une part, l'importance accordée à la sécurité avec l'hygiène et la traçabilité, mais aussi le fait de pouvoir fonctionner en mode dégradé. Le fonctionnement hospitalier est continu (24h/24 et 365 jours par an) et implique de pouvoir réaliser à tout moment l'approvisionnement des unités de soins ; d'autre part, le fonctionnement avec des moyens finis. Ces deux particularités font que les moyens logistiques en milieu hospitalier sont souvent surdimensionnés par rapport à ceux de l'industrie afin de conserver une marge de sécurité.

Il est important que les activités logistiques soient planifiées, organisées et gérées de manière efficace, pour assurer un bon fonctionnement des opérations de l'hôpital et une qualité des soins optimale. Cela nécessite la participation et la collaboration étroite de tous les acteurs impliqués, tant à l'intérieur qu'à l'extérieur de l'établissement de santé. Il est également important de noter que les activités logistiques en milieu hospitalier doivent être régulièrement évaluées et améliorées pour s'adapter aux changements dans les pratiques de soins, les technologies et les réglementations.

L'optimisation de la logistique hospitalière est considérée comme une opportunité significative pour améliorer la qualité des soins. En effet, les activités logistiques sont souvent effectuées par le personnel médical, ce qui peut empiéter sur le temps consacré aux soins aux patients. En libérant le personnel soignant de ces activités, il est possible d'améliorer la qualité des soins [160]. Arora et al. présentent les principaux avantages de la gestion de la chaîne logistique [9]. Ils citent les avantages suivants :

- Rationalisation du flux de travail entre les différentes équipes et personnes impliquées
- Avoir une gestion serrée des stocks, afin d'être économiquement viable et facilement disponible en cas de besoin.
- Réduire les défaillances et les pertes monétaires dues à des équipements défectueux, des médicaments périmés, etc.

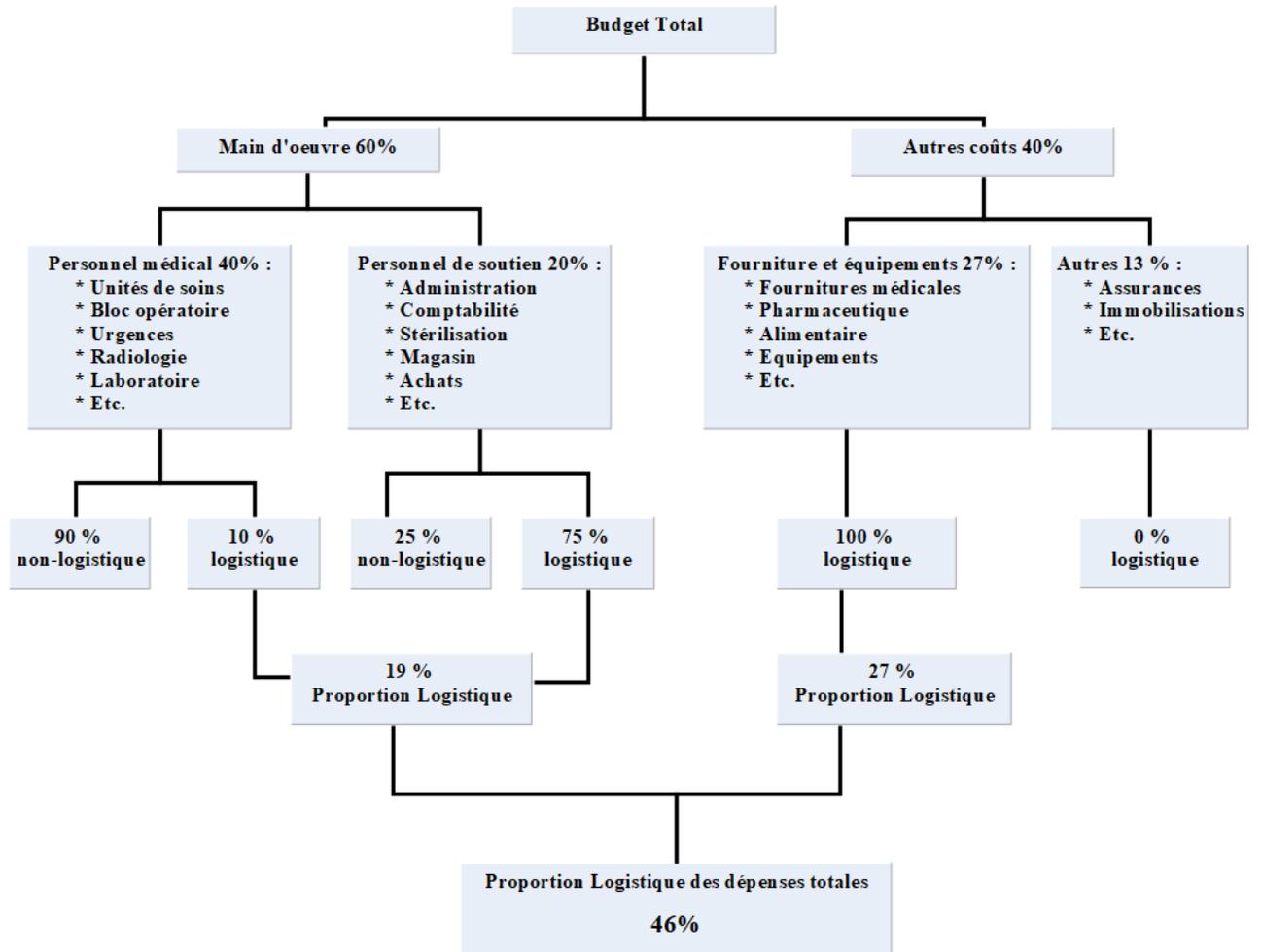


FIGURE 1.4 – Les proportions des coûts de la logistique hospitalière (Source : [72, 49])

- Optimiser le coût des équipements/médicaments en commandant en gros ou en pièces détachées en fonction de la tendance de la consommation, et améliorer la gestion des fournisseurs en numérisant toutes les communications et en assurant le suivi de la consommation.

L'importance de la logistique hospitalière est établie depuis les années 80 comme le rappellent les travaux de [133, 72, 49]. Dans leur travail, Landry et al [72], en reprenant la démarche de Housley [54], ont décomposé les différentes dépenses d'un hôpital en leur attribuant un pourcentage associé aux activités logistiques. Outre les achats de fournitures et d'équipements, 75% des dépenses salariales du personnel de soutien et, plus étonnant, 10% des dépenses salariales du personnel soignant sont consacrés à la logistique hospitalière (Voir schéma 1.4).

Selon les auteurs, cet exercice permet d'estimer le poids des activités logistiques à 46% des dépenses total effectuées par un centre hospitalier. Bourgeon et al. [22] dans leur étude intitulée « *Évaluation des coûts logistiques hospitaliers en France et aux Pays-Bas* », arrivent aux mêmes conclusions que les établissements de santé Nord américains. La recherche a été menée sur des échantillons d'hôpitaux en France et aux Pays-Bas, et les auteurs ont conclu que les coûts logistiques représentaient entre 31% et 34% du budget total de ces établissements.

Cette part considérable des coûts logistiques explique pourquoi les stratégies logistiques sont devenues des éléments clés dans la réduction des dépenses de santé au sein des hôpitaux. Les travaux de Nachtmann et al. de l'université d'Arkansas, viendront confirmer ces assertions en 2009 en indiquant que 31% des coûts de fonctionnement annuels des hôpitaux sont consacrés au soutien de la chaîne logistique [101]. Ce qui permet de dire que les dépenses liées aux activités logistiques sont comprises entre 30 à 46 %. Ce qui fait de la logistique hospitalière la deuxième composantes de dépenses après les dépenses de personnel. Les produits pharmaceutiques représentent 70 à 80% des coûts d'approvisionnement, tandis que le matériel médico-chirurgical en représente entre 20 et 25 % [128, 66]. Il existe donc un potentiel de gains intéressant. Cela constitue une motivation de cette recherche, surtout la quête d'un bon dimensionnement et gestion des stocks.

En réduisant les coûts logistiques, cela entraîne comme corollaire une amélioration des dépenses dans les hôpitaux sans pour autant affecter les soins apportés aux patients. Dans notre contexte, cette réduction passe par un bon dimensionnement et une gestion efficace des stocks de produits de santé. Mais qu'est-ce qu'un produit de santé et comment procède-t-on à son identification? De ce fait, la section suivante est dédiée au rappel de notions essentielles sur les produits de santé, en tenant compte de leur traçabilité, avant d'aborder le dimensionnement et la gestion des stocks de produits de santé, ainsi que les méthodes qui y sont associées.

## 1.2 Les produits de santé

Dans cette section, nous procédons à une description des produits de santé, ainsi qu'à l'examen des régulations et de la traçabilité associées.

### 1.2.1 Définitions et types de produits de santé

D'après Serre et al. [140], il n'existe pas de définition juridique des produits de santé. En revanche plusieurs approches peuvent être retenues pour cerner les produits entrant dans la catégorie sanitaire. Les auteurs soulignent cinq approches qui peuvent être regroupées en trois catégories :

- *selon la réglementation des produits par le Code de la santé publique*
- *selon les produits entrant dans le champ de compétence de l'Agence nationale de sécurité du médicament et des produits de santé (ANSM)*
- *selon les circuit de distribution.*

Les deux premières approches de définition des produits de santé, soulignés par Serre et al. [140], relèvent de la régulation et des dispositions du code de la santé publique. Le code de santé publique française, dans la cinquième partie en ses articles L5111-1 à L5542-2<sup>2</sup>

---

2. Légifrance : Service public de la diffusion du droit.

<https://www.legifrance.gouv.fr/codes/id/LEGISCTA000006125349/>

permet de définir les produits de santé, les dispositions générales y afférant, ainsi que l'organe chargé de la régulation et du contrôle. En son article L5311-1, il est énuméré les catégories de produits relevant des compétences de l'ANSM et on y distingue "*les produits à finalité humaine*" et "*les produits à finalité cosmétique*". Dans cette thèse, l'intérêt est porté sur les produits à finalité humaine.

Dans l'approche selon le circuit de distribution, trois types de circuits de distribution sont envisagés : l'officine pharmaceutique, les circuits spécialisés de distribution sélective (magasins spécialisés dans la « parapharmacie ») et la grande distribution. Et trois types de produits sont distingués :

- Les produits qui sont obligatoirement distribués en officine (champ du monopole officinal : article L. 4211-1 et arrêté du 15 février 2002 modifié le 19 juin 2021<sup>3</sup>);
- Les produits hors monopole qui peuvent être distribués en officine (Liste de l'arrêté du 15 février 2002 en soustraction à l'article L. 4211-1. Exemples : Les articles d'hygiène bucco-dentaire, Les produits cosmétiques, Les compléments alimentaires, etc.)
- Les autres produits.

Parmi les officines pharmaceutiques on y retrouve par exemple les pharmacies de ville, hospitalières, militaires, etc. Étant donné que nous analysons la consommation de produits de santé dans des hôpitaux, l'officine pharmaceutique dans notre contexte est la pharmacie hospitalière.

Relativement au code de santé publique française, les produits de santé sont classés en deux grandes catégories :

- Les produits pharmaceutiques ou médicaments,
- Les dispositifs comprenant les dispositifs médicaux, les dispositifs médicaux de diagnostic in vitro et autres produits et objets.

Et compte tenu des perspectives de cette thèse, nous nous focalisons sur les médicaments. A travers les deux sous-sections suivantes nous revenons, de façon synthétisée, sur les définitions des deux types de produits de santé.

### 1.2.1.1 Les produits pharmaceutiques

Selon l'Organisation mondiale de la santé (OMS), la définition de produit pharmaceutique se présente comme suit : « *Pharmaceutical products - also known as medicines or drugs - are special preparations used in modern and traditional medicine. They are essential for the prevention and treatment of diseases, and protection of public health.* »<sup>4</sup>

Une définition qui relate que le produit pharmaceutique et le médicament sont deux termes interchangeables. Une définition plus détaillée dans le contexte français du médicament est fournie dans les articles L5111-1 et L5111-2 du code de santé publique française. De ces articles, pour qualifier un produit de médicament deux critères sont essentiels :

---

3. Arrêté du 15 février 2002 fixant la liste des marchandises dont les pharmaciens peuvent faire le commerce dans leur officine <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000593784/>

4. OMS : <http://www.emro.who.int/health-topics/pharmaceutical-products/index.html>

- *Sa présentation* : Les éléments de présentation ne se limitent pas seulement à l'aspect matérielle (forme galénique, dosage, conditionnement) mais intègrent des critères comme la présence de conseils d'utilisation évoquant une indication thérapeutique, la posologie, la précaution d'emploi etc.
- *la fonction du produit* : Le fait que le produit soit administré à l'homme, l'objectif poursuivi et le mode d'action (pharmacologique, immunologique ou métabolique), ces éléments retenus permettent de décrire la fonction d'un médicament et mieux de faire un distinguo entre un médicament et un dispositif médical.

### 1.2.1.2 Les dispositifs médicaux

L'article L2511-1 du code de la santé publique définit le dispositif médical comme suit :

*« On entend par dispositif médical tout instrument, appareil, équipement, matière, produit, à l'exception des produits d'origine humaine, ou autre article utilisé seul ou en association, y compris les accessoires et logiciels nécessaires au bon fonctionnement de celui-ci, destiné par le fabricant à être utilisé chez l'homme à des fins médicales et dont l'action principale voulue n'est pas obtenue par des moyens pharmacologiques ou immunologiques ni par métabolisme, mais dont la fonction peut être assistée par de tels moyens. Constitue également un dispositif médical le logiciel destiné par le fabricant à être utilisé spécifiquement à des fins diagnostiques ou thérapeutiques. »*

*Les dispositifs médicaux qui sont conçus pour être implantés en totalité ou en partie dans le corps humain ou placés dans un orifice naturel, et qui dépendent pour leur bon fonctionnement d'une source d'énergie électrique ou de toute source d'énergie autre que celle qui est générée directement par le corps humain ou la pesanteur, sont dénommés dispositifs médicaux implantables actifs. »*

De cette définition, deux critères sont pris en compte pour définir le champ du dispositif médical à savoir la *destination voulue par le fabricant* et la manière dont est obtenue l'action principale. Deux critères qui vont compter dans la sérialisation de ce produit de santé.

Au delà des définitions des deux types de produits de santé, l'article L53311-1 du code de la santé publique fournit 21 groupes (dont 6 abrogés sur *Ordonnance n°2022-1086 du 29 juillet 2022 - art. 11*) de produits qui sont placés sous le contrôle de l'autorité chargée de la régulation. En exemple, nous avons

- Les médicaments, y compris les insecticides, acaricides et antiparasitaires à usage humain, les préparations magistrales, hospitalières et officinales, les substances stupéfiantes, psychotropes ou autres substances vénéneuses, les huiles essentielles et plantes médicinales, les matières premières à usage pharmaceutique ;
- Les produits contraceptifs et contragestifs ;
- Les biomatériaux et les dispositifs médicaux ;
- Les dispositifs médicaux de diagnostic in vitro ;
- Les produits sanguins labiles ;

- Les organes, tissus, cellules et produits d'origine humaine ou animale, y compris lorsqu'ils sont prélevés à l'occasion d'une intervention chirurgicale ;
- Les logiciels d'aide à la prescription et les logiciels d'aide à la dispensation.
- Etc.

Ces définitions ainsi parcourues vont faciliter la compréhension, l'exploration et l'analyse des données étant données que nous travaillons sur la consommation de médicaments dans les hôpitaux.

### 1.2.2 Régulations et sérialisations des produits de santé

Après avoir rappelé succinctement la situation de régulations des produits de santé, dans cette sous section nous revenons sur les notions de sérialisation (traçabilité) des deux types de produits de santé en France. Ces notions sont d'autant plus importantes, car elles nous permettent d'identifier les produits de santé dans chaque hôpital et de nous faciliter la fouille des données dans les chapitres de la deuxième partie de cette thèse.

#### 1.2.2.1 Régulations des produits de santé en France

En général, les produits de santé sont soumis à une régulation et à une réglementation strictes. En France, l'autorité chargée de ces questions est l'Agence Nationale de Sécurité du Médicament et des Produits de Santé (ANSM). C'est un établissement public de l'État placé sous la tutelle du ministre chargé de la santé. Elle a été créée par la loi du 29 décembre 2011, relative au renforcement de la sécurité sanitaire des médicaments et des produits de santé. L'ANSM s'est substituée le 1<sup>er</sup> mai 2012 à l'Agence française de sécurité sanitaire du médicament et des produits de santé (AFSSAPS) dont elle a repris les missions. Les missions assignées à cette institution figurent dans les articles L5311-1 à L5311-3 du code de la santé publique. Il convient de noter que parmi ces missions, l'agence est chargée du contrôle du respect des dispositions des autorisations délivrées en application de l'article L1161-5. Elle assure également la mise en œuvre des systèmes de vigilance, interpellation, de contrôle publicitaire. L'agence procède à l'évaluation des bénéfices et des risques liés à l'utilisation des produits à finalité sanitaire destinés à l'homme et des produits à finalité cosmétique. Elle surveille le risque lié à ces produits et effectue des réévaluations des bénéfices et des risques.

#### 1.2.2.2 Sérialisation des médicaments

Étant donné que nous nous concentrons sur les produits pharmaceutiques, nous abordons uniquement la sérialisation de ces produits dans cette sous-section.

La sérialisation, aussi appelée « *traçabilité à la boîte* » peut se définir comme étant un système de vérification de l'authenticité d'un produit (médicament ou dispositif médical) entre sa mise en distribution et sa dispensation effective à un patient.

Pour les médicaments, ce système consiste à apposer un identifiant unique sur chaque boîte de médicament sous forme de « *Datamatrix* », lors de sa mise en distribution et à charger l'ensemble des codes des boîtes dans une base de données centrale. Au moment de la dispensation, ce code sera scanné et comparé à ceux enregistrés dans la base de données centrale pour vérifier l'authenticité du médicament ou détecter s'il s'agit d'un médicament falsifié. De plus, chaque médicament sérialisé sera muni d'un dispositif anti-effraction. Les obligations liées à la sérialisation des médicaments et au dispositif anti-contrefaçon sont entrées en vigueur le samedi 9 février 2019, en France, conformément au règlement délégué 2016/161 de la Commission Européenne.

D'après le FMVO<sup>5</sup>, chaque boîte de médicament munie de dispositif soumise à prescription devra comporter les éléments suivants :

- Un dispositif d'inviolabilité
- Des informations lisibles en clair constitués de :
  - Le code du produits
  - le numéro de lot
  - La date de péremption
  - Le numéro de série
- Un identifiant unique encodé contenant le numéro de série (sérialisation) et encodé sous la forme d'un datamatrix (comme le présente la figure 1.5).



FIGURE 1.5 – Exemple sérialisation de médicament (source FMVO)

Après avoir scanné le code Datamatrix, on vérifie la concordance des informations contenues dans ce code avec celles de la base de données nationale (NMVS), qui est inter-connectée à une base de données européenne (EMVS) :

- Si les informations concordent, l'identifiant unique est désactivé dans la base de données, et la boîte peut être dispensée.

5. FMVO (France Medicines Verification Organisation) est l'organisme national de gouvernance de la sérialisation pour la France. [https://www.france-mvo.fr/wp-content/uploads/2019/09/Fiche-Technique-PUI\\_n%C2%B01\\_V2.pdf](https://www.france-mvo.fr/wp-content/uploads/2019/09/Fiche-Technique-PUI_n%C2%B01_V2.pdf)

- Si les informations ne concordent pas ou si l'identifiant unique a déjà été désactivé, une alerte est générée et automatiquement transmise à la FMVO et au titulaire de l'AMM (Autorisation de Mise sur le Marché) en vue d'une enquête. En cas de falsification avérée, les autorités de santé sont informées.

### **Traçabilité des médicaments :**

Pour les médicaments à usage humain, il existe plusieurs codes pour l'identification et de présentation. Les trois codes principaux les plus couramment utilisés sont les suivants :

- Anatomical Therapeutic Chemical (ATC),
- Unité Commune de Dispensation (UCD),
- Code Identifiant de Présentation (CIP).

**Le code ATC** est un système de classification internationale utilisé pour classer les médicaments en fonction de leur action pharmacologique, de leurs indications thérapeutiques et de leur utilisation anatomique spécifique dans le corps humain. Il est développé et géré par OMS en collaboration avec l'Association internationale de la sécurité sociale (AISS) et est largement utilisé dans le domaine de la pharmacie et de la santé publique. Le code ATC se compose de cinq niveaux hiérarchiques qui permettent de classer les médicaments en groupes et sous-groupes en fonction de leur mode d'action, de leurs indications thérapeutiques et de leur anatomie cible. Les 5 différents niveaux de hiérarchies sont :

- 1<sup>er</sup> niveau : classe anatomique principale qui indique le système anatomique cible du médicament.
- 2<sup>ème</sup> niveau : sous-classe thérapeutique qui indique le groupe thérapeutique auquel appartient le médicament.
- 3<sup>ème</sup> niveau : sous-classe pharmacologique
- 4<sup>ème</sup> niveau : sous-classe chimique qui indique une catégorie plus spécifique.
- 5<sup>ème</sup> niveau : substance active qui représente le médicament spécifique

En exemple, si on prend le PARACETAMOL dont le code ATC est N02BE01, les niveaux de hiérarchies sont les suivantes :

- **Classe anatomique principale** : N : *Système nerveux*
- **Sous-classe thérapeutique** : N02 : *Analgésiques*
- **Sous-classe pharmacologique** : N02B : *Autres analgésiques et antipyrétiques*
- **Sous-classe chimique** : N02BE : *Anilides*
- **Principe actif** : N02BE01 : *Paracétamol*

Dans notre exemple, ce code signifie que le paracétamol est un analgésique et un antipyrétique non opiacé.

Un lien est établi entre le code ATC et le principe actif. Le code ATC est attribué à un médicament spécifique en fonction de son principe actif et de son utilisation thérapeutique. Il

permet de regrouper différents médicaments contenant le même principe actif dans la même classe thérapeutique, facilitant ainsi la recherche et la comparaison des médicaments ayant des actions pharmacologiques similaires. Par exemple, si plusieurs médicaments contiennent le principe actif "paracétamol", ils seront tous classés dans la même classe thérapeutique du code ATC (groupe N02BE01) pour les analgésiques et antipyrétiques. Cela permet aux professionnels de la santé de savoir qu'ils appartiennent à la même catégorie de médicaments, même s'ils ont des noms commerciaux différents.

**Le code UCD**, fourni par l'association dénommée Club Inter Pharmaceutique sous le couvert de l'ANSM en France, identifie de manière unique un médicament. Il joue un rôle essentiel dans les systèmes d'information et les échanges entre clients et fournisseurs, contribuant ainsi à la fluidité et à la fiabilité des échanges interprofessionnels. De plus, le code UCD est utilisé pour identifier et tracer les spécialités pharmaceutiques rétrocédées par les pharmacies hospitalières et celles facturées en sus de la tarification à l'activité (T2A). En d'autres termes, il s'agit d'une codification utilisée par les établissements de soins pour suivre les spécialités dispensées et faciliter la facturation des médicaments hors GHS<sup>6</sup>. Le code UCD existe sous deux formes, UCD 13 et UCD 7, avec le premier incluant le deuxième (voir figure 1.6).

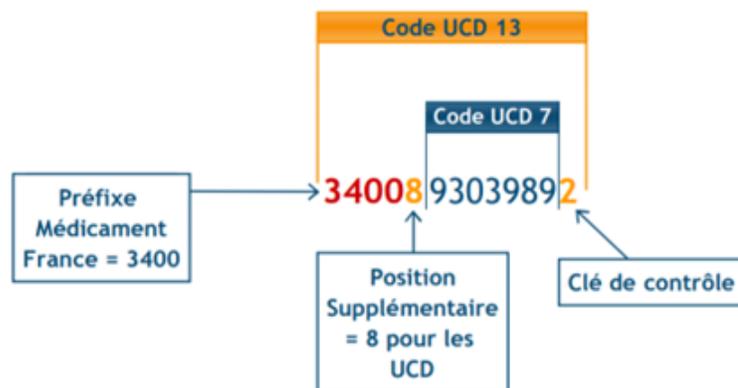


FIGURE 1.6 – Exemple Code UCD (source : Club Inter Pharmaceutique [117])

En lien avec le code ATC, un code ATC contient en général plusieurs codes UCD. La table 1.1 permet d'illustrer cet aspect avec l'exemple du Paracétamol. Cette table présente une sélection de médicaments avec différents codes UCD et des noms différents, mais tous ayant le même code ATC. Les deux versions des codes UCD (13 et 7 chiffres) sont mentionnées.

Cet extrait de tableau (table 1.1) est issue de la base de données de PHAST<sup>7</sup>.

6. Le Groupe homogène de séjours (GHS) est un système de classification utilisé dans la tarification à l'activité des hôpitaux en France pour regrouper les séjours hospitaliers en fonction de la pathologie, de la complexité des soins et des ressources utilisées.

7. PHAST organise en son sein la communauté Système d'Information Pharmaceutique dédiée notamment à l'interopérabilité des applications sur le circuit du médicament à l'hôpital avec trois référentiels : médicaments, dispositifs médicaux et biologie médicale. Composés d'un ensemble de données informatique, ils décrivent toutes les spécialités pharmaceutiques commercialisées en France bénéficiant d'une codification

UCD	Libellé
<b>3400891705256</b> 9170525	CLARADOL 500 mg, cpr séc
<b>3400892488073</b> 9248807	DAFALGAN 1 000 mg, cpr
<b>3400893545737</b> 9354573	DAFALGAN 1 000 mg, cpr efferv
<b>3400890000109</b> 9000010	DAFALGAN 1 000 mg, gélule
<b>3400893561065</b> 9356106	DAFALGAN 500 mg, cpr
<b>3400893545966</b> 9354596	DAFALGAN 500 mg, cpr efferv séc
<b>3400890238243</b> 9023824	DAFALGAN 500 mg, gélule
<b>3400891044225</b> 9104422	DAFALGAN 600 mg, suppo
<b>3400894570202</b> 9457020	DAFALGANCAPS 1 000 mg, gélule
<b>3400890000024</b> 9000002	DAFALGANCAPS 500 mg, gélule
<b>3400893556382</b> 9355638	DAFALGANHOP 1 000 mg, cpr efferv
<b>3400894570370</b> 9457037	DAFALGANTABS 1 000 mg, cpr
<b>3400891297706</b> 9129770	DOLIPRANE 1 000 mg, ad, suppo
<b>3400892390918</b> 9239091	DOLIPRANE 1 000 mg, cpr

TABLE 1.1 – Extrait de codes UCD en lien avec le code ATC N02BE01

**Le code CIP** est pour les pharmacies d'officines ce que le code UCD est pour les pharmacies hospitalières. Le code est un identifiant unique attribué à chaque médicament commercialisé en France. Il s'agit d'un identifiant unique attribué à chaque médicament commercialisé en France. Ce code est utilisé pour spécifiquement identifier un médicament donné et est généralement utilisé dans le cadre de la gestion des stocks, de la facturation et de la dispensation des médicaments. Le code CIP fait partie des codes proposés par le club inter-pharmaceutique correspondant à l'AMM d'une présentation d'un médicament utilisé dans les normes d'échanges entre les officines et l'assurance maladie. Comme le code UCD, le code CIP est un code numérique à 7 ou 13 chiffres qui permet d'identifier une présentation (ou conditionnement) d'un médicament (voir la Figure 1.7). Un code CIP identifie une présentation et une seule, définie par les éléments suivants :

- Sa dénomination (nom de marque + laboratoire)
- Sa forme galénique

sous le couvert de l'AMM et l'ATU. <http://www.phast.fr/>

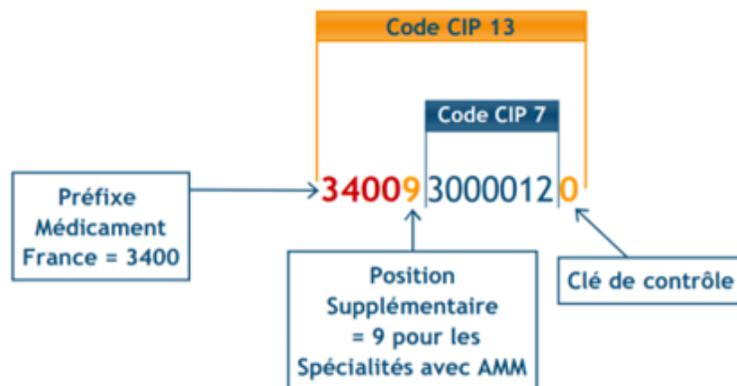


FIGURE 1.7 – Exemple de codification CIP (source : Club Inter Pharmaceutique [117])

- Son dosage
- Son conditionnement
- La contenance du conditionnement.

Tout comme il existe plusieurs codes UCD pour un code ATC, il est à noter qu'un code UCD regroupent 1 ou plusieurs codes CIP relativement aux conditionnements. La table 1.2 fournit des exemples de lien entre les trois codes de traçabilité des médicaments.

Code ATC	Code UCD	Libellé	Code CIP	Libellé conditionnement
PARACETAMOL N02BE01	<b>3400892488073</b> 9248807	DAFALGAN 1000 mg, cpr	3400930170441	Tube(s) polypropylène de 8 comprimés
			3400936158832	Plaquette(s) thermoformée(s) PVC-Aluminium de 8 comprimés
			3400956411528	Plaquette(s) thermoformée(s) PVC-Aluminium de 100 comprimés
			3400958341069	Plaquette(s) thermoformée(s) PVC-Aluminium de 40 comprimés
	<b>3400892390918</b> 9239091	DOLIPRANE 1000 mg, cpr	3400935955838	Plaquette(s) thermoformée(s) PVC-aluminium de 8 comprimés
			3400956369553	Plaquette(s) thermoformée(s) PVC-aluminium de 100 comprimés

TABLE 1.2 – Exemples entres les trois codes CIP, UCD et ATC (Source PHAST)

Du point de vue logistique, il peut être intéressant d'apprécier les besoins de la structure

sanitaire en médicaments par les codes CIP. Les trois codes de traçabilité faciliteront l'identification des médicaments, la fouille et l'analyse des données.

La traçabilité des produits de pharmaceutique nous est indispensable pour des identifications uniques quelque soit l'établissement de santé et cela, eu égard au jeu de données à notre disposition (jeux de données qui seront explicité dans le chapitre 3 dans la seconde partie de ce manuscrit).

## 1.3 Dimensionnement et gestion de stocks des produits de santé

En logistique hospitalière, le stock fait référence aux produits de santé tels que les médicaments, les dispositifs médicaux, qui sont entreposés dans les installations de stockage de l'hôpital ou de l'établissement de santé en vue d'une utilisation ultérieure. Le stock en logistique hospitalière doit être géré de manière à assurer une disponibilité constante des produits de santé pour les patients tout en évitant les coûts excessifs de stockage et les risques de perte due à des produits périmés. Cela implique des processus de gestion des stocks tels que la planification et estimation de la demande, l'analyse de la consommation, la gestion de la rotation des stocks, la mise en place de méthodes de réapprovisionnement appropriées, la gestion des niveaux de stock de sécurité et la gestion des produits périmés. Les fonctions et catégorisation des stocks s'appliquent également dans le milieu hospitalier. Cependant, les hôpitaux mettent l'accent sur la fonction de sécurité qui garantit que les articles critiques sont toujours disponibles en quantité suffisante pour répondre aux besoins des patients, et la fonction de gestion de la date de péremption, qui consiste à s'assurer que les articles ne sont pas utilisés après leur date d'expiration. Et donc tout ceci nécessite une bonne gestion et un dimensionnement suffisant des stocks.

### 1.3.1 Dimensionnement et gestion des stocks

L'ASCM (Association for Supply Chain Management), dans leur dictionnaire définit la gestion des stocks comme « *The branch of business management concerned with planning and controlling inventories* ». John Toomey, dans son livre [156], en s'appuyant à cette définition ajoute que le rôle de la gestion des stocks est de maintenir un niveau de stock souhaité pour des produits ou des articles spécifiques. Les systèmes qui planifient et contrôlent les stocks doivent être basés sur le produit, le client et le processus (fabriqué ou acheté) qui rend le produit disponible. Le coût du maintien des stocks tout au long du processus est un coût caché, mais il fait néanmoins partie du coût du produit.

Les principales questions couvertes par la gestion des stocks dans la littérature sont : Que va-t-on stocker ? Faut-il stocker ? Quand et combien stocker ? [118, 146].

- Quoi stocker ? Il s'agit de déterminer quels articles ou matières premières doivent être stockées en fonction des besoins de l'entreprise et des demandes des clients. Cette

décision peut être basée sur des considérations telles que la demande prévue, le temps de cycle de production, les coûts de stockage et de commande, la saisonnalité, etc.

- Faut-il stocker ? Il s’agit de décider si le stockage est la meilleure option pour répondre aux besoins de l’entreprise (l’hôpital) et des clients (patients).
- Quand et combien stocker ? Il s’agit de déterminer les niveaux de stock optimaux pour chaque article ou matière première en fonction de la demande prévue, du temps de cycle de production, des coûts de stockage et de commandes, etc.

Le dimensionnement des stocks et la gestion des stocks sont deux concepts étroitement liés et essentiels pour les entreprises, y compris les hôpitaux et les pharmacies. Parmi les trois questions, celle qui renvoie au dimensionnement des stocks est celle formulée comme suit : « Quand et combien stocker ? ». Cependant, une clarification est nécessaire car cette question en englobe en réalité deux. La première partie, le « *quand* », c’est-à-dire le moment optimal pour stocker les produits en tenant compte des délais de livraison, des besoins et des coûts de stockage, relève davantage de la gestion des stocks. Il s’agit de déterminer le timing optimal pour passer une commande. Le dimensionnement des stocks concerne surtout le « *combien* », c’est-à-dire la quantité de produits à stocker pour répondre à la demande. Il vise à déterminer la quantité optimale de produits à maintenir en stock. La question « Faut-il stocker ? » peut également être liée au dimensionnement des stocks, car elle contribue à déterminer si un article doit être stocké ou non, ce qui, à son tour, aide à établir la quantité optimale de produits à conserver en stock.

Dans la littérature, diverses variantes du problème de dimensionnement des stocks ont été étudiées. Par exemple, des études ont examiné l’augmentation ou la réduction de la capacité de l’entrepôt en allouant un espace supplémentaire, que ce soit de manière statique ou dynamique [142, 74, 124, 129], ainsi que le nombre de types de produits à stocker [39, 162]. Ces variations ont été explorées pour trouver des solutions optimales permettant d’optimiser l’utilisation de l’espace de stockage, de minimiser les coûts de détention de stocks et d’améliorer la gestion globale des stocks dans différents contextes industriels. Partant de ces travaux, l’on tire que le dimensionnement des stocks est une étape cruciale qui précède la mise en place de stratégies et de pratiques concrètes de gestion des stocks. Il fournit les paramètres de base pour la planification et la gestion des stocks, permettant ainsi une prise de décision éclairée et une mise en œuvre efficace des politiques de gestion des stocks pour répondre aux besoins. Par exemple, avant de commencer à gérer les stocks, il est essentiel d’avoir une vision claire des prévisions de la demande future. Le dimensionnement des stocks prend en compte ces prévisions pour déterminer les niveaux optimaux de stock à maintenir : il s’agit de la planification et la prévision. C’est d’ailleurs l’aspect prévisionnel du dimensionnement des stocks qui nous intéresse.

Dans les hôpitaux et les pharmacies, le dimensionnement des stocks de produits de santé est crucial pour garantir que ces derniers soient disponibles en quantité suffisante pour répondre aux besoins des patients, tout en minimisant les coûts de stockage et de commande ainsi qu’en optimisant de l’espace de stockage. La gestion des stocks est également importante pour s’assurer que les produits sont stockés correctement et que les niveaux de stock sont surveillés et gérés efficacement.

### 1.3.2 Méthodes et techniques du dimensionnement et gestion des stocks

La littérature est très riche en ce qui concerne les méthodes et techniques du dimensionnement et gestion des stocks. Nous choisissons, sans être exhaustif, d'orienter ces méthodes en fonctions des trois questions émises : « Que va-t-on stocker ? », « Faut-il stocker ? », « Quand et combien stocker ? ». Dans la table 1.3, nous présentons des méthodes les plus répandues chez les praticiens et dans la littérature.

Méthodes/Techniques	Que va-t-on stocker ?	Faut-il stocker ?	Quand et combien stocker ?	Références
FIFO	✓		✓	[156]
FILO	✓		✓	[156]
MRP	✓	✓	✓	[147]
Six Sigma		✓		[25]
Lean Six Sigma		✓		[84, 25]
Principe de Pareto	✓	✓		[32, 126]
Méthodes ABC	✓	✓		[147, 126, 114]
Méthode ABC/XYZ	✓	✓	✓	[114, 152]
Vendor-managed inventory (VMI)	✓	✓		[82]
L'inventaire	✓	✓	✓	[147]
Évaluation du risque de péremption ou d'obsolescence	✓	✓		[90]
Méthode du coût économique total		✓	✓	[147, 56, 156]
Modèle de Wilson		✓	✓	[147, 156]
Modèles de réapprovisionnement		✓	✓	[147]
Juste à Temps		✓	✓	[147]
Méthodes de la prévision de la demande		✓	✓	[97, 48, 57, 8, 147, 161]

TABLE 1.3 – Méthodes de gestion des stocks en fonction des questions posées

Notons que ce classement n'est pas exhaustif et peut varier selon les interprétations et les utilisations des différentes méthodes. Bien évidemment, dans le tableau (Table 1.3), il y a un chevauchement entre les différentes méthodes, et certains aspects de la gestion des stocks peuvent être abordés par plusieurs méthodes. Cependant, nous avons choisi de les classer de cette manière pour donner une idée générale de leur objectif principal. Les méthodes FIFO, FILO, ABC, ABC/XYZ, et les méthodes d'évaluation du risque de péremption ou d'obsolescence par exemple, visent à déterminer quels articles ou matières premières doivent être stockées en priorité en fonction de leur importance ou de leur valeur. Les méthodes comme les modèles de réapprovisionnement, juste à temps, vendor-managed inventory permettent

de savoir s'il est nécessaire de stocker des articles ou s'il est possible de les commander ou de les produire à la demande. Quand aux méthodes comme le MRP, l'inventaire, l'analyse des coûts de rupture de stock, Analyse des coûts de passation de commande, Méthode du coût économique total, Modèle de Wilson, Méthodes de la prévision de la demande permettent de déterminer le moment optimal pour commander ou produire les articles en fonction de la demande et de minimiser les coûts liés à la gestion des stocks.

En tenant compte de la problématique que nous abordons, ce sont les méthodes de prévision de la demande qui retiennent notre attention dans cette thèse, puisque notre objectif est d'estimer avec précision les besoins quantitatifs en médicaments des hôpitaux en se basant sur leurs caractéristiques. Des précisions seront apportées dans le chapitre suivant concernant ces méthodes de prévision de la demande, lesquelles seront centrées sur les produits de santé.

## 1.4 Conclusion

Dans ce chapitre, nous avons abordé plusieurs aspects. Tout d'abord, nous avons introduit le concept de logistique hospitalière, en commençant par une revue des définitions ainsi que des spécificités et de l'importance de la logistique hospitalière. Cette première partie nous a permis de poser le cadre général de notre étude.

Ensuite, nous avons approfondi notre compréhension des produits de santé en abordant leur traçabilité, ce qui est essentiel pour définir clairement l'objet de notre étude : les médicaments. Nous avons examiné les notions fondamentales liées à ces produits de santé, ce qui a renforcé notre contextualisation.

Dans un troisième volet, nous avons revisité des notions importantes relatives au dimensionnement et à la gestion des stocks, ainsi que les techniques associées, ce qui est le domaine particulier de nos travaux. Nous avons mis en lumière ces concepts pour mieux cerner le champ d'application de notre recherche.

Parmi ces concepts, ce sont spécifiquement les méthodes de prévision de la demande qui ont retenu notre attention dans cette thèse, étant donné leur pertinence par rapport à la problématique que nous abordons. Nous allons fournir des détails plus approfondis sur les méthodes de prévision de la demande des produits de santé dans les hôpitaux dans le chapitre suivant.

# Chapitre 2

## Prévision de la demande des produits pharmaceutiques

### Sommaire

---

<b>2.1 Méthodes classiques pour l'estimation des besoins en produits de santé</b> . . . . .	<b>28</b>
2.1.1 Méthode de consommation . . . . .	29
2.1.2 Méthode de morbidité . . . . .	30
2.1.3 Méthodes de consommation par procuration . . . . .	30
2.1.4 Projection du besoin budgétaire au niveau du service . . . . .	31
2.1.5 Méthode de la demande des hôpitaux . . . . .	31
2.1.6 Méthode basée sur la population . . . . .	32
2.1.7 Comparaison des méthodes classiques de quantification . . . . .	32
<b>2.2 Méthodes statistiques de prévision de la demande</b> . . . . .	<b>36</b>
2.2.1 Méthodes quantitatives de prévision de la demande . . . . .	38
2.2.2 Machine Learning pour la prévision de la demande . . . . .	39
<b>2.3 Conclusion</b> . . . . .	<b>47</b>

---

Une fois avoir centré le domaine et le type d'approche qui nous intéresse dans le chapitre précédent, dans ce chapitre, nous faisons le point des méthodes de prévision de la demande des produits de santé dans les hôpitaux. Dans un premier temps, nous abordons les méthodes classiques pour l'estimation des besoins en produits de santé. Une comparaison de ces méthodes classiques de quantification est également fournie. En second lieu, nous revenons également sur les méthodes statistiques de prévision de la demande, des approches qui vont nous intéresser pour la suite de nos travaux.

## 2.1 Méthodes classiques pour l'estimation des besoins en produits de santé

Dans les hôpitaux, la mise en place effective d'un plan d'action pour la consommation de produits de santé nécessite un bon système de gestion des approvisionnements. Ce système est composé de la sélection, de la quantification, de l'approvisionnement, de la distribution et de l'utilisation [109, 33].

La sélection consiste à déterminer quels sont les produits nécessaires. C'est un processus qui commence par la définition d'une liste de maladies courantes pour chaque niveau de soins. Elle a un impact considérable sur la qualité des soins et le coût des traitements, ce qui en fait l'un des piliers d'un système de santé efficace. La quantification est le processus qui consiste à estimer les quantités et les coûts des produits de santé nécessaires pour une période donnée, ainsi qu'à déterminer quand les livraisons de ces produits doivent être effectuées pour assurer un approvisionnement optimal et ininterrompu. L'approvisionnement est le processus qui consiste à choisir les fournisseurs, à passer et à suivre les commandes, à contrôler la quantité et la qualité des produits de santé et à payer les fournisseurs. Ensuite, vient la distribution qui comprend la réception, le stockage, l'inventaire, le transport et la tenue des registres, c'est-à-dire la collecte et le contrôle des informations. La prescription, la délivrance et l'utilisation des produits de santé ; l'observance du patient concerne la phase d'utilisation du système.

Les questions soulevées par le processus de quantification concernent le dimensionnement et la gestion des stocks. Dans le cadre de notre recherche, nous nous intéressons particulièrement à ce processus. Même si certaines des méthodes qui en découlent incluent la sélection, elles sont basées sur le principe que les produits de santé doivent d'abord être déterminés par l'institution ou l'organisation sanitaire concernée.

Selon l'OMS, l'efficacité de la quantification dépend de la qualité des informations et des ressources disponibles, ainsi que l'exactitude des données sur la morbidité et la consommation antérieure de produits de santé [109]. Selon l'organisation *Management Science Health (MSH)*, la quantification n'est pas un simple calcul, c'est la première étape du processus d'approvisionnement. La quantification comporte de nombreux aspects tels que : Le calcul des quantités estimées des commandes, des coûts et des dates de livraison des envois ; La planification, la mobilisation et la sécurisation des ressources financières ; L'estimation des besoins de stockage ; L'évaluation de l'utilisation rationnelle des produits ; La facilitation de la coordination des achats et de la logistique avec les donateurs, les fournisseurs, les établissements de santé et les autres parties prenantes ; L'information des fabricants et des fournisseurs sur la demande future pour la fabrication, les achats et les décisions de gestion logistique, etc.

Le bon sens ne permettant pas de faire face à certaines complexités de la quantification, des méthodes, recommandées par l'OMS, ont été développées depuis les années 1980. Parmi ces méthodes, nous pouvons citer :

- Méthode de consommation (Consumption method)
- Méthode de morbidité (Morbidity method)
- Méthodes de consommation par procuration (Proxy consumption methods)

- Projection du besoin budgétaire au niveau du service (Service-level projection of budget requirement)
- Méthode de la demande des hôpitaux (Hospital Request method)
- Méthode basée sur la population (Population-based method)

Les méthodes les plus rencontrées dans la littérature sont les deux premières de la liste [149, 109, 111, 38, 33, 35]. Les produits de santé peuvent être quantifiés en utilisant une ou une combinaison de ces méthodes. Nous allons brièvement expliquer chacune d'entre elles dans les paragraphes suivants en vue de les comparer (voir le tableau 2.1). De plus, l'une de ces méthodes fera l'objet d'une mise en œuvre et d'une comparaison dans le chapitre 5.

### 2.1.1 Méthode de consommation

Cette méthode est basée sur la consommation existante de produits de santé d'un établissement de santé. Une liste de tous les produits pouvant être commandés ou achetés est préparée, et un inventaire plus précis de la consommation passée, généralement sur une période récente de six à douze mois, est utilisé pour calculer les quantités nécessaires de chaque produit pour chaque établissement. Une analyse de la consommation de l'établissement est nécessaire et chaque fois qu'elle apparaît anormale pour un produit donné, une correction est effectuée en l'ajustant à la hausse ou à la baisse jusqu'à ce qu'elle atteigne un niveau adéquat. L'ajustement est déjà effectué en fonction des ruptures de stock afin d'obtenir la consommation mensuelle moyenne. Ensuite, la consommation mensuelle moyenne est multipliée par le nombre de mois à couvrir par l'achat. Les niveaux de stock de sécurité et les délais de livraison sont également multipliés par la consommation mensuelle prévue. Ces trois chiffres sont additionnés pour obtenir les besoins bruts de la période. On soustrait de l'estimation brute le stock utilisable en main et tout stock en commande pour en déduire la quantité à acheter. Un autre ajustement est ensuite effectué pour les pertes. Si le coût est un facteur dans la quantification, le coût unitaire prévu pour chaque produit est multiplié par le nombre d'unités à acheter pour obtenir la valeur d'achat prévue pour la quantité totale qui est une addition des valeurs d'achat de chaque produit sur la liste. Enfin, un ajustement final est effectué si ce coût est supérieur au budget. C'est pourquoi cette méthode est également appelée méthode de la consommation ajustée.

Des conditions préalables doivent être remplies pour l'utilisation de cette approche. Il faut disposer de données de consommation précises et fiables, disponibles ou pouvant être obtenues relativement, facilement. Il en va de même pour les listes de produits avec leurs conditionnements et leurs prix. Les pertes dues à la péremption, aux dommages et au vol ne doivent pas être excessives. De même, l'approvisionnement en produits dans les établissements "*types*" devrait être suffisant (dans la pratique, où il n'y a pas eu de pénurie de produits pendant plus de trois mois de l'année). Il faudrait également connaître les délais de livraison des fournisseurs et les budgets à allouer. Si cette méthode de quantification est inadéquate pour les nouveaux hôpitaux, elle est généralement préférable pour les programmes stables où le financement, la gestion pharmaceutique et la prescription sont raisonnablement

satisfaisants. Elle est également plus facile à utiliser dans des environnements tels que les hôpitaux où les problèmes de santé sont nombreux et les traitements sont complexes (Voir [109, 33] pour plus de détails).

### 2.1.2 Méthode de morbidité

En se basant sur le nombre de personnes souffrant d'une maladie donnée à un moment donné dans une population donnée, selon la prévalence ou l'incidence, la méthode dite de morbidité a été établie. Cette méthode utilise les données relatives à la fréquentation des établissements de santé et à la morbidité des patients pour prédire les besoins en produits de santé en se fondant sur des hypothèses quant à la manière dont les problèmes seront traités. En ce sens, elle nécessite des ensembles de données : une liste de problèmes de santé courants, une liste de produits de santé comprenant la thérapie pour les problèmes, et un ensemble de traitements standard pour la quantification, soit sur la base de régimes de traitement standard convenus pour chaque problème de santé défini, soit sur la base de la pratique actuelle.

Bien que très exigeante en termes de données, cette méthode est généralement préférable pour les services nouveaux ou en évolution rapide ou lorsque les services sont radicalement réorganisés. Elle est également préférable si les pratiques de prescription sont coûteuses et irrationnelles, car elle fournit une base systématique pour l'amélioration. Pour mener à bien son application et obtenir l'estimation la plus précise de la quantité de produits et de leurs coûts, MSH présente cette méthode en 12 étapes allant de la spécification des listes à la comparaison des coûts et du budget en passant par les corrections (Voir [109, 33] pour plus de détails).

### 2.1.3 Méthodes de consommation par procuration

Les approches précédentes nécessitent des données souvent très précises, et certains systèmes sont confrontés à des déficits d'information. De ce fait, elles ne sont pas réalisables, et la méthode de consommation par procuration se présente comme une alternative. Dans cette méthode, les données de consommation sont extrapolées à partir d'une base de données d'un autre système de santé ou d'une autre région présentant des similitudes. Pour cette raison, elle est utilisée pour les nouveaux établissements qui peuvent manquer de données. Cette méthode peut être basée sur deux approches. La première approche basée sur la population consiste à définir la consommation de produits pour 1 000 habitants et la seconde approche basée sur les services consiste à établir la consommation de produits par cas spécifiques ou par admission de patients hospitalisés. Il est également conseillé, pour une quantification complète, d'utiliser une combinaison des deux approches. Pour son application, 9 étapes sont établies pour une quantification réussie, depuis la sélection du système standard pour la comparaison et l'extrapolation, la définition des listes de produits jusqu'à l'estimation des coûts de chaque médicament et du total [33]. Il s'agit d'une méthode assez soucieuse des indicateurs et des facteurs de consommation. En outre, les installations standards doivent ressembler étroitement à la région ou au pays pour lequel l'estimation est faite, ce qui est

une condition principale pour obtenir une bonne prédiction. La ressemblance doit se faire en termes de géographie et de climat, de population de patients desservis, de schémas de morbidité, de pratiques de prescription, de directives de traitement standard, de listes de produits de santé et d'état de l'approvisionnement pharmaceutique (Voir [33] pour plus de détails).

#### 2.1.4 Projection du besoin budgétaire au niveau du service

La méthode décrite ici est généralement utilisée pour estimer les besoins financiers, et non les quantités spécifiques de produits de santé, pour l'achat de produits pharmaceutiques sur la base des coûts par patient traité à différents niveaux du même système de santé ou sur la base de données provenant d'autres systèmes de santé. Comme la méthode de consommation par procuration, cette méthode est une méthode d'extrapolation qui conduit à des estimations approximatives étant donné les grandes variations qui peuvent exister entre le système de santé cible et le système utilisé comme source de données standard. Les sources d'erreur rencontrées sont les suivantes : les prescripteurs du système cible utilisent une combinaison de produits de santé différente de celle du système source, la variabilité de la fréquence des maladies et des visites de patients par établissement, et les différences d'efficacité des systèmes d'approvisionnement et de gestion financière dans les deux contextes. Il n'anticipe pas les besoins spécifiques en produits de santé, mais fournit une justification claire et logique des besoins en financement pharmaceutique [33]. Pour son application, la principale exigence de cette méthode est une estimation assez fiable du coût moyen des produits de santé par patient et du nombre moyen de consultations à différents niveaux du système de santé standard. Les données qui doivent être compilées sont : le nombre moyen de visites ambulatoires curatives et non curatives et de journées-lits et/ou d'autres types de contact avec les patients pour chaque type d'établissement du système de santé d'origine ; et le coût moyen par visite ambulatoire curative et non curative et journée-lit et/ou autre type de contact avec les patients dans chaque type d'établissement du système de santé d'origine [33] pour plus de détails).

#### 2.1.5 Méthode de la demande des hôpitaux

Cette méthode est considérée comme l'une des plus anciennes approches pour quantifier et estimer les besoins. Il s'agit d'une méthode de calcul des besoins en produits de santé qui se base sur les demandes des hôpitaux pour les pharmacies nationales, régionales ou d'entrepôt. Le libre choix des produits de santé aux prescripteurs individuels pourrait être mentionné comme un avantage de cette méthode. Nous pouvons également ajouter la réduction des besoins de planification et la garantie qu'un ensemble de produits sera disponible, mais ils sont limités dans le type et la quantité de produits de santé et entraînent un gaspillage et des ruptures de stock importants. Cependant, les problèmes peuvent inclure : des demandes hospitalières pour plus de produits que nécessaire, des demandes hospitalières inadéquates (en tenant compte du rapport coût/thérapeutique) ou des produits de santé inutiles par des préférences personnelles ou l'influence des compagnies pharmaceutiques, le personnel de santé fait souvent des pratiques de prescription non optimales (polypharmacie, surprescription et

prescription inutile) [149, 110].

### 2.1.6 Méthode basée sur la population

La dernière méthode décrite dans ce document est la méthode basée sur la population, qui repose sur la prévalence de diverses maladies dans la population. La méthode basée sur la population dépend du besoin total, qui est différent de la demande. Les informations nécessaires comprennent des données démographiques et des données de suivi des maladies, par exemple le nombre de nouveaux patients, le nombre de patients qui poursuivent leur traitement, ceux qui doivent changer de traitement, ainsi que les taux de prévalence des maladies dans la population-cible. Une formule est requise pour chaque produit, ce qui peut être irréaliste pour les produits à usages multiples [110]. Cependant, comme on peut s'y attendre, cette méthode théoriquement intéressante n'est pas utilisée en pratique car les données requises sont très difficiles à obtenir, ce qui implique très peu de références dans la littérature [149]. Parmi ces données, nous pouvons noter les données socio-économiques telles que l'âge, le sexe, le statut marital, la co-morbidité, la position géographique des patients. Ces difficultés étaient encore plus prononcées dans les premières années d'application de l'approche, car les outils informatiques manquaient dans les structures de santé et les outils de données n'étaient pas encore à l'ordre du jour des politiques.

### 2.1.7 Comparaison des méthodes classiques de quantification

Sur la base de ces résultats, une comparaison a été faite dans ce document en mettant en évidence les avantages et les limites de chaque méthode résumée dans le tableau 2.1.

Parmi les six méthodes mises en évidence dans la littérature, les plus utilisées sont les méthodes de consommation, de morbidité et de consommation par procuration. Les raisons qui peuvent être avancées sont la disponibilité et la spécificité des données qui entrent dans le champ d'application de la quantification et les facteurs qui peuvent avoir un impact sur la consommation dans les établissements de santé.

La méthode considérée comme étant la plus précise pour quantifier les besoins des produits de santé est l'approche basée sur la consommation. Pour appliquer cette méthode, les données doivent être complètes, précises et correctement ajustées pour les périodes de rupture de stock et les changements anticipés dans la demande et l'utilisation. Toutefois, le principal problème de cette méthode est qu'elle ne tient normalement pas compte de la pertinence des habitudes de consommation passées, qui peuvent ou non correspondre aux priorités et aux besoins de santé publique en matière de morbidité. Cela peut conduire à une consommation irrationnelle de produits. Si les ruptures de stock ont été généralisées pendant de longues périodes, il peut être impossible d'appliquer cette méthode avec précision, c'est pourquoi il est essentiel de saisir la demande réelle pour obtenir l'approche la plus précise.

L'estimation basée sur la morbidité est la méthode la plus complexe et la plus longue à appliquer, selon ce qui est dit dans la littérature et le type de données saisies. En effet, dans de nombreux pays, il est très difficile de collecter des données de morbidité valides pour plusieurs maladies ; par conséquent, certains besoins seront négligés dans la quantification. Les données

sur les consultations des patients sont souvent incomplètes et imprécises, et il est difficile de prévoir quel pourcentage de prescripteurs suivra réellement les schémas thérapeutiques standard utilisés pour la quantification. Malgré ces contraintes, cette méthode peut encore être la meilleure alternative pour la planification des achats ou pour l'estimation des besoins budgétaires dans un système ou un établissement de santé. La méthode est également souvent utilisée pour estimer le coût des produits dans un grand nombre d'établissements, comme un petit système de soins primaires, un hôpital spécialisé ou un nouvel hôpital.

La consommation par procuration est généralement utilisée lorsque ni la méthode basée sur la consommation ni celle basée sur la morbidité ne sont réalisables. Cette méthode est plus susceptible de fournir des projections précises lorsqu'elle est utilisée pour extrapoler un ensemble d'installations à partir d'un autre ensemble d'installations desservant le même type de population dans le même type d'environnement géographique et climatique. Si la méthode est appliquée en utilisant des données standards d'un autre pays, les résultats ne seront que des estimations approximatives. Même lorsque la cible et les établissements de santé sont similaires, il est très difficile de supposer que l'incidence des maladies, les modes d'utilisation et les habitudes de prescription seront essentiellement les mêmes dans les deux cas. Néanmoins, cette méthode peut être la meilleure alternative en l'absence de données appropriées requises pour la méthode basée sur la consommation ou la morbidité. La méthode de la consommation de substitution est également utile pour recouper les projections faites avec d'autres méthodes. Toutes les méthodes rapportées sont basées sur un certain nombre de facteurs et d'indicateurs liés à la consommation de produits. Par conséquent, la détermination des facteurs qui influencent la consommation devrait contribuer à la fois à un choix judicieux de la méthode et à la quantification la plus précise.

	<b>Avantages</b>	<b>Limites</b>
<b>Méthode de consommation</b>	<ul style="list-style-type: none"> <li>— Les données détaillées sur la morbidité et les régimes de traitement standard ne sont pas nécessaires.</li> <li>— Il nécessite des calculs moins détaillés.</li> <li>— Elle est utile pour les institutions telles que les hôpitaux, où les problèmes sont nombreux et les traitements médicamenteux complexes.</li> <li>— Elle est fiable si la consommation est bien enregistrée et stable et ne risque pas de différer de manière significative de l'approvisionnement actuel.</li> <li>— Elle identifie les problèmes de gestion des stocks et encourage les améliorations.</li> <li>— Méthode facile, ne nécessitant qu'une étude par quelques personnes.</li> <li>— Améliore la disponibilité des produits essentiels et améliore le rapport coût-efficacité</li> </ul>	<ul style="list-style-type: none"> <li>— Doit disposer de données de consommation précises.</li> <li>— Il peut être difficile d'obtenir des données fiables sur la consommation pharmaceutique, en particulier dans les services nouveaux ou en évolution rapide.</li> <li>— Elle ne fournit pas une base détaillée ou systématique pour examiner l'utilisation des produits de santé et améliorer la prescription ; si le modèle de prescription est insatisfaisant et n'est pas corrigé, cette méthode risque de le perpétuer.</li> <li>— Elle n'est pas fiable s'il y a eu de longues ruptures de stock (plus de trois mois) ou des pertes ou gaspillages importants de produits de santé.</li> <li>— Elle n'est pas propice à un bon enregistrement de la morbidité.</li> <li>— Il n'interagit pas avec le système d'information existant Les pénuries de produits de santé précédentes peuvent entraîner des résultats biaisés.</li> <li>— Les pratiques de prescription existantes (non optimales) sont acceptées.</li> <li>— Changements dans la demande et l'utilisation, y compris l'utilisation irrationnelle des produits de santé.</li> </ul>
Continue à la page suivante		

Table 2.1 – suite de la table précédente

	Avantages	Limites
<b>Méthode de morbidité</b>	<ul style="list-style-type: none"> <li>— Il n'est pas nécessaire de disposer de données sur la consommation de produits de santé; la méthode peut être utilisée pour les nouveaux services qui ne disposent pas de ces données.</li> <li>— Basée sur un système de prescription rationnelle, elle fournit une base systématique pour examiner la consommation et la prescription de produits de santé, en particulier au niveau des soins primaires où les traitements médicamenteux sont moins fréquents et plus simples.</li> <li>— Améliore les habitudes de prescription et interagit avec le système d'information existant. Améliore la disponibilité des produits essentiels, réduit la consommation de produits non essentiels et améliore le rapport coût-efficacité.</li> <li>— Favorise un enregistrement fiable de la morbidité.</li> </ul>	<ul style="list-style-type: none"> <li>— Des problèmes peuvent survenir avec les données détaillées sur la morbidité et les régimes de traitement standard convenus.</li> <li>— Elle nécessite des calculs plus détaillés.</li> <li>— Les résultats peuvent être très différents de l'approvisionnement réel en produits de santé.</li> <li>— L'approvisionnement ne correspondra pas à l'utilisation si les traitements standard ne sont pas suivis.</li> <li>— N'évalue que les quantités nécessaires pour traiter les patients; les pertes et gaspillages doivent être considérés séparément.</li> <li>— Les hospitalisations exactes sont difficiles à prévoir.</li> </ul>
<b>Méthode de consommation par procuration</b>	Peu exigeant en termes de quantité de données sur la morbidité et la consommation pharmaceutique	<ul style="list-style-type: none"> <li>— Comparabilité des installations, des schémas de morbidité et des pratiques de traitement entre les sites/pays,</li> <li>— Données de consommation incomplètes ou inexactes provenant de l'installation de référence.</li> </ul>
Continue à la page suivante		

Table 2.1 – suite de la table précédente

	Avantages	Limites
<b>Projection du besoin budgétaire au niveau du service</b>	<ul style="list-style-type: none"> <li>— Estimation fiable du coût moyen des produits de santé par patient.</li> <li>— Pas besoin de détails sur la consommation et de données sur la morbidité.</li> </ul>	<ul style="list-style-type: none"> <li>— Ne permet pas de prévoir les besoins en produits de santé spécifiques,</li> <li>— Utilisation variable des installations (établissements)</li> <li>— Modes de traitement de la fréquentation,</li> <li>— Efficacité du système d’approvisionnement.</li> </ul>
<b>Méthode basée sur la demande des hôpitaux</b>	Le libre choix des produits de santé par les prescripteurs, notamment pour les systèmes de santé centralisés (pays, régions, etc.).	<ul style="list-style-type: none"> <li>— Les hôpitaux demandent plus de produits de santé que nécessaire, sachant que les quantités seront de toute façon réduites.</li> <li>— Les hôpitaux demandent des produits de santé qui sont inappropriés (coût/thérapeutique) ou inutiles en raison de préférences personnelles ou de l’influence des entreprises pharmaceutiques.</li> <li>— Le personnel soignant adopte souvent des pratiques de prescription sous-optimales</li> <li>— Des produits de santé coûteux et non essentiels sont nécessaires.</li> <li>— Les pratiques de prescription non optimales sont acceptées.</li> <li>— Les prescripteurs sont sensibles aux pratiques de marketing des entreprises pharmaceutiques et à la pression des consommateurs.</li> </ul>
<b>Méthode basée sur la population</b>	Intéressant pour les produits de santé qui n’ont pas d’usages multiples	<ul style="list-style-type: none"> <li>— Les données sont très difficiles à obtenir, notamment les informations socio-économiques sur les patients.</li> <li>— Difficile de quantifier les produits de santé à usages multiples.</li> </ul>

TABLE 2.1 – Avantages et limites des méthodes classiques de quantification

## 2.2 Méthodes statistiques de prévision de la demande

En économie, la demande peut être définie comme la quantité d’un produit ou d’un service que les gens sont désireux et capables d’acheter pendant une période donnée. On en déduit que la prévision de la demande est l’art de prédire le niveau de la demande qui pourrait se

produire à un moment ou à une période future [6].

La prévision de la demande ou de la consommation a été et continue d'être largement commentée dans la littérature. Cette pratique a été longuement considérée comme incontournable dans les domaines industriels et hospitaliers surtout dans des situations d'instabilité économique et d'accroissement de l'incertitude. Plusieurs auteurs considèrent que la problématique de prévision reste d'actualité et que cette pratique est indispensable pour les prises de décision dans les structures. Cela justifie encore plus à l'ère de la "Big Data". Dans le cadre de nos travaux, la prévision de la demande revêt une importance particulière. Elle constitue l'un des leviers les plus conséquents, couramment déployés pour réduire les situations de "overstocking" et "understocking", c'est-à-dire pour la maîtrise et la réduction des incertitudes propres aux opérations de planification et aux pratiques logistiques. Les auteurs comme Armstrong ne manquent de mettre l'accent également sur l'importance des prévisions, et notamment les prévisions à court et moyen terme, dans le contexte particulier de la gestion et le dimensionnement de stock et des réapprovisionnements [8, 147].

Dans la littérature, On identifie plusieurs techniques de prévision de la demande qui peuvent être regroupées en deux catégories (voir figure 2.1), comme le soulignent [8, 81]. Dans la première catégorie, il se trouve les méthodes qualitatives ou d'enquête. Ces techniques sont essentiellement subjectives et s'appuient sur des jugements et des opinions ou imitent les choix des consommateurs pour faire une prédiction comme le mentionne Ebrahimi Meimand et al. [91]. Elles peuvent être utilisées lorsque les données historiques ne sont pas disponibles ou lorsque les circonstances évoluent si rapidement qu'un modèle mathématique de prévision basé sur des données passées peut devenir non pertinent ou discutable [161]. La seconde catégorie regroupe les méthodes dites méthodes quantitatives. Et c'est sur cette catégories de méthodes que nous nous concentrerons.

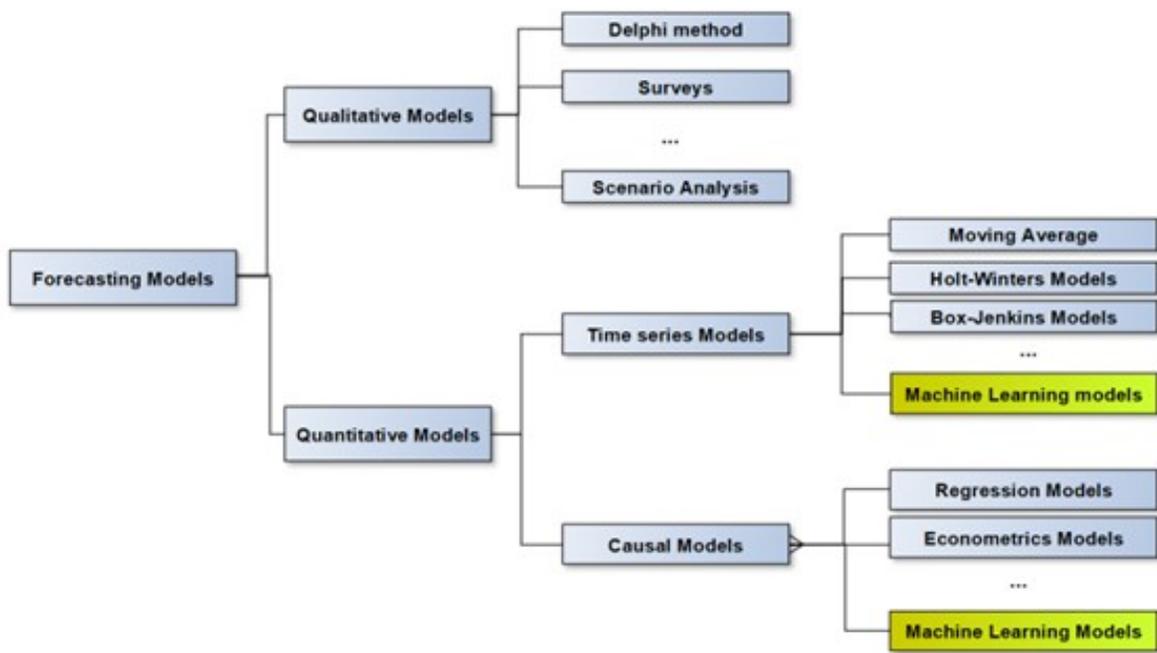


FIGURE 2.1 – Méthodes quantitatives de la prévision de la demande

## 2.2.1 Méthodes quantitatives de prévision de la demande

Les méthodes quantitatives, également appelées méthodes statistiques, constituent la famille de méthodes la plus utilisée dans le milieu industriel et s'adaptent à notre problématique. Cette famille de méthodes comprend essentiellement deux types de méthodes (Voir figure 2.1 une réadaptation de la catégorisation des modèles de prévision [161]) :

- Dépendantes fortement de l'historique de la demande, de la variable temps : il s'agit de méthodes des séries temporelles
- Dépendantes des variables explicatives : Il s'agit des méthodes causales.

### 2.2.1.1 Méthodes de séries temporelles

Les méthodes de séries temporelles utilisent principalement la dimension temporelle pour prédire la demande. Elles se fondent sur l'analyse des données historiques pour identifier les tendances et les motifs saisonniers, puis utilisent ces informations pour anticiper la demande future. Avec cette famille de méthodes, on part du postulat que les tendances futures sont similaires aux tendances historiques.

Les quatre propriétés fondamentales des séries temporelles sont la tendance, la saisonnalité, les cycles et le caractère aléatoire. La tendance fait référence à la stabilité de l'augmentation ou de la diminution des données globales. On peut décrire la tendance comme la direction, qui peut être linéaire ou non linéaire. La saisonnalité fait référence aux modèles observés dans les séries temporelles à des intervalles fixes et répétés. Les cycles ont un concept similaire à celui de la saisonnalité, à une exception près : le modèle est observé à des intervalles variables. Enfin, le caractère aléatoire fait référence au résidu et parfois aux valeurs aberrantes qui rendent les modèles difficiles à établir [143].

Ces méthodes se divisent en deux catégories : univariées et multivariées. Les méthodes univariées, considérées comme simples et pratiques, se concentrent deux variables : le temps et la quantité consommée ou demandée d'un produit. En revanche, les méthodes multivariées prennent en compte plusieurs variables pour modéliser la demande en lien avec le temps.

Parmi les techniques bien connues de cette catégorie, on peut citer les méthodes des moyennes mobiles, le lissage exponentiel, les modèles Holt-Winters, les modèles Box-Jenkins, des modèles de machine learning etc. Pour plus de détails, nous vous referons [19, 65, 80, 81, 161]. Au cas où il existerait plusieurs variables explicatives en lien avec la quantité et dépendantes du temps, il existe une deuxième famille de méthodes appelées méthodes de séries temporelle multivariées. Ces méthodes incluent des techniques telles que la régression vectorielle autorégressive (VAR), Modèle d'auto-régression vectorielle à moyenne mobile (VARMA), Modèle autorégressif multivarié (MAR), des modèles de machine learning, etc.

### 2.2.1.2 Méthodes causales

Les méthodes causales aussi appelées méthodes explicatives, dans leurs approches, elles cherchent à établir une relation entre l'évolution de la consommation ou la demande (qui est la variable à prévoir, couramment appelée variable expliquée) et une ou plusieurs variables

explicatives, autres que la série de données en elle-même. Les modèles causaux sont plus utiles lorsque des relations de corrélations existent, c'est-à-dire une corrélation entre la variable à prévoir et les variables explicatives, ou lorsque des changements importants sont attendus dans les variables explicatives sur l'horizon de prévision, comme le mentionne [8, 161]. Les méthodes causales pour la prévision de la demande ne doivent pas être confondues avec l'inférence causale, qui vise à déterminer si une variable est une cause directe ou indirecte d'une autre variable, en utilisant des méthodes expérimentales ou quasi-expérimentales pour tester l'effet causal [53, 116]. Les méthodes phares de cette famille de méthodes sont par exemple la régression linéaire (simple ou multiple), les modèles économétrique, etc. Le modèle causal de prévision est une relation mathématique entre les variables dépendantes et indépendantes, exprimée sous la forme  $Y = f(X_1, X_2, \dots, X_n)$ . Dans le modèle, les  $X$  sont les variables indépendantes (les facteurs) et  $Y$  est la variable dépendante (la quantité de produits de santé à estimer, dans notre cas). La principale difficulté est d'obtenir les bonnes données. Les méthodes causales sont plus complexes que les méthodes de séries temporelles, mais elles offrent une meilleure compréhension de la relation entre la demande et les facteurs. A l'ère d'une production massive de données hospitalières faire appel aux techniques de machine Learning est incontournable.

De ces deux groupes de méthodes de prévision de la demande et au égard à la problématique exposée en introduction générale, ce sont les techniques de machine learning qui seront mises en application.

### 2.2.2 Machine Learning pour la prévision de la demande

Le Machine Learning (ML) trouve son origine dans les travaux du psychologue Donald Hebb en 1949, qui décrit l'apprentissage Hebbien, un ancêtre du ML. Cette théorie est basée sur le renforcement des connexions synaptiques entre les neurones lorsqu'ils sont activés simultanément [51]. En 1950, Alan Turing propose le test de Turing pour évaluer l'intelligence d'une machine dans sa publication intitulée "Computing Machinery and Intelligence". En 1956, la Conférence de Dartmouth marque le début officiel de l'intelligence artificielle en tant que domaine de recherche. En 1957, Frank Rosenblatt invente le *perceptron*, une première application du machine learning. En 1959, Arthur Samuel utilise le terme "machine learning" pour décrire la capacité des ordinateurs à apprendre à partir de données. Depuis lors, le domaine s'est développé avec l'expansion des champs d'application et l'amélioration des algorithmes. Au cours des dernières années, le machine learning a connu une croissance exponentielle, notamment dans la vision par ordinateur, le traitement du langage naturel, la reconnaissance vocale, la recommandation de contenu, la détection de fraude, etc. Le domaine sanitaire est également concerné par ces avancées. Mais qu'est-ce que le Machine learning ?

On rencontre plusieurs définitions formelles dans la littérature, depuis la description faite par Arthur Samuel en 1959 [134] à nos jours. Nous avons retenu deux définitions :

- Selon Tom Mitchell en 1997, le machine learning est : "*A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure*

*P, if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ " [93].*

- Ethem Alpaydin en 2014, dira que "*Machine learning is programming computers to optimize a performance criterion using example data or past experience*" [2].

Les différentes définitions mettent en avant l'aspect "apprentissage" des machines, c'est-à-dire leur capacité à améliorer leur performance à travers l'expérience. Elles soulignent également que cette amélioration se fait à travers l'analyse et l'apprentissage à partir de données, sans que l'on doive spécifier explicitement les règles à suivre.

Dans la littérature, différentes méthodes de machine learning sont utilisées, choisies en fonction de critères tels que la nature du problème, la taille et la complexité des données, la performance, la compréhensibilité, la robustesse et la scalabilité des modèles. La performance peut varier en termes de précision et de temps de calcul, tandis que la compréhensibilité peut influencer la facilité de compréhension et d'interprétation des résultats. La robustesse peut être un critère important pour résister aux données aberrantes ou aux erreurs de mesure. Enfin, la scalabilité peut être un élément clé pour s'adapter à des ensembles de données plus volumineux. Sans être exhaustive, les catégories de méthodes de machine learning les rencontrées dans la littérature sont :

- La classification : Elle est un type de tâche qui consiste à prédire la classe d'un objet ou une observation à partir de caractéristiques ou attributs mesurés. La classification peut être binaire, c'est-à-dire qu'il y a seulement deux classes possibles, ou multiclassée lorsque plusieurs classes peuvent être prédites.
- La Régression : Elle consiste à modéliser la relation entre une variable cible continue et un ou plusieurs variables explicatives (prédicteurs) qui peuvent être continus ou catégoriels. L'objectif de la régression est de trouver une fonction qui puisse prédire la valeur de la variable cible en fonction des valeurs des prédicteurs.
- Le Clustering (ou regroupement) : Elle vise à regrouper un ensemble de données en sous-groupes (ou clusters) homogènes selon leur similarité ou distance. Les méthodes de clustering peuvent être basées sur des approches hiérarchiques ou non hiérarchiques, ainsi que sur des méthodes de partitionnement ou de densité. Le choix de la méthode dépend de la nature des données et des objectifs de l'analyse.

A ces derniers on peut ajouter d'autres catégories de méthodes visant à mieux comprendre les données comme :

- La Réduction de la dimensionnalité : Elle permet de transformer des données complexes et riches en informations en données plus simples et compréhensibles, tout en conservant l'essentiel de l'information. Cette technique est particulièrement utile dans le cas où le nombre de variables dans les données est très élevé, ce qui peut compliquer l'analyse et la compréhension des résultats. La réduction de la dimensionnalité est souvent utilisée en combinaison avec d'autres tâches de machine learning, comme la classification ou la régression, pour simplifier les données et améliorer les performances des modèles.
- La détection d'anomalies (ou détection de valeurs aberrantes) : Elle est une tâche qui consiste à identifier des données qui sont significativement différentes de la majorité des

autres données dans un ensemble de données donné. Les anomalies sont des données qui sont inattendues ou qui ne suivent pas les modèles de données normaux.

- Exploration de données (Data mining en anglais) : Il s'agit d'un processus itératif et exploratoire qui vise à extraire des connaissances et des modèles à partir de données brutes. Elle consiste à analyser les données pour découvrir des relations, des modèles ou des tendances qui pourraient être utiles pour résoudre des problèmes ou prendre des décisions éclairées. C'est une technique qui peut inclure des étapes telles que la préparation et le nettoyage des données, la sélection de variables, la transformation des données, l'application de techniques de modélisation et l'évaluation des résultats.

Les principales méthodes d'apprentissage rencontrées sont l'apprentissage supervisé (qui regroupe les catégories de méthodes de classification et de régression), l'apprentissage non-supervisé (clustering, réduction de dimensionnalité, détection d'anomalies), l'apprentissage semi-supervisé (combinaison de supervision et non-supervision), l'apprentissage par renforcement (apprentissage interactif avec l'environnement) et l'apprentissage en ligne (mise à jour du modèle au fur et à mesure des nouvelles données). D'autres concepts importants pour résoudre des problèmes de données massives incluent l'apprentissage de représentation, l'apprentissage actif, l'apprentissage profond (Deep learning), le transfert d'apprentissage, l'apprentissage distribué et parallèle, l'apprentissage basé sur les noyaux et l'apprentissage fédéré (Federated learning). Pour plus de détails, il est conseillé de consulter les références suivantes [2, 145, 60].

L'apprentissage supervisé consiste à apprendre à partir de données étiquetées (labellisées), c'est-à-dire des données pour lesquelles la réponse souhaitée est connue à l'avance. L'objectif est de construire un modèle qui peut généraliser à de nouvelles données en faisant des prédictions précises pour des données non étiquetées. L'apprentissage supervisé utilise des algorithmes qui apprennent à partir de données étiquetées en ajustant les paramètres du modèle de sorte que les prédictions du modèle correspondent aux étiquettes d'apprentissage autant que possible. Une fois que le modèle est formé, il peut être utilisé pour prédire les étiquettes pour de nouvelles données qui n'ont pas été utilisées pour l'entraînement. C'est une approche appropriée pour la prévision de la demande des produits de santé dans les hôpitaux en particulier la régression au égard à notre objectif principal. D'autres concepts comme le Deep learning seront également pris en compte en fonction des besoins de nos travaux. Nous reviendrons sur certains définitions de concept dans la suite de nos travaux.

### 2.2.2.1 Le Machine Learning dans le domaine de la santé

Au cours des dernières années, l'application de techniques de machine learning dans le domaine de la santé a connu une croissance exponentielle. Le machine learning est devenu une partie intégrante de certaines pratiques médicales, permettant de réaliser des diagnostics de plus en plus précis. Les techniques de ML sont utilisées pour aider à diagnostiquer des maladies, pour prédire l'évolution de la maladie, pour aider à concevoir des médicaments et pour aider à améliorer l'efficacité et la qualité des soins de santé.

Cette thématique suscite également un intérêt croissant au sein de la communauté académique. À la date du 11 avril 2023, une recherche croisée des mots-clés ("machine learning" OR

"deep learning" OR "Artificial intelligence") AND ("Health care" OR "Healthcare") AND ("Hospital") sur la plate-forme *Science Direct*<sup>1</sup> sur la période de 2010 à 2023 a révélé 12 245 articles de recherche. La figure 2.2 présente la tendance et la répartition des publications d'articles de recherche sur la thématique du machine learning dans le domaine de la santé.

De cette recherche, menée en anglais, montre que les travaux liés au machine learning touchent plusieurs secteurs. Bien que ce travail ne soit pas exhaustive, il offre un aperçu des mots-clés et des thèmes abordés, comme le montre la figure 2.3. On peut noter que, après les mots-clés tels que "machine learning", "artificial intelligence" et "deep learning", des termes comme "healthcare", "Covid-19" se présentent comme les plus cités. Nous sommes tentés de déduire que l'avènement de la pandémie à Covid-19 dès la fin de l'année 2019 a contribué à stimuler les travaux de recherches sur le machine learning dans le domaine médical. D'autres facteurs, et non des moindres sont la production massive et l'accessibilité de données dans le secteur médical, la hausse des coûts des soins de santé et l'accent sur les soins personnalisés qui motivent cette croissance exponentielle des recherches [153].

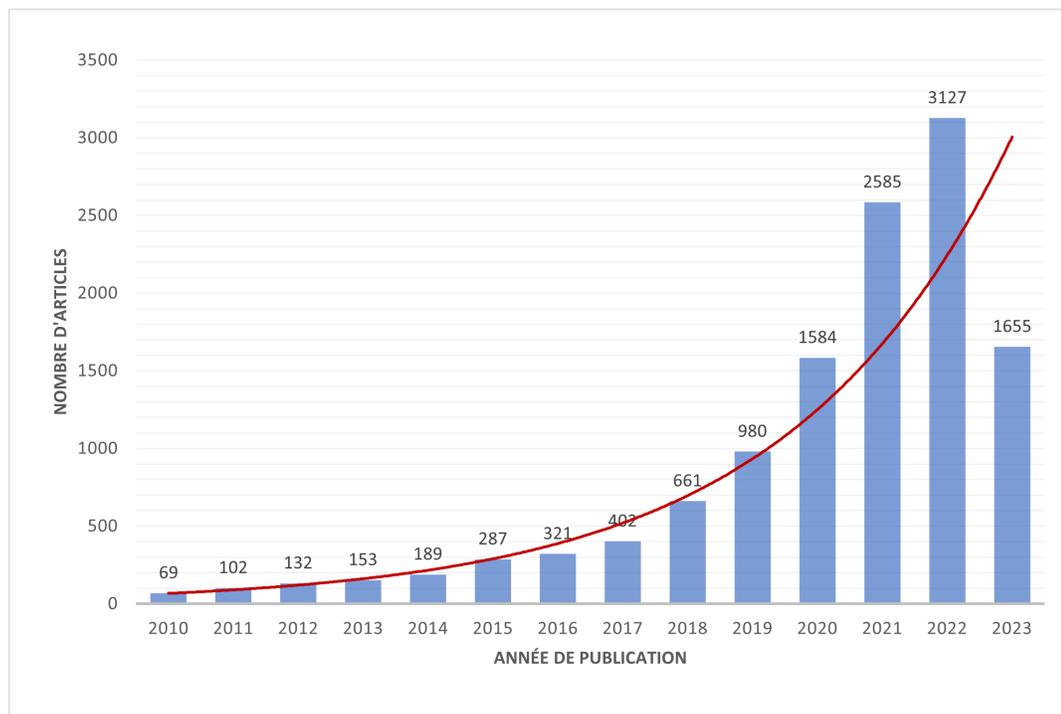


FIGURE 2.2 – Tendance des publications sur le ML dans le milieu sanitaire

Ce réseau de mots-clés met en évidence la présence de termes liés à l'imagerie médicale, à la télémédecine, au diagnostic, aux cancers, aux médicaments, etc. Cela indique que les champs d'application du machine learning dans le domaine de la santé sont diverses et incluent également les produits de santé. Cependant, la présence des termes *medicine* et *prediction* dans le réseau n'implique pas nécessairement qu'il s'agit de la prévision de la demande des médicaments. Il peut s'agir des recherches pour la découverte de médicaments, le développement de thérapies personnalisées, l'optimisation de la formulation des médica-

1. <https://www.sciencedirect.com/>

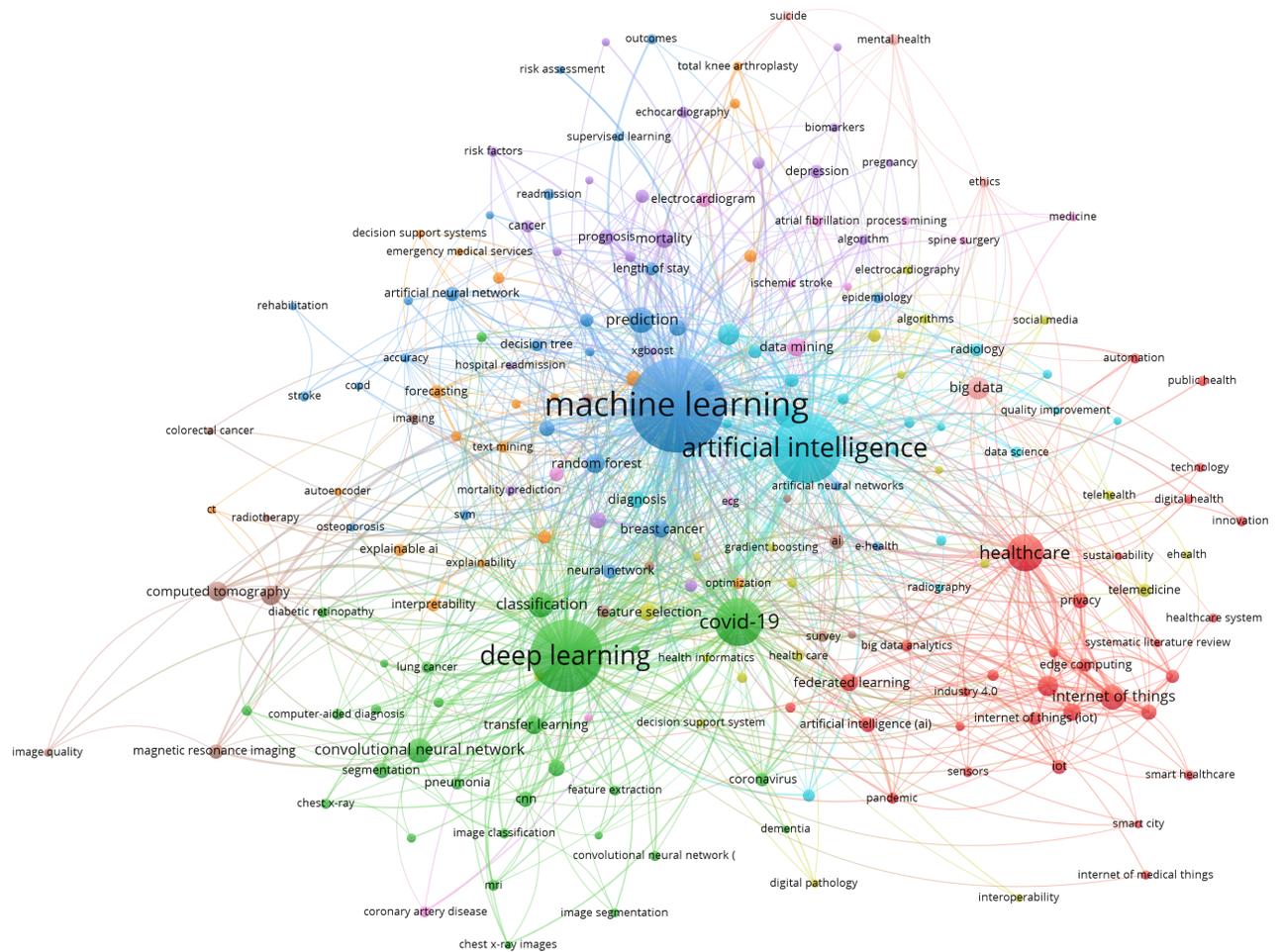


FIGURE 2.3 – Visualisation du réseau de mots-clés sur le ML dans la santé

ments, la prédiction d'effets secondaires, la surveillance de l'efficacité des médicaments, la détection de contrefaçons, etc. Étant donné que la prévision des besoins en produits pharmaceutiques est l'objet de cette thèse, une recherche croisée des mots-clés précédents avec ("Demand forecasting" OR "Demand prediction") AND ("Medicine" OR "Drug") a été effectuée dans les mêmes conditions, ce qui a donné 52 articles de recherche. Ce qui est bien inférieur au nombre précédent dans le secteur médical, ce qui suscite un intérêt particulier. La sous-section suivante explore cette question.

### 2.2.2.2 Machine Learning pour la prévision de demande des produits de santé

Dans la littérature, on trouve des travaux portant sur la prévision de la demande pour des produits de santé. En ce qui concerne les méthodes de séries temporelles, plusieurs techniques d'apprentissage automatique, telles que les réseaux de neurones récurrents (RNN), les mémoires à court et long terme (LSTM), les unités récurrentes à portes (GRU), ainsi que des approches floues, ensemblistes et hybrides ont été appliquées aux séries temporelles pour améliorer les prévisions [97, 158]. Janardhanan & Barrett [59] ont montré que les modèles

d'apprentissage automatique traitant de la prévision des séries temporelles sont plus efficaces que les méthodes statistiques conventionnelles. Leur étude s'est concentrée sur la prévision de l'utilisation de l'unité centrale des machines dans les centres de données en utilisant LSTM et l'a comparée aux modèles traditionnels de moyenne mobile intégrée autorégressive (ARIMA), largement utilisés pour la prévision. Les résultats ont conclu que l'algorithme LSTM présentait de plus faibles erreurs par rapport au modèle ARIMA (17-23% contre 37-42%). Plusieurs travaux corroborent cette efficacité, notamment dans le monde industriel.

Concernant les produits pharmaceutiques dans le domaine hospitalier, Guo et al. [44] ont exploré l'utilisation des méthodes de séries temporelles classiques pour la prévision de la demande. Leur objectif était de prédire la demande mensuelle en globules rouges pour les patients pédiatriques, en se basant sur les données cliniques de transfusion sanguine du National Center for Children's Health à Beijing, en Chine. Ils ont utilisé des modèles SARIMA pour cette prévision. De manière similaire, Hafiz et al. [46] ont cherché à anticiper la demande de médicaments et à prendre des mesures préventives pour ajuster les besoins dans tous les districts du Bangladesh. Leur approche s'appuyait sur un modèle de mémoire à long terme (LSTM).

Motamedi et al. [98] ont, quant à eux, utilisé des méthodes univariées et multivariées pour prédire la demande en plaquettes dans quatre hôpitaux à Hamilton, Ontario. Ils ont employé ARIMA et Prophet dans leurs approches univariées, en se basant uniquement sur les données historiques de la demande. Pour les approches multivariées, ils ont utilisé la régression Lasso et LSTM en tenant compte des indicateurs cliniques. Leurs résultats ont montré que la régression Lasso et LSTM surpasse les méthodes des approches univariées, soulignant ainsi l'importance de considérer plusieurs déterminants (facteurs) dans les modèles, en plus du temps.

Récemment, Mbonyinshuti et al. ont démontré l'efficacité des approches d'apprentissage automatique pour 10 médicaments au Rwanda. Des modèles de séries temporelles univariées ont été utilisés sur des données historiques de 2015 à 2019, et Random Forest a montré la meilleure précision (76% de score métrique  $R^2$ ), par rapport aux modèles de réseau neuronal et de régression linéaire [89]. Les données de cette étude se limitent aux attributs tels que le mois, l'année, le district de consommation, le nom et la quantité du médicament. Les deux seules variables prises en compte dans les modèles de prédiction sont la date et la quantité de consommation. Contrairement à Mbonyinshuti et al. [89], dans ce travail, nous considérons de nombreux facteurs intrinsèques à la consommation, en particulier les attributs caractérisant l'hôpital pour lequel la consommation du médicament doit être prédite (par exemple, le nombre de lits, le nombre d'unités fonctionnelles, etc.).

Sarvestani et al., Rathipriya et al. ainsi que Schilling et al. ont également utilisé des méthodes de séries temporelles pour prévoir la demande de produits sanguins et de leurs dérivés. Ils ont opté pour des méthodes d'apprentissage automatique en tant qu'alternatives aux méthodes statistiques classiques telles qu'ARIMA [137, 125, 139].

Les méthodes de séries temporelles sont largement utilisées dans la littérature pour prédire la demande de médicaments, et les chercheurs ont récemment montré plus d'intérêt dans ce domaine. Malgré des résultats intéressants, les techniques de séries temporelles sont plus

appropriées lorsque nous pouvons décrire des modèles ou des tendances générales sans tenir compte des facteurs affectant la variable prédite, c'est-à-dire la consommation ou la demande. En outre, les méthodes de séries temporelles ne sont pas toujours adaptées à la prévision des fluctuations soudaines ou des changements rapides de la demande de médicaments dans les hôpitaux, tels que les épidémies ou les urgences médicales.

Les méthodes de séries temporelles s'appliquent bien aux jeux de données structurés, notamment lorsque les données sont agrégées (sommées) sur des périodes de consommation. Cependant, l'utilisation de données de consommation hospitalière non agrégées sur les dates de consommation peut introduire davantage de bruit (résidu) dans les données. Cela peut rendre le modèle de prévision basé sur les séries temporelles moins précis, car il est plus difficile de distinguer les tendances réelles des variations aléatoires. Créer une série temporelle distincte avec des valeurs mensuelles de consommation pour chaque hôpital, bien que possible, s'éloignerait de l'objectif principal de cette thèse.

Dans ce travail, notre objectif est de prédire la demande de médicaments pour un hôpital sans historique de consommation en utilisant des modèles qui intègrent de multiples facteurs. Nous nous appuyons sur les données historiques de différents hôpitaux, en prenant en compte plusieurs facteurs sans agrégation. Notre approche vise à élaborer des modèles indépendants des chronologies de consommation, mais dépendants des caractéristiques de chaque hôpital. Dans ce contexte, les méthodes causales s'avèrent les plus appropriées, bien que leur collecte et leur analyse de données puissent présenter des défis, surtout lorsqu'il s'agit de données historiques sur des variables externes potentiellement complexes.

En ce qui concerne les approches causales dans la prévision des produits de santé, il est à noter l'étude de Qingkui et Junhu [122]. Ces chercheurs ont cherché à estimer la demande de seringues de perfusion en prenant en compte neuf facteurs dans un hôpital en Chine. Ils ont utilisé un réseau neuronal de rétro-propagation (BP) pour obtenir une précision de 87,9%. Parmi les neuf facteurs influençant la consommation, le nombre de patients était le plus significatif, subdivisé en cinq catégories en fonction du lieu de traitement. Cependant, cette étude est limitée dans son applicabilité à d'autres hôpitaux et à d'autres produits de santé, tels que les médicaments. De plus, elle n'a pas pris en compte des indicateurs cruciaux liés à la capacité de l'hôpital, comme le nombre de lits et le nombre de membres du personnel.

Plus récemment, Amalnick et al. [5] ont proposé des algorithmes de réseaux neuronaux tels que MLP, RBF et GMDH pour anticiper la demande dans l'industrie pharmaceutique. Leur étude s'est appuyée sur l'historique de 519 échantillons de données provenant d'une usine pharmaceutique iranienne, comprenant 43 types de produits avec 14 caractéristiques, notamment le type de produit, le code produit, la tranche d'âge, le sexe, les maladies liées au produit, etc. Avant d'appliquer les trois algorithmes d'apprentissage automatique pour la prévision, ils ont effectué une classification et un regroupement en quatre catégories de produits. Les modèles MLP ont donné des résultats impressionnants, avec des scores  $R^2$  allant de 0,97 à 1. Cette étude illustre bien l'approche causale dans la prévision de la demande, même si elle se concentre sur l'industrie pharmaceutique plutôt que sur les hôpitaux.

En résumé, la littérature existante sur l'utilisation de la méthode causale pour anticiper la demande dans les hôpitaux est limitée. Cette lacune peut s'expliquer par diverses raisons,

notamment la difficulté d'accéder aux données et les préoccupations liées à la confidentialité des données des patients, qui sont primordiales dans le domaine des soins de santé. La demande de médicaments est influencée par de multiples facteurs, comme détaillés au chapitre 4. À notre connaissance, il n'existe pas de travaux prévoyant la demande hospitalière de médicaments en utilisant des données provenant d'autres établissements hospitaliers et en incluant des facteurs de consommation sans nécessairement dépendre des tendances chronologiques. Ainsi, cette étude expérimentale acquiert une signification particulière, justifiant ainsi l'introduction de la formulation mathématique (équation (1)) et le choix de l'approche causale.

Sans être exhaustif, nous proposons dans la tables 2.2 les méthodes populaires et qui ont démontré leur efficacité dans les milieux industriel, manufacturier et hospitalier dans la prévision de la demande. Nous faisons usage aux techniques relatives aux méthodes causales pour les applications dans le cadre de cette thèse.

Modèles	Description brève	Méthodes de séries chronologiques	Méthodes causales
Régression linéaire multiple	Un modèle statistique qui cherche à établir une relation linéaire entre une variable dépendante et plusieurs variables indépendantes	✓	✓
Réseaux de neurones artificiels	Des modèles inspirés du fonctionnement du cerveau humain, utilisés pour résoudre des problèmes de classification et de régression.	✓	✓
Random Forest	Un ensemble d'arbres de décision qui agrège leurs prédictions pour améliorer la précision et réduire le surapprentissage	✓	✓
Gradient Boosting	Une technique d'apprentissage par ensemble qui construit des modèles de manière séquentielle pour corriger les erreurs des modèles précédents.	✓	✓
XGBoost	Une implémentation optimisée de l'algorithme de gradient boosting, connue pour sa haute performance prédictive.	✓	✓
AdaBoost	Un algorithme d'apprentissage par ensemble qui ajuste les poids des échantillons d'entraînement pour mettre l'accent sur les erreurs précédentes.	✓	✓
Support Vector Regression	Une extension de la "Support Vector Machine" pour les problèmes de régression.	✓	✓
Réseaux de neurones récurrents	Des réseaux de neurones conçus pour traiter des données séquentielles en maintenant une mémoire interne.	✓	
LSTM	Des réseaux de neurones récurrents avec des portes d'oubli, d'entrée et de sortie pour mieux gérer les longues dépendances séquentielles.	✓	
Continue à la page suivante			

Table 2.2 – Suite de la table précédente			
Modèles	Description brève	Méthodes de séries chronologiques	Méthodes causales
GRU	Des réseaux de neurones récurrents avec des portes d'actualisation et de ré-initialisation pour faciliter l'apprentissage sur de longues séquences.	✓	
Modèles de machine learning hybrides	Des modèles qui combinent des éléments de différents algorithmes pour améliorer la performance prédictive.	✓	✓

TABLE 2.2 – Présentation succincte de quelques modèles de machine learning

Pour plus de détails sur les différentes méthodes, nous conférons les lecteurs aux travaux et documents suivants [97, 48, 60, 45, 57, 76, 73, 2, 93]. Cependant nous reviendrons sur ces méthodes dans nos applications dans les chapitres qui suivent.

## 2.3 Conclusion

L'objectif de ce chapitre a été de passer en revue les méthodes de prévision de demande dans le monde hospitalier. De ce fait, nous avons présenté en premier lieu, un ensemble de sept méthodes classiques d'estimation des produits de santé les plus connues dans le milieu hospitalier depuis les années 1980. Une comparaison de ces méthodes est fournie en donnant les avantages et les limites de chaque méthode avant de passer en revue les méthodes statistiques de prévision de la demande des produits pharmaceutiques. Dans la seconde partie de ce chapitre, nous avons examiné les méthodes statistiques de prévision de la demande, en mettant l'accent principalement sur le machine learning qui nous intéresse dans cette thèse. Deux catégories de méthodes quantitatives résultent de la littérature. Il s'agit des méthodes à séries chronologiques et les méthodes causales. Les méthodes d'apprentissage automatique liées à l'approche causale sont celles qui seront explorées dans les chapitres suivants. Cependant, leur mise en œuvre nécessite la collecte, le traitement et la structuration des données. Le chapitre suivant nous offre l'occasion de décrire ce processus afin d'obtenir des données fiables pour la poursuite de nos travaux.

Deuxième partie  
Contributions de la thèse

# Chapitre 3

## Description, structuration et pré-traitement des données

### Sommaire

---

<b>3.1</b>	<b>Descriptions des sources de données</b>	<b>50</b>
3.1.1	Bases de données internes	50
3.1.2	Bases de données externes	58
3.1.3	Étude de terrain	62
<b>3.2</b>	<b>Organisation et structuration des données</b>	<b>64</b>
3.2.1	Pré-traitements des données	64
3.2.2	Premier croisement des données	73
3.2.3	Second croisement des données	74
<b>3.3</b>	<b>Conclusion</b>	<b>78</b>

---

Les différents travaux entrant dans le cadre de cette thèse sont principalement axés sur des expérimentations. Dans ce sens, la disponibilité des données s'avère essentielle, surtout que nous désirons utiliser les techniques de machine learning qui requièrent un volume de données substantiel. Dans cette thèse, plusieurs sources de données sont exploitées, à la fois internes aux hôpitaux, partenaires et externes en relation avec ces hôpitaux.

Ainsi, cette partie de la thèse décrira les bases de données émanant de l'entreprise partenaire et d'autres organismes officiels. S'en suit les descriptions des procédures d'extraction, de collecte ainsi que les démarches de pré-traitements et structurations des données qui vont servir aux différentes expérimentations. À ces étapes, s'ajoute la description des différents questionnaires réalisés. Ce travail nous permettra également de lever les équivoques de traçabilité des produits de santé, des unités fonctionnelle ainsi que l'aspect organisationnel des hôpitaux et dans notre cas, les Centres Hospitaliers Universitaires (CHU).

Dans le déroulé de ce chapitre, la section 3.1 est dédiée à la description des bases de données internes constituées des bases de produits de santé, des unités fonctionnelles des hôpitaux et des historiques de consommation ainsi que des bases de données externes utilisées. Les questionnaires et interviews réalisés dans le cadre de cette thèse seront également décrits.

La section 3.2 est consacrée à des procédures de l'organisation et de structuration des données obtenues après pré-traitements pour usages finaux. Enfin, la dernière section, la section 3.3, propose une synthèse, des conclusions et des recommandations.

## 3.1 Descriptions des sources de données

### 3.1.1 Bases de données internes

L'entreprise partenaire a mis à notre disposition trois types de bases de données : une base de données des articles, une base de données des unités fonctionnelles, et une base de données relative à l'historique de consommation de six hôpitaux français. Il s'agit d'hôpitaux de types CHRU, CHU et CH que nous nommerons de H1 à H6. Nous allons présenter chacune de ces trois bases de données dans cette section.

#### 3.1.1.1 Base de données des articles

La base de données des articles est un répertoire de tous les produits de santé enregistrés sur la plate-forme de gestion WMS des hôpitaux, partenaires. Elle est subdivisée en deux catégories : Une catégorie nommée « Pharmacie » pour les consommables médicamenteux et une seconde catégorie dénommée « DMS » pour les dispositifs médicaux, les pansements et autres produits de santé. En rappel, les différents concepts de ces produits de santé sont rappelés dans le chapitre 1 à la section 1.2. Cette base de données représente environ 34.2 MB de données contenant 121 343 lignes et 42 colonnes d'enregistrement (attributs). Chaque ligne est supposée être un vecteur de 42 informations sur un produit de santé enregistré par un hôpital. Les 42 colonnes font partie des variables dédiées à la description et à la caractérisation de chaque produit de santé dans le système de gestion des stocks des hôpitaux proposé par l'entreprise KLS.

A l'aide de requête SQL, une extraction est effectuée pour collecter les données des six hôpitaux mis à notre disposition. De cette base de données, on déduit la table 3.1 qui fournit le nom de chaque colonne et nombre d'enregistrement par caractéristique dans la base. La variable `ID_SITE_RATTACHE` permet d'identifier le site, c'est-à-dire l'hôpital où l'article est enregistré et consommé. `ID_REF` désigne l'identifiant de l'article dans un hôpital ce qui sous-entend que : le même `ID_REF` n'identifie pas forcément le même article dans des hôpitaux distincts. Pour un identifiant unique, il convient de se rabattre sur les codes de traçabilité unique, décrits dans le chapitre 1 section 1.3, comme les codes ATC, UCD, CIP, etc., des codes représentés ici par `HOPI_CODE_UCD` et `CODE_CIP`. Dans cette base initiale, il ne figure pas de colonne relative aux codes ATC.

Parmi ces 42 caractéristiques, on retrouve également celles dédiées aux libellés. Il s'agit des variables `DESIG`, `DESIG_SEC`, `DESIGN_TER`, qui varient souvent d'un hôpital à l'autre. Cela rend nécessaire une uniformisation et donc un pré-traitement des données. Cette uniformisation revêt une importance cruciale compte tenu de l'objectif principal de cette thèse,

qui repose sur l'analyse et les études expérimentales à partir de bases de données provenant de plusieurs hôpitaux pouvant avoir des approches d'enregistrement différentes.

### 3.1.1.2 Unité fonctionnelle et base de données des unités fonctionnelles

Dans la littérature scientifique, la consommation des produits de santé est fonction des types d'activités, d'unités médicales fonctionnelles, de services ou département de la structure sanitaire. Dans le cadre de notre thèse plusieurs questions demeurent en suspens, parmi lesquelles :

- Quels sont les types d'établissement de santé et comment sont-ils organisés ?
- Au sein de l'organisation hospitalière, quelle est l'unité qui permet de mieux apprécier la consommation au sein d'un hôpital ?
- Existe t-il une nomenclature ou traçabilité standard nous permettant d'effectuer des études comparatives concernant cette unité ?

De plus, parmi les données mises à notre disposition, figure celles regroupant les unités fonctionnelles ou les destinations de consommation des produits de santé dans les différents hôpitaux. Une prise en main de ces données s'impose relativement à l'objectif global de la thèse.

Dans cette section, un rappel succinct sur les unités fonctionnelles émanant de l'organisation des hôpitaux en France est effectué avant de procéder à une description des données en lien avec les unités de destination des produits de santé de chaque hôpital.

#### a) Qu'est ce qu'une unité fonctionnelle dans l'hôpital ?

En France, selon le rapport de la Direction de la Recherche, des Études, de l'Évaluation et des Statistiques (DREES) sur les établissements de santé édition 2022, le secteur hospitalier en France est divisé en trois secteurs :

- Public avec 1347 établissements, dont 180 CHR/CHU, 936 CH (y compris les anciens hôpitaux locaux), 92 centres hospitaliers spécialisés en psychiatrie (CHS) ; et 139 autres établissements publics, correspondant majoritairement à des unités de soins de longue durée (USLD).
- Privés à but non-lucratif constitués de 670 établissements dont 20 centres de lutte contre le cancer (CLCC).
- Et de 972 cliniques privées à but lucratif.

Ce qui totalise 2989 établissements de santé de statut public ou privé, autrement appelé *entités géographiques*. Toutes les données de consommation mises à notre disposition relèvent du premier secteur.

Les établissements de santé peuvent avoir des statuts juridiques, des activités et des modes de financement différents, selon qu'ils sont publics ou privés (à but lucratif ou non). Mais par-dessus ces aspects une hiérarchisation est instituée en terme organisationnel. L'organisation

	Colonne	Nombre d'enregistrement de la colonne
<b>0</b>	<b>ID_SITE_RATTACHE</b>	<b>121343</b>
<b>1</b>	<b>ID_REF</b>	<b>106551</b>
<b>2</b>	<b>DESIG</b>	<b>106135</b>
<b>3</b>	DESIG_SEC	82937
<b>4</b>	DESIG_TER	48078
<b>5</b>	FAMILLE_PROD	39766
<b>6</b>	EAN13	3536
<b>7</b>	COMMENTAIRE	8412
<b>8</b>	ID_REF_SUBST	3263
<b>9</b>	TYPE_FIFO	105433
<b>10</b>	CLASSE_ABCDEF	101512
<b>11</b>	SAISIE_LOT_RECEP	105389
<b>12</b>	SAISIE_LOT_INTERNE	105394
<b>13</b>	MODE_SORTIE	105485
<b>14</b>	GESTION_NUM_CONTENANT	105394
<b>15</b>	GESTION_SSCC	105271
<b>16</b>	GESTION_LOT	105362
<b>17</b>	MULTI_LOTS_EMP	105289
<b>18</b>	DUREE_PEREMPTION	103280
<b>19</b>	DUREE_QUARANTAINE	103240
<b>20</b>	PRIX_UNITAIRE	105320
<b>21</b>	UNITE_STOCK	104171
<b>22</b>	PRECONISATION_EMP	105266
<b>23</b>	ID_FLUX_APPRO	46993
<b>24</b>	ID_FLUX_CONSO	41229
<b>25</b>	UM_NIV_1	57884
<b>26</b>	QTMAX_NIV_1	103476
<b>27</b>	UV	7
<b>28</b>	DIM_X	100598
<b>29</b>	DIM_Y	100599
<b>30</b>	DIM_Z	100600
<b>31</b>	VOLUME	100703
<b>32</b>	HOSPI_PRIX_MOYEN	61943
<b>33</b>	HOSPI_DOMAINE	14819
<b>34</b>	HOSPI_CODE_DCI	36362
<b>35</b>	<b>HOSPI_CODE_UCD</b>	<b>16650</b>
<b>36</b>	HOSPI_MODE_GESTION	104888
<b>37</b>	HOSPI_REF_FOUR	76567
<b>38</b>	<b>CODE_CIP</b>	<b>7103</b>
<b>39</b>	ID_FAMILLE_ACTION	76030
<b>40</b>	HOSPI_STERILE	70977
<b>41</b>	HOSPI_POINT_COMMANDE_FOUR	59101

TABLE 3.1 – Table des caractéristiques des produits de santé de la base

des établissements de santé en France comprend plusieurs niveaux (confirmés lors de nos entretiens avec les hôpitaux partenaires), de l'entité juridique à l'unité fonctionnelle (Voir la figure 3.1). Voici une description de chaque niveau :

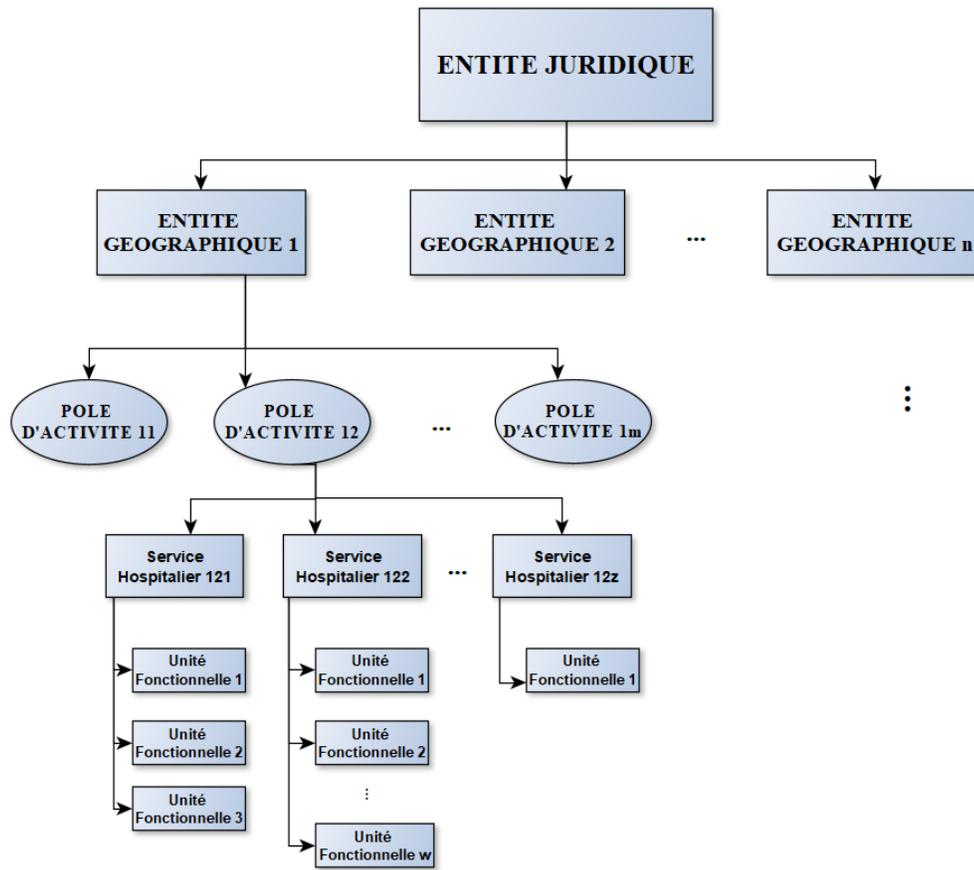


FIGURE 3.1 – Organisation et hiérarchisation des Hôpitaux

- **Entité juridique (EJ)** : Elle correspond à la définition de l'entité institutionnelle de la comptabilité publique qui possède un conseil d'administration ou un conseil de surveillance, une direction, et elle est maîtresse de sa décision. Elle peut englober plusieurs entités géographiques, comme l'exemple Assistance publique – Hôpitaux de Paris (AP-HP), qui regroupe plusieurs entités géographiques sur différents sites parfois relativement éloignés. Dans le secteur privé, elle représente souvent la société d'exploitation, même si des regroupements se produisent. Au sommet de la hiérarchie des établissements de santé, elle dirige la gestion globale et les choix stratégiques, typiquement sous forme d'établissement public de santé, de CHU ou de clinique privée.
- **Entité géographique (EG)** : Elle correspond en général au site de production, et éventuellement au site porteur du budget. Chaque EG dépend d'une unique entité juridique et une même entité juridique peut donner lieu à plusieurs EG, à condition que ces dernières aient des implantations géographiques ou des budgets différents.
- **Pôle d'activité cliniques ou médico-techniques** : La réforme de 2005 (Ordonnance n° 2005-406 du 2 mai 2005) a introduit les "pôles d'activités" dans les établissements

publics de santé, sauf les hôpitaux locaux. Ils regroupent des secteurs, services et unités fonctionnelles partageant des activités de soins, prévention, enseignement et recherche. Les pôles d'activité favorisent la coordination des soins et la gestion des ressources au sein de l'établissement <sup>1</sup>.

- **Service** : Les services hospitaliers sont un regroupement d'unités fonctionnelles au sein d'un pôle d'activité. Ils regroupent des équipes médicales et soignantes qui fournissent des soins spécifiques, tels que le service de cardiologie, le service de chirurgie ou le service de pédiatrie. Chaque service est dirigé par un chef de service.
- **Unité fonctionnelle (UF)** : Les UF sont des structures élémentaires de prise en charge des malades par une équipe soignante ou médico-technique, identifiées par leurs fonctions et leur organisation. Une définition appuyée par la Haute Autorité de Santé, qui stipule qu'une Unité Fonctionnelle (UF) est la plus petite unité de production identifiée dans un établissement de santé <sup>2</sup>. Les UF forment les bases des services ou départements, et une seule UF peut constituer un service indépendant si elle n'a pas de complémentarité directe avec d'autres unités similaires. Plusieurs UF de même discipline peuvent constituer un service médical. Le code de la santé publique en son Article L6146-1 et la BNDS <sup>3</sup> fournissent davantage de clarification. Les unités sont mises sur pied relativement aux activités et spécialités de soins et l'organisation de l'hôpital. Il existe plusieurs activités hospitalières que l'on peut regrouper en 5 catégories à savoir : MCO(Médecine, Chirurgie, Obstétrique), SSR (soins de suite et de réadaptation), SLD (Soins de Longue Durée), Psychiatrie (PSY) et HAD (Hospitalisation à domicile).

Les unités fonctionnelles ne sont pas à confondre avec les unités médicales (UM) qui sont un ensemble individualisé de moyens matériels et humains assurant des soins à des malades hospitalisés.

L'unité fonctionnelle étant la plus petite entité organisationnelle d'un établissement de santé, la quantification de la consommation de médicaments, produits de santé en général, est plus appréciable à ce niveau. De ce fait nous allons nous focaliser sur cette entité dans le cadre de cette thèse, pour apprécier et agréger les consommations par hôpital. Il nous revient également des différents interviews et questionnaires effectués auprès des hôpitaux que :

- Lorsqu'il s'agit d'apprécier la consommation de médicaments et de produits de santé, l'unité fonctionnelle peut jouer un rôle important. En regroupant les professionnels de santé et les services impliqués dans un domaine spécifique, elle facilite la collecte et l'analyse des données relatives à la consommation de médicaments.

---

1. Plus de détail consulter le BNDS (Bibliothèque Numérique de Droit de la Santé et d'Ethique Médicale) <https://www.bnds.fr/edition-numerique/dictionnaire/pa.html>  
Accessed on : June 29, 2023.

2. UF : [https://www.has-sante.fr/jcms/c\\_1011366/fr/unite-fonctionnelle-uf](https://www.has-sante.fr/jcms/c_1011366/fr/unite-fonctionnelle-uf)

3. BNDS : Bibliothèque Numérique de Droit de la Santé et d'Ethique Médicale <https://www.bnds.fr/edition-numerique/dictionnaire/uf.html>

Accessed on : June 29, 2023.

- Au niveau de l'unité fonctionnelle, il est possible de suivre de près les prescriptions médicales, les stocks de médicaments, les utilisations spécifiques des produits de santé, et d'effectuer des évaluations de l'efficacité et de la sécurité des traitements.
- L'appréciation de la consommation de médicaments au niveau de l'unité fonctionnelle permet de surveiller les pratiques, de détecter d'éventuelles variations ou anomalies, et d'ajuster les protocoles de prescription et de gestion des médicaments pour garantir une utilisation rationnelle et sécurisée.
- Elle peut également contribuer à l'optimisation des ressources en permettant une gestion plus précise des stocks, des approvisionnements et des coûts associés à la consommation de médicaments et de produits de santé.

#### b) Description des bases de données des unités fonctionnelles

La base de données des unités fonctionnelles, nommée « *Tiers* » et mise à notre disposition, contient les informations relatives aux destinations de l'approvisionnement (potentiellement des unités fonctionnelles) des produits de santé dans les différents hôpitaux partenaires. Cette base représente environ 7,5 MB de données, dont 127 486 lignes et 4 colonnes. Avant toute exploitation, une exploration et une fouille de données sont nécessaires pour identifier les informations utiles pour les futurs travaux.

A l'aide de requête SQL, une extraction est effectuée pour collecter les données relatives aux unités fonctionnelles des six hôpitaux. Suite à l'exécution d'un script correspondant pour l'obtention des données, la table récapitulative 3.2 est générée. Dans ce tableau récapitulatif, la répartition des informations par colonne pour chaque hôpital est indiquée.

Les quatre variables (colonnes) de la base de données sont : "ID\_SITE\_RATTACHE", "ID\_TIER", "TYPE\_TIER", "DESCRIPTION". La variable "ID\_SITE\_RATTACHE" a la même description que celle fournie dans la description de la base de données des articles et qui indique l'hôpital où la consommation a eu lieu. La variable "ID\_TIER" désigne le code de l'identifiant unique, propre à chaque hôpital, pour la destination du produit de santé dans la distribution. La variable "DESCRIPTION" fournit, comme son nom l'indique, le nom ou la description de l'"ID\_TIER" dans chaque hôpital. La variable "TYPE\_TIER" désigne le type de destinataire. Pour cette variable, quatre types de valeurs sont rencontrés à savoir : 'CLIENT', 'PROPRIETAIRE', 'TRANSPORTEUR', 'FOURNISSEUR'. Ici, ce qui nous intéresse est la valeur 'CLIENT'. Il s'agit d'une clé pour signifier la destination de consommation au sein de l'hôpital (y compris les unités fonctionnelles). Au stade de l'extraction, les unités fonctionnelles sont difficilement identifiables, un prétraitement des données recueillies est nécessaire.

##### 3.1.1.3 Bases de données historiques

La base de données appelée « *Historique* » est dédiée à l'historique de consommation des produits de santé dans les différents hôpitaux partenaires qui nous a été fournie. Cette

ID_SITE	ID_TIER	TYPE_TIER	DESCRIPTION
H1	19 645	19 645	19 558
H2	20 020	20 020	20 018
H3	44 395	44 395	44 366
H4	1 660	1 660	1 660
H5	4 904	4 904	4883
H6	16 632	16 632	16 623

TABLE 3.2 – Répartition des informations par hôpital

base a une taille d'environ 5,37 Go et est constituée de 27 colonnes et de 30 722 265 lignes. L'extraction des données, comme pour les autres bases, est réalisée à l'aide de scripts SQL Oracle. Une première exploration nous a permis de résumer les informations dans le tableau 3.3. Dans ce tableau, figure le nom de chaque colonne ainsi que le nombre d'informations manquantes dans ces colonnes de la base de données extraites. Le fait d'avoir un nombre significatif de données manquantes dans certaines colonnes entraînera des exclusions.

Après avoir obtenu la base de données, des échanges ont eu lieu avec plusieurs acteurs de l'entreprise KLS concernant les données. Une exclusion de 10 caractéristiques jugées non utiles dans le cadre de notre travail a été réalisée, ce qui nous a conduit à retenir les variables suivantes : TYPE\_FLUX, ID\_FLUX, DESIG\_FLUX, ID\_LIGNE\_MOUV, ID\_REF, ID\_CLIENT, QTE\_PREVUE, QTE\_MOUV, DATE\_MOUV, ID\_SERVICE, DATE\_FABRICATION, STATUT\_STOCK\_INITIAL, STATUT\_STOCK\_FINAL, ID\_ZONE\_ORIG, ID\_ZONE\_DEST, ID\_SITE\_RATTACHE. Cependant, au vu de la table 3.3, les variables "ID\_SERVICE" et "DATE\_FABRICATION" sont très peu renseigné, elles ont également été exclues.

Ces informations proviennent de la plate-forme WMS implantée dans les hôpitaux pour la gestion des stocks. La variable "TYPE\_FLUX" désigne le type de mouvement de produits de santé, avec deux valeurs possibles : "CONSOMMATION" et "APPROVISIONNEMENT". "CONSOMMATION" indique que le produit est distribué à une unité fonctionnelle pour consommation. L'unité fonctionnelle est identifiée par la variable "ID\_CLIENT", équivalent de "ID\_TIER" dans la base de données des unités fonctionnelles. "APPROVISIONNEMENT" signifie que le produit a été approvisionné et stocké suite à une commande de l'hôpital. La valeur de la variable "TYPE\_FLUX" retenue est "CONSOMMATION". Chaque produit est identifié par un code "ID\_REF" propre à chaque hôpital. Les flux de consommation et d'approvisionnement sont identifiés par les variables "ID\_FLUX" pour l'identifiant du flux et "DESIG\_FLUX" pour le libellé du flux.

Les différents mouvements d'approvisionnement et de consommation sont identifiés par la variable "ID\_LIGNE\_MOUV". Les variables "QTE\_PREVUE" et "QTE\_MOUV" représentent respectivement la quantité de produits de santé demandée par le destinataire et la quantité réellement fournie. La date de fabrication d'un produit est renseignée à travers la variable "DATE\_FABRICATION". Les statuts des stocks sont indiqués par les variables "STATUT\_STOCK\_INITIAL" et "STATUT\_STOCK\_FINAL" (en attente, colisage, disponible ou bloqué). Les variables "ID\_SITE\_RATTACHE", "ID\_ZONE\_ORIG",

	Nom de la colonne	Nombre d'informations manquantes
1	TYPE_ORDRE	4,098,277
2	TYPE_FLUX	0
3	STATUT_STOCK_INITIAL	1 027 939
4	STATUT_STOCK_FINAL	0
5	QTE_PREVUE	0
6	QTE_MOUV	0
7	ID_ZONE_ORIG	3 437
8	ID_ZONE_DEST	0
9	ID_SITE_RATTACHE	0
10	ID_SERVICE	10 946 918
11	ID_REF	0
12	ID_ORDRE	351 289
13	ID_OPERATEUR	0
14	ID_LOT_FOURNISSEUR	16 789 661
15	ID_LOT	14 165 315
16	ID_LISTE_MOUV	0
17	ID_LIGNE_ORDRE	0
18	ID_LIGNE_MOUV	0
19	ID_FOURNISSEUR	11 743 035
20	ID_FLUX	0
21	ID_CREATEUR	0
22	ID_CLIENT	11 913
23	DLV	16 835 039
24	DLUO	14 193 763
25	DESIG_FLUX	0
26	DATE_MOUV	0
27	DATE_FABRICATION	16 818 224

TABLE 3.3 – Synthèse des caractéristiques de la base de données

et "ID\_ZONE\_DEST" désignent respectivement l'identifiant de l'hôpital, l'origine du produit et la destination du produit, comme les unités fonctionnelles ou la destruction, etc.

Suite à l'extraction des données, la table récapitulative 3.4 a été générée, montrant la répartition des lignes d'informations par hôpital. Chaque ligne représente une consommation de produits de santé dans un hôpital à un moment donné de la journée.

Cette table est une vue globale de l'historique de consommation de tous les produits de santé par hôpital.

Hôpital	Nombre de lignes	Date Min	Date Max
H1	5 012 689	14/01/2013	30/12/2019
H2	3 028 544	30/03/2017	30/12/2019
H3	14 331 589	20/05/2014	30/12/2019
H4	2 606 925	01/06/2018	29/03/2021
H5	439 370	17/05/2019	02/05/2020
H6	5 303 148	01/01/2016	05/03/2019

TABLE 3.4 – Répartition des données de consommation par hôpital dans la base

### 3.1.2 Bases de données externes

En sus aux bases de données des hôpitaux partenaires fournies par l'entreprise KLS, plusieurs bases des données ont également été explorées et exploitées tout en demeurant dans la sphère de la problématique que nous traitons. Il s'agit notamment de trouver des bases de données en lien avec les facteurs impactant la consommation des produits de santé identifiés dans la littérature et suite à des analyses avec des experts. Les études sur les facteurs impactant la consommation des produits pharmaceutiques sont conduites dans le chapitre 4. Les informations externes relatives à la caractérisation des hôpitaux sont également recherchées en complément à ce que nous disposons déjà.

Lors de nos recherches de données, nous avons été confrontés à deux types de bases de données : les "*Open data*" ou « données open source » et les "*Private data*".

#### 3.1.2.1 Données open sources :

Le système national des données de santé (SNDS) constitue une plateforme pour chaîner les données de l'assurance maladie (base SNIIRAM), base de données des hôpitaux (base PMSI), les causes médicales de décès etc., (Pour plus de détail consulter le site web de la SNDS<sup>4</sup>). Il est à noter que toute personne ou structure, publique ou privée, à but lucratif ou non-lucratif, peut accéder aux données du SNDS sur autorisation de la Commission Nationale de l'Informatique et des Libertés (CNIL)<sup>5</sup>, en vue de réaliser une étude, une recherche ou une évaluation présentant un intérêt public. Dans cette section, nous relatons les open sources.

##### i) Données de la DREES :

La Direction de la recherche, des études, de l'évaluation et des statistiques (DREES<sup>6</sup>) est une structure agissant sous la tutelle des ministères : des solidarités et de la santé, du travail de l'emploi et de l'insertion, ainsi que du ministère de l'Économie, des finances et de la relance depuis 1998. La DREES est le service statistique public du Ministère des solidarités et de la santé. Elle fait partie du service statistique public piloté par l'INSEE. Parmi leur mission, la DREES a pour but de réaliser des études et produire des statis-

4. Lien SNDS : <https://www.snds.gouv.fr/SNDS/Qu-est-ce-que-le-SNDS>

5. Lien CNIL : <https://www.cnil.fr/fr/cnil-direct/question/la-cnil-cest-quoi>

6. Lien de description de la DREES : <https://drees.solidarites-sante.gouv.fr/drees>

tiques entrant également dans le cadre de la santé. Un espace dénommé Data.DREES<sup>7</sup> permet la mise à disposition de tableaux de données dynamiques abordant les thèmes de la santé et de la protection sociale. La DREES réalise donc des statistiques annuelles au niveau des établissements de santé appelées "Statistique Annuelle des Établissements de santé (SAE)". Il s'agit d'une enquête administrative exhaustive obligatoire, principale source de données sur tous les établissements de santé de France<sup>8</sup>.

En ce qui nous concerne nous y avons trouvé :

- Des informations et données générales sur les établissements de santé<sup>9</sup> dans cette base de données nous retrouvons, sans être exhaustifs, le nombre de personnels médicaux et non médicaux, le nombre de places et lits par établissement, le nombre de séjours par activité et groupes d'activités hospitalières (MCO, SSR, SLD, PSY, HAD) et biens d'autres activités autorisées ou spécifiques qui sont fournis annuellement par établissement de santé.
- Hormis les bases de données en lien avec les statistiques des établissements de santé (SAE), (à la dernière date de consultation 01 juillet 2023), près de 126 autres bases de données sont fournies. On y retrouve des tables dynamiques sur le nombre d'hospitalisations selon les motifs de recours (pathologie ou morbidité), le sexe, la classe d'âge, l'année et le niveau géographique de résidence (département ou région). Des données qui sont fournies par la DREES depuis 2010<sup>10</sup> sont analysées dans la prise en compte des facteurs de consommation.

## ii) Open Data de la CNAMTS :

La Caisse nationale d'assurance maladie des travailleurs salariés (CNAMTS) publie dans sa rubrique « *Statistiques et Publications* » des données relatives aux dépenses d'assurance maladie, à la consommation de soins et à l'activité des professionnels de santé, ainsi que les derniers rapports, analyses et études statistiques ou médico-économiques. Des études produites par la Direction de la Stratégie, des Études et des Statistiques (DSES) et issues du Système National d'Information Inter Régimes de l'Assurance Maladie (SNIIRAM). Parmi les opens data du SNIIRAM, nous pouvons noter celles relatives aux médicaments nommées « Open Medic ». Ces données présentent les classes anatomiques et thérapeutiques, les codes CIP, les caractéristiques des bénéficiaires et les spécialités des prescripteurs, notamment le nombre de consommateurs, les montants remboursés et remboursables, ainsi que le nombre de boîtes délivrées.

Le jeu de données complet « Open Medic » permet d'étudier les dépenses annuelles liées aux médicaments (montants remboursés et remboursables) ainsi que le nombre de

---

7. Lien espace de Data.DREES : <https://data.drees.solidarites-sante.gouv.fr/pages/accueil/>

8. Lien de description de la SAE : [https://data.drees.solidarites-sante.gouv.fr/explore/dataset/708\\_bases-statistiques-sae/information/?sort=-annee](https://data.drees.solidarites-sante.gouv.fr/explore/dataset/708_bases-statistiques-sae/information/?sort=-annee)

9. SAE Diffusion : <https://www.sae-diffusion.sante.gouv.fr/sae-diffusion/recherche.htm>

10. Visualisation des données : [https://odin.sante.gouv.fr/dataviz\\_drees/?defaultPage=Morbidit%C3%A9\\_Hospitali%C3%A8re\\_Tableau\\_2\\_\\_D%C3%A9partements&user=dataviz\\_sante&pass=dataviz\\_sante#2](https://odin.sante.gouv.fr/dataviz_drees/?defaultPage=Morbidit%C3%A9_Hospitali%C3%A8re_Tableau_2__D%C3%A9partements&user=dataviz_sante&pass=dataviz_sante#2)

boîtes délivrées, en fonction de différents descriptifs sur les bénéficiaires (tranche d'âge, sexe, région de résidence) ou de l'information sur la spécialité du prescripteur. Cette base de données est utilisée pour l'identification et l'uniformisation des références des produits de santé issus des bases de données des hôpitaux (Voir le lien des données <sup>11</sup>). Toutefois, ces données sont utilisées en complément aux données de PHAST (une base de données privée qui sera évoquée dans la sous-section suivante).

### iii) Données de l'ATIH

L'Agence technique de l'information sur l'hospitalisation (ATIH), un établissement public de l'État à caractère administratif, dispose de plateforme de restitution de données brutes et des indicateurs relatives aux établissements de santé appelé ScanSanté <sup>12</sup>. Cette plateforme met à disposition des acteurs du système de soins à la fois des données brutes et des indicateurs « prêts à l'emploi » nécessaires à la connaissance et à la gestion du secteur hospitalier. C'est l'une des grandes bases de données des activités hospitalières. On y trouve des données et informations relativement à 6 thèmes :

- Les indicateurs synthétiques caractérisant un établissement de santé, un panier d'établissements, une région, ou l'ensemble de l'activité d'un champ sanitaire
- Les activités : Les activités sont subdivisées en 5 catégories sur ScanSanté : MCO, HAD, SSR, Psychiatrie. Des fichiers fournissent les indicateurs des consommations et productions selon les champs d'activités. Des analyses des activités sont également restituées.
- Territoires et parcours de soins : Ce thème regroupe les outils permettant d'analyser finement l'offre de soins des établissements pour répondre aux besoins en soins de la population.
- Coûts-finances : On y trouve des données et analyses économiques des établissements de santé.
- Amélioration des prises en charge : On y restitue les indicateurs de qualité des soins et de sécurité des patients.
- Médico Social : Restitution des indicateurs de l'offre de soins du secteur médico-social.

Pour notre part, à travers la rubrique activité, nous avons identifié :

- Des tables relatives à la nomenclature générale des unités médicales qui nous aide à hiérarchiser les structures de santé en France. Ces tables ont contribué à nous aider dans l'analyse de l'identification des unités fonctionnelles et des services.
- Les données sur les médicaments et les dispositifs médicaux : Ces tableaux présentent par code UCD pour les molécules onéreuses et par code LPP pour les dispositifs médicaux. Des données qui nous ont servi dans l'identification, l'analyse

---

11. Données CNAMTS : [https://assurance-maladie.ameli.fr/etudes-et-donnees/donnees/liste-bases-de-donnees-open-data#Open\\_Medic](https://assurance-maladie.ameli.fr/etudes-et-donnees/donnees/liste-bases-de-donnees-open-data#Open_Medic)

12. Lien Plateforme ScanSanté : <https://www.scansante.fr/>

et pré-traitement des informations en lien avec les produits de santé des Hôpitaux (Voir <https://www.scansante.fr/applications/synthese-dmi-mo-sus>).

- Le nombre de séjours, de patients et d'établissements par an et par région ainsi que des tables par type de séjour, âge, sexe et catégorie d'établissement <sup>13</sup>.

iv) **Les données en provenance de data.gouv.fr :**

La plateforme « data.gouv.fr <sup>14</sup> », administrée par Etalab, rassemble les données publiques produites ou reçues dans le cadre d'une mission de service public. Elle encourage la réutilisation de ces données au-delà de leur utilisation première par l'administration. Différentes restitutions anonymes de données du SNDS sont hébergées sur cette plateforme. Étant donnée qu'il existe une connexion de cette plateforme avec les structures dans le domaine de la santé, nous pouvons retrouver des bases de données comme celles de la DREES.

Nous disposons désormais de la table *t\_finess* qui est le référentiel des structures sanitaires, sociales et médico-sociales de la base de données DataSanté <sup>15</sup>. Ce référentiel recense toutes les structures enregistrées dans le répertoire Finess, qu'elles soient entités juridiques ou entités géographiques, qui existent ou qui ont existé depuis 2004 et leur associent leurs principales caractéristiques (adresse, code-commune, statut juridique, catégorie, etc.). Ces données vont nous servir à déterminer les caractéristiques (type, position géographique, nombre d'établissement etc.) de nos hôpitaux.

v) **INSEE :**

L'Institut national de la statistique et des études économiques (INSEE) collecte, produit, analyse et diffuse des informations sur l'économie et la société françaises. Même s'il est possible de retrouver des données de l'INSEE sur la plateforme data.gouv.fr, à travers les données de l'INSEE nous parvenons à obtenir les données sur la mortalité, des revenus, des inégalités territoriales de niveau de vie en France, les statistiques sur le personnel et l'équipement de santé, les estimations de la population de 1975 à nos jours des régions et des départements (France métropolitaine et DOM) déclinées par sexe et par âge <sup>16</sup>. Ces informations socio-démographiques ont bonnes réputations quant aux déterminants de la consommation de produits de santé dans la littérature scientifique. Ces tables de données nous servent dans l'analyse des facteurs impactant la consommations de produits de santé.

---

13. Lien de données ATIH sur les patients <https://www.scansante.fr/applications/caracteristiques-des-sejours-par-region>

14. Lien du site web de data.gouv.fr : <https://www.data.gouv.fr/fr/pages/spd/reference/>

15. Base *t\_finess* disponible au : <https://www.data.gouv.fr/fr/datasets/referentiel-finess-t-finess/>

16. Données sur la population : <https://www.insee.fr/fr/statistiques/1893198>

#### 3.1.2.2 Base de données nécessitant une autorisation ou payantes

En France, plusieurs bases de données exigent une autorisation ou un paiement pour accès à leurs informations, y compris dans le domaine médical. Le Health Data Hub <sup>17</sup> créé en 2019 dans le cadre de la transformation du système de santé en France en est un exemple. Il agit comme une plateforme centrale pour la collecte, le stockage, l'accès et l'utilisation sécurisée et éthique des données de santé. Le Health Data Hub facilite la recherche et l'innovation en utilisant ces données pour améliorer les soins aux patients, promouvoir la recherche médicale et soutenir les politiques de santé publique. Il permet aux chercheurs, professionnels de la santé, institutions académiques et entreprises de demander l'accès aux données de santé pour des projets spécifiques après avoir obtenu les autorisations nécessaires et mis en place des garanties de protection des données. Le Hub rassemble des données de diverses sources, comme l'Assurance maladie, les hôpitaux, les laboratoires, etc., provenant notamment du SNDS. Cependant, toutes les données ne sont pas en open source ; une demande doit être soumise et approuvée par la CNIL. Les données sensibles, telles que les informations médicales, les données agrégées, génétiques, d'imagerie médicale et de recherche clinique, nécessitent des autorisations spécifiques pour y accéder. Nous n'avons pas eu accès à ces catégories de données pour nos analyses, dans le cadre de cette thèse.

PHAST est la source de données privées à laquelle nous avons eu accès par l'intermédiaire de l'entreprise KLS. PHAST est une entreprise de service numérique qui développe et déploie des services et produits dédiés à l'interopérabilité sémantique dans le domaine de la santé en France. Des bases de données de cette entreprise figurent l'ensemble des spécialités pharmaceutiques, médicaments, des dispositifs médicaux commercialisés en France bénéficiant d'une codification. Nous intéressons particulièrement à ces données qui nous permettent les identifications uniques des produits de santé issues des bases de données des hôpitaux à notre disposition. Les informations recueillies nous ont permis de vérifier l'authenticité et d'uniformiser les informations relatives aux produits de santé enregistrées dans les bases de données des hôpitaux.

#### 3.1.3 Étude de terrain

Même si la pandémie de la Covid-19 n'a pas facilité nos interactions avec les acteurs du monde hospitalier des hôpitaux partenaires, entre novembre 2021 et mars 2022, nous avons réussi à effectuer des visites, ainsi que des discussions en vidéo-conférences et par téléphone avec des experts. Les experts au niveau hospitalier sont constitués de trois pharmaciens hospitaliers, de deux cadres supérieurs de pôle pharmaceutique, de deux responsables du domaine administratif et logistique ainsi que quatre responsables de service opération. Cette phase, parmi les tâches de cette thèse, est appelée « étude de terrain ». Cependant, en ce qui concerne la thématique de la mise en place d'une nouvelle plate-forme de logistique hospitalière, nous n'avons pas eu l'opportunité d'y assister au cours de nos recherches.

---

17. Lien site web de Health Data Hub : <https://www.health-data-hub.fr/>

Le format que nous avons privilégié est une enquête sous forme de questionnaire. Les objectifs de ce questionnaire sont autour de trois grands points :

- (i) Le premier point concerne la structuration de l'hôpital. L'objectif est de déterminer avec la plus grande précision l'unité la plus pertinente permettant de mieux juger de la consommation des médicaments et produits de santé en général, c'est-à-dire de déterminer le niveau d'agrégation de la donnée de consommation. C'était le lieu également de s'assurer de l'hiérarchisation et organisation proposées lors de la description des unités fonctionnelles. Une confirmation a été obtenue à l'unanimité. Selon la hiérarchisation et le découpage des hôpitaux, ici des CHU, CHRU et CH (qui sont aussi des entités juridiques qui regroupent plusieurs établissements), il résulte que seule la consommation des produits de santé est mieux appréciée au niveau des unités fonctionnelles ou par groupes d'activités hospitaliers (MCO, SLD, SSR, HAD, PSY). De plus, nous avons entrepris de rechercher des informations sur la nomenclature standard des unités fonctionnelles et des services de soins hospitaliers. Il en résulte qu'il n'existe pas de nomenclature standard pour ces services ni pour ces unités fonctionnelles, chaque hôpital étant libre de créer sa propre nomenclature. Il est également important de noter que la nomenclature des unités médicales (UM) proposée par l'Agence technique de l'information sur l'hospitalisation (ATIH) diffère de la nomenclature des unités fonctionnelles (UF) de l'hôpital. Celle de l'ATIH est utilisée pour le suivi des médicaments référencés et pour les déclarations auprès d'organismes tels que la sécurité sociale et les organismes de Tarification à l'activité (T2A). L'absence d'une nomenclature standard pour les UF complique la traçabilité et la comparaison entre les hôpitaux. Pour résoudre ce problème, nous avons entrepris la catégorisation des services en fonction des activités qui leur sont associées. Cette catégorisation était soumise à l'appréciation des experts (voir la proposition de regroupement des services d'activités et le questionnaire en annexe B). Cependant, cette proposition de regroupement a suscité diverses opinions et n'a pas pu faire l'unanimité. Cette tâche, en plus du temps à lui consacrer, nécessiterait la mise en place et la collaboration d'une équipe pluridisciplinaire spécialisée dans le domaine médical, ce qui dépasse le cadre de cette thèse.
  
- (ii) Le deuxième grand point se résume en une question à savoir : *Quels sont les facteurs qui rentrent en compte dans la quantification des besoins de l'hôpital ?* Dans cette partie du questionnaire, la détermination des variables impactant la consommation des produits pharmaceutiques est recherchée. Munis des facteurs identifiés de la littérature scientifique, il s'agit d'obtenir des avis des experts et d'enquêter sur les critères privilégiés, en pratique, au sein des structures sanitaires. Les résultats seront évoqués dans le chapitre suivant, un travail qui nous a permis de retenir les potentiels facteurs impactant la consommation des produits pharmaceutiques. A ce travail, s'est ajoutée l'identification du code de traçabilité utilisé dans le monde hospitalier et de comprendre la relation entre le conditionnement et la consommation. Il en ressort que, pour la sérialisation des médicaments, à tenir compte dans les estimations, le code CIP est le plus approprié pour la détermination du médicament à sérialiser. En effet, outre son utilisation pour

l'identification des médicaments tout au long de la chaîne d'approvisionnement, de la fabrication à la dispensation et au remboursement, ce code permet également d'identifier le produit en fonction de son conditionnement spécifique. Le code UCD est une information utilisée pour le référencement national, tandis que le Code ATC n'est pas pertinent dans ce contexte.

- (iii) La troisième partie de l'enquête était de s'assurer de la disponibilité des données comme le nombre de lits, le nombre de personnel (médical et non-médical), le nombre de séjours par services et/ou par service ou UF, au sein de chaque hôpital. Cependant, cette phase n'a pas été très fructueuse, car l'un des arguments avancés par nos interlocuteurs était la confidentialité de ces informations. Il est à noter toutefois que ces données, agrégées annuellement, ont été extraites de sources externes fiables, comme nous l'avons précédemment mentionné.

Les différents points évoqués nous ont permis d'avoir une vision plus claire de l'orientation de nos travaux et des différentes hypothèses à formuler dans nos contributions.

## 3.2 Organisation et structuration des données

Une fois que les données brutes essentielles ont été rassemblées à l'aide de Python, SQL et Excel, il est nécessaire de les organiser et de les structurer avant d'aborder les séries d'analyses expérimentales. Cette étape est d'autant plus cruciale du fait que les données provenant de deux catégories de bases de données distinctes (internes et externes) doivent être fusionnées pour obtenir des données propices aux analyses et aux expérimentations en lien avec nos objectifs. Les procédures d'organisation et de structuration sont représentées dans la figure 3.2.

Dans cette démarche, en plus des processus d'extraction, on retrouve les phases de pré-traitements des bases de produits de santé et des unités fonctionnelles des hôpitaux. Les tables résultantes des pré-traitements sont ensuite croisées avec les données des historiques de consommation dans un premier temps. Ensuite, après avoir effectué la sélection et les pré-traitements des données externes, la première base de données croisée et structurée est de nouveau combinée avec les jeux de données externes pour aboutir à la base de données finales.

Les sous-sections suivantes détailleront davantage les procédures de pré-traitement et les résultats obtenus.

### 3.2.1 Pré-traitements des données

Au sein de cette sous-section, nous présentons en détail les pré-traitements appliqués aux bases de données internes des fournies pas l'entreprise partenaire ainsi que les résultats obtenus à partir de ces différentes étapes de pré-traitement.

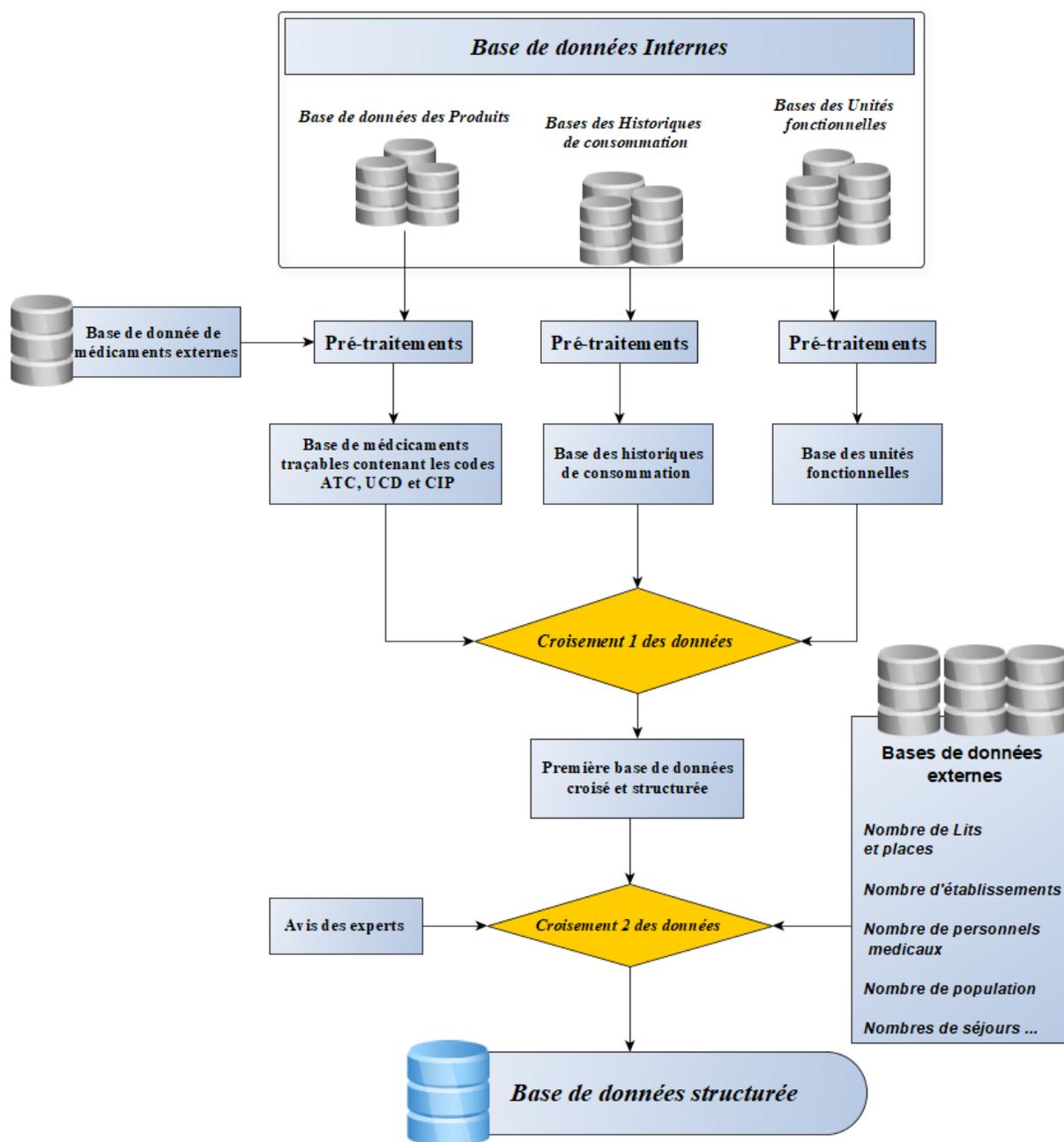


FIGURE 3.2 – Process global d’organisation des données

### 3.2.1.1 Pré-traitements des données des produits

#### a) Démarche employée

La démarche employée dans cette partie de nos travaux (illustrée dans la figure 3.3) se subdivise en trois phases. La première phase consiste à explorer les données dans son ensemble et d’établir des filtres afin d’en extraire les enregistrements comportant au moins une codification. En plus du code ID\_REF pour l’hôpital, il faut fournir au moins un code CIP ou UCD. Dans la deuxième phase, il s’agit de corriger et de nettoyer la base obtenue

dans la phase précédente. Cette étape est subdivisée en quatre étapes :

- Identification des potentiels mauvais codes par le biais des principes et standards de codifications (UCD et CIP).
- Une correction manuelle est effectuée pour certains codes. Par exemple, on retrouvait des codes CIP 13 qui avaient des 0 au début, ce qui pouvait être considéré comme mauvais.
- Vérification des codes potentiellement incorrects (identifiés manuellement) à partir de la base externe PHAST. À l'issue de cette étape, certains codes qui avaient été initialement considérés comme corrects se sont révélés incorrects.
- Correction et uniformisation des formats des codes. Parfois, nous pouvons avoir des codes composés de 7 ou 13 chiffres. Ces codes peuvent être utilisés pour les UCD ou les CIP, ou bien pour les deux à la fois. Il peut également y avoir des différences de codes entre différents hôpitaux.

La dernière étape, la troisième, est dédiée aux compléments d'informations à partir d'un code CIP ou UCD. Ces compléments d'informations peuvent être les codes CIP ou UCD, le code ATC, et les libellés du médicament. Ce travail nous fournira ce qu'on appelle *la table de référence des médicaments* qui nous aidera dans nos différentes expérimentations. La figure 3.3 résume le processus de pré-traitement des données de produits de santé.

Suite à l'extraction des données de la base de données des articles, on constate que tous les champs ne sont pas renseignés (voir table 3.1) et même si, ils sont renseignés, il faut vérifier la véracité des informations en particulier pour leurs codes de traçabilité. Il faut explorer et scruter aux détails près, la base pour tirer les informations exploitables et générer une base de référence pouvant nous servir dans le cadre de nos études expérimentales.

Relativement aux codes de traçabilité, comme l'atteste la table de répartition des codes par hôpital (Table 3.5), nous constatons à première vue que les hôpitaux, de façon générale, renseignent peu ou pas certains codes. Le code UCD est un peu mieux renseigné que les autres.

ID_SITE_RATTACHE	Nombre de ID_REF	Nombre de CODE_UCD	Nombre de CODE_CIP
<b>H1</b>	26 194	4 271	3 893
<b>H2</b>	5 273	4 232	3 095
<b>H3</b>	27 963	2 692	115
<b>H4</b>	21 113	2 904	0
<b>H5</b>	11 024	312	0
<b>H6</b>	14 380	2 239	0

TABLE 3.5 – Répartition des codes de traçabilité par hôpital

De même, suite à une première exploration, nous avons constaté qu'il y avait 367 mauvaises informations enregistrées dans la colonne (ID\_SITE\_RATTACHE), qui renseigne sur l'hôpital. 604 codes ID\_REF (supposés permettre d'identifier les produits dans un hôpital)

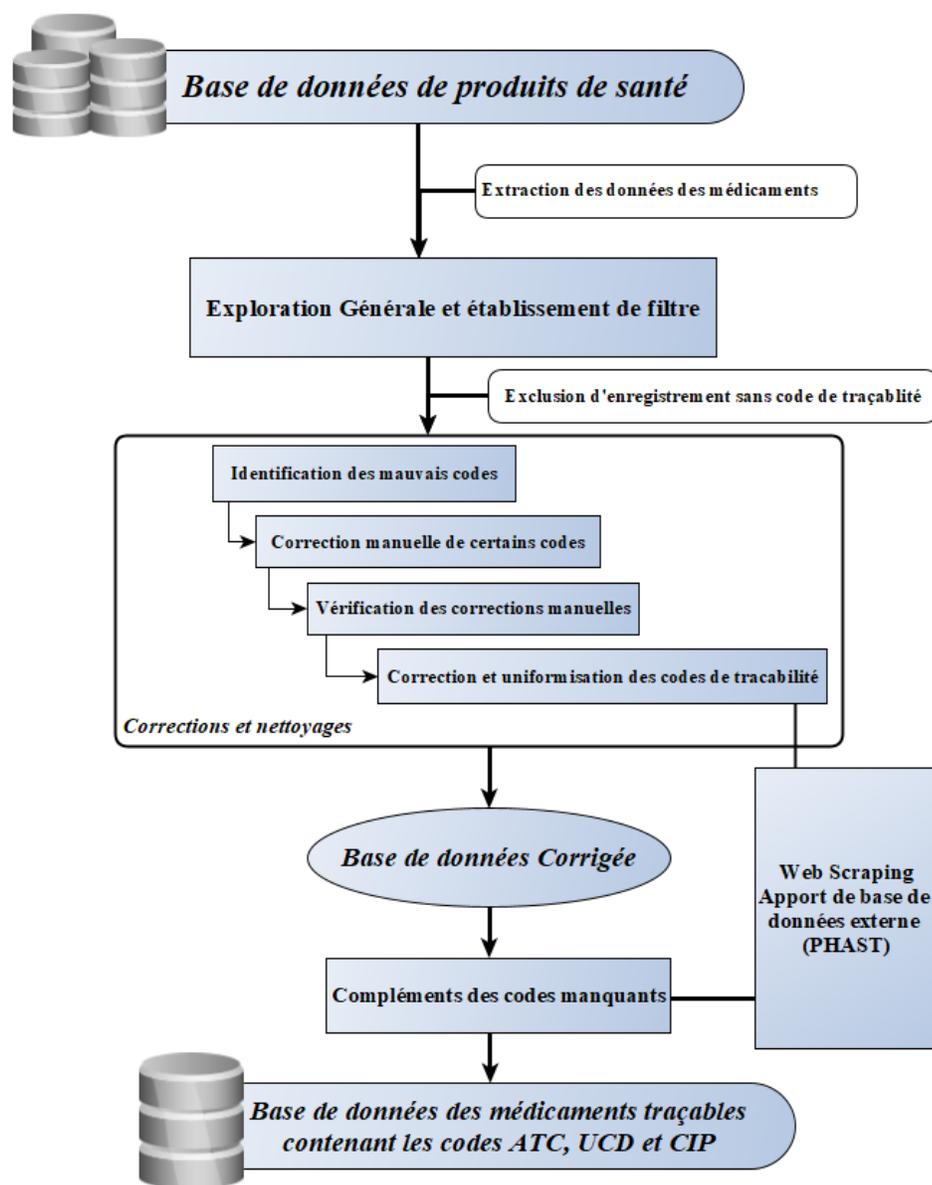


FIGURE 3.3 – Démarche employée pour le pré-traitement des données de produits de santé

ont été rattachés à ces faux hôpitaux. Une exclusion de ces données incorrectes est effectuée à cet effet tout en se focalisant sur les médicaments au lieu des dispositifs médicaux et autres produits de santé. Dans les paragraphes qui suivent, les analyses se feront sur les médicaments relativement aux codes ID\_REF, UCD et CIP des médicaments.

Les approches adoptées pour le pré-traitement des données comprennent l'utilisation de la fouille de données ainsi que de la technique de Web-scraping. Dans le cadre de la fouille de données, notre démarche consiste à explorer, nettoyer et corriger les données afin de les rendre utilisables. Parallèlement, le web-scraping est employé pour collecter des données externes en vue de vérifications et de corrections ultérieures.

La principale source externe de données médicamenteuses que nous avons utilisée est la base de médicaments de PHAST. Par ailleurs, la table des médicaments de CNAMTS a

également été mobilisée dans ce contexte. Lorsqu'un code UCD ou CIP est disponible dans la base de produits de santé des hôpitaux, les tables externes nous ont permis de retrouver les informations manquantes relatives à ce produit. Par ailleurs, les informations concernant les codes ATC des médicaments qui ne figuraient pas dans la base des hôpitaux ont été entièrement obtenues à partir de la base de données de PHAST.

Le web scraping, également appelé extraction de données web, est une technique permettant d'extraire automatiquement des informations à partir de sites web[165]. Cela implique l'utilisation de programmes informatiques pour collecter et analyser le contenu des pages web, puis extraire les données souhaitées dans un format structuré et exploitable. Le processus de web scraping comprend généralement les étapes suivantes :

- Analyse de la structure du site web
- Récupération du contenu HTML
- Extraction des données
- Transformation et stockage des données.

En rappel, le but de PHAST est de faciliter et de sécuriser les échanges de données dans un contexte d'interopérabilité des systèmes d'information (SI). PHAST produit notamment des référentiels de "Codification Inter-Opérable" (CIO) de pharmacie (Spécialités Pharmaceutiques et Dispositifs Médicaux) et a mis en place en 2011 le premier service de distribution des terminologies de santé.

## b) Résultats du traitement

Sur l'ensemble des 121 343 lignes d'enregistrements issues de la base de données des Articles, en se focalisant sur les médicaments qui comportent au moins un code UCD ou CIP, il reste 14740 sans doublons de d'identifiant ID\_REF. Ce qui sous-entend qu'après exclusion d'enregistrements sans codes de traçabilité (UCD ou CIP) on se retrouve avec environ 11,14% des données supposées exploitables.

La phase de l'identification des mauvais codes a permis de constater que :

- Il y a 2 types de mauvais codes CIP à savoir :
  - Mauvaise saisie des codes ne respectant pas les normes de sérialisation : "0051002121", "685020", "340093008096", "7286761", "SA374003000000", "34009311527054" et "1176042".
  - 3030 codes qui avaient un chiffre zéro de trop au début de leurs codes CIP 13.
- Il y a 764 de mauvais codes UCD. En exemple nous avons : "X", "0", "0802290", "92506326", "0800265", "XCVBN", "1631-FRE", "9402059\xa0", "60000233", "0800060", "F0000065D", "2045-40000008", "870006", "M", "F0000066", etc.

Suite aux différentes vérifications et corrections, 1216 lignes d'enregistrements ont été exclues. Ce qui porte désormais le nombre d'enregistrements de médicaments comportant au moins un code à 13524 (voir les répartitions dans la table 3.6) qui va constituer la base de données corrigées de référence des articles (la table de médicaments traçables). Les corrections et vérifications des informations des médicaments de chaque hôpital se font sur la

base des informations fournies par la base externe PHAST par le bien d'un certain nombre d'algorithmes que nous avons mis sur pied.

ID SITE RATTACHE	ID_REF	CODE_UCD	CODE_CIP	CODE_ATC
H1	2281	2281	2223	2281
H2	3687	3687	3549	3687
H3	2652	2652	2092	2652
H4	2421	2421	1897	2421
H5	303	303	261	303
H6	2180	2180	1685	2180

TABLE 3.6 – Répartition des codes de médicaments par hôpital après traitements

A cette table de médicaments traçables, on complète les codes manquants (voir la table illustrative tableau 3.7).

	SITE	ID_REF	DESIG	CODE_UCD	CODE_CIP	CODE_ATC	Libele ANSM	Nombre de CIP
0	H1	960412	FOLINATE DE CALCIUM 350MG AGUETTANT,	3400892330365	3400935728142	V03AF03	FOLINATE DE CALCIUM 350 mg AGUETTANT, pdr pr	1
1	H1	960407	VOLTARENE EMULGEL 1%, GEL, TUBE 50 G	3400891252002	3400932886975	M02AA15	VOLTARENE EMULGEL 1%, gel, tube 50 g	1
2	H1	960405	PAMIDRONATE DE SODIUM 15MG/5ML SOL ET	3400893044285	3400938169669	M05BA03	PAMIDRONATE DE SODIUM 15 mg/5 mL HOSPIRA, sol ...	1
...	...	...	...	...	...	...	...	...
13521	H6	123116	TIVICAY 50MG COMPRIME	3400893988664	3400927714634	J05AJ03	TIVICAY 50 mg, cpr	1
13522	H6	123121	MAVIRET 100MG/40MG COMPRIME	3400894287391	3400930108765	J05AP57	MAVIRET 100 mg/40 mg, cpr	1
13523	H6	126188	DUODOPA 20MG/ML GEL INTEST SACH +4	3400892680002	3400936511002	N04BA02	DUODOPA 20 mg/5 mg/mL, gel intestinal, sachet ...	1

TABLE 3.7 – Esquisse de la table de référence des médicaments traçables dans les 6 hôpitaux

Dans cette table, il arrive que le code CIP ne soit pas renseigné. Dans ces situations, il y a un manque d'information sur le code CIP dans la base d'origine. L'utilisation du code UCD pour retrouver le CIP correspondant conduit à plusieurs propositions de CIP. Dans ces

cas de figures, aucun code CIP n'est retenu pour éviter les ambiguïtés. Cependant, dans la table récapitulative, la colonne « Nombre de CIP » précise le nombre de codes CIP trouvés correspondant au code UCD utilisé. Nous déduisons de cette table également :

- 5 674 médicaments distincts ont été identifiés, si on tient compte uniquement des codes UCD. Seuls 70 codes UCD sur 5 674 sont en commun aux 6 hôpitaux.
- Relative au code ATC, 115 codes sur 1615 sont en commun aux 6 hôpitaux seuls.

Il est essentiel dans notre étude que les médicaments qui vont servir pour les expérimentations soient des médicaments en commun. Le médicament doit être identique dans tous les hôpitaux. Le nombre de médicaments en commun, au regard des codes UCD est très faible. Cela peut s'expliquer par plusieurs facteurs :

- La diversité de fournisseurs et distributeurs de médicaments.
- Le faible ou le manque d'enregistrement des médicaments avec au moins un code de traçabilité dans le système de gestion des hôpitaux

La base de référence de médicaments, obtenue à ce stade, sera utilisée pour les croisements, notamment avec la base de données de l'historique de consommation.

### 3.2.1.2 Pré-traitements des données des unités fonctionnelles

#### a) Démarche employée

Suite à l'extraction des données relatives aux unités fonctionnelles, de la répartition par hôpital des informations fournies dans table 3.2, une question est de savoir si les informations sont tous exploitables pour la suite de nos travaux. Aussi, tout comme pour les médicaments, il faut identifier les UF en communs pour favoriser les analyses futures. De ce fait, la démarche illustrée dans le graphique 3.4 est proposée.

Après l'extraction des données relatives aux unités fonctionnelles, il s'agit d'identifier ces unités en incluant un filtre sur la variable "TYPE\_TIER". Comme stipulé dans la description de cette base, nous nous focaliserons sur la valeur "CLIENT" de cette variable. Une nouvelle table est ainsi obtenue.

La deuxième phase de notre démarche consiste à vérifier que toutes les lignes de cette nouvelle table contiennent des informations correctes. Il s'agit principalement de vérifier, par hôpital, les informations des colonnes "ID\_TIER" et "DESCRIPTION". Nous ne souhaitons pas avoir de valeurs NULL (pas de donnée), en particulier dans la colonne "ID\_TIER".

L'étape suivante consiste à lier les informations des UF obtenues avec d'éventuelles informations standards sur les UF, au niveau hospitalier et national. Il s'agit également d'identifier les UF communes et de générer une table de référence des unités fonctionnelles pour les futures analyses.

#### b) Résultats du traitement

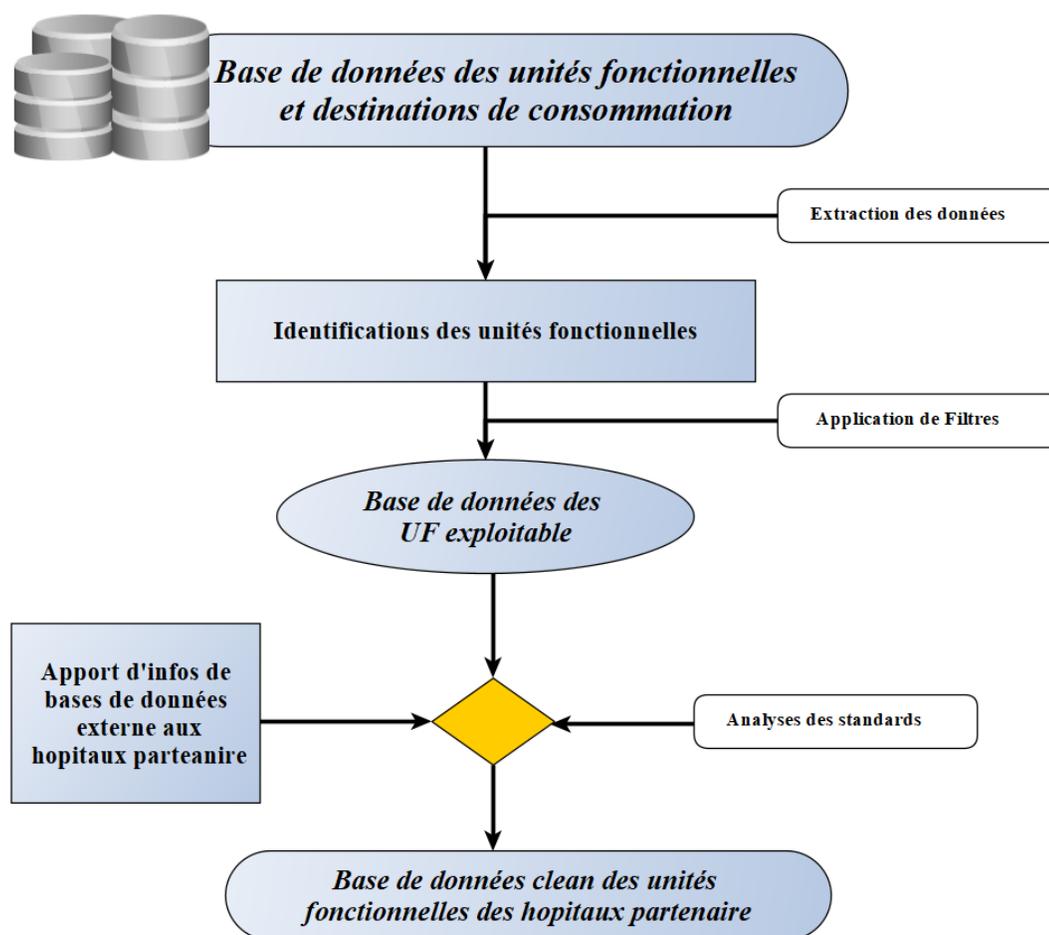


FIGURE 3.4 – Processus de pré-traitement des données des unités fonctionnelles

Sur l'ensemble des 127 486 lignes d'enregistrements issues de la base de données, après l'application des filtres (`TYPE_TIER = "CLIENT"`, code `ID_TIER` non Null et `DESCRIPTION` de valeur non Null), il reste 13 684 lignes, soit environ 10,73% des données exploitables. La répartition par hôpital de ces données est fournie dans la table 3.8. Nous disons exploitables, car ces destinations de consommation ne sont pas uniquement pour les médicaments. De plus rien ne nous garantit encore que chaque ligne d'informations par hôpital représente effectivement une unité fonctionnelle à ce stade.

L'identification effective des unités fonctionnelles peut se faire à travers trois points :

1. Croisement des informations de notre nouvelle base de données des UF avec des bases de données externes. Il peut s'agir de bases de données d'organismes officiels, ou des données fournies par l'hôpital qui ne nous sont pas soumises. Soit des bases de données regroupant les informations sur les standards nationaux sur les unités fonctionnelles comme c'est le cas pour les médicaments comme les codes UCD, CIP, etc.
2. Vérifications élémentaires des informations de cette nouvelle base avec des spécialistes de chaque hôpital.

ID SITE RAT-TACHE	ID_TIER	DESCRIPTION
H1	1 726	1 726
H2	990	990
H3	3 840	3 840
H4	830	830
H5	626	626
H6	5 672	5 672

TABLE 3.8 – Répartition des destinations de consommations par hôpital

3. Croisement des informations des UF de cette base avec la base de données de l'historique de consommation des médicaments.

Pour le premier point, en ce qui concerne le croisement des informations, nous avons lancé en premier lieu des recherches sur des nomenclatures et codifications standards des unités fonctionnelles. Ce qui nous a conduit à visiter les sites-web officiels de chaque hôpital.

Cela nous a permis d'avoir une vision des activités, services et souvent les pôles d'activités. Les descriptions des unités fonctionnelles ne sont pas fournies sur ces sites. La technique du webscraping nous a servi uniquement à obtenir une tables des services. En sus, nous avons visité la plateforme de ATIH<sup>18</sup>. Cette agence propose, au lieu de tables des unités fonctionnelles, des table de nomenclatures sur les unités médicales autorisées. Il s'agit de tables de références de classification des unités. Ces unités sont relatives au MCO et SSR avec chacune une codification. La codification de ces unités médicales<sup>19</sup> est en trois caractères (alphanumérique) au maximum. Des codes qui sont largement différents des codes (ID\_TIER) fournis dans nos données. Il en est de même pour les descriptions. Ce qui rend impossible l'établissement de lien. Avec la base de données Finess (t\_finess) issue de data.gouv.fr, le constat a été pareil, nous n'avons pas pu établir le lien avec les unités fonctionnelles de notre base de données avec les données externes. Ces faits conduisent en première intention de déduire qu'il n'existe pas de standard national de nomenclatures et codification des unités fonctionnelles en France. Cependant, il reste à vérifier cette affirmation auprès des spécialistes des hôpitaux, d'où le deuxième point. L'étude sur le terrain confirme qu'il n'y a pas de nomenclature standard. Nous devons donc nous appuyer sur les bases de données disponibles. Cependant, il est difficile de trouver des unités fonctionnelles communes. En effet, bien que les codes ID\_TIER puissent être similaires, les unités fonctionnelles correspondantes sont différentes. Pour déterminer l'efficacité des unités fonctionnelles, nous avons donné la priorité au croisement des informations des UF avec les bases de données de l'historique de la consommation des médicaments.

À la question de introductive de savoir : « Existe-t-il une nomenclature ou traçabilité

18. ATIH : <https://www.atih.sante.fr/>

19. Nomenclature des Unités médicales :

<https://www.atih.sante.fr/nomenclatures-de-recueil-de-l-information/autorisations-des-unites-medicales>

standard nous permettant d'établir une étude comparative relativement à cette unité fonctionnelle ? », la réponse est non.

### 3.2.2 Premier croisement des données

Suite aux pré-traitements des données sur les médicaments et sur les unités fonctionnelles, un croisement des informations recueillies est effectué avec les données de la base de données de consommation. Ce croisement nous a permis d'identifier les médicaments effectivement consommés et les unités fonctionnelles (ou les destinations de consommation) en lien avec la consommation de médicaments.

Suite à ce croisement, on obtient une base de données dont nous résumons les informations à travers la table 3.9. Cette table fournit un récapitulatif des répartitions des données de la nouvelle base historique de consommation.

Hôpital	Nombre de lignes	Date Min	Date Max	Nombre de Médicaments	Nombre d'unités fonctionnelles
H1	1 089 786	01/01/2016	30/12/2019	1 716	342
H2	1 559 578	30/03/2017	30/12/2019	2 424	326
H3	2 455 519	04/01/2016	30/12/2019	2 383	463
H4	471 065	01/06/2019	29/03/2021	1 738	335
H5	30 725	17/05/2019	02/05/2020	232	154
H6	640 695	01/01/2016	05/03/2019	1 843	679

TABLE 3.9 – Répartition des données de consommation après croisement avec les autres bases de données pré-traitées

La table contient le nombre de lignes de données dans chaque hôpital, le nombre de médicaments et des unités fonctionnelles (ou destinations de consommation) de chaque hôpital. Il y figure également les périodes des historiques de consommation des médicaments. De cette table, nous tirons que suite au croisement seuls 20,33% des données de la base de données des historiques de consommations de médicaments peuvent être exploitées. L'obtention de cette nouvelle base de données pré-traitée nous a permis, en sus, d'avoir une idée claire sur les unités fonctionnelles. Selon chaque hôpital, les potentiels unités fonctionnelles obtenu dans cette phase sont analysés manuellement. Des tables contenant les variables comme le nom de l'établissement de santé, le code et le nom de l'UF, le nombre de mouvements de médicaments, la quantité ainsi que le nombre de médicaments à destination de l'UF, sont établies pour analyser les UF. Deux constats sont faits :

1. Parmi les potentielles UF, sur toutes les périodes de consommation, certaines combinent un enregistrement des mouvement faibles, peu de médicaments pour une quantité faible. Après échanges avec les experts, il s'agit pour la plupart des destinations en lien avec la recherche scientifique. Un filtre est établi pour exclure une partie de cette catégorie

de destinations. Il s'agit de ceux ayant un nombre de mouvement inférieur au nombre d'années de consommation.

2. Certains UF semblent être de gros consommateurs selon les identifiants des UF. Ce sont des apparences trompeuses car il s'agit souvent d'identifiant de point de stockage d'une entité géographique chargée de refaire les distributions aux UF sous leurs coupes. N'oublions pas que les établissements partenaires sont la plupart de CHUs qui regroupent plusieurs établissements.

Suite à ce travail, nous avons pu identifier de manière précise les unités fonctionnelles de chaque hôpital et obtenir une base de données sur l'historique de consommation exploitable, contenant des détails sur chaque médicament tracé, ses unités fonctionnelles de consommation, etc.

Dans la première base de données croisée et structurée, nous avons inclus des informations telles que l'identifiant, la date et la quantité de produits consommés, ainsi que l'établissement et l'unité fonctionnelle. En général, nous avons conservé les attributs de la base de l'historique de consommation. La distribution des médicaments par hôpital se présente comme suit : H1 a 1716 médicaments, H2 a 2424 médicaments, H3 a 2383 médicaments, H4 a 1738 médicaments, H5 a 232 médicaments et H6 en a 1843. À partir de cette table, nous avons identifié 45 médicaments communs consommés, en tenant compte des périodes de consommation des 6 hôpitaux. La table 3.10 fournit des informations sur ces médicaments.

Il est important de noter que la diversité des fournisseurs dans les hôpitaux peut expliquer la faible quantité de médicaments en commun consommés. De plus, l'hôpital de H5, l'un des plus petits parmi les 6 hôpitaux et l'un des partenaires les plus récents de l'entreprise KLS, contribue également à cette limitation. Il convient également de souligner que le médicament le plus consommé de la base, le Doliprane 500mg, gélule, avec le code UCD 3400891996128 et le code ATC N02BE01 (le PARACETAMOL), ne figure pas dans la liste des médicaments communs aux 6 hôpitaux.

Désirant d'intégrer les facteurs impactant la consommation des médicaments, un second croisement s'impose, d'où le point suivant.

### 3.2.3 Second croisement des données

Le second croisement dans notre démarche d'organisation et structurations des données consiste à croiser la base de données issue du premier croisement avec des informations en provenance de bases de données externes.

De ce fait, suite à plusieurs analyse et pré-traitements, plusieurs jeux de données issues de 160 tables d'organismes officiels (comme de l'INSEE, data.gouv.fr, DREES, etc.), décrits dans la section précédente, sont utilisés, souvent agrégés, pour être mis en lien par croisement avec la première base de données croisée. Dans la base de données qui en découle figurent les informations :

- Nombre de lits par an, par hôpital, et par hospitalisation complète et partielle,

	<b>Libellé du médicament</b>	<b>CODE UCD</b>	<b>CODE ATC</b>
1	AMOX-AC CLAV 1 g/200 mg SANDOZ,	3400892975047	J01CR02
2	MABTHERA 500 mg/50 mL,	3400891977196	L01XC02
3	MIDAZOLAM 50 mg/10 mL MYLAN	3400892359281	N05CD08
4	DEXAMETHASONE 4 mg/1 mL MYLAN	3400892230849	H02AB02
5	MIDAZOLAM 5 mg/5 mL MYLAN	3400892359113	N05CD08
6	ORBENINE 1 g	3400891084436	J01CF02
7	MIDAZOLAM 5 mg/1 mL MYLAN	3400892376950	N05CD08
8	AMOX-AC CLAV 2 g/200 mg SANDOZ	3400892975108	J01CR02
9	BRILIQUE 90 mg	3400893804414	B01AC24
10	CATAPRESSAN 0.15 mg/1 mL	3400890168243	C02AC01
11	DECORENONE 50 mg	3400892461397	C01EB09
12	VESANOID 10 mg	3400891821840	L01XF01
13	REMICADE 100 mg	3400892137377	L04AB02
14	REVLIMID 25 mg	3400892981420	L04AX04
15	REVLIMID 10 mg	3400892981130	L04AX04
16	AMBISOME LIPOSOMAL 50 mg	3400892182612	J02AA01
17	AMOX-AC CLAV 500 mg/50 mg SANDOZ	3400892976808	J01CR02
18	RIVOTRIL 1 mg/1 mL	3400891996586	N03AE01
19	DOBUTAMINE 250 mg/20 mL PANPHARMA	3400892336398	C01CA07
20	CEFOXITINE 2 g PANPHARMA	3400892124421	J01DC01
21	MIMPARA 30 mg	3400892665214	H05BX01
22	REVLIMID 15 mg	3400892981369	L04AX04
23	HYPNOMIDATE 20 mg/10 mL	3400891203783	N01AX07
24	AMIKACINE 500 mg MYLAN	3400892841861	J01GB06
25	NOVORAPID 100 u/mL 3 mL FLEXPEN	3400892402437	A10AB05
26	REVLIMID 5 mg	3400892981598	L04AX04
27	ORGARAN 750 iu antiXa/0.6 mL	3400891839470	B01AB09
28	SODIUM BICARBONATE 1.4% LAVOISIER	3400891088519	B05XA02
29	INVANZ 1 g	3400892505725	J01DH03
30	HALDOL 5 mg/1 mL	3400890414722	N05AD01
31	LOXAPAC 50 mg/2 mL	3400891368994	N05AH01
32	PIPERACILLINE TAZOBACTAM 2 g/250 mg MYLAN	3400893022573	J01CR05
33	AMIKACINE 50 mg/1 mL MYLAN	3400892085937	J01GB06
34	DIFICLIR 200 mg	3400893843734	A07AA12
35	NOZINAN 25 mg/1 mL	3400890640718	N05AA02
36	NOVORAPID 100 u/mL	3400892229959	A10AB05
37	ORENCIA 250 mg	3400893001813	L04AA24
38	ROACTEMRA 80 mg/4 mL	3400893318911	L04AC07
39	BRIDION 200 mg/2 mL	3400893231012	V03AB35
40	HYDROCORTANCYL 125 mg/5 mL	3400890450911	H02AB06
41	ROACTEMRA 200 mg/10 mL	3400893318799	L04AC07
42	TUBERTEST 1 mL	3400892372297	V04CF01
43	URSOFALK 50 mg/mL 250 mL	3400892993195	A05AA02
44	FIRAZYR 30 mg/3 mL	3400893201664	B06AC02
45	NATRIUMTHIOSULFAT 25 g/100 mL, sol	3400893133446	V03A

TABLE 3.10 – Visualisation des médicaments en commun consommées dans les hôpitaux

- Nombre de séjours par an, par hôpital et par groupe d'activité (MCO, SLD, SSR, HAD, PSY),
- Nombre de personnel médical non interne par spécialité médicale, regroupée ici, par année et par hôpital,
- Nombre de personnel non médical par activité, regroupée ici, par an et par hôpital
- Nombre d'habitants par département (où se trouve l'hôpital) sauf pour l'hôpital de Roubaix.
- Nombre d'établissements rattachés à l'hôpital. Pour rappel, les hôpitaux de cette étude

sont en majorité des CHU , c'est-à-dire des hôpitaux universitaires qui sont des entités juridiques regroupant plusieurs établissements.

En plus de ces données dans la base finale, nous avons inséré le regroupement des UF en services. Cela permet de faciliter l'identification des types de services où les médicaments ont été consommés.

Dans notre contexte, la consommation fait référence à la livraison de médicaments aux unités fonctionnelles dans les hôpitaux concernées. En réalité il existe un autre type de consommation. Il s'agit de la Prescription-Délivrance et Administration à patient par le personnel médical. Ce type de consommation permet d'avoir en temps réel et au plus fin les quantités de produits consommés. Mais les bases de données y afférant ont été hors de notre porté.

Pour faciliter les travaux ultérieurs, nous avons procédé à une agrégation de la base structurée finale. Il s'agit de regrouper la consommation par mois. Le mois est un horizon temporel décidé avec les experts de l'entreprise KLS. De ce fait, une variable supplémentaire sera ajoutée, le nombre d'unités fonctionnelles où le produit a été consommé lors de la période indiquée (mois). Un aperçu de la table finale basée sur les 45 médicaments consommés dans les différents hôpitaux est fourni à travers la table 3.11. L'ensemble des variables découlant de ce jeu de données sont fournis dans le tableau A.1 de l'annexe A. Dans cette table finale nous retrouvons finalement avec 6218 lignes (sur 204550 lignes pour tous les médicaments traçables) et 100 colonnes. Dans les chapitres suivants, nous indiquerons les types de données utilisées.

ID SITE RAT-TACHE	ID_REF	DESIG	CODE ATC	ANNEE	HOSPI UCD	CODE CIP	MOIS	N ETB	...	PERS MED GLE	PN MEDICAL	P MEDICAL
0	H1	960266	NOVORAPID 100 IU/ML, SOL INJ, STYLO 3	A10AB05	2018	3400892402437	04	12	...	94	5362	776
1	H1	960266	NOVORAPID 100 IU/ML, SOL INJ, STYLO 3	A10AB05	2018	3400892402437	12	12	...	94	5362	776
2	H1	960239	CLOXACILLINE (ORBENINE) 1 G, PDR PR SO	J01CF02	2018	3400891084436	11	12	...	94	5362	776
3	H1	960239	CLOXACILLINE (ORBENINE) 1 G, PDR PR SO	J01CF02	2018	3400891084436	08	12	...	94	5362	776
4	H1	960162	AMOX-AC CLAV 500 MG/50 MG SANDOZ, NOUR	J01CR02	2018	3400892976808	02	12	...	94	5362	776
5	H1	894287	LENALIDOMIDE (REVLIMID) 25 MG, GÄLULE	L04AX04	2019	3400892981420	01	12	...	82	5303	813
6	H1	894243	ABATACEPT (ORENCIA) 250 MG, PDR PR SOL	L04AA24	2017	3400893001813	09	12	...	95	5303	777
7	H1	893088	NOVORAPID 100 IU/ML, SOL INJ, FLAC 10	A10AB05	2018	3400892229959	11	12	...	94	5362	776
8	H1	893088	NOVORAPID 100 IU/ML, SOL INJ, FLAC 10	A10AB05	2017	3400892229959	06	12	...	95	5303	777
9	H1	893088	NOVORAPID 100 IU/ML, SOL INJ, FLAC 10	A10AB05	2018	3400892229959	12	12	...	94	5362	776

TABLE 3.11 – Un aperçu de la table de données finale

### 3.3 Conclusion

Dans ce chapitre, nous avons tout d'abord fourni une description des sources de données. Trois principales sources de données ont été explorées : les bases de données internes des hôpitaux partenaires, les bases de données externes provenant principalement d'organisations officielles françaises, ainsi que des informations recueillies dans le cadre d'une étude de terrain en recueillant les avis d'experts des hôpitaux pour lesquels nous disposons de données. Ensuite, l'accent a été mis sur l'organisation et la structuration des données. Ce travail nous a permis de mettre en évidence les prétraitements et les croisements de données nécessaires pour obtenir des tables de données exploitables dans le cadre d'études expérimentales. Les données résultantes de ces divers traitements seront utilisées dans les chapitres à venir. Elles serviront notamment dans le chapitre suivant à l'analyse de corrélations des facteurs impactant la consommation des médicaments dans les hôpitaux. Il est important de noter que l'application des approches de machine learning dépend fondamentalement de la qualité des données, ce qui a demandé un temps considérable pour cette phase d'analyse.

En considérant la typologie et la qualité des données, l'analyse de données nous a conduits à nous concentrer sur les médicaments, avec 45 médicaments en commun consommés dans les 6 hôpitaux. Par ailleurs, un constat majeur a été fait concernant la faible qualité des enregistrements issus des bases de données hospitalières. En ce qui concerne les produits de santé, notamment les médicaments, seuls 11,14 % sont exploitables (c'est-à-dire traçables), et en ce qui concerne les destinations de consommation, notamment les unités fonctionnelles, nous avons constaté que seulement 10,73 % des enregistrements sont exploitables. Dans cette optique, nous recommandons aux décideurs de mettre davantage l'accent sur les informations clés de traçabilité afin de faciliter les utilisations futures.

# Chapitre 4

## Étude et analyse des facteurs impactant la consommation des produits pharmaceutiques

### Sommaire

---

<b>4.1</b>	<b>Introduction</b>	<b>79</b>
<b>4.2</b>	<b>Méthodologie</b>	<b>80</b>
<b>4.3</b>	<b>Les Facteurs impactant la consommation de médicaments</b>	<b>82</b>
4.3.1	Identifications des facteurs selon la littérature	82
4.3.2	Choix des facteurs dans le cadre de notre étude	86
<b>4.4</b>	<b>Description des données et méthodes d'analyse de corrélation</b>	<b>91</b>
4.4.1	Description des données	91
4.4.2	Méthodes d'analyse de corrélation des variables	92
<b>4.5</b>	<b>Résultats et analyses</b>	<b>95</b>
4.5.1	Corrélations pour tous les médicaments confondus	95
4.5.2	Corrélations par médicament	99
4.5.3	Etude complémentaire sur la variable "MOIS"	102
<b>4.6</b>	<b>Conclusion</b>	<b>104</b>

---

### 4.1 Introduction

Les consommations des produits pharmaceutiques sont influencées par des facteurs qui ne sont pas tous pris en compte dans les méthodes de quantification, comme les méthodes classiques de consommation, de morbidité, EOQ ou les méthodes univariées de séries temporelles. Par conséquent, en ce qui concerne l'objectif principal de cette thèse, il est impératif de réaliser une revue de la littérature sur les facteurs qui ont un impact sur la consommation de médicaments dans les hôpitaux.

Dans la littérature, les facteurs qui influencent la consommation de produits de santé sont discutés par plusieurs auteurs, notamment Alkan et al.[167]; Muhia et al.[100], Henricson

et al.[52], Sarwar et al.[138], Jiménez-Rubio et al. [62], Rocío et al.[130], Mishuk et al. [92], Howard et al.[55], Borrell et al. [21]. Cependant, ces études sont souvent basées sur les aspects socio-économiques et sanitaires des pays où elles ont été réalisées, voire limitées à une seule structure de soins. Par exemple, certaines études menées aux États-Unis, en Suède ou en Espagne [62, 52] montrent une corrélation entre la consommation de médicaments et la race ou l'origine du patient. Il est évident que tous ces facteurs ne sont pas applicables dans tous les pays.

Dans le contexte français, il y a un manque de travaux établissant les corrélations des facteurs potentiels de consommation au sein des hôpitaux. De plus, pour l'utilisation de l'approche causale, la caractérisation des hôpitaux à travers ces facteurs est également essentielle. Les données recueillies et structurées dans le chapitre précédent, le chapitre 3, vont servir à cet effet. Les résultats de corrélation pourraient être utilisés dans des modèles de prédiction.

Dans ce chapitre, dans un premier temps nous mettons en évidence les facteurs potentiels impactant la consommation des produits de santé selon la littérature. S'en suit une mise en contexte dans le système de santé Français en intersection avec des avis d'experts du monde hospitalier. Une fois les données et informations recueillies, le second objectif est de procéder à une analyse et une étude de corrélations des facteurs potentiels avec la quantité de produits de santé consommés notamment les médicaments.

La suite du chapitre est subdivisé comme suit : la section 4.2 est consacrée à la description de la méthodologie et des approches de cette étude. La section 4.3 traite des facteurs influençant la consommation des médicaments émanant de la littérature et des avis des experts. La section suivante, la section 4.4, donne une description de l'étude de corrélations des facteurs sélectionnés tout en décrivant les données utilisées pour l'analyse. Avant de conclure et de dégager des perspectives de recherches dans la section 4.6, la section 4.5 relate les résultats obtenus et leurs interprétations.

## 4.2 Méthodologie

La méthodologie employé dans le cadre de cette étude peut être résumé par la figure 4.1. C'est une méthodologie qui peut être subdivisée en trois grandes phases.

La première phase regroupe les étapes de recherche bibliographique, d'identification des facteurs pertinents à la littérature et de sélection de ces facteurs selon notre étude. L'étape de recherche bibliographique consiste à identifier la littérature pertinente en utilisant des plate-formes telles que Google Scholar, Science Direct, Web of Science et IEEE. En premier lieu, Google Scholar et Science Direct sont utilisés pour identifier les mots-clés appropriés. Ensuite, une deuxième recherche est effectuée sur les quatre plate-formes à l'aide des mots-clés pertinents identifiés précédemment. Une référence croisée est également effectuée à ce stade et la recherche est limitée aux documents en anglais et en français publiés entre 1980 et 2020. Seuls les articles et les rapports officiels tels que ceux de l'OMS et de Management Sciences for Health (MSH<sup>1</sup>) sont pris en compte. Une fois les documents regroupés en fonction des filtres,

---

1. MSH est une organisation à but non lucratif mondiale qui fournit aux gouvernements, aux organisations de santé et au secteur privé les stratégies, les outils et le soutien en gestion nécessaires pour offrir efficacement

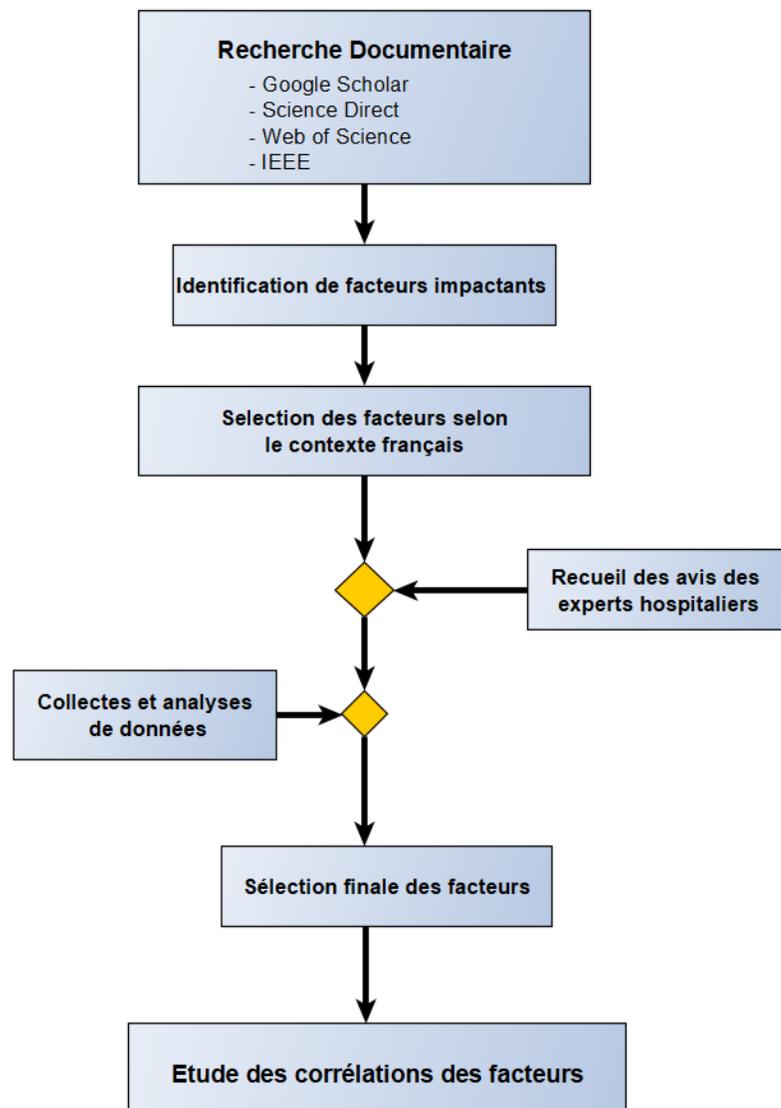


FIGURE 4.1 – Résumé de la méthodologie employée

les doublons sont éliminés et les documents qui ne répondant pas à la problématique de l'étude sont également éliminés. Ces différentes étapes ont permis d'obtenir des documents relatifs aux facteurs de consommation de produits de santé, en particulier les médicaments. La lecture de ces travaux a permis d'identifier les facteurs potentiels qui impactent la consommation. Toutefois, tous ces facteurs ne sont pas universels à tous les systèmes de santé, y compris en France. Cela conduit à une sélection de déterminants spécifiques au système de santé français, avec des exclusions éventuelles.

La deuxième phase englobe les étapes de collecte d'avis d'experts dans le domaine médical, de collecte et d'analyse de données, ainsi que la sélection des facteurs pour l'étude de corrélation. En effet, les facteurs potentiels identifiés à partir de la littérature sont présentés

et efficacement des systèmes de santé performants.

<https://www.msh.org/>

Accessed on : May 10, 2023.

aux pharmaciens et aux cadres pharmaceutiques des hôpitaux partenaires de l'étude, via un questionnaire. Ensuite, des informations sur la consommation de produits de santé sont ensuite collectées auprès de plusieurs hôpitaux, principalement des CHU français, pour rester en accord avec le contexte de l'étude. La collecte est complétée par l'utilisation de données provenant de structures officielles, souvent en utilisant la méthode du web-scraping. Une fois la collecte terminée, une exploration de données, une correction et un nettoyage sont effectués pour obtenir des données traçables et uniformes pour l'étude. À partir de ces données, une dernière sélection des facteurs potentiels est effectuée. Toutefois, certaines données peuvent ne pas être disponibles, ce qui peut entraîner des exclusions. Une partie substantielle de ce travail est expliquée dans le chapitre 3.

La dernière phase est dédiée à l'étude de corrélations des facteurs. En effet, la pertinence des facteurs identifiés doit être validée statistiquement. Cette phase consiste dans ce sens à étudier la corrélation entre les quantités consommées et les facteurs potentiels sélectionnés. Ces analyses s'appuient sur des outils statistiques des études relationnelles.

## 4.3 Les Facteurs impactant la consommation de médicaments

### 4.3.1 Identifications des facteurs selon la littérature

Suite à la première phase, 59 documents relatifs aux facteurs de consommation de médicaments ont été identifiés. Parmi ces 59 documents, différentes catégories de facteurs considérés dans la quantification des besoins pharmaceutiques ont été recensées. Plusieurs articles, tels que ceux de Jiménez-Rubio et al [62], Rocío et al [130], Mishuk et al [92], Howard et al [55], Henricson et al [52], Pokela et al [119] font référence à des facteurs liés au statut des patients, ainsi qu'à la population à prendre en compte dans le système de santé. Ces facteurs incluent généralement des aspects socio-économiques en relation avec le système médical. On trouve également des études portant sur des produits spécifiques ainsi que sur des catégories de produits, en prenant en compte la classification et la catégorisation de l'OMS. En résumé, ces facteurs ont été subdivisés en trois catégories et sont présentés dans le tableau 4.1.

La première catégorie comprend des facteurs socio-économiques tels que : le sexe, l'âge, l'origine du patient, le revenu annuel, la classe sociale, etc. Certaines études menées aux États-Unis sur la corrélation entre l'utilisation de médicaments et la race et l'origine ne sont pas applicables dans tous les pays. Dans la deuxième catégorie, viennent les facteurs liés à la santé du patient. Ceux-ci comprennent des facteurs tels que la co-morbidité, l'assurance santé, le mode de vie (tabagisme ; consommation d'alcool, obésité : masse corporelle, etc.). La dernière catégorie comprend des facteurs en lien avec l'établissement et le personnel de santé. Tous ces facteurs pourraient être utilisés pour estimer les besoins en médicaments pour une nouvelle structure hospitalière, ou pour faciliter une bonne gestion des stocks pour les structures de santé existantes. La dernière catégorie permet de caractériser les hôpitaux. Cette catégorie est particulièrement intéressante pour notre étude en général, car nous combinons les données de plusieurs hôpitaux dans l'optique d'une estimation.

Catégories des facteurs	N°	Facteurs	Sources
Socio-démographiques et socio-économiques	1	Genre (Sexe)	Pappa et al, 2008 [115]; Jiménez-Rubio et al, 2010 [62]; Rocío et al, 2018 [130]; Mishuk et al, 2018 [92]; Howard et al, 2018 [55]; Henricson et al, 1998 [52]; Sarwar et al, 2017 [138]; Borrell et al, 2010 [21]
	2	Âge	Pappa et al 2008 [115]; Jiménez-Rubio et al, 2010 [62]; Rocío et al, 2018 [130]; Mishuk et al, 2018 [92]; Howard et al, 2018 [55]; May et al, 1974 [88]; Henricson et al, 1998 [52]; Pokela et al, 2014 [119]
	3	Race/ethnicité/pays de naissance	Mishuk et al, 2018 [92]; Howard et al, 2018 [55]; Henricson et al, 1998 [52]
	4	Revenu (annuel)	Jiménez-Rubio et al, 2010 [62]; Martin et al, 2015 [86]; Mishuk et al, 2018 [92]; Howard et al, 2018 [55]; Henricson et al, 1998 [52]
	5	Statut matrimonial/civil	Pappa et al 2008 [115]; Rocío et al, 2018 [130]; Sarwar et al, 2017 [138]
	6	Éducation	Pappa et al 2008 [115]; Jiménez-Rubio,2010 [62]; Henricson et al, 1998 [52]; Pokela et al,2014 [119]
	7	Résidence (rurale/urbaine)	Pappa et al 2008 [115]; Martin et al, 2015 [86]; Henricson et al, 1998 [52]
	8	Statut d'emploi ou d'activité (employé, retraité, chômeur, inactif)	Jiménez-Rubio, 2010 [62]; Henricson et al, 1998 [52]; Sarwar et al, 2017 [138]
	9	Classe sociale	Rocío et al, 2018 [130]; Borrell et al, 2010 [21]

Continue à la page suivante

Table 4.1 – suite de la table précédente

Catégories des facteurs	N°	Facteurs	Sources
Liés à la santé	10	Assurance santé	Jiménez-Rubio, 2010 [62]; Howard et al, 2018 [55]; Als-paugh et al, 2021 [4], Gorin, 1997[40]
	11	Risque de santé ou mode de vie (tabagisme, consommation d'alcool, obésité : indice de masse corporelle)	Rocío et al, 2018 [130]
	12	Maladies chroniques	Pappa et al 2008 [115]; Jiménez-Rubio et al, 2010 [62]; Howard et al, 2018 [55]
	13	Comorbidité	Howard et al, 2018 [55]
	14	Expérience antérieure de l'utilisation de médicaments	Howard et al, 2018 [55]
Liés à l'établissement de santé et au personnel	15	L'indice des prix des médicaments	Berndt, 2002 [18]; Pappa et al 2008 [115]
	16	Nombre de médecins pour les consultations et l'hospitalisation	Pappa et al 2008 [115]
	17	Nombre de lits d'hôpital	Pappa et al 2008 [115]; Martin et al, 2015 [86]
	18	Nombre de visites (séjours) médicales	Martin et al, 2015 [86]
	19	Motifs de morbidité (fréquence des problèmes de santé)	Martin et al, 2015 [86]
	20	Nombre d'établissement	Muhia et al, 2017 [100]; Henricson et al, 1998 [52]; Sarwar et al, 2017 [138]
	21	Type d'établissement	WHO, 1988 [109]; Alkan et al, 2015 [167]
	22	Position géographique des établissements (région, district)	WHO, 1988 [109]; Alkan et al, 2015 [167]
	23	Type et nombre de personnel non médical (infirmières, sage-femmes, ...)	WHO, 1988 [109]

Continue à la page suivante

Table 4.1 – suite de la table précédente

Catégories des facteurs	N°	Facteurs	Sources
	24	Niveaux de service (lits-jour)	Muhia et al, 2017 [100]
	25	Nombre et type de services (ou UF) de santé	Muhia et al, 2017 [100]
	26	Facteurs saisonniers (et remplacement d'un médicament plus ancien)	MSH, 2012 [33]
	27	Périodes (délai de livraison)	MSH, 2012 [33]
	28	Spécialiste médical (caractéristiques des travailleurs de la santé)	Mishuk et al, 2018 [92]; MSH, 2012 [33]
	29	Point de stockage (capacité)	Julius et al, 2018 [64]; Rocío et al, 2018 [130]
	30	Processus d'approvisionnement	MSH, 2012 [33]
	31	Processus de budgétisation	Julius et al, 2018 [64]
	32	Exigences légales	Julius et al, 2018 [64]
	33	Nouveaux programmes, expansion de services existants	Julius et al, 2018 [64]
	34	Technologies des installations	Howard et al, 2018 [55]

TABLE 4.1 – Facteurs affectant la consommation de médicaments dans les hôpitaux

A ces 34 facteurs énumérés dans la table 4.1, on peut également souligner l'existence d'autres déterminants identifiés dans la littérature comme :

- *Croyances et pratiques culturelles* : Les croyances culturelles sur la maladie, la médecine traditionnelle et les thérapies alternatives peuvent affecter la consommation de certains produits de santé [105].
- *Accessibilité et disponibilité des services de santé* : La proximité et la disponibilité des établissements et des services de santé peuvent influencer la consommation de soins de santé, notamment en cas d'accès limité ou de délais d'attente prolongés [13].
- *Connaissances en matière de santé* : Le niveau de connaissances en matière de santé peut influencer la consommation de soins de santé, car des individus bien informés sont plus enclins à adopter des comportements préventifs et à suivre les recommandations médicales. [79].

- *Réseaux de soutien social* : La présence de réseaux de soutien social tels que la famille et les amis peut influencer la consommation de soins de santé en fournissant un soutien émotionnel et pratique pour rechercher et suivre les traitements [148, 37].
- *Barrières culturelles et linguistiques* : Les barrières linguistiques et culturelles peuvent affecter la consommation de produits de santé, surtout parmi les populations immigrées ou ayant une maîtrise limitée de la langue locale [95].
- *Stigmatisation et discrimination* : La stigmatisation liée à certains problèmes de santé peut décourager la recherche de soins, notamment pour des questions de santé mentale ou d'infections sexuellement transmissibles [155].
- *Médias et publicité* : Les médias et la publicité jouent un rôle dans la consommation de soins de santé en influençant les perceptions et les choix des consommateurs, ce qui souligne l'importance de l'information précise [71].
- *Facteurs psychologiques* : Les facteurs psychologiques tels que la peur, l'anxiété, la motivation et l'efficacité personnelle peuvent influencer l'utilisation des soins de santé [37].
- *Politiques et réglementations gouvernementales* : Les politiques de santé, les réglementations et les réformes gouvernementales peuvent influencer la consommation de soins de santé en modifiant l'accès, la couverture et d'autres aspects du système de santé [42].

Cependant ces derniers facteurs soulignés ci-dessus peuvent être difficiles à prendre en compte lors du dimensionnement des stocks de médicaments dans les hôpitaux. Ces facteurs comprennent des aspects culturels, sociaux, psychologiques, politiques et réglementaires qui peuvent varier considérablement d'une population à une autre, d'un pays ou une région à une autre. L'incorporation de ces facteurs dans les modèles de machine learning nécessite une collecte et une intégration de données complexes et diverses. Par exemple, la prise en compte des croyances et pratiques culturelles peut nécessiter des informations spécifiques sur la population concernée et les différentes approches médicales auxquelles elle adhère. De même, l'accessibilité et la disponibilité des services de santé peuvent varier selon les régions et les systèmes de santé spécifiques, ce qui rend difficile la généralisation des modèles de prévision des stocks. De plus, les facteurs psychologiques tels que la peur, l'anxiété et la motivation peuvent être difficiles à quantifier et à intégrer dans des modèles, car ils sont liés à des expériences individuelles et subjectives. Cependant avec l'éclosion de certaines techniques de machine learning comme l'*analyse de sentiment* aussi appelé *analyse des opinions* ou "*opinion mining*", il est possible de déterminer l'opinion, le sentiment ou l'attitude exprimée dans un texte donné, tel qu'une phrase, un paragraphe, un commentaire ou un document complet [154, 104]. Ce qui n'entre pas dans le cadre de notre étude.

### 4.3.2 Choix des facteurs dans le cadre de notre étude

Émanant de la littérature, sans être exhaustif, les facteurs répertoriés et subdivisés en 3 catégories (socio-démographiques incluant également l'aspect socio-économique, des facteurs

liés à la santé du patient et des facteurs liés à l'établissement de soins et au personnel de santé) dans la table 4.1 sont de facteurs potentiels pour notre étude. Ces facteurs identifiés pourraient expliquer la consommation de certains médicaments dans les hôpitaux et peuvent constituer des paramètres pour estimer ou quantifier les besoins en produits pharmaceutiques. Cependant, une mise en contexte et une enquête auprès d'experts des hôpitaux partenaires afin de valider et d'affiner cette liste de facteurs s'avèrent nécessaires. Ce qui entraînera dans le cadre de notre étude, à des exclusions parmi les 34 répertoriés dans la table. Les critères soutenant cette exclusion sont fondé sur la non conformité aux contexte du système de santé français, les difficultés ou manques de données et des avis des experts recueillis lors de notre questionnaire.

#### 4.3.2.1 Critère de non conformité de contexte français

Le système de santé varie d'un pays à l'autre, et un exemple notable est le système de sécurité sociale en France, qui a été référencé dans plusieurs rapports, notamment ceux de l'OCDE [106, 107, 108]. Ce système vise à assurer les moyens essentiels de subsistance et de décence pour chaque citoyen, et repose sur le principe de solidarité en se basant sur les cotisations des travailleurs pour financer diverses prestations sociales, telles que les soins médicaux, la maternité, la retraite, les accidents du travail et le chômage.

En effet, le système de santé français garantit un accès universel aux soins médicaux pour tous les résidents, indépendamment de leur âge, de leur revenu ou de leur statut professionnel, ce qui constitue la base de la couverture universelle [4, 40, 14]. Cette couverture est réalisée par l'obligation pour tous les résidents d'avoir une assurance maladie, financée par des cotisations sociales prélevées sur les salaires et les revenus. Cette assurance couvre les dépenses médicales, les médicaments et les hospitalisations, offrant ainsi une protection financière aux patients [151].

De ce fait, dans les catégories de facteurs socio-démographiques et ceux liées à la santé, les facteurs tels que la race ou l'ethnicité ou le pays de naissance, le revenu, l'éducation, le lieu de résidence (rural ou urbain), le statut matrimonial, le statut d'emploi ou d'activité, la classe sociale et l'assurance santé sont exclus eu égard à l'hypothèse de non discrimination du système de sécurité social Français. Dans la dernière catégorie de la table 4.1, c'est à dire les facteurs liés à l'établissement de santé, les facteurs tels que les exigences légales, les technologies des installations bien qu'ayant un impact selon des études, sont également exclus. Pour les technologies des installations, nous supposons que les hôpitaux respectent les exigences légales et sont équipés des ressources matériels nécessaires, notamment avec l'utilisation de la plate-forme WMS, du moins tous les hôpitaux dans cette étude en disposent.

#### 4.3.2.2 Critère lié à la disponibilité de données

La confidentialité et protection des données des patients constituent un majeur frein pour la disponibilité des informations relatives. En réalité, nous n'avons pas eu à notre disposition des données sur la consommation réelle par patient, c'est-à-dire nous n'avons pas eu à notre disposition une base de données de suivi médical (Prescription-Livraison et Administration

par patient). En rappel, dans notre contexte, la consommation fait référence à la livraison de médicaments dans les unités fonctionnelles dans les hôpitaux concernées. Et la recherche de données externes pour compenser ces manques n'ont pas été entièrement absorbés.

Partant de là, les facteurs tels que la comorbidité, les risques sanitaires ou le mode de vie des patients sous traitement, le statut de pathologie chronique, ainsi que l'expérience antérieure de l'utilisation de médicaments des patients sont exclus. Ce sont des informations qui ne figuraient pas dans nos données et dont nous n'avons pas eu accès. Dans la première catégorie de la table 4.1, les facteurs potentiels tels que le sexe et l'âge sont également exclus, car en plus de ne pas figurer dans les historiques de consommation à notre disposition, les hôpitaux pris en compte dans cette étude s'occupent de toutes les catégories de patients, bien que ces deux facteurs puissent être utiles pour caractériser le type de population de la région où se trouve l'hôpital.

Les processus d'approvisionnement, les processus de budgétisation relèvent aussi d'un caractère confidentielle au sein des hôpitaux vis à vis des acteurs externes, de ce fait nous ne pouvons pas intégrer ces facteurs parmi les sélectionnés. Toujours parmi les facteurs de la catégorie 3, les informations telles que la date de livraison des produits, la morbidité et la capacité de stockage de ces établissements de santé ne sont pas incluses dans les données et sont donc exclus de nos analyses. Il en est de même pour l'indice des prix de médicaments (prix unitaire) qui est une variable presque pas renseignée dans nos données (voir table 3.1).

En conséquence, la liste non ordonnée de facteurs regroupés dans la table 4.2 est utilisée, avec l'ajout de la taille de la population.

	<b>Facteur</b>
1	Nombre de médecins pour les consultations et hospitalisations
2	Nombre de lits
3	Nombre de visites (séjours) médicales ou de patients
4	Position géographique des établissements de santé
5	Type d'établissements de santé
6	Nombre d'établissements de santé
7	Facteurs saisonniers
8	Nombre et type de services (ou UF) de santé
9	Nombre de personnel non médical (infirmières, administration, ...)
10	Spécialistes médicaux (caractéristiques des travailleurs de la santé)
11	Niveaux de service (lits-jour)
12	Taille de la population

TABLE 4.2 – Les facteurs sélectionnés

- 1) Le Nombre de médecins pour les consultations et hospitalisations influe directement sur la capacité de fournir des soins médicaux et sur la consommation de médicaments. Un ratio insuffisant de médecins par patient peut entraîner des consultations plus courtes et moins de temps pour évaluer les symptômes et prescrire des médicaments appropriés.

Cela peut conduire à une utilisation excessive ou inadéquate de médicaments à travers l'automédication, par exemple.

- 2) Le nombre de lits reflète la capacité d'accueil de l'hôpital. Un nombre limité de lits peut conduire à des hospitalisations plus courtes, nécessitant une gestion plus intensive des médicaments pour un rétablissement rapide. Cela peut également entraîner une rotation plus rapide des patients et potentiellement une augmentation de la prescription de médicaments.
- 3) Le nombre de visites (séjours) médicales ou de patients : Un grand nombre de visites ou de patients peut augmenter la demande de médicaments pour les traitements, les prescriptions et les soins post-hospitaliers, ce qui peut influencer la consommation globale de médicaments.
- 4) La position géographique des établissements de santé peut impacter la consommation de médicaments. Par exemple, les hôpitaux situés dans des zones rurales peuvent éprouver des difficultés à accéder à certains médicaments, ce qui peut entraîner une augmentation de l'utilisation de médicaments génériques ou moins coûteux. Les établissements de santé situés dans des régions éloignées ou mal desservies peuvent inciter les patients à stocker ou à consommer davantage de médicaments pour faire face à d'éventuelles ruptures de disponibilité de soins. Les réalités pathologiques ne sont pas nécessairement uniformes selon les localisations géographiques.
- 5) Type d'établissements de santé : Différents types d'établissements de santé, tels que les hôpitaux de soins aigus et les cliniques de soins primaires, auront des pratiques de prescription différentes en fonction de leur domaine de spécialisation. Les hôpitaux de référence peuvent proposer un éventail plus large de médicaments, tandis que les cliniques de proximité peuvent se concentrer sur des traitements plus courants.
- 6) Nombre d'établissements de santé : Plus le nombre d'établissements de santé est élevé, plus la demande en médicaments peut être importante pour répondre aux besoins des patients dans différents endroits.
- 7) Facteurs saisonniers : Les variations saisonnières, comme les épidémies de grippe ou d'allergies, peuvent entraîner une augmentation temporaire de la consommation de médicaments pour traiter les symptômes associés. Par exemple, la consommation de médicaments contre le rhume et la grippe augmente généralement pendant les mois d'hiver. Une gestion inadéquate des stocks de médicaments saisonniers peut entraîner des pénuries ou un gaspillage.
- 8) Nombre et type de services (ou UF) de santé : Des services de santé spécialisés, tels que les unités de soins intensifs, peuvent nécessiter des médicaments spécifiques en plus grande quantité, ce qui peut influencer la consommation globale de médicaments.
- 9) Le Nombre de personnel non médical peut impacter la consommation de médicaments dans les hôpitaux en influençant la gestion des stocks, la distribution correcte des médicaments, le suivi des patients pour ajuster les traitements, l'éducation des patients sur la posologie et les effets des médicaments, la planification efficace des rendez-vous médicaux, ainsi que la collaboration interdisciplinaire pour une meilleure coordination

des soins. Un personnel non médical bien formé et organisé contribue à une utilisation plus appropriée et efficace des médicaments, minimisant ainsi la surconsommation ou l'utilisation inadéquate de médicaments.

- 10) Spécialistes médicaux : La disponibilité de spécialistes peut conduire à une meilleure identification des traitements spécifiques, optimisant ainsi la prescription et la consommation de médicaments.
- 11) Niveaux de service (lits-jour) : Les niveaux de service reflètent également la capacité de l'hôpital à accueillir les patients. Des niveaux inadéquats peuvent influencer les décisions de traitement et la prescription de médicaments.
- 12) Taille de la population : Une population plus nombreuse peut entraîner une demande accrue de médicaments pour répondre aux besoins de santé de la communauté, ce qui peut avoir un impact significatif sur la consommation de médicaments dans les hôpitaux.

Relativement à l'objectif principale de la thèse, identifiés des facteurs qui reflètent les caractéristiques et capacité de chaque hôpital. Et la plupart de des facteurs retenus prennent en compte ces aspects.

##### **4.3.2.3 Avis des experts sur le choix des facteurs**

Même si nous avons bénéficié des points de vue des experts au sein de l'entreprise KLS pour le choix des facteurs impactant la consommation des médicaments, nos interactions avec les experts des hôpitaux partenaires, comme souligné dans la section 3.1 du chapitre 3, nous ont également permis de valider nos choix. En réalité, les facteurs sélectionnés dans la table 4.2 sont soumis à l'appréciation des experts tels que les pharmaciens et les cadres pharmaceutiques des hôpitaux partenaires par le biais d'un questionnaire, comme décrit dans le chapitre 3.

Bien que les avis puissent varier légèrement en ce qui concerne les facteurs liés au nombre de personnel (médical et non médical), tous les autres facteurs sont considérés comme ayant un impact sur la consommation des médicaments. En outre, il est souligné de manière insistante que pour effectuer de nouvelles estimations, les acteurs se réfèrent principalement aux consommations antérieures pour les molécules (médicaments) déjà connues.

En ce qui concerne les nouveaux médicaments, les acteurs prennent en compte les posologies des produits et le potentiel nombre de séjours liés aux pathologies traitées. Les experts n'ont pas omis de mentionner la nature volatile des consommations au sein des hôpitaux. L'ensemble de ces informations ouvre la voie à la collecte de données et à la réalisation d'une dernière sélection des facteurs avant d'entamer l'étude de corrélation.

## 4.4 Description des données et méthodes d'analyse de corrélation

### 4.4.1 Description des données

La collecte des données pour cette étude tient compte des divers facteurs sélectionnés dans la section précédente (consultables dans la table 4.2). Comme souligné dans le précédent chapitre (chapitre 3), les données utilisées dans cette étude sont extraites du système de gestion d'entrepôt (WMS) de six hôpitaux situés dans six départements distincts en France. Ces données sont enrichies par des données externes telles que la taille de la population et le nombre d'établissements etc. Comme on peut le constater dans l'esquisse de la table finale 3.11 décrite dans le chapitre 3 et la table des variables A.1 dans l'annexe.

Cette base de données utilisée pour effectuer l'étude contient implicitement certains facteurs. Les attributs de cette base sont relative aux facteurs sont référencés dans le tableau 4.3. Relativement au facteurs sélectionnés dans la table 4.2, dans la table de description des variables de données, on constate que :

- Le type d'établissement est identifié par les identifiants des hôpitaux qui sont généralement des CHU.
- La variable CODE\_DEPT dans le jeu de données fourni la position géographique de chaque hôpital
- Les variables ANNEE et MOIS fournissent les périodes de consommation des médicaments qui sont en lien avec les facteurs de saisonnalité.
- Nombre d'établissements rattachés à chaque hôpital, dans notre jeu de donnée est identifié par la variable N\_ETB.
- Pour les unités fonctionnelles, leur nombres sont identifiés à travers la variable N\_UFS
- Le nombre de lits, au regard des données collectées, est subdivisé en deux : le nombre de lits pour l'hospitalisation complète LIT\_HC, et le nombre de lits pour hospitalisation partielle LIT\_HP.
- Le nombre de médecins pour les consultations et hospitalisations ainsi que les spécialistes médicaux est fourni dans notre table par la variable P\_MEDICAL
- Le nombre de personnel non médical est quant à lui déterminé par la variable PN\_MEDICAL.
- Le nombre de séjours, relativement à la collecte de données est fourni en fonctions des groupes d'activités au sein de hôpitaux.
- Le nombre de population couvert par chaque hôpital est obtenu par la variable POPULATION dans notre jeu de données.

Outre les variables du tableau 4.3, l'ensemble de données contient également *les services hospitaliers* (regroupement d'unités fonctionnelles proposée) où les médicaments sont consommés. Le type de ces données est booléen. Dans la base de données, les valeurs de PN\_MEDICAL et de P\_MEDICAL sont fournis par an et par hôpital. Cependant, il figure également dans la base les effectifs des personnels par *les services hospitaliers*. Il est à

Nom de la colonne	Description
ID_SITE	Identifiant de l'hôpital
ID_REF	Identifiant du médicament par hôpital
DESIG	Nom du médicament
CODE_UCD	Code UCD du médicament
CODE_DEPT	Situation géographique (Département)
ANNEE	Année de consommation du médicament
MOIS	Mois de consommation
N_ETB	Nombre d'établissements de l'hôpital
N_UFS	Nombre d'unités fonctionnelles
LIT_HC	Nombre de lits pour hospitalisation complète
LIT_HP	Nombre de lits pour hospitalisation partielle
SEJ_MCO	Nombre de séjours relative aux activités MCO
SEJ_HAD	Nombre de séjours relative aux activités HAD
SEJ_SLD	Nombre de séjours relative aux activités SLD
SEJ_SSR	Nombre de séjours relative aux activités SSR
SEJ_PSY	Nombre de séjours relative aux activités PSY
PN_MEDICAL	Nombre de personnels non-médicaux (infirmiers, sages-femmes, etc.)
P_MEDICAL	Nombre de personnel médicaux
POPULATION	Nombre de population couvert par les hôpitaux
QUANTITE	Quantité de médicament consommée

TABLE 4.3 – Description des colonnes du jeu de données

mentionner que la période de collecte des informations est de l'année 2016 à l'année 2019 en ne prend pas en compte la période COVID-19. Ces données contiennent l'historique de consommation des 45 médicaments en commun entre les 6 hôpitaux. Aussi, comme la table le montre, les consommations sont agrégées mensuellement dans la base de données.

Suite à une fouille de données avant de procéder aux analyses de corrélation, un constat est fait. La variable SEJ\_SSR n'est pas régulièrement renseigné. Dans ce sens cet attribut ne sera pas pris en compte dans les analyses. Et puisque notre proposition de regroupements des unités fonctionnelles (*les services hospitaliers*) n'a pas fait l'objet de consensus, les variables les concernant sont également exclues des analyses.

#### 4.4.2 Méthodes d'analyse de corrélation des variables

Les méthodes de prédiction de la demande nécessitent une corrélation entre les variables dépendantes (ici, les facteurs) et la variable à estimer, c'est-à-dire la quantité de médicaments dans notre cas. Un autre problème, d'un point de vue de la détermination du modèle, est de connaître la relation entre les variables dépendantes afin d'éliminer les redondances qui

peuvent affecter la qualité de la prédiction.

Dans la littérature scientifique, il existe différentes méthodes pour une analyse de corrélation en fonction des types de données (quantitatives, qualitatives ou mixtes) ou du type de relation (linéaire ou non linéaire). Les auteurs comme Benesty et al [16], ; Havlicek et al [50], O'Brien [113], Borgognone et al [20], Hamdi et al [47] traitent de ces questions. Les objectifs de l'analyse avec ces méthodes sont de :

- Déterminer s'il existe une relation entre les variables ;
- Caractériser la forme de la relation entre les variables (positive ou négative, linéaire ou non linéaire, monotone ou non monotone) ;
- Tester si la relation est statistiquement significative ;
- Quantifier l'intensité de la relation.

Une méthode rapide d'analyse est la visualisation des données, également appelée *analyse graphique*. C'est un bon moyen de comprendre les caractéristiques des données. Cependant, elle ne permet pas de valider les points mentionnés ci-dessus plus particulièrement l'intensité de la relation.

Dans le contexte de notre travail et tenant compte du format de nos données, l'accent est mis sur les méthodes quantitatives. Parmi ces méthodes, il y a la corrélation de Bravais-Pearson (ou test de corrélation de Pearson) qui permet de mesurer la dépendance linéaire entre deux variables [10, 16]. La mesure de l'intensité de la corrélation, également appelée coefficient de corrélation, est fournie par le coefficient  $r_{XY}$  :

$$r_{XY} = \frac{Cov(X, Y)}{\sigma_X \cdot \sigma_Y}, \quad r_{XY} \in [-1, 1] \quad (4.4.1)$$

où  $Cov(X, Y)$  est la covariance entre deux variables et  $\sigma_X, \sigma_Y$  sont les écarts-types des variables.

Le test de corrélation de Pearson évalue la relation linéaire entre deux variables continues, mesurant dans quelle mesure les variations d'une variable sont associées aux variations de l'autre. Cependant, il nécessite que les données soient normalement distribuées pour obtenir des résultats fiables. Ainsi, il est recommandé de vérifier la normalité des données avant d'utiliser ce test, à l'aide de méthodes telles que le test de Shapiro-Wilk, le test de Kolmogorov-Smirnov etc. Lorsque les données suivent une distribution normale, les résultats du test de corrélation de Pearson sont plus fiables. Cependant, si vos données ne sont pas normalement distribuées, cela ne signifie pas nécessairement qu'on ne peut pas utiliser du tout le test de corrélation de Pearson. Dans de nombreux cas, le test de corrélation de Pearson est encore utilisé lorsque les données sont approximativement normales ou lorsque la violation de la normalité n'est pas extrême. Cependant, il est important d'interpréter les résultats avec prudence et de considérer d'autres méthodes de corrélation qui sont moins sensibles aux violations de la normalité.

Lorsque l'assurance de la normalité des variables n'est pas établie, les tests de corrélation non paramétriques, tels que les tests de corrélation de Kendall et de Spearman, sont préconisés. Ces méthodes offrent des alternatives robustes. De plus, la régression linéaire (simple

ou multiple) ainsi que l'analyse en composantes principales (ACP) sont recommandées pour analyser les relations entre variables quantitatives. Dans des scénarios non linéaires, des techniques d'apprentissage automatique comme les arbres de décision, les méthodes ensemblistes comme le Random Forest, sont également pertinentes, comme évoqué par Mu et al [99], et Kijewski et al [68].

Dans cette étude, notre objectif est de vérifier l'existence de liens entre les quantités de médicaments consommées et les facteurs potentiels identifiés grâce à une revue de la littérature et aux avis d'experts, en utilisant des tests de corrélation. Étant donné que la majorité de nos données ne sont pas de nature catégorielle, nous appliquons les tests de corrélation de Pearson, de Kendall et de Spearman afin d'obtenir des résultats robustes pour évaluer les corrélations. Le test de Kolmogorov-Smirnov est également retenu pour évaluer la normalité des données. Le choix de ce test de normalité est justifié par la grande taille des échantillons et de sa capacité à résister aux valeurs aberrantes, contrairement au test de Shapiro-Wilk. Nous considérons qu'une relation de corrélation est significative lorsque la valeur de signification, ou *p\_value*, est inférieure à 5%.

Il est important de rappeler qu'en statistique, la *p\_value* est une mesure de la probabilité d'obtenir des résultats aussi extrêmes que ceux observés dans une étude, si l'hypothèse nulle est vraie. Généralement exprimée en pourcentage, une *p\_value* inférieure à un seuil donné, souvent 0,05, est considérée comme significative. En ce qui concerne les tests de corrélation, la *p\_value* mesure la probabilité d'obtenir une corrélation aussi élevée ou supérieure à celle observée entre deux variables, en supposant que les variables ne sont pas corrélées. Une *p\_value* inférieure à 0,05 est considérée comme significative, ce qui signifie qu'il y a moins de 5% de chances que les variables ne soient pas corrélées.

Dans le cas où il y a une variable quantitative, la méthode ANOVA (ANalysis Of VAriance) [113] s'avère propice pour l'analyse de lien entre cette variable et la variable QUANTITE en lien et place des trois tests de corrélations. Il existe plusieurs variantes de l'ANOVA parmi lesquelles on peut citer :

- *L'ANOVA à un facteur* : L'ANOVA à un facteur ("One-Way ANOVA") est utilisée lorsqu'on a une seule variable indépendante catégorielle (facteur) et une variable dépendante continue.
- *L'ANOVA à deux facteurs* : L'ANOVA à deux facteurs ("two-way ANOVA") est utilisée lorsqu'on a deux variables indépendantes et une variable dépendante. Cette méthode permet de déterminer si les effets de chacun des facteurs sont indépendants ou s'ils interagissent entre eux pour influencer la variable continue.
- *L'ANOVA à plusieurs facteurs* : L'ANOVA à plusieurs facteurs est utilisée lorsque vous avez plus de deux variables indépendantes et une variable dépendante.

Cependant, lorsque les données ne sont pas normalement distribuées ou que les variances ne sont pas homogènes, il est possible d'utiliser en alternative le test de Kruskal-Wallis [112]. Le test de Kruskal-Wallis est une alternative non paramétrique à l'ANOVA et il est plus robuste aux violations des hypothèses de normalité et d'homogénéité des variances. Relativement à nos données, si la normalité n'est pas établie, c'est le test de Kruskal-Wallis

qui sera retenu pour les tests avec les variables catégorielles. Ce test nous évite de nous focaliser sur les hypothèses de normalité et d'homoscédasticité (variances homogènes) des données.

## 4.5 Résultats et analyses

Dans cette section, nous présentons les résultats obtenus suite à l'application des tests de corrélation à chaque facteur sélectionné en fonction de la quantité de médicament consommé. Étant donné que nous avons combiné les consommations de 45 médicaments provenant de 6 hôpitaux, nous présentons deux types de résultats et d'analyses de corrélations.

Tout d'abord, nous avons réalisé des analyses de corrélation pour chaque facteur en relation avec la consommation globale de tous les médicaments, sans distinction. Ensuite, nous avons effectué des analyses de corrélation par médicament, ce qui nous a permis d'examiner comment les facteurs potentiels interagissent avec les quantités de médicaments consommées, à la fois dans une perspective globale (sans distinction) et lorsque les produits pharmaceutiques sont considérés séparément. *Python* et le module *stats* du package *scipy* sont les outils principaux utilisés dans ce travail.

### 4.5.1 Corrélations pour tous les médicaments confondus

Pour comprendre la corrélation entre les facteurs potentiels identifiés et la quantité de médicaments consommés, sans distinction, nous avons appliqué en première instance trois tests de corrélation : le test de corrélation de Pearson-Bravais, le test de Kendall et le test de Spearman. Les tests non paramétriques de Kendall et de Spearman ne nécessitent pas d'hypothèse de normalité des données. L'objectif était de mesurer la relation entre la consommation de médicaments et les facteurs potentiels (à l'exception des services et des unités fonctionnelles) et de vérifier si cette relation est significative.

Pour effectuer les tests de corrélations de Pearson, le test de normalité (le test de Kolmogorov) des données est préalablement appliqué. On obtient dans ce cas de figure la table 4.4.

Dans le tableau 4.4, la variable "Statistique" représente la valeur calculée lors du test de Kolmogorov-Smirnov, mesurant la différence entre la distribution observée de nos données et la distribution théorique (normale dans ce cas). La variable "P\_value" est la probabilité associée à la statistique de test, indiquant la probabilité d'obtenir des données aussi éloignées de la distribution normale que celles que nous avons observées, si les données étaient réellement normalement distribuées. Dans le test de Kolmogorov-Smirnov, une *p\_value* très faible ( $< 0.001$ ) suggère que les données ne suivent pas une distribution normale. En d'autres termes, il y a des preuves significatives pour rejeter l'hypothèse que les données sont normalement distribuées. Dans notre cas, pour tous les attributs, la *p\_value* est très faible ( $< 0.001$ ), ce qui indique que les données ne suivent pas une distribution normale. Par conséquent, nous avons des preuves pour rejeter l'hypothèse de normalité pour ces attributs. Ainsi, la fiabilité des résultats des tests de Pearson, fournis dans la table 4.5, est mise en doute.

Facteur	Statistique	P_value
N_ETB	1.0	< 0.001
N_UFS	0.84	< 0.001
MOIS	0.89	< 0.001
POPULATION	1.0	< 0.001
LIT_HC	1.0	< 0.001
LIT_HP	1.0	< 0.001
SEJ_MCO	1.0	< 0.001
SEJ_HAD	0.5	< 0.001
SEJ_PSY	0.7	< 0.001
SEJ_SLD	0.78	< 0.001
PN_MEDICAL	1.0	< 0.001
P_MEDICAL	1.0	< 0.001

TABLE 4.4 – Résultats du test de Kolmogorov-Smirnov pour la normalité des données (médicament sans distinction)

Facteur	Coefficient de corrélation	P_value	Significatif
N_ETB	0.0287	$2.434 \times 10^{-02}$	Vrai
N_UFS	0.1472	$4.361 \times 10^{-31}$	Vrai
MOIS	-0.0055	$6.669 \times 10^{-01}$	Faux
POPULATION	0.0191	$1.351 \times 10^{-01}$	Faux
LIT_HC	0.0649	$3.519 \times 10^{-07}$	Vrai
LIT_HP	0.0564	$9.904 \times 10^{-06}$	Vrai
SEJ_MCO	0.0667	$1.664 \times 10^{-07}$	Vrai
SEJ_HAD	-0.0063	$6.205 \times 10^{-01}$	Faux
SEJ_PSY	0.0170	$1.833 \times 10^{-01}$	Faux
SEJ_SLD	0.0386	$2.479 \times 10^{-03}$	Vrai
PN_MEDICAL	0.0679	$9.969 \times 10^{-08}$	Vrai
P_MEDICAL	0.0676	$1.131 \times 10^{-07}$	Vrai

TABLE 4.5 – Corrélations de Pearson

Cette situation justifie l'application des tests de Kendall et de Spearman qui constituent des alternatives à Pearson pour plus de fiabilité. Pour ces tests non paramétriques, nous avons obtenu les résultats présentés dans les tables 4.6 et 4.7 respectivement.

En analysant les tableaux 4.6 et 4.7, les résultats des tests de Kendall et de Spearman indiquent une corrélation faible mais statistiquement significative entre chaque facteur sélectionné et la quantité de consommation de médicaments, à l'exception de la variable *MOIS*. Par exemple, pour le facteur N\_ETB lors du test de Kendall, le coefficient de corrélation est de 0,137. Cela signifie que les variables N\_ETB et QUANTITE sont liées, mais que la relation n'est pas très forte. Sa p\_value, étant de  $1.801 \times 10^{-48}$ , indique qu'il y a moins d'une

Facteur	Coefficient de corrélation	P_value	Significatif
N_ETB	0.1370	$1.801 \times 10^{-48}$	Vrai
N_UFS	0.3390	$0.00 \times 10^0$	Vrai
MOIS	-0.0170	0.0559	Faux
POPULATION	0.1121	$4.451 \times 10^{-37}$	Vrai
LIT_HC	0.1814	$2.760 \times 10^{-94}$	Vrai
LIT_HP	0.1335	$6.370 \times 10^{-52}$	Vrai
SEJ_MCO	0.1742	$4.687 \times 10^{-87}$	Vrai
SEJ_HAD	0.0622	$4.028 \times 10^{-10}$	Vrai
SEJ_PSY	0.0749	$1.215 \times 10^{-16}$	Vrai
SEJ_SLD	0.1599	$1.053 \times 10^{-71}$	Vrai
PN_MEDICAL	0.1874	$2.564 \times 10^{-99}$	Vrai
P_MEDICAL	0.1806	$2.003 \times 10^{-93}$	Vrai

TABLE 4.6 – Corrélations de Kendall

Facteur	Coefficient de corrélation	P_value	Significatif
N_ETB	0.1876	$9.776 \times 10^{-50}$	Vrai
N_UFS	0.4692	$0.000 \times 10^0$	Vrai
MOIS	-0.0244	0.0558	Faux
POPULATION	0.1594	$3.095 \times 10^{-36}$	Vrai
LIT_HC	0.2596	$4.025 \times 10^{-95}$	Vrai
LIT_HP	0.1959	$3.545 \times 10^{-54}$	Vrai
SEJ_MCO	0.2501	$3.180 \times 10^{-88}$	Vrai
SEJ_HAD	0.0797	$3.925 \times 10^{-10}$	Vrai
SEJ_PSY	0.1059	$8.90 \times 10^{-17}$	Vrai
SEJ_SLD	0.2259	$6.946 \times 10^{-72}$	Vrai
PN_MEDICAL	0.2664	$2.883 \times 10^{-100}$	Vrai
P_MEDICAL	0.2590	$1.088 \times 10^{-94}$	Vrai

TABLE 4.7 – Corrélations de Spearman

chance sur  $10^{48}$  que la corrélation observée soit due au hasard. Généralement, on utilise un seuil de signification de 0,05 (soit 5%) pour décider si la corrélation est statistiquement significative. Si la  $p\_value$  est inférieure à ce seuil, on rejette l'hypothèse nulle de non-corrélation et on conclut qu'il y a une corrélation statistiquement significative. Par conséquent, on peut dire qu'il y a très probablement une corrélation statistiquement significative entre les deux variables. Cependant, la faible valeur du coefficient de corrélation indique que l'intensité de la relation entre les deux variables est faible. Cette interprétation est identique pour les autres facteurs des deux tableaux (4.6 et 4.7). Selon ces deux tests empiriques, les variables MOIS et QUANTITE ne sont pas corrélées.

Une analyse plus poussée sur la variable "MOIS" est à prendre en compte. Rappelons, avant de poursuivre, que la variable "MOIS", bien que numérique (01, 02, ..., 12) est une variable catégorielle, c'est-à-dire une variable qualitative dont la distribution n'est pas normale, de sorte que le résultat pour la variable "MOIS" n'est pas fiable.

Des différents résultats une hypothèse se dégage. La faiblesse des intensités de corrélation est due à la différences de consommation des médicaments (comme on peut le voir dans les figures 4.2 et 4.3, en illustration). Dans cette première étude, aucune distinction n'a été faite concernant les médicaments. La section suivante approfondira l'investigation par médicament.

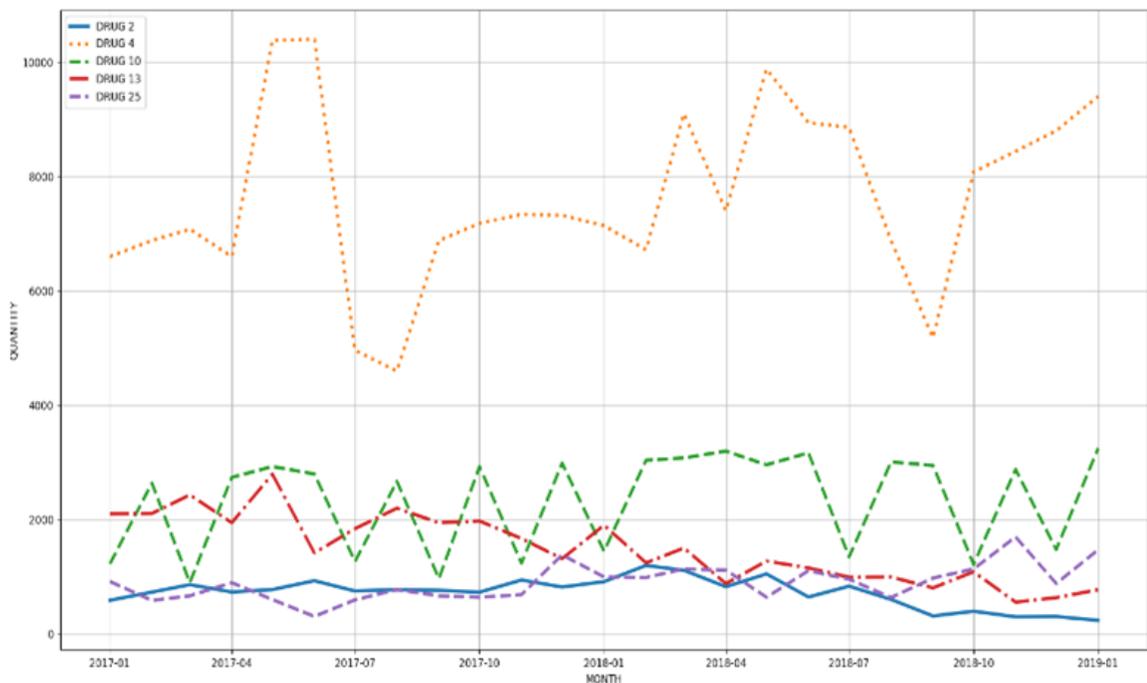


FIGURE 4.2 – Exemples de tendances de consommation de 5 médicaments à l'Hôpital H3 (De janvier 2017 à janvier 2019)

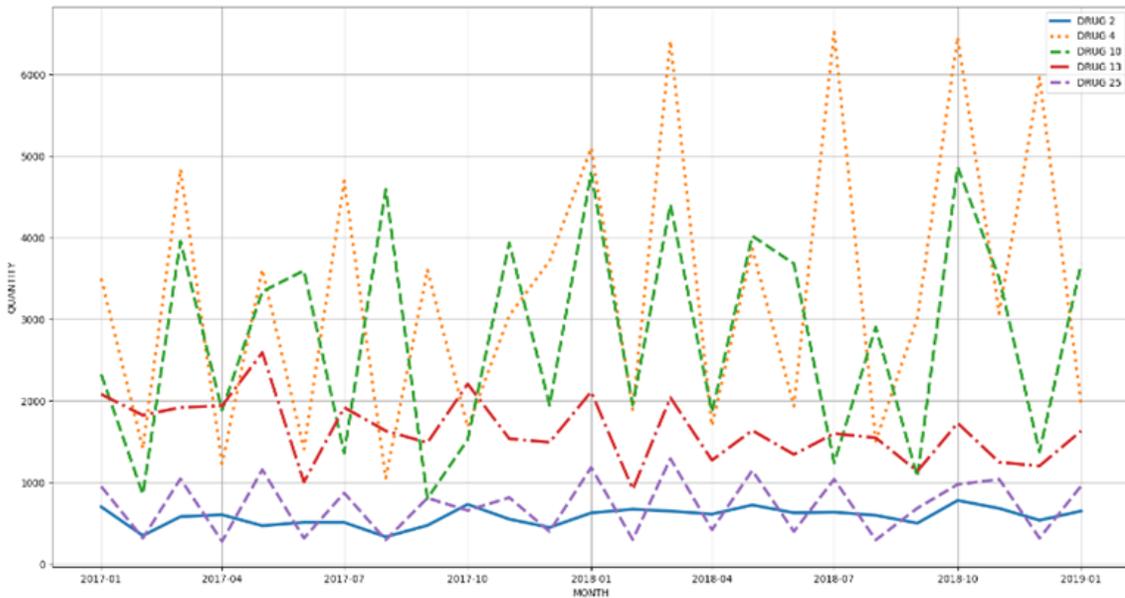


FIGURE 4.3 – Exemples de tendances de consommation de 5 médicaments à l’Hôpital H6 (De janvier 2017 à janvier 2019)

#### 4.5.2 Corrélations par médicament

Dans cette sous-section, nous examinons les liens entre chaque facteur et les quantités de consommation pour chacun des 45 médicaments communs. Afin d’analyser la faible corrélation, qui a été évoquée comme hypothèse dans la sous-section précédente, nous avons utilisé deux tests de corrélation (Kendall, Spearman) sur les données de chaque médicament. Par conséquent, nous obtenons les deux tableaux suivants : Tableau 4.8 et Tableau 4.9. Ces tableaux présentent des résumés des coefficients de corrélation issus des deux tests par médicament, dont la plupart sont statistiquement significatifs.

De ces tables, pour la majorité des médicaments, les tests de corrélations montrent qu’il existe des corrélations (souvent fortes, au-delà de 0.5) et significatives au seuil de 5% entre la consommation et les facteurs cités. Parmi les 45 médicaments, quatre médicaments, à savoir les médicaments 2, 10, 16 et 21, se sont distingués avec le plus grand nombre de coefficients faibles en fonction des facteurs énumérés. En effet, les médicaments 2 et 16 présentent des corrélations faibles (souvent très proches de zéro) sur l’ensemble des facteurs, tandis que les médicaments 10 et 21 n’affichent qu’une corrélation forte chacun avec le nombre d’unités fonctionnelles. Une explication possible est que ces produits ont un faible taux d’utilisation dans les différents hôpitaux. Par exemple, le médicament 2 est un traitement associé à une chimiothérapie, indiqué pour les patients atteints de lymphome folliculaire de stade III-IV n’ayant jamais été traités auparavant. Le lymphome folliculaire représente environ 3200 nouveaux cas par an en France [30].

Il convient de noter qu’une corrélation faible dans le cas de Pearson, Kendall ou Spearman ne signifie pas nécessairement qu’il y a absence de corrélation entre les variables. Les coefficients de corrélation de Pearson, Kendall et Spearman varient tous dans la plage de -1 à

MED	N_ETB	N_UFS	MOIS	POPULATION	LIT_HC	LIT_HP	SEJ_MCO	SEJ_HAD	SEJ_PSY	SEJ_SLD	PN_MEDICAL	P_MEDICAL
1	0.4604	0.1556	-0.0574	0.5279	0.4745	0.3992	0.5809	0.2131	0.3896	0.4455	0.5938	0.5934
2	0.0198	0.2807	-0.0889	-0.0907	-0.0006	0.0946	0.0283	0.1439	0.1011	-0.0109	0.0367	0.0208
3	0.3819	0.2124	-0.0740	0.3188	0.4745	0.3286	0.4301	0.2087	0.1350	0.3591	0.4432	0.4022
4	0.3362	0.4024	0.0206	0.4488	0.4584	0.3821	0.5505	0.2055	0.1152	0.4404	0.5988	0.5686
5	0.4196	0.4454	-0.0408	0.4949	0.7052	0.2953	0.6673	0.0894	0.0458	0.6920	0.6935	0.6677
6	0.7643	0.3523	-0.0721	0.1029	0.2792	0.6190	0.2664	0.5855	0.6613	0.1759	0.2954	0.2800
7	-0.3013	0.5494	-0.0229	-0.0406	-0.0399	-0.4039	-0.0346	-0.3881	-0.5726	0.0841	0.0609	-0.0063
8	0.4743	0.5466	0.0043	0.2462	0.3916	0.4640	0.4054	0.3699	0.2796	0.2606	0.3697	0.3715
9	0.4126	-0.1450	-0.0086	0.4021	0.5705	0.3889	0.5320	0.2174	0.2143	0.4162	0.5123	0.4991
10	-0.3008	0.5887	-0.0174	-0.1902	-0.0847	-0.1134	-0.0330	-0.2861	0.0247	0.0196	-0.0098	-0.0066
11	-0.3137	0.5690	-0.0199	-0.3553	-0.2431	-0.0104	-0.2123	-0.0811	-0.0247	-0.1703	-0.1596	-0.1377
12	0.1019	-0.1128	0.0577	0.3899	0.3231	0.2732	0.4051	-0.0446	0.0083	0.3171	0.4750	0.4517
13	0.3072	0.7363	-0.1051	-0.1189	0.0621	0.3983	-0.0165	0.5332	0.6384	-0.0702	0.0634	0.0527
14	0.5627	0.3547	-0.1208	0.3393	0.5021	0.5112	0.4830	0.4179	0.3028	0.4513	0.5351	0.4999
15	-0.0162	0.4629	-0.0393	0.1356	0.1898	0.0343	0.1633	-0.2226	-0.0364	0.2117	0.2142	0.2155
16	0.0358	0.1930	-0.0689	-0.0412	0.0753	0.0805	0.0311	0.0759	0.1250	0.0139	0.0718	0.0690
17	0.3313	0.1273	-0.0511	0.5302	0.6595	0.3513	0.6880	0.0754	0.1003	0.5357	0.6696	0.6473
18	0.0695	0.5739	-0.0173	0.2782	0.3730	0.0581	0.4059	-0.3023	-0.1981	0.4769	0.4564	0.4147
19	0.1990	0.3921	-0.0846	0.2733	0.4465	0.2244	0.4166	0.0089	0.1030	0.3769	0.4655	0.3957
20	0.1636	-0.1177	-0.0515	0.2993	0.4941	0.0467	0.4339	-0.1955	-0.1402	0.5060	0.4904	0.4273
21	0.3119	0.4087	-0.0473	0.0891	0.1493	-0.0049	0.0817	0.1643	-0.0746	0.2224	0.0985	0.0504
22	0.2677	0.0847	-0.0359	0.3826	0.5558	0.1142	0.5008	-0.1256	-0.1260	0.5659	0.5071	0.4934
23	0.0215	-0.0812	0.0226	0.2691	0.3253	0.1894	0.3731	-0.1062	-0.0835	0.3097	0.4607	0.4723
24	0.3362	0.1179	-0.2400	0.2560	0.1387	0.1729	0.1936	0.3020	0.0529	0.3039	0.1996	0.2010
25	0.3333	0.1619	-0.0486	0.2213	0.4466	0.2710	0.3827	0.1822	0.0954	0.3516	0.4166	0.3776
26	0.5096	0.2524	-0.0482	0.4008	0.3528	0.6040	0.3984	0.4202	0.4268	0.3519	0.4329	0.4417
27	0.1944	0.0539	-0.0966	0.2834	0.5458	0.0764	0.4954	-0.1497	-0.0977	0.5028	0.5382	0.4865
28	0.7334	0.3970	-0.0511	0.2385	0.3829	0.4290	0.3465	0.5247	0.3167	0.2876	0.3370	0.3126
29	0.3418	-0.3274	-0.0307	0.4668	0.6382	0.4624	0.6357	0.1942	0.2026	0.4819	0.6325	0.6364
30	0.5301	0.1179	-0.1314	0.1932	0.4965	0.4060	0.4417	0.3674	0.2909	0.3177	0.4420	0.4044
31	0.3215	0.4278	0.1038	0.2007	0.1561	0.1473	0.1723	0.2372	0.0614	0.2431	0.2512	0.2037
32	0.4828	-0.1630	-0.1050	0.2469	0.6203	0.4026	0.4774	0.2975	0.3042	0.4188	0.4846	0.4606
33	0.2102	0.6178	-0.0652	0.4456	0.4627	0.2192	0.4603	-0.0460	-0.0225	0.4786	0.5286	0.5415
34	0.1256	0.3164	0.1002	0.3783	0.2385	0.0394	0.2756	-0.2097	-0.1973	0.3910	0.3468	0.3548
35	0.0565	0.4494	-0.0088	0.1685	0.3247	0.1089	0.3319	-0.0972	0.0120	0.2863	0.3755	0.3027
36	0.5427	0.1477	0.0367	0.2880	0.2559	0.5836	0.3031	0.5095	0.4257	0.2488	0.3304	0.3346
37	0.4772	0.1821	-0.1000	0.5204	0.6803	0.3295	0.6602	0.2266	0.1330	0.6839	0.7358	0.6454
38	0.1609	0.4229	-0.0213	0.2283	0.3424	0.2648	0.3781	0.1260	0.3379	0.2564	0.4215	0.4247
39	0.3457	-0.0158	-0.0742	0.1319	0.4357	0.2232	0.3175	0.4064	0.3201	0.2733	0.4384	0.3307
40	0.3908	0.7441	0.0002	0.7000	0.6432	0.4047	0.7305	0.1708	0.0889	0.6343	0.7579	0.7442
41	0.2257	0.4523	-0.0133	0.3011	0.4948	0.1842	0.4143	-0.0732	0.0057	0.4078	0.4225	0.3895
42	0.5977	0.5481	-0.0193	0.4096	0.4432	0.4241	0.4183	0.3786	0.2500	0.3322	0.4373	0.4279
43	0.2056	0.3715	-0.0819	0.4101	0.5227	0.0886	0.5172	-0.1359	-0.1513	0.6411	0.6010	0.5164
44	-0.0776	0.2913	0.0677	-0.0315	0.0381	0.2231	0.1334	0.0158	0.1214	-0.0305	0.0928	0.1605
45	0.2474	0.2501	-0.2037	0.0527	0.3186	0.2293	0.3295	0.1262	0.2000	0.3158	0.3828	0.3784

TABLE 4.8 – Correlations de Kendall par médicament

MED	N_ETB	N_UFS	MOIS	POPU	LIT_HC	LIT_HP	SEJ_MCO	SEJ_HAD	SEJ_PSY	SEJ_SLD	PN_MEDICAL	P_MEDICAL
1	0.6385	0.1251	-0.0778	0.6773	0.7066	0.6137	0.7788	0.2587	0.5757	0.6241	0.8037	0.8023
2	0.0389	0.3552	-0.1376	-0.1465	-0.0246	0.1439	0.0158	0.1860	0.1405	-0.0159	0.0322	0.0067
3	0.5295	0.2072	-0.1135	0.4107	0.6647	0.4755	0.6069	0.2501	0.2038	0.5059	0.5928	0.5573
4	0.5085	0.5259	0.0296	0.5506	0.6463	0.5697	0.7306	0.2309	0.1867	0.5947	0.7559	0.7438
5	0.5965	0.5533	-0.0564	0.6012	0.8738	0.4968	0.8430	0.1167	0.0408	0.8588	0.8677	0.8427
6	0.8966	0.5076	-0.1012	0.2532	0.4990	0.8098	0.4890	0.7427	0.8348	0.2825	0.5033	0.4994
7	-0.3832	0.6852	-0.0447	-0.1000	-0.0610	-0.5437	-0.0651	-0.5168	-0.7271	0.1348	0.0109	-0.0627
8	0.6237	0.6925	0.0075	0.3773	0.5584	0.6419	0.5872	0.4722	0.3730	0.3780	0.5347	0.5395
9	0.6098	-0.2224	-0.0145	0.5011	0.7437	0.5836	0.7015	0.2772	0.3059	0.5772	0.6865	0.6818
10	-0.3931	0.7450	-0.0245	-0.2828	-0.1074	-0.1572	-0.0652	-0.3578	0.0207	-0.0037	-0.0228	-0.0110
11	-0.4175	0.7225	-0.0151	-0.4923	-0.3013	-0.0315	-0.2746	-0.1156	-0.0230	-0.2695	-0.2011	-0.1773
12	0.1738	-0.2428	0.0807	0.4825	0.4975	0.3904	0.5608	-0.0958	0.1101	0.4713	0.5943	0.6153
13	0.4052	0.8860	-0.1489	-0.1348	0.0775	0.5831	0.0234	0.6784	0.8099	-0.0990	0.0843	0.0876
14	0.7371	0.4714	-0.1715	0.4536	0.6975	0.7065	0.6837	0.5180	0.4433	0.6322	0.7197	0.6960
15	-0.0254	0.6213	-0.0542	0.1921	0.2823	0.0307	0.2351	-0.2818	-0.0495	0.3052	0.3069	0.3123
16	0.0537	0.2690	-0.0972	-0.0679	0.1039	0.1145	0.0394	0.0955	0.1834	0.0235	0.0944	0.0856
17	0.5363	0.1430	-0.0735	0.6390	0.8406	0.5623	0.8655	0.1018	0.1340	0.7107	0.8437	0.8352
18	0.0410	0.7622	-0.0278	0.3601	0.5770	0.0604	0.5864	-0.4008	-0.3079	0.6389	0.6148	0.5887
19	0.3159	0.5378	-0.1166	0.3714	0.6161	0.3498	0.5792	0.0126	0.1777	0.5197	0.6299	0.5612
20	0.2081	-0.2821	-0.0806	0.3971	0.6765	0.0886	0.6069	-0.2500	-0.2428	0.6823	0.6788	0.6107
21	0.4333	0.5803	-0.0692	0.2209	0.2633	-0.1093	0.1696	0.2282	-0.1268	0.3471	0.2215	0.1105
22	0.3633	0.0631	-0.0545	0.5510	0.7443	0.1952	0.7158	-0.1567	-0.2733	0.7446	0.7251	0.6857
23	0.0676	-0.1720	0.0243	0.3229	0.5006	0.2478	0.5286	-0.1658	-0.0351	0.4541	0.6147	0.6345
24	0.4298	0.1465	-0.3188	0.3516	0.1764	0.2561	0.2483	0.3710	0.0812	0.3874	0.2701	0.2721
25	0.4622	0.1980	-0.0694	0.2945	0.6127	0.3975	0.5356	0.2279	0.1465	0.5001	0.5731	0.5195
26	0.7336	0.3038	-0.0678	0.6123	0.6274	0.8060	0.6554	0.5264	0.5982	0.5436	0.6898	0.7003
27	0.2795	-0.0589	-0.1400	0.3912	0.7363	0.1585	0.6836	-0.1848	-0.1918	0.6901	0.7289	0.6729
28	0.8640	0.5220	-0.0783	0.4203	0.5585	0.6346	0.5370	0.6982	0.4534	0.4532	0.5384	0.5317
29	0.5889	-0.4704	-0.0482	0.5427	0.8166	0.6650	0.8180	0.2379	0.3148	0.6668	0.8164	0.8224
30	0.6869	0.1614	-0.1892	0.2681	0.6666	0.5861	0.6127	0.4851	0.4088	0.4945	0.6129	0.5856
31	0.4336	0.5534	0.1424	0.2843	0.2399	0.2189	0.2449	0.2952	0.0968	0.3387	0.3510	0.3062
32	0.6807	-0.1964	-0.1505	0.3428	0.7822	0.6071	0.6515	0.4086	0.4018	0.5909	0.6687	0.6488
33	0.2918	0.7293	-0.0846	0.5761	0.6209	0.3077	0.6198	-0.0700	-0.0318	0.6471	0.6933	0.7030
34	0.1095	0.4142	0.1379	0.5858	0.4025	0.0414	0.4312	-0.2737	-0.2931	0.5703	0.5243	0.5266
35	0.0903	0.5670	-0.0125	0.2227	0.4423	0.1765	0.4552	-0.1235	0.0286	0.3878	0.5131	0.4274
36	0.7055	0.1706	0.0548	0.4314	0.4525	0.7378	0.4967	0.6190	0.5917	0.3879	0.5174	0.5188
37	0.6838	0.2317	-0.1446	0.6000	0.8585	0.5896	0.8320	0.2743	0.2304	0.8522	0.8936	0.8445
38	0.2633	0.5330	-0.0322	0.1799	0.4491	0.4586	0.4437	0.1374	0.5322	0.3272	0.4847	0.5099
39	0.4860	-0.0473	-0.1059	0.1743	0.6006	0.3270	0.4670	0.5229	0.4151	0.4022	0.5956	0.4750
40	0.6100	0.8670	-0.0012	0.7916	0.8471	0.6269	0.8892	0.1724	0.1657	0.7997	0.8981	0.9003
41	0.2971	0.5759	-0.0234	0.4162	0.6546	0.2749	0.5813	-0.0901	0.0090	0.5613	0.5750	0.5417
42	0.7501	0.7114	-0.0215	0.5931	0.6226	0.6298	0.5859	0.4850	0.3641	0.4601	0.6082	0.6115
43	0.3040	0.4377	-0.1162	0.5040	0.7578	0.1420	0.7129	-0.2022	-0.2779	0.8208	0.7916	0.7149
44	-0.1101	0.3521	0.0989	-0.0492	0.0651	0.3443	0.1782	0.0176	0.2109	-0.0640	0.1110	0.2245
45	0.3489	0.3328	-0.2713	0.0372	0.4420	0.3351	0.4538	0.1485	0.3069	0.4469	0.5023	0.4964

TABLE 4.9 – Correlations de Spearman par médicament

1. Une valeur proche de 0 indique une corrélation faible, mais cela ne signifie pas que les variables sont complètement indépendantes. Les tests de Kendall et de Spearman s'intéressent principalement à la tendance générale des variables à évoluer ensemble, indépendamment de la forme spécifique de cette relation. Lorsqu'une corrélation faible est obtenue avec les tests de Kendall ou de Spearman, cela peut indiquer que les valeurs des variables ne suivent pas de manière cohérente une tendance ascendante ou descendante conjointe. Cependant, il est important de noter que d'autres types de relations pourraient toujours exister entre les variables. Par exemple, il pourrait y avoir une relation non monotone ou une relation qui n'est pas capturée par la mesure de corrélation utilisée.

En général, les résultats sont acceptables même s'il existe des liens faibles souvent statistiquement significatifs. En pratique, pour une étude de machine learning, la question du lien faible n'est pas rejetée dans le sens où même s'il y a des liens faibles, la combinaison de plus de deux variables indépendantes peut être très explicative de la variable dépendante.

### 4.5.3 Etude complémentaire sur la variable "MOIS"

L'observation faite sur la variable "MOIS" reste valable dans ces tables. Il n'y a pas de lien significatif pour les différents tests de Kendall et de Spearman, en raison du fait que cette variable est de type qualitative. Pour l'étude du lien entre une variable qualitative (MOIS) et une variable quantitative (QUANTITE), le test de Kruskal-Wallis est utilisé à la place de l'ANOVA (un test paramétrique nécessitant l'hypothèse de normalité) pour examiner le lien entre les deux variables. Après l'application de ce test, la table 4.10 est obtenue.

Les hypothèses du test de Kruskal-Wallis sont les suivantes :

- Hypothèse nulle (H0) : Les distributions des groupes (les mois 1, 2, 3, ..., 12 dans notre cas) sont égales (c'est-à-dire qu'il n'y a pas de différence significative entre les groupes).
- Hypothèse alternative (H1) : Au moins une des distributions des groupes est différente des autres.

Comme interprétation des résultats, avec l'exemple de la première ligne :

- *Statistic* : La statistique de test de Kruskal-Wallis, a une valeur de 179.294. Cette valeur est utilisée pour calculer la p-value.
- *p\_value* : La p\_value est une mesure de l'évidence contre l'hypothèse nulle. Dans l'exemple, la p\_value est très proche de zéro ( $6.91e - 41$ ), ce qui signifie qu'il y a une très faible probabilité d'observer ces données si l'hypothèse nulle était vraie (c'est-à-dire s'il n'y avait pas de différence entre les groupes).
- *Interprétation de la p\_value* : Si la p-value est inférieure au niveau de signification (généralement fixé à 0.05), on rejette l'hypothèse nulle en faveur de l'hypothèse alternative. Dans cet exemple, la p\_value est beaucoup plus petite que 0.05, ce qui signifie qu'il y a des preuves statistiquement significatives pour conclure qu'il y a au moins une différence entre les distributions des groupes.

En définitive dans notre exemple, avec une p\_value aussi faible, on peut conclure que les différences entre les groupes sont statistiquement significatives, et donc, il y a un lien

MED	STATISTIC	P_VALUE
1	179.294	6.91 e-41
2	215.334	9.42e-49
3	206.323	8.71e-47
4	218.454	1.976e-49
5	208.918	2.36e-47
6	174.808	6.59e-40
7	105.045	1.19e-24
8	29.991	4.34e-08
9	213.693	2.15e-48
10	217.01	4.06e-49
11	111.971	3.63e-26
12	219.944	9.30e-50
13	201.016	1.25e-45
14	189.141	4.89e-43
15	201.97	7.76e-46
16	216.943	4.2e-49
17	182.994	1.07e-41
18	214.535	1.41e-48
19	197.566	7.09e-45
20	221.446	4.37e-50
21	179.44	6.42e-41
22	212.224	4.49e-48
23	199.516	2.66e-45
24	32.839	1.001e-08
25	222.477	2.61e-50
26	216.805	4.50e-49
27	216.589	5.017e-49
28	218.443	1.977e-49
29	219.944	9.30e-50
30	221.446	4.38e-50
31	206.805	6.84e-47
32	214.202	1.66e-48
33	1.696	<b>0.193</b>
34	216.953	4.18e-49
35	214.297	1.59e-48
36	216.989	4.1e-49
37	101.621	6.72e-24
38	171.21	4.03e-39
39	147.568	5.89e-34
40	2.682	<b>0.101</b>
41	203.979	2.827e-46
42	143.085	5.63e-33
43	172.606	1.99e-39
44	166.009	5.51e-38
45	141.636	1.167e-32

TABLE 4.10 – Test statistique pour la corrélation entre la consommation et le "MOIS"

significatif entre la variable catégorielle ("MOIS") et la variable continue ("QUANTITE") étudiées. Les autres informations sont interprétées identiquement.

A l'exception des médicaments 33 et 40, dont les p\_values sont supérieures à 0.05 ( $> 0.05$ ), on peut déduire que les variables "MOIS" et "QUANTITÉ" sont associées pour la plupart des médicaments. Mais les résultats du test ne permettent pas d'avoir un avis définitif sur la question de la saisonnalité.

## 4.6 Conclusion

La détermination des paramètres de dimensionnement et d'estimation des besoins de médicaments dans les hôpitaux implique la connaissance des potentiels facteurs de consommation. L'objectif de ce chapitre est de passer en revue les facteurs pertinents qui ont un impact sur la consommation de médicaments dans les hôpitaux français. L'ensemble des facteurs identifiés dans la littérature sont analysés dans le contexte français tout en bénéficiant d'analyses d'experts dans le domaine hospitalier. Une analyse de corrélation a été réalisée entre ces facteurs et la consommation de médicaments afin d'identifier les facteurs pertinents pouvant être utilisés dans les méthodes de prévision. Ce travail a été réalisé avec l'historique de consommation de 45 médicaments en commun au six hôpitaux.

Les analyses démontrent l'existence d'une corrélation, souvent forte et statistiquement significative, entre la quantité de médicament consommée et certains des facteurs retenus. On observe également que les corrélations sont appréciables par médicament. Ceci implique qu'il faudrait un modèle de prévision pour chaque médicament.

L'étude pourrait être améliorée en prenant en compte certains des facteurs exclus. Le principal frein est la disponibilité et la qualité des données. L'absence d'une nomenclature standard des services hospitaliers et des unités fonctionnelles ne permet pas des analyses globales de lien entre consommation de médicament, de produit de santé en général, dans divers structures hospitalières.

Avec la contextualisation et les liens établis entre les facteurs et la consommation de médicaments, ces résultats seront utilisés dans les chapitres suivants pour la construction de modèles de prévision de la demande dans le contexte hospitalier français. Ces modèles pourront être utilisés pour gérer efficacement les besoins en médicaments d'un hôpital donné ou pour dimensionner les stocks d'une nouvelle structure de santé en utilisant des outils d'apprentissage automatique (machine learning).

Les travaux de ce chapitre ont fait l'objet de publication dans deux conférences scientifiques :

1. CIGI-Qualita 21 : Conférence Internationale Génie Industriel QUALITA du 5 au 7 mai 2021 à Grenoble en France. Le papier publié est intitulé : Koala, D., Yahouni, Z., Alpan, G., Frein, Y. (2021). ***Factors influencing Drug Consumption and Prediction Methods***. arXiv preprint arXiv :2109.11779 [69].
2. MIM 2022 : 10<sup>th</sup> IFAC Conference on manufacturing modelling, management and control – Nantes du 22 au 24 juin 2022 à Nantes en France. Le papier publié est intitulé : Koala, D., Yahouni, Z., Alpan, G., MOHAND, D. S. (2022). ***Correlation analysis of factors impacting health product consumption in French hospitals***. IFAC-PapersOnLine, 55(10), 895-900 [70].

# Chapitre 5

## Étude comparative des modèles de machine learning vs méthodes classiques

### Sommaire

---

<b>5.1</b>	<b>Introduction</b>	<b>105</b>
<b>5.2</b>	<b>Méthodologie et démarche</b>	<b>109</b>
5.2.1	Extraction et collecte des données	109
5.2.2	Exploration et pré-traitement des données	110
5.2.3	Sélection des données	111
5.2.4	Application des modèles	112
5.2.5	Évaluation et comparaison	116
<b>5.3</b>	<b>Résultats expérimentaux et analyses</b>	<b>117</b>
5.3.1	Évaluation des performances des techniques d'apprentissage automatique sélectionnées	118
5.3.2	Comparaison des modèles de Random Forest, du DNN et de la méthode de consommation	120
<b>5.4</b>	<b>Conclusion et perspectives</b>	<b>122</b>

---

### 5.1 Introduction

L'objectif de ce chapitre est d'évaluer certaines techniques d'apprentissage automatique et méthodes classiques de quantification, pour estimer la demande de médicaments. Les performances des méthodes d'apprentissage automatique sélectionnées sont comparées à un modèle d'estimation classique, la méthode de consommation en utilisant les données réelles de consommation pré-traitées dans le chapitre 3. Dans ce chapitre, nous expérimentons les modèles d'apprentissage automatique en combinant les historiques de consommations des hôpitaux et prenant en compte les facteurs étudiés dans le chapitre 4. La combinaison des données des hôpitaux ainsi que la diversité des facteurs de consommation nous amènent à utiliser des méthodes causales.

Parmi les méthodes classiques mentionnées dans le chapitre 2, les méthodes de morbidité et de consommation sont les plus fréquemment utilisées et sont également considérées comme les plus précises pour estimer les besoins en médicaments [75, 38, 31, 35].

Dans leur étude empirique rétrospective, Sanubari et al. [136] comparent les performances de la méthode de consommation et de la méthode de morbidité pour estimer les stocks mensuels de médicaments de chimiothérapie. Les résultats montrent que la méthode de morbidité a une valeur moyenne du mois de stock inférieure à la méthode de consommation (1,853 mois contre 2,125 mois). La planification avec une combinaison des deux méthodes a donné une valeur du mois de stock de 1,41 mois, plus rentable que la livraison normale. Les résultats montrent que la méthode de morbidité a recommandé 67 médicaments de chimiothérapie à acheter, contre 55 pour la méthode de consommation. La méthode de morbidité a donc une meilleure précision que la méthode de consommation.

Johnson et al. [63], dans une étude statistique récente, ont mené une enquête sur les pratiques de prévision de la demande de médicaments ARV, ayant pour objectifs d'identifier les pratiques de gestion des stocks utilisées, d'évaluer la performance de la chaîne d'approvisionnement et de déterminer les défis affectant la gestion des stocks de médicaments ARV dans les hôpitaux. Il est apparu que les méthodes de prévision prédominantes utilisées pour déterminer les quantités de commandes étaient la méthode de consommation (100 %) et la méthode de morbidité (80,95 %). Quatre méthodes ont été soumises aux enquêtés, à savoir la méthode de consommation, la méthode de morbidité, le "Judgmental method" et le "causal method". Cela montre que, dans la pratique, au sein des hôpitaux, les méthodes classiques, en particulier la méthode de consommation, sont toujours courantes. Une hypothèse qui est appuyée par l'étude de Iqbal et al. [58]. Dans leurs travaux, les auteurs montrent que la méthode de consommation, ainsi que les données de distribution, les variations saisonnières, etc., sont utilisées pour évaluer la quantification et la prévision des besoins en médicaments dans les établissements de santé étudiés.

Dans de nombreux cas, la méthode classique la plus précise pour quantifier les besoins en médicaments est l'approche basée sur la consommation, compte tenu de la simplicité de son application et du type de données à collecter. Ainsi, dans le cadre de ce chapitre, cette méthode est retenue à des fins de comparaison. De la littérature, certaines études comparatives impliquent les méthodes classiques par rapport à d'autres méthodes. Par exemple, Dewi et al. [31], dans une étude visant à déterminer la gestion efficace des médicaments subdivisés en trois groupes (stagnants, manquants et normaux) en utilisant trois méthodes (la quantité de commande économique, le niveau de stock minimum-maximum), montrent que la méthode de consommation est la meilleure pour déterminer les médicaments les moins manquants. As-thariq et al. [11] comparent également la méthode de consommation à une méthode hybride appelée ABC-EOQ-ROP-SS qui combine la méthode ABC, EOQ, ROP (Reorder point) et SS (Safety Stock). Cependant, une comparaison entre les méthodes classiques et les méthodes de prévision de la demande utilisant l'apprentissage automatique reste un champ à explorer, d'où l'intérêt de ce travail.

Même si le caractère imprévisible de la demande hospitalière amène certains auteurs à

dire qu'il est très difficile, voire impossible, de prédire les besoins en médicaments [123, 12], d'autres auteurs tels que Volland et al. [160] et Brennan [24] précisent que la prévision de la demande revêt une grande importance, notamment pour les responsables des achats et de l'approvisionnement hospitalier, car les contrats les liant aux fournisseurs reposent parfois sur des quantités minimales. Et comme le soulignent Varghese et al. [157] et Makridakis et al. [81], pour faire face à l'incertitude de la prévision de la demande, des outils statistiques et d'apprentissage automatique peuvent être utilisés.

De ces faits, et en croisant l'état de l'art sur la prévision de la demande de médicaments dans la section 2.2 du chapitre 2, les modèles proposés dans la plupart des études ne sont pas généralisables à plusieurs médicaments et sont souvent centrés sur un seul hôpital. L'approche choisie permet de prendre en compte cet aspect et est justifiée par l'intégration de multiples facteurs ayant un impact sur la consommation de médicaments et la prise en compte de données historiques provenant de plusieurs hôpitaux. Ces facteurs caractérisent les hôpitaux. De même, il a été noté que ces modèles ne sont pas comparés à une méthode classique, d'où l'intérêt supplémentaire de ce travail.

Le tableau ci-dessous, la table 5.1, résume les principales gaps sur le sujet dans la littérature.

Le déroulé de ce chapitre se présente comme suit : La section 5.2 est consacrée à la description de la méthodologie utilisée, de la collecte des données à l'évaluation des performances et à la comparaison des modèles. La section 5.3 est dédiée aux expériences numériques sur des données réelles et à l'analyse des résultats. L'étude se termine par une note de conclusions et des suggestions pour des recherches futures dans la section 5.4.

	Auteurs	Données des modèles		Modèle	Comparaison des modèles			Nombre de médicaments	Modèle par médicament	Modèle unique pour tous les médicaments	Modèle quantitatif de prévision de la demande	Intégration des facteurs	Modèle généralisable
		Données réelles	Données simulées		ML versus méthodes classiques	ML versus méthodes statistiques	Entre les modèles de ML						
1	MSH[33]		✓	C			F	✓					
2	Soeters[149]		✓	C			109	✓					
3	Osore [111]		✓	C			45	✓					
4	Mbonyinshuti et al[89]		✓	ML			10	✓					
5	Qingkui et al [122]	✓		ML			1	✓				N	
6	Guo [44]	✓		ST			1	✓			✓		
7	Motamedi et al [98]	✓		M		✓	1	✓			✓		
8	Sarvestani et al [137]	✓		M		✓	8	✓			✓		
9	Rathipriya et al[125]		✓	M		✓	57*	✓			✓		
10	Schilling et al[139]	✓		M		✓	1	✓			✓		
11	Dewi et [31]	✓		C			126						✓
12	Asthariq et al[11]	✓		C			269						✓
	<b>Notre travail</b>	✓		<b>C/ML</b>	✓	✓	<b>10</b>	✓			✓	✓	✓

\*57 médicaments regroupés en 8 ; C : méthode classique d'évaluation hospitalière ; ST : méthode statistique ; ML : modèle d'apprentissage automatique ; M : Modèles statistiques et d'apprentissage automatique ; N : Utilisation principalement du nombre de patients comme facteur ; F : Données fictives pour expliquer les méthodes.

TABLE 5.1 – Résumé des gaps de la littérature et nouveauté de ce travail.

## 5.2 Méthodologie et démarche

Cette section est dédiée à la description détaillée de notre méthodologie en cinq étapes appliquée dans cette étude (voir Figure 5.1).

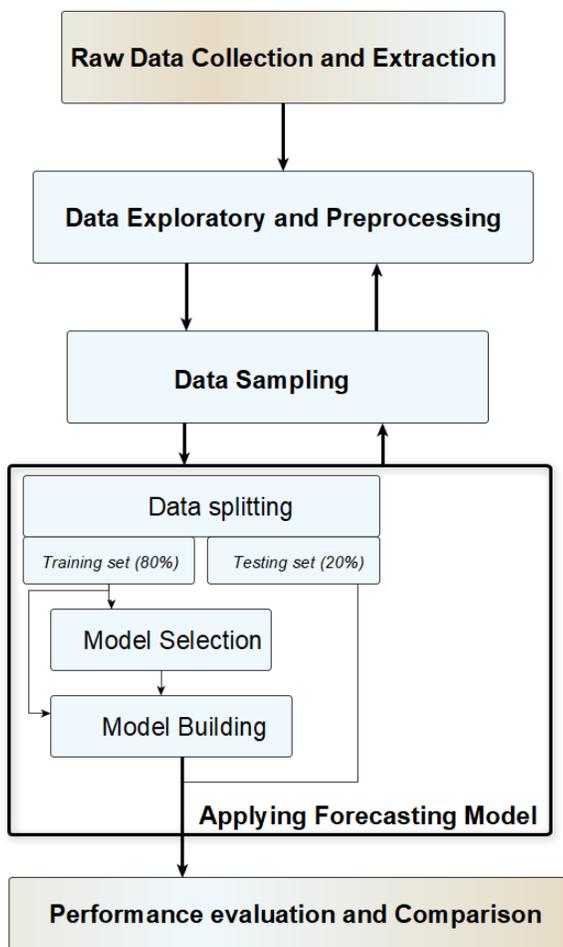


FIGURE 5.1 – Approche proposée

### 5.2.1 Extraction et collecte des données

Les données utilisées dans cette étude relèvent de celles collectées, extraites et organisées dans le chapitre 3 de ce manuscrit. La procédure de cette phase est amplement détaillée dans le chapitre en question. Dans ce travail, ce sont les données de consommation de médicaments de six hôpitaux français situés dans six régions différentes, qui sont utilisées. Pour rappel, dans nos travaux, la consommation se réfère à la livraison de médicaments aux unités fonctionnelles des hôpitaux. Les données hospitalières proviennent essentiellement de trois bases de données interdépendantes : les médicaments, les historiques de consommation et les unités fonctionnelles. Ces données hospitalières sont pré-traitées et croisées avec des données provenant d'organismes officiels. La base de données utilisée pour cette étude est similaire à celle de l'étude précédente et contient les attributs regroupés dans la table 5.2 :

	<b>Nom de la colonne</b>	<b>Description</b>
<b>0</b>	ID_SITE	Identifiant de l'hôpital
<b>1</b>	ID_REF	Identifiant du médicament par hôpital
<b>2</b>	DESIG	Nom du médicament
<b>3</b>	CODE_UCD	Code UCD du médicament
<b>4</b>	CODE_DEPT	District de l'hôpital
<b>5</b>	ANNEE	Année de consommation
<b>6</b>	MOIS	Mois de consommation
<b>7</b>	N_ETB	Nombre d'établissements hospitaliers
<b>8</b>	N_UFS	Nombre d'unités fonctionnelles
<b>9</b>	LIT_HC	Nombre de lits pour hospitalisation complète
<b>10</b>	LTT_HP	Nombre de lits pour hospitalisation partielle
<b>11</b>	SEJ_MCO	Nombre de visites dans les services MCO
<b>12</b>	SEJ_HAD	Nombre de visites dans les services HAD
<b>13</b>	SEJ_SLD	Nombre de visites dans les services SLD
<b>14</b>	SEJ_SSR	Nombre de visites dans les services SSR
<b>15</b>	SEJ_PSY	Nombre de visites dans les services PSY
<b>16</b>	PN_MEDICAL	Nombre personnel non-médicaux (infirmiers, sages-femmes, etc.)
<b>17</b>	P_MEDICAL	Nombre de personnel médicaux
<b>18</b>	POPULATION	Nombre de population couvert par les hôpitaux
	<b>QUANTITÉ</b>	<b>Quantité de médicament consommée mensuellement</b>

TABLE 5.2 – Les caractéristiques principales l'ensemble de données

Les données sont collectées de janvier 2016 à février 2019 afin d'exclure la période de Covid-19. Le processus décrit ci-dessus conduit à une table finale qui est utilisée pour l'analyse. Les données utilisées sont principalement quantitatives.

### 5.2.2 Exploration et pré-traitement des données

Pour toute application et pour obtenir de meilleures performances d'un modèle d'apprentissage automatique (ML), il est toujours impératif de s'assurer de la qualité des données. L'exploration des données est donc nécessaire. Dans ce travail, nous appliquons principalement des techniques de fouille des données telles que la visualisation des données, le nettoyage des données et la correction des données.

Notre jeu de données suite aux pré-traitements dans le chapitre 3, comportaient six hôpitaux au lieu de cinq. Avant la réalisation de ce chapitre, un examen minutieux a été effectué avec l'entreprise partenaire KLS. Il se trouve que l'hôpital H6 est un grand regroupement de plusieurs hôpitaux, l'équivalent d'un groupement hospitalier de territoire. Les principales unités fonctionnelles de cette hôpital était en réalité des points de stockage des produits de

santé pour une redistribution. De ce fait, nous avons convenus d'exclure cette hôpital dans les analyses futures.

Dans la base, La variable "SEJ\_SSR" ne contenait pas de données pour la période allant de 2016 à 2018 inclus, pour les différents hôpitaux. Cela nous a conduit à exclure cette variable pour la suite de notre travail. Par ailleurs, il a été nécessaire de procéder au typage des données. Ainsi, par exemple, la variable "QUANTITE", qui était de type "Object", a été convertie en type float. De même, lors de cette phase, la visualisation a permis d'observer les tendances de consommation de chaque médicament par hôpital et d'exclure les valeurs aberrantes.

Cette phase a également été consacrée à la sélection des variables à prendre en compte dans les modèles. La sélection des variables s'est faite en deux temps : la collecte de l'avis des experts hospitaliers et une étude de corrélation proposée dans le chapitre précédent et le papier [70]. Ainsi, à partir du Tableau 5.2, 12 variables, à savoir "POPULATION", "P\_MEDICAL", "MONTH", "N\_ETB", "N\_UFS", "LIT\_HC", "LTT\_HP", "SEJ\_MCO", "SEJ\_HAD", "SEJ\_SLD", "SEJ\_PSY", "PN\_MEDICAL", "P\_MEDICAL", ont été retenues comme les 12 facteurs potentiels ayant un impact sur la consommation de médicaments que nous utiliserons dans nos modèles.

### 5.2.3 Sélection des données

Durant la visualisation, nous avons constaté que certains médicaments, bien que communs aux cinq hôpitaux, présentaient une consommation assez différente en termes de quantité et flux. Il se trouve que d'autres médicaments ont une faible consommation en termes de flux (faible historique d'enregistrement) pour être utilisés dans les modèles ML. Ainsi, sur les 45 médicaments communs aux cinq hôpitaux, nous avons retenu, pour les besoins de ce travail, les 10 médicaments les plus consommés, présentés dans le tableau 5.3 :

<b>MED</b>	<b>CODE ATC</b>	<b>Nom du ATC</b>	<b>CODE</b>	<b>CODE UCD</b>
<b>01</b>	L01XC02	RITUXIMAB		3400891977196
<b>02</b>	H02AB02	DEXAMETHASONE		3400892230849
<b>03</b>	N05CD08	MIDAZOLAM		3400892376950
<b>04</b>	C02AC01	CLONIDINE		3400890168243
<b>05</b>	L04AB02	INFLIXIMAB		3400892137377
<b>06</b>	L04AX04	LENALIDOMIDE		3400892981420
<b>07</b>	J01DC01	CEFOXITINE		3400892124421
<b>08</b>	A10AB05	INSULINE PARTE	AS-	3400892402437
<b>09</b>	J01GB06	AMIKACINE		3400892085937
<b>10</b>	L04AC07	TOCILIZUMAB		3400893318911

TABLE 5.3 – Les médicaments sélectionnés

Dans la suite de ce travail, notre étude se concentre sur ces 10 médicaments. Suite à cette décision et après avoir bien exécuté l'étape du pré-traitement des données, nous obtenons 1031 lignes de données, un nombre limité en raison d'agrégation par mois.

## 5.2.4 Application des modèles

Cette phase est divisée en deux parties principales : l'application de la méthode de consommation et l'application des modèles d'apprentissage automatique.

### 5.2.4.1 Méthode de consommation

Le choix de cette méthode est basé sur deux critères principaux :

1. L'opinion dans la littérature : qui présente le modèle de consommation (voir l'équation (5.2.1)) comme la méthode la plus efficace et populaire par rapport aux autres méthodes classiques.
2. Disponibilité des données : Les données historiques sur la consommation de médicaments sont disponibles dans notre contexte, tandis que la méthode de morbidité nécessite des informations sur les pathologies traitées, les dosages, etc., qui ne sont pas disponibles pour notre étude.

Cette méthode est basée sur la consommation des 12 derniers mois pour déterminer celle du mois suivant. Elle est appliquée pour chaque médicament et pour chaque hôpital. En tant que méthode déterministe, l'application de la méthode de consommation est simple. Notre base de données ne contient que les informations de consommation mensuelle. La quantité de consommation projetée est calculée comme suit :

$$Q_A = \left[ \frac{C_T}{R_M - \frac{D_{os}}{30.5}} \times (1 + A_u) \times (P_P + L_T) + S_s - (S_I + S_o) \right] \times (1 + A_L) \quad (5.2.1)$$

$C_T$  : Consommation totale pendant la période considérée, en unités de base (telles que millilitres ou grammes)

$R_M$  : Période totale considérée de la consommation (en mois)

$D_{os}$  ou  $M_{os}$  : Nombre de jours ou de mois où un produit était en rupture de stock pendant la période considérée. Dans l'équation, 30.5 jours correspondent à la moyenne des jours d'un mois.

$A_u$  : Ajustement de l'utilisation (environ 5 % si l'utilisation est supposée augmenter dans la période à venir)

$P_P$  : Période d'approvisionnement (nombre de mois à couvrir par la commande)

$L_T$  : Délai de livraison moyen (pour l'approvisionnement projeté ou le pire des cas), en mois

$S_s$  : Quantité nécessaire pour le stock de sécurité. Le stock de sécurité est la quantité de stock conservée en réserve au cas où un article serait indisponible chez le fournisseur ou en cas d'augmentation soudaine de la demande.

$S_I$  : Stock réel en inventaire, en unités de base

$S_o$  : Stock réel en commande mais pas encore reçu, en unités de base.

$A_L$  : Ajustement des pertes entre 5 % et 30 % (en raison de vols, d'expiration, etc.).

Les paramètres  $S_s$ ,  $S_I$  et  $S_o$  ont été négligés pour rester dans l'élan de la consommation. La valeur de 5 % a été attribuée aux taux d'ajustement  $A_L$  et  $A_u$ . Dans le contexte des hôpitaux français sélectionnés, ces chiffres sont raisonnables pour les pertes dues à l'expiration et à l'évolution de la consommation de médicaments.

#### 5.2.4.2 Modèles d'apprentissage automatique

Dans ce travail, nous utiliserons l'apprentissage supervisé pour prédire la consommation de médicaments dans les hôpitaux. Un modèle d'apprentissage automatique supervisé apprend à faire correspondre des variables d'entrée à des variables de sortie en s'appuyant sur un ensemble de données étiquetées contenant à la fois les données d'entrée et les valeurs de sortie correspondantes (étiquettes). L'objectif du modèle est de prédire avec précision la sortie pour de nouvelles données d'entrée non vues. L'apprentissage supervisé comporte deux approches principales : la régression et la classification, qui ont toutes deux pour objectif de construire un modèle capable de prédire la valeur de l'attribut dépendant (sortie) à partir des variables attributaires (entrée). La différence entre ces deux tâches est que l'attribut dépendant est un nombre continu pour la régression et des étiquettes catégorielles ou des classes pour la classification. Pour plus de détails, voir [2]. La régression est plus appropriée pour traiter notre problème. Il existe plusieurs types d'algorithmes de régression et dans le cadre de ce travail, nous avons retenu la régression linéaire (LR), les méthodes d'ensemble telles que Random Forest (RF), AdaBoost, Gradient Boosting, ainsi que le Deep Neural Network (DNN).

**La méthode de régression linéaire (LR)** est une technique statistique et l'un des modèles d'apprentissage automatique supervisé les plus simples qui permet de prédire la valeur d'une variable cible en utilisant un certain nombre de variables explicatives. L'objectif est d'établir une relation linéaire entre les variables explicatives indépendantes  $X$  et la variable cible dépendante  $Y$  :  $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_p X_p$ , où  $\alpha_0$  est l'intercepteur et  $\alpha_i$  sont les coefficients. Pour plus de détails, voir [2, 87].

**Le modèle Random Forest (RF)** est une approche d'ensemble pour les problèmes de régression et d'autres tâches (comme la classification) qui construit un grand nombre de choix discrétionnaires lors de l'apprentissage. Comme son nom l'indique, RF est un ensemble basé sur des arbres, chaque arbre dépendant d'un ensemble de variables aléatoires. Mathématiquement, pour un vecteur aléatoire  $X$  à  $p$  dimensions  $X = (X_1, \dots, X_p)^T$  représentant les variables d'entrée ou prédicteurs réels et une variable aléatoire  $Y$  indiquant la réponse réelle, nous supposons une distribution conjointe inconnue  $P_{XY}(X, Y)$ . L'objectif est de trouver une fonction de prédiction  $f(X)$  pour prévoir  $Y$ . Pour plus de détails, voir [163]. En pratique, l'approche des forêts aléatoires est utilisée avec plusieurs paramètres tels que " $n\_estimators$ "

(le nombre d'arbres dans la forêt), *max\_depth*, *random\_state*, etc. (voir la documentation de scikit-learn pour plus de détails<sup>1</sup>).

Le **AdaBoost** également appelé *Adaptive Boosting* est une technique de boosting utilisée comme méthode ensembliste en apprentissage automatique. Le principe de base de AdaBoost est d'ajuster une séquence de modèles faibles sur des données d'entraînement pondérées différentes. Il commence par prédire l'ensemble de données d'origine et attribue un poids égal à chaque observation. Si la prédiction est incorrecte avec le premier modèle, il donne un poids plus élevé aux observations qui ont été prédites de manière incorrecte. Étant un processus itératif, il continue d'ajouter des modèles jusqu'à ce qu'une limite soit atteinte en termes de nombre de modèles ou de précision. Pour plus de détails, voir [2].

Comme le modèle Random Forest, le modèle de **Gradient Boosting** est un modèle ensembliste. Dans ce modèle, la procédure d'apprentissage ajuste successivement de nouveaux modèles pour fournir une estimation plus précise de la variable de réponse. L'idée principale est de construire les nouveaux apprenants de base de manière à être corrélés avec le gradient négatif de la fonction de perte associée à l'ensemble complet. Les fonctions de perte appliquées peuvent être arbitraires, mais pour donner une meilleure intuition, si la fonction d'erreur est la fonction de perte classique de l'erreur quadratique, la procédure d'apprentissage donnerait des ajustements successifs de l'erreur. Pour plus de détails, voir [34].

Le **DNN** est un réseau de neurones artificiels (ANN) avec au moins deux couches cachées entre la couche d'entrée et la couche de sortie, et un certain niveau de complexité. En d'autres termes, un DNN est un ANN composé d'au moins quatre couches, comprenant la couche d'entrée, les couches cachées et la couche de sortie [78]. L'ANN est une technique de modélisation basée sur le système nerveux humain qui permet l'apprentissage par l'exemple à partir de données représentatives pour expliquer un événement réel ou un processus conduisant à la prise de décision. Un aspect déterminant de l'ANN est sa capacité à établir des liens empiriques entre les variables indépendantes et dépendantes, ainsi qu'à extraire des informations nuancées et des connaissances sophistiquées à partir d'ensembles de données représentatifs. Pour plus de détails, voir [73, 28].

Notre structure de base du DNN (indiquée à la figure 5.2) est composée de 3 couches denses, la dernière étant la sortie. La première couche dense est composée de 128 unités avec la fonction d'activation ReLU (Rectified Linear Unit) et le nombre de colonnes pour la forme d'entrée. La deuxième couche dense est composée de 64 unités avec à nouveau ReLU comme fonction d'activation. La dernière couche dense n'a qu'une seule unité pour la prévision. Étant donné que nous traitons le problème de régression visant à prédire une valeur positive (QUANTITÉ), les fonctions d'activation les plus appropriées sont ReLU et Linear. La fonction de perte utilisée pour l'ensemble de l'étude est l'erreur quadratique moyenne (MSE) et l'erreur absolue moyenne (MAE) pour la métrique d'apprentissage. L'optimiseur Adam avec un taux d'apprentissage de 0,001 est également utilisé pour la construction du modèle DNN de chaque médicament avec une taille de lot de 8 et le paramètre d'époques de 300.

---

1. Lien scikit-learn sur Random Forest : <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html?highlight=random%20forest#sklearn.ensemble.RandomForestRegressor>

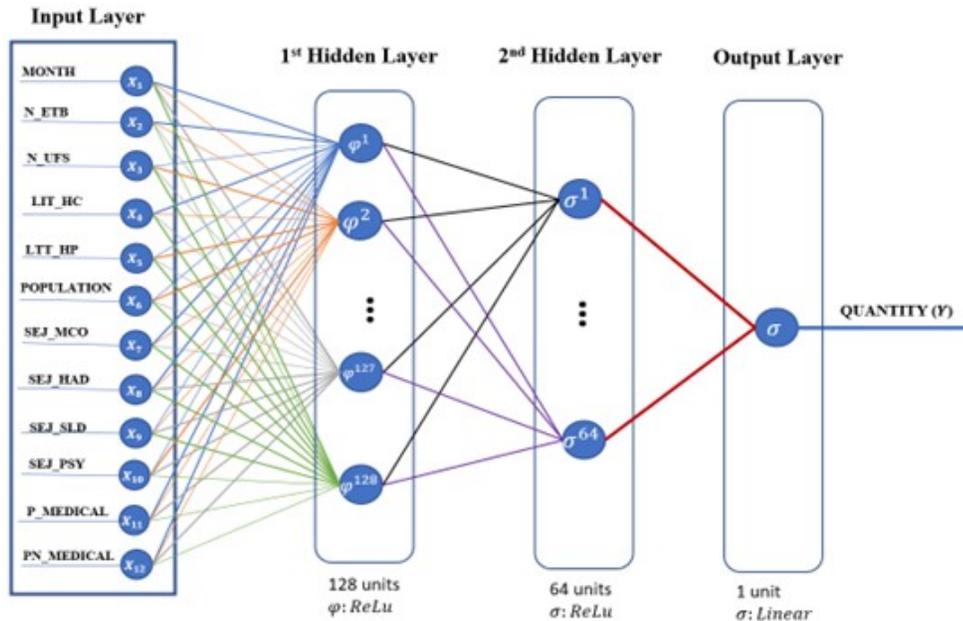


FIGURE 5.2 – Structure basique de notre DNN

Pour éviter le surajustement, nous avons utilisé une fonction de rappel. Les lecteurs intéressés par les différents paramètres et hyperparamètres impliqués dans l'application pratique de l'apprentissage profond peuvent consulter la plateforme Keras TensorFlow <sup>2</sup>.

Dans les modèles d'apprentissage automatique, l'entraînement est effectué sur 80% (données d'entraînement) des données par médicament et les 20% restants (données pour les tests) sont utilisés pour le test du modèle.

Pour les méthodes d'apprentissage automatique autres que le DNN, une première application est réalisée à des fins de comparaison et pour sélectionner le meilleur modèle en utilisant la technique de validation croisée de Scikit-learn <sup>3</sup> sur l'ensemble d'entraînement. Scikit-learn est une bibliothèque d'apprentissage automatique open-source populaire en Python. Elle fournit une gamme d'outils pour différentes tâches d'apprentissage automatique telles que la classification, la régression, le regroupement et la réduction de dimension, entre autres. Scikit-learn est construit sur NumPy, SciPy et Matplotlib, qui sont également des bibliothèques Python populaires pour le calcul scientifique et la visualisation des données. La validation croisée n'ayant pas de correspondance dans le package TensorFlow, à la date de ce travail cela entraîne la non présence du DNN lors de cette phase sélective.

La validation croisée est une méthode statistique d'évaluation et de comparaison des algorithmes d'apprentissage en divisant les données en deux segments : l'un utilisé pour apprendre ou entraîner un modèle et l'autre utilisé pour valider le modèle. La validation croisée permet d'éviter le surajustement (*overfitting*) en fournissant une estimation de la manière dont le modèle se comportera sur de nouvelles données non vues auparavant. Dans une validation croisée typique, les ensembles d'apprentissage et de validation doivent se chevaucher dans

2. Lien d'accès à TensorFlow : [https://www.tensorflow.org/api\\_docs/python/tf/keras](https://www.tensorflow.org/api_docs/python/tf/keras)

3. Lien vers la validation croisée : [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.cross\\_validate.html#sklearn.model\\_selection.cross\\_validate](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.cross_validate.html#sklearn.model_selection.cross_validate)

des séquences successifs, de sorte que chaque point de données ait une chance d'être validé. La forme basique de la validation croisée est la validation croisée  $k - fold$ . D'autres formes sont des cas particuliers de la validation croisée  $k - fold$  ou impliquent des séquences répétées de validation croisée  $k - fold$ . Dans la validation croisée  $k - fold$ , les données sont d'abord partitionnées en  $k$  segments ou plis de taille égale (ou presque) [127]. Ensuite,  $k$  itérations d'apprentissage et de validation sont effectuées de sorte que, dans chaque itération, un pli différent des données est réservé à la validation, tandis que les  $k - 1$  autres plis sont utilisés pour l'apprentissage. Dans notre cas,  $k$  sera égal à 5 et les paramètres seront les paramètres par défaut pour chaque modèle.

Le modèle sélectionné comme étant le meilleur, s'il ne s'agit pas du modèle de régression linéaire (LR), est optimisé en utilisant le module GridSearchCV<sup>4</sup> de scikit-learn sur le paramètre  $n\_estimator$ , que nous avons fait varier entre 2 et 100. Le modèle obtenu est comparé avec le modèle DNN et le modèle de consommation.

### 5.2.5 Évaluation et comparaison

Pour l'évaluation et la comparaison des modèles, plusieurs propositions sont faites pour les modèles de régression dans la littérature [26]. Dans ce travail, nous avons sélectionné les métriques  $R^2$ , RMSE, MAE, et MAPE qui sont toutes utilisées pour évaluer la qualité d'un modèle de régression. Elles mesurent la différence entre les valeurs prédites par le modèle et les valeurs réelles.

**R-carré** ( $R^2$ ), appelé le coefficient de détermination, est utilisé pour indiquer la proximité entre une ligne de régression (valeurs prédites) et les valeurs réelles des données. Il est obtenu à l'aide de la formule suivante :

$$R^2 = 1 - \frac{\sum_i^n (reel_i - Pred_i)^2}{\sum_i^n (reel_i - moy_{Reel})^2} \quad (5.2.2)$$

où  $moy_{Reel} = \frac{1}{n} \sum_i^n reel_i$ .

$R^2$  mesure la proportion de la variance totale dans les données de la variable cible qui est expliquée par le modèle. Lorsque la métrique  $R^2$  tend vers 1, le modèle est considéré comme performant.

**L'erreur quadratique moyenne (RMSE, pour Root Mean Square Error)** : avec cette métrique, les erreurs sont mises au carré avant d'être moyennées. Cela signifie que le RMSE donne plus de poids aux erreurs importantes. Ainsi, lorsque de grandes erreurs sont présentes, elles ont une influence considérable sur les performances du modèle. Plus la valeur du RMSE est faible, meilleure est la performance du modèle.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_i^n (reel_i - Pred_i)^2}{n}} \quad (5.2.3)$$

4. Lien pour GridSearchCV : [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html#sklearn.model\\_selection.GridSearchCV](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html#sklearn.model_selection.GridSearchCV)

RMSE est lié au  $R^2$  car les valeurs mises au carré dans son calcul sont utilisées pour calculer la somme totale des carrés des erreurs, qui est également utilisée dans le calcul du  $R^2$ .

**L'erreur absolue moyenne (MAE, pour Mean Absolute Error)** mesure la distance moyenne entre les valeurs prédites et les valeurs réelles, en tenant compte de la valeur absolue de l'erreur. Elle est moins sensible aux valeurs aberrantes que le RMSE. Le MAE est calculé comme suit :

$$MAE = \frac{1}{n} \sum_i^n |reel_i - Pred_i| \quad (5.2.4)$$

**L'erreur moyenne absolue en pourcentage (MAPE, pour Mean Absolute Percentage Error)** mesure l'erreur moyenne en pourcentage entre les valeurs réelles (ou observées) et les valeurs prédites par le modèle. Le MAPE est calculé par l'équation (5.2.5) :

$$MAPE = \left( \frac{1}{n} \sum_{i=1}^n \left| \frac{reel_i - Pred_i}{reel_i} \right| \right) \times 100 \quad (5.2.5)$$

MAPE est une métrique qui est souvent utilisée dans les cas où les erreurs ne sont pas proportionnelles aux valeurs réelles. Elle est plus sensible aux outliers que RMSE et MAE, mais elle offre une meilleure compréhension de la précision des prédictions du modèle dans les cas où les erreurs sont importantes. Son utilisation est recommandée dans des tâches où il est plus important d'être sensible aux variations relatives qu'aux variations absolues. Cependant, il présente également un certain nombre d'inconvénients, les plus critiques étant la restriction de son utilisation à des données strictement positives par définition et son biais en faveur de prévisions faibles, ce qui le rend inadapté aux modèles prédictifs où de grandes erreurs sont attendues [26].

Comme pour le RMSE, plus la valeur du MAE ou MAPE est faible, meilleure est la performance du modèle.

## 5.3 Résultats expérimentaux et analyses

Dans cette section, nous présentons deux groupes d'expériences numériques : nous comparons d'abord les performances des techniques d'apprentissage supervisé sélectionnées. L'ensemble des données d'entraînement est utilisé à cette fin. Une seule de ces techniques sera sélectionnée pour le deuxième groupe de tests, qui consiste en une comparaison de la meilleure technique d'apprentissage supervisé, du modèle DNN et de la méthode de consommation. Notre ordinateur dispose des capacités suivantes : 16 Go de RAM, et le processeur est un Intel Core i5-1145G7 de 11e génération fonctionnant à une fréquence de base de 2,60 GHz et une fréquence de boost de 1,50 GHz.

### 5.3.1 Évaluation des performances des techniques d'apprentissage automatique sélectionnées

Suite à l'application de la validation croisée sur les quatre algorithmes d'apprentissage automatique, la régression linéaire, Random Forest, AdaBoost et Gradient Boosting, nous obtenons les scores moyens de  $R^2$ , RMSE, MAE et MAPE par médicament et par modèle, référés dans le tableau 5.4.

Comme vous pouvez le voir dans ce tableau, chaque algorithme est évalué sur un ensemble de données de 10 médicaments différents. Chaque colonne représente une mesure d'évaluation différente pour un algorithme spécifique. L'objectif de l'évaluation est de comparer les performances des quatre algorithmes sur les différents ensembles de données en utilisant les trois métriques d'évaluation afin de sélectionner le meilleur pour l'étape suivante. Dans le tableau, les meilleures performances sont représentées en gras.

Selon  $R^2$ , Random Forest surpasse les autres méthodes pour tous les médicaments, sauf pour les médicaments 5, 7 et 10. Pour le médicament 5, la meilleure méthode est Gradient Boosting, et pour le médicament 10, c'est AdaBoost. Cependant, la différence de scores avec le modèle RF n'est pas très significative. Les métriques RMSE, MAE et MAPE soutiennent également le choix de RF, car dans la plupart des cas, RF donne la meilleure prédiction. Cela nous amène à retenir RF parmi les quatre méthodes pour l'étape suivante de notre étude comparative. Les paramètres sont optimisés pour les expériences suivantes.

Med	Régression Li-néaire			Random Forest			AdaBoost			Gradient Boosting						
	R <sup>2</sup>	Rmse	Mae	Mape	R <sup>2</sup>	Rmse	Mae	Mape	R <sup>2</sup>	Rmse	Mae	Mape	R <sup>2</sup>	Rmse	Mae	Mape
<b>01</b>	0,78	56,66	40,21	155,82	<b>0,88</b>	<b>50,35</b>	<b>26,31</b>	<b>45,10</b>	0,83	57,77	27,96	101,24	0,87	57,16	27,55	98,98
<b>02</b>	0,25	462,15	388,4	45,96	<b>0,69</b>	<b>333,04</b>	<b>225,9</b>	<b>22,07</b>	0,57	384,68	253,9	44,3	0,64	348,27	233,4	42,49
<b>03</b>	0,86	300,97	225,2	56,95	<b>0,87</b>	<b>241,67</b>	<b>193,6</b>	<b>45,44</b>	0,87	261,06	199,9	53,32	0,82	264,12	202,8	52,6
<b>04</b>	0,76	284,75	177,7	92,95	<b>0,80</b>	<b>260,1</b>	<b>161,1</b>	<b>61,75</b>	0,70	283,9	180,2	114,58	0,66	308,16	170,5	64,39
<b>05</b>	0,5	156,21	118,4	123,24	0,77	129,02	<b>91,42</b>	<b>98,93</b>	0,74	134,68	93,86	105,06	<b>0,78</b>	<b>128,26</b>	93,12	105,13
<b>06</b>	0,68	133,16	106,1	50,44	<b>0,70</b>	<b>124,87</b>	<b>94,23</b>	<b>46,22</b>	0,60	145,77	103,3	49,16	0,67	135,02	102,6	48,75
<b>07</b>	0,58	42,02	30,58	42,71	<b>0,64</b>	37,6	<b>28,11</b>	40,05	0,61	38,14	28,15	42,71	0,65	<b>36,56</b>	28,38	<b>38,04</b>
<b>08</b>	0,83	61,42	41,67	49,97	<b>0,87</b>	<b>50,21</b>	<b>31,18</b>	28,74	0,84	51,88	34	35,43	0,83	52,52	33,01	<b>27,7</b>
<b>09</b>	0,40	29,04	23,37	80,09	<b>0,49</b>	<b>26,88</b>	<b>22,15</b>	57,37	0,44	28,869	22,48	66,06	0,42	28,856	22,73	<b>50,72</b>
<b>10</b>	0,23	10,92	8,57	42,42	0,76	10,65	8,04	61,42	<b>0,82</b>	<b>8,81</b>	<b>6,49</b>	<b>31,03</b>	0,77	10,59	8,32	46,14

TABLE 5.4 – Comparaison des performances des modèles suite à la validation croisée

### 5.3.2 Comparaison des modèles de Random Forest, du DNN et de la méthode de consommation

Dans cette section, nous comparons d'abord les performances des méthodes RF, DNN et de la méthode de consommation pour chaque médicament en utilisant l'ensemble de données test, qui correspond à 20% de l'ensemble de données provenant des cinq hôpitaux. Le Tableau 5.5 résume les résultats expérimentaux avec les métriques de précision sélectionnées :  $R^2$ , RMSE, MAE et MAPE. Pour les méthodes d'apprentissage automatique, les précisions pour les phases de test sont indiquées. Pour la méthode de consommation,  $R^2$  n'est pas pertinent car les différences entre les valeurs réelles et prédites sont énormes, ce qui conduit à des valeurs  $R^2$  négatives. Cependant, les métriques RMSE et MAE sont suffisantes pour une comparaison avec les modèles d'apprentissage automatique.

Med	Random Forest				DNN				Méthode de consommation			
	$R^2$	RMSE	MAE	MAPE	$R^2$	RMSE	MAE	MAPE	$R^2$	RMSE	MAE	MAPE
01	<b>0,97</b>	<b>23,15</b>	<b>14,25</b>	<b>18,62</b>	0,91	41,44	26,82	40,48	-0,51	88,71	67,15	346
02	<b>0,82</b>	<b>189,08</b>	<b>149,75</b>	<b>12,97</b>	0,61	277,47	233,78	22,22	0,45	366,27	291,46	8,8e+18
03	0,94	184,78	155,19	37,28	<b>0,95</b>	<b>157,83</b>	<b>125,43</b>	<b>28,40</b>	-0,09	227,1	172,23	37,99
04	<b>0,84</b>	<b>238,27</b>	<b>143,89</b>	<b>24,65</b>	0,75	257,47	191,98	41,41	-0,12	260,5	218,33	4,38e+18
05	<b>0,90</b>	<b>94,93</b>	<b>80,83</b>	<b>19,52</b>	0,75	146,41	115,65	33,84	-1,66	172,13	149,25	38,11
06	<b>0,82</b>	120,02	91,42	32,15	0,73	146,31	118,78	44,65	0,69	<b>68,32</b>	<b>58,88</b>	<b>16,49</b>
07	<b>0,71</b>	<b>24,20</b>	<b>17,96</b>	<b>92,27</b>	0,55	30,49	24,93	104,44	-0,18	44,26	35,62	1,09e+18
08	<b>0,96</b>	<b>31,57</b>	20,68	35,46	0,95	32,07	<b>19,84</b>	<b>34,66</b>	-0,168	68,15	59,09	4,05e+18
09	0,53	31,28	<b>20,12</b>	<b>39,31</b>	<b>0,54</b>	<b>31,02</b>	22,62	44,3	-0,23	21,79	18,33	1,18e+18
10	<b>0,60</b>	14,05	<b>7,81</b>	<b>43,90</b>	0,39	17,48	12,66	67,65	0,11	<b>13,63</b>	11,03	9,85e+17

TABLE 5.5 – Présentation des résultats par médicament et par modèle sur l'ensemble de données de validation

Nous pouvons tirer plusieurs déductions intéressantes du Tableau 5.5. En complément, le lecteur peut se référer aux graphiques dans l'annexe C qui illustrent les performances des modèles avec l'ensemble de données de test.

Premièrement, en regardant les scores des métriques, à la fois RF et DNN montrent des résultats nettement meilleurs par rapport à la méthode de consommation. On peut déjà déduire que les modèles d'apprentissage automatique sont meilleurs que la méthode de quantification classique pour les médicaments et les facteurs considérés.

Deuxièmement, en se concentrant sur les méthodes d'apprentissage automatique, nous observons que la RF surpasse le DNN pour tous les médicaments, à l'exception des médicaments n°3 et n°8, selon la métrique MAE et MAPE. Les métriques  $R^2$  et RMSE soutiennent également cette hypothèse, Random Forest est meilleur que DNN dans notre contexte. En effet, seulement deux des dix médicaments ont de meilleurs scores de  $R^2$  et RMSE pour la méthode DNN, mais la différence de scores n'est pas significative. Nous observons également une bonne qualité de prédiction pour RF. Seuls les médicaments 7, 9 et 10 montrent une précision inférieure à 75%, et le médicament numéro 1 a le score le plus élevé de 97% (métrique  $R^2$ ). De cela, nous pouvons déduire que Random Forest est la meilleure méthode des deux. Pour expliquer pourquoi certaines méthodes montrent une précision inférieure à

75%, plusieurs raisons peuvent être identifiées. Ces raisons incluent les algorithmes utilisés, les hyperparamètres qui doivent être optimisés et le caractère aléatoire du processus d'apprentissage automatique. Chaque algorithme d'apprentissage automatique a ses points forts et ses faiblesses, et les hyperparamètres doivent être correctement réglés pour optimiser les performances du modèle. Le caractère aléatoire du processus d'apprentissage automatique peut également entraîner des variations des performances du modèle. En outre, les fortes disparités dans la consommation de chaque médicament peuvent affecter la précision.

Dans les expériences précédentes, l'apprentissage et la validation ont été effectués sur les ensembles de données des cinq hôpitaux par médicament. Pour une gestion efficace des stocks, les modèles peuvent être utilisés pour la prédiction par hôpital et de chaque médicament. C'est l'occasion pour nous d'évaluer visuellement les performances des modèles par hôpital et par médicament. Pour illustrer, les figures 5.3, 5.4 et 5.5 montrent la valeur de la demande réelle et les prédictions pour le médicament n°3 (l'un de ceux avec  $R^2 > 90\%$  dans le Tableau 5.5), le médicament n°6 (avec  $R^2$  entre 75% et 90%) et le médicament n°9 (avec un score  $R^2 < 75\%$ ) pour l'hôpital n°1.

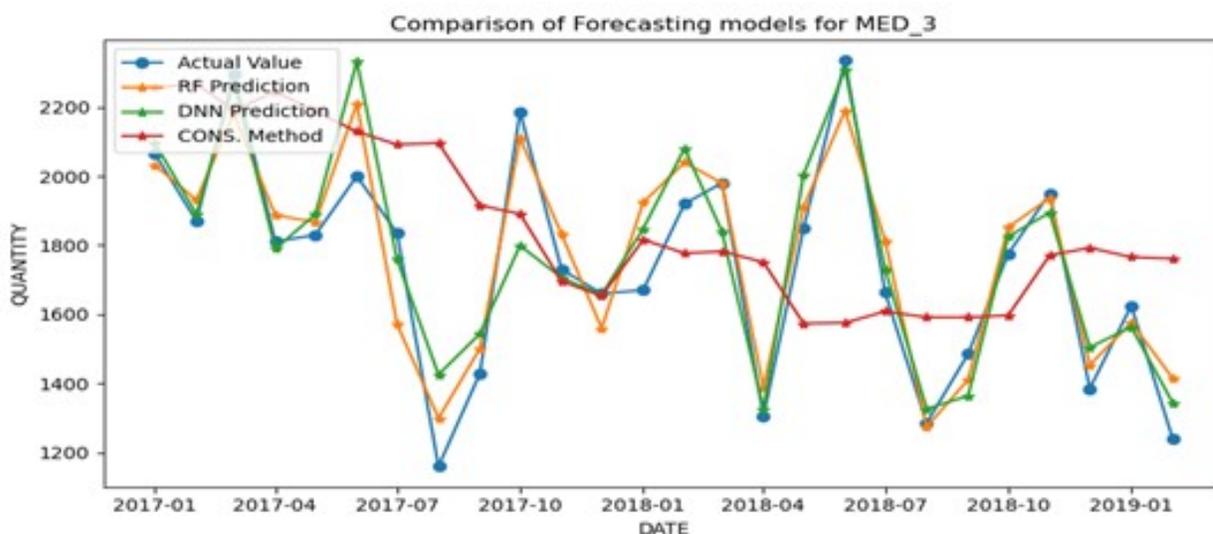


FIGURE 5.3 – Prédiction pour le médicament 3 à l'hôpital H3

Il apparaît également que les modèles Random Forest et DNN se rapprochent le plus de la consommation réelle, ce qui corrobore les résultats de la table 5.5 et les graphiques de validation dans l'annexe C.

Dans nos expériences, nous observons que la méthode de consommation est inefficace dans tous les cas pour estimer la consommation à relativement court terme (c'est-à-dire mensuelle). Nous observons également que la méthode Random Forest est un meilleur outil pour prédire la demande mensuelle de médicaments.

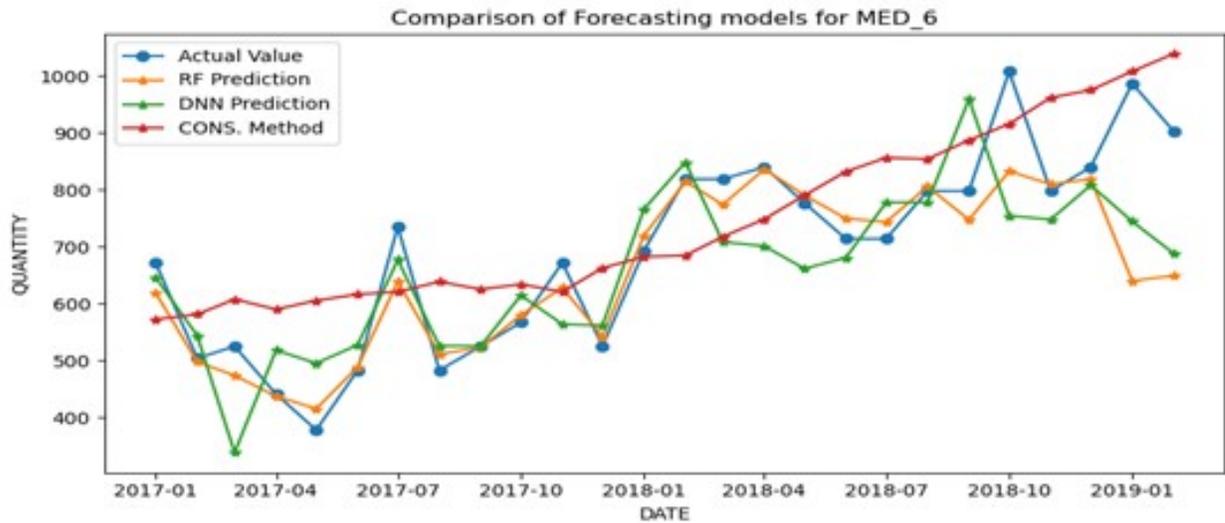


FIGURE 5.4 – Prédiction pour le médicament 6 à l'hôpital de H3

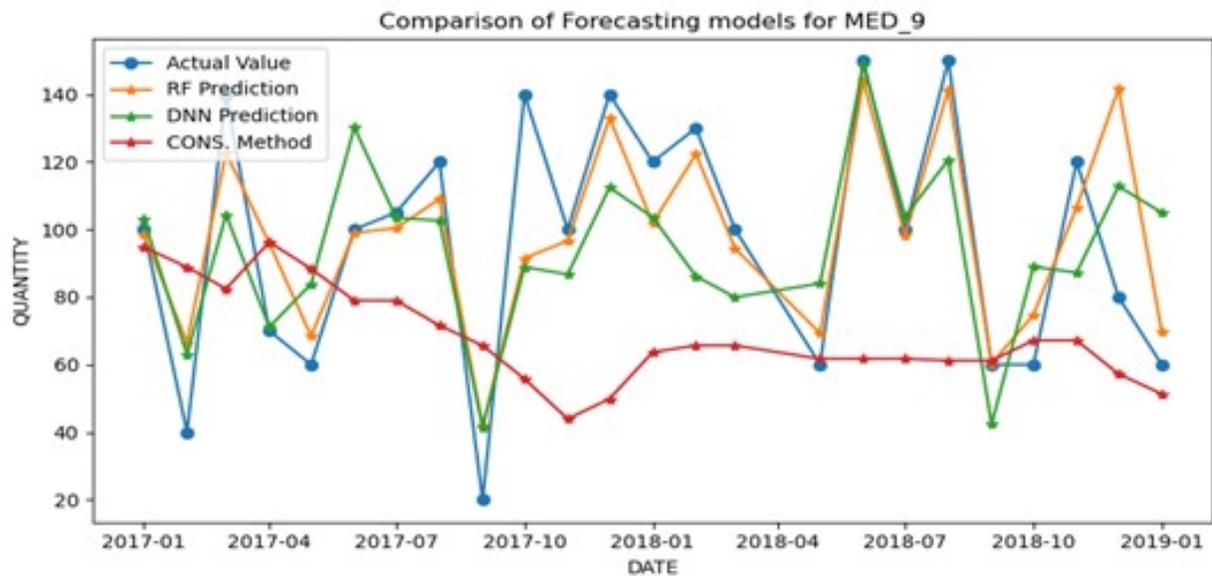


FIGURE 5.5 – Prédiction pour le médicament 9 à l'hôpital de H3

## 5.4 Conclusion et perspectives

Ce travail s'est concentré sur l'application de modèles d'apprentissage automatique pour estimer la demande mensuelle de médicaments par rapport à la méthode de consommation proposée par l'OMS en 1982 et améliorée par le groupe MSH en 2012. Des expériences ont été réalisées avec des données provenant de cinq hôpitaux français intégrant des facteurs ayant un impact sur la consommation.

Les résultats montrent que les modèles d'apprentissage automatique fournissent les meilleures prédictions, notamment Random Forest, qui atteint une précision de plus de 90 % selon  $R^2$  pour une sélection des médicaments les plus consommés parmi les 45 médicaments traçables

et courants dans les cinq hôpitaux. Cette recherche apporte une triple contribution. Tout d'abord, nous montrons que les méthodes d'apprentissage automatique en général peuvent être très utiles pour prévoir la consommation de médicaments et, par extension, faciliter une gestion efficace des stocks pharmaceutiques dans les hôpitaux.

Nous illustrons également que les outils d'intelligence artificielle représentent de meilleures alternatives à la méthode de consommation. La deuxième contribution est que nous avons intégré des facteurs ayant un impact sur la consommation de médicaments dans nos modèles d'apprentissage automatique, tels que le nombre de lits, le personnel, le nombre d'hospitalisation des patients, les populations où se trouvent les hôpitaux. Ces caractéristiques n'ont pas été prises en compte dans les travaux antérieurs dans ce domaine d'application. Par conséquent, contrairement aux méthodes classiques d'estimation qui ne reposent que sur les données historiques, ces modèles de prédiction sont réactifs aux évolutions futures en capturant tous les changements dans l'environnement. Enfin, nous notons que, en l'absence de très grandes quantités de données, les outils classiques d'apprentissage automatique, comme Random Forest par exemple, s'avèrent être de meilleures alternatives aux réseaux neuronaux.

Ce travail peut être étendu aux dispositifs médicaux ou aux médicaments rares (spécialisés) pour en évaluer l'efficacité et passer à une généralisation. Vu la performance des modèles d'apprentissage automatique pour la prévision de la demande de médicaments, avec des données combinées de plusieurs hôpitaux sans agrégation et une approche causale, ce travail ouvre la porte à l'exploration de prévisions relativement à l'objectif principal de thèse. Et il sera également intéressant d'élargir l'expérimentation sur un grand nombre de médicament. Le chapitre suivant aborde toutes ces questions.

# Chapitre 6

## Dimensionnement des besoins pour un nouvel hôpital

### Sommaire

---

<b>6.1</b>	<b>Introduction</b>	<b>124</b>
<b>6.2</b>	<b>Méthodes et Matériels</b>	<b>125</b>
6.2.1	Description, pré-traitement et séparation des données	125
6.2.2	Modèles d'apprentissage automatique de base et métriques de performance	128
6.2.3	Modèle hybride	128
<b>6.3</b>	<b>Résultats expérimentaux et analyses</b>	<b>130</b>
6.3.1	Conditions expérimentales	130
6.3.2	Évaluations de la performance des prévisions après l'entraînement	132
6.3.3	Prévision pour d'autres hôpitaux	138
<b>6.4</b>	<b>Conclusions</b>	<b>149</b>

---

### 6.1 Introduction

Dans ce chapitre, l'objectif est d'utiliser des techniques d'apprentissage automatique pour prédire la demande en médicaments d'un hôpital ne disposant pas d'historique de consommation ou d'un nouvel hôpital, en se basant sur des données réelles provenant d'autres hôpitaux. L'approche proposée consiste à utiliser les données d'hôpitaux précédemment établis (expériences de consommation) pour prédire la consommation de médicaments dans un autre hôpital ne disposant pas d'historiques de consommation. Cette idée se formalise à travers l'équation 1 que nous rappelons :

$$Y_i^{H_0} = f_i^{H_1, H_2, \dots, H_m}(X_1, X_2, \dots, X_n) \quad (1)$$

Ici,  $H_1, H_2, \dots, H_m$  représentent les  $m$  hôpitaux connus avec  $n$  caractéristiques  $X_1, X_2, \dots, X_n$  (les facteurs de consommation ici) et  $H_0$  est l'hôpital pour lequel nous dimensionnons le stock

pharmaceutique.  $Y_i^{H_0}$  représente l'estimation du besoin d'un médicament  $i$  pour l'hôpital  $H_0$ . La fonction  $f_i^{H_1, H_2, \dots, H_m}$  est déterminée à l'aide de techniques d'apprentissage automatique. Cette approche devrait permettre de dimensionner et de gérer les stocks de médicaments dans les nouveaux établissements, ou unités médicales. Comme dans le chapitre précédent, chapitre 5, la combinaison de données provenant de plusieurs hôpitaux et de divers facteurs de consommation nous amène à utiliser des méthodes causales. En plus d'évaluer les performances des méthodes classiques d'apprentissage automatique, comme le Random Forest qui s'est révélé efficace dans les conclusions du chapitre 5, nous utiliserons également des méthodes d'apprentissage automatique hybrides, des techniques populaires pour résoudre des problèmes complexes. Ces méthodes prennent en compte les facteurs tels que détaillés dans le chapitre 4.

Le reste de ce chapitre est organisé comme suit : la section 6.2 décrit la méthodologie et les matériels utilisés pour atteindre notre objectif. La section 6.3 présente les résultats expérimentaux de l'étude de cas. Ces résultats sont également analysés avant de conclure et d'identifier les limites et les perspectives dans la section 6.4.

## 6.2 Méthodes et Matériels

Dans cette section, nous présentons la méthodologie appliquée tout au long de cette étude (voir Figure 6.1). Elle se compose de trois parties principales :

- L'aperçu des données, leur pré-traitement et leur séparation.
- La construction et la validation du modèle.

Pour les deux parties (l'aperçu et pré-traitement des données, et la construction du modèle), nous utilisons le même ensemble de données que celui traité et analysé dans le précédent chapitre. En rappel, cet ensemble de données est dérivé du système WMS (Warehouse Management System) des 5 hôpitaux français, utilisés pour l'expérimentation dans le chapitre précédent, (appelés ici H1, H2, H3, H4 et H5) de 5 régions différentes et complété par des données externes.

### 6.2.1 Description, pré-traitement et séparation des données

Les données sont collectées de janvier 2016 à février 2019. Dans cette étude, nous nous étendons sur 45 médicaments en commun consommés dans les hôpitaux afin de pouvoir apprendre de ces hôpitaux et prédire la consommation de ces médicaments dans un nouvel établissement. Pour une bonne performance d'un modèle d'apprentissage automatique, il est toujours impératif de garantir la qualité des données, d'où le pré-traitement des données. La plupart du temps, les ensembles de données contiennent des valeurs manquantes et des valeurs extrêmes appelées valeurs aberrantes.

L'attribut nombre de séjours relative aux activités de soins de suite et de réadaptation (SEJ\_SSR) du jeu de données n'avait pas de données pour la période de 2016 à 2018 inclus pour les hôpitaux individuels. La variable dédiée "Mois de consommation" de type catégorique

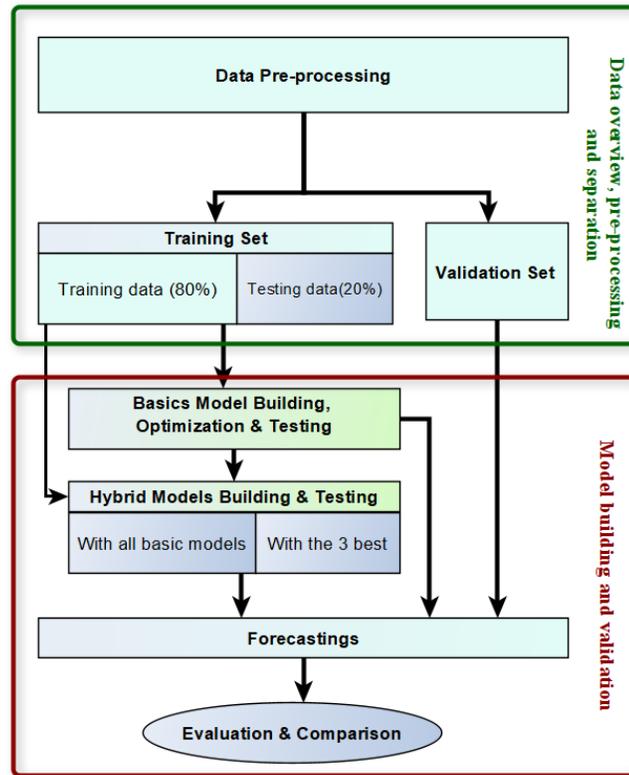


FIGURE 6.1 – Diagramme de flux de la méthodologie proposée.

est encodée avec le module *OneHotEncoder* de la bibliothèque *scikit-learn*. Cela nous conduit à 12 colonnes supplémentaires, chaque mois ayant pour valeur soit 1 soit 0. Ensuite, nous standardisons les données, ce qui est nécessaire pour la plage de variables indépendantes ou caractéristiques de notre jeu de données avant d'utiliser des algorithmes d'apprentissage automatique. Nous utilisons pour cela le module *MinMaxScaler* de la bibliothèque *scikit-learn*. Le module *MinMaxScaler* transforme les caractéristiques en les mettant à l'échelle à une valeur donnée de 0 à 1 dans notre cas.

Après la mise à l'échelle des données, vient la séparation en tenant compte du contexte de notre recherche. La table 6.1 présente la répartition des 4560 lignes de données (vecteur d'informations relatives à une consommation mensuelle) des cinq hôpitaux inclus dans cette étude. On remarque dans le tableau que la répartition des données est inégale : les hôpitaux H4 et H5 disposent de moins d'un an de données (12 mois). Étant donné que la phase d'apprentissage des modèles de machine learning nécessite suffisamment de données pour un ajustement optimal des paramètres, les données relatives aux 45 médicaments de ces cinq hôpitaux sont divisées en deux groupes principaux :

- Le premier groupe est composé des données de trois hôpitaux (H1, H2, H3) et est utilisé pour la construction et le test des modèles. Nous utilisons 80% de ces données pour l'entraînement du modèle et 20% pour le test. Nous nous assurons que les données sont bien mélangées avant la séparation.
- Dans le second groupe de données, nous avons les informations sur les deux autres

hôpitaux restants (H4, H5). Étant donné que les données de ces hôpitaux sont limitées, les inclure dans la construction des modèles pourrait biaiser l'apprentissage. Ainsi, les données de ces deux hôpitaux sont utilisées à des fins de validation. Le modèle de prédiction, construit à partir des données du premier groupe, est évalué en utilisant les données de validation provenant de H4 et H5. L'objectif est de trouver un modèle capable de prédire la consommation dans H4 et H5.

Med	H1	H2	H3	H4	H5	Total H1 à H3	Total
1	28	27	39	7	8	94	109
2	28	26	39	7	8	93	108
3	29	26	39	7	8	94	109
4	28	26	39	7	8	93	108
5	28	26	39	7	8	93	108
6	27	27	39	7	8	93	108
7	27	26	39	7	8	92	107
8	26	24	39	7	8	89	104
9	15	27	39	7	8	81	96
10	29	26	39	7	8	94	109
11	27	24	38	7	8	89	104
12	28	24	36	7	4	88	99
13	28	27	39	7	8	94	109
14	28	25	39	7	8	92	107
15	28	26	38	7	8	92	107
16	28	26	39	7	6	93	106
17	27	26	39	7	8	92	107
18	29	27	39	7	8	95	110
19	29	26	39	7	8	94	109
20	29	26	38	7	8	93	108
21	28	26	39	6	8	93	107
22	28	26	39	7	8	93	108
23	1	1	39	7	8	41	56
24	29	27	38	7	8	94	109
25	28	2	39	7	8	69	84
26	28	24	39	7	8	91	106
27	30	27	36	7	8	93	108
28	28	26	39	7	8	93	108
29	28	26	39	7	7	93	107
30	28	26	39	7	8	93	108
31	28	26	39	7	6	93	106
32	28	26	38	7	6	92	105
33	27	26	37	7	6	90	103
34	28	18	39	7	8	85	100
35	25	19	39	1	1	83	85
36	28	2	39	7	8	69	84
37	28	26	38	7	8	92	107
38	28	26	38	7	6	92	105
39	27	23	34	6	8	84	98
40	21	26	36	7	7	83	97
41	28	25	38	7	7	91	105
42	26	24	39	7	7	89	103
43	26	23	38	7	4	87	98
44	21	15	33	7	5	69	81
45	14	13	12	4	7	39	50

TABLE 6.1 – Répartition du nombre de lignes d'informations par médicament par hôpital et total

## 6.2.2 Modèles d'apprentissage automatique de base et métriques de performance

Dans cette étude, nous explorons différentes méthodes telles que Support Vector Regression (SVR), Multilayer Perceptron (MLP), Multiple Linear Regression (MLR), General Regression Neural Network (GRNN) et Random Forest (RF). Ces modèles ont été choisis en raison de leur succès dans les tâches de régression, une approche d'apprentissage supervisée, et ont été efficaces dans la prévision de la demande de médicaments pour aider à la gestion des stocks. Cela les rend aptes à une investigation et une expérimentation approfondies dans notre recherche. SVR est une méthode d'apprentissage supervisé pour la régression, la classification et la détection d'anomalies, gérant divers défis avec la sélection de la fonction de noyau et des paramètres de régularisation [41, 67]. MLP est un réseau neuronal artificiel à propagation avant avec apprentissage par rétro-propagation pour les tâches de régression [41, 67]. MLR établit une relation linéaire avec des variables explicatives pour estimer la variable cible [87]. GRNN est une variante de réseau neuronal probabiliste avec des fonctions d'activation gaussiennes [150, 1, 27]. Le Random Forest qui a montré son efficacité dans le chapitre précédent est également utilisé. En rappel il s'agit d'un modèle d'ensemble combinant des arbres de décision avec une attribution de nœuds aléatoire et une élagage pour améliorer la précision de la classification et de la régression [77, 121].

Pour évaluer l'efficacité des modèles de prévision, nous utilisons les métriques de régression comme celles utilisées dans le chapitre 5. Il s'agit notamment des métriques populaires :  $R^2$ , RMSE, MAE et MAPE. Ces métriques fournissent des mesures quantitatives de la performance des modèles, évaluant l'exactitude et l'ajustement des valeurs prédites par rapport aux valeurs réelles.

Les modèles basiques sélectionnés qui possèdent des caractéristiques distinctes sont combinés pour avoir d'autres modèles, des modèles hybrides.

## 6.2.3 Modèle hybride

En se basant sur les divers algorithmes de prévision sélectionnés, avec leurs avantages et leurs limites, il est possible de développer des modèles hybrides en combinant différents régresseurs d'apprentissage automatique afin d'améliorer les performances de modélisation. Cette approche gagne en popularité, comme le soulignent Ardabili et al. dans leur revue de littérature [7]. Pour illustrer cette idée, les méthodes hybrides peuvent être comparées à une entreprise où des employés aux compétences variées collaborent pour atteindre un objectif commun. De cette manière, les méthodes hybrides résultent de la combinaison de deux ou plusieurs méthodes d'apprentissage automatique et/ou d'algorithmes dans le but d'obtenir de meilleures performances et des résultats optimaux.

Les modèles hybrides peuvent être utilisés pour fusionner différents types de modèles ou de techniques, tels que la combinaison d'un réseau neuronal avec un arbre de décision, ou l'intégration de plusieurs réseaux neuronaux présentant des architectures différentes. L'objectif d'un modèle hybride est d'exploiter les points forts de différentes techniques et/ou algorithmes pour obtenir de meilleures performances que celles de chaque modèle pris individuellement.

Pour la prédiction de la consommation de médicaments dans les hôpitaux, où le comportement présente des variations et des incertitudes, ce type d'approche s'avère particulièrement adapté. L'idée fondamentale derrière l'utilisation de modèles hybrides dans ce contexte est que même si un régresseur particulier peut faire des prévisions moins précises, d'autres régresseurs, même s'ils sont moins performants, peuvent contribuer à améliorer les prédictions globales. Dans cette étude, nous avons également opté pour une approche hybride basée sur l'apprentissage automatique, en plus des modèles de base sélectionnés. Vous pouvez voir la structure de ce modèle hybride dans la figure 6.2.

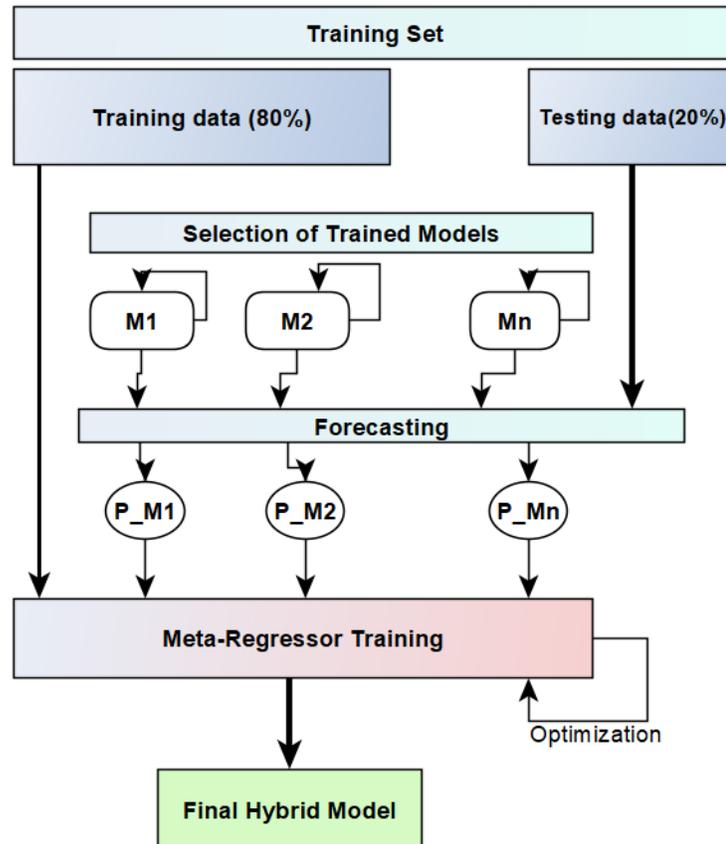


FIGURE 6.2 – Structure du modèle hybride proposé

Les modèles d'apprentissage automatique entraînés (M1 à Mn dans la figure 6.2) sont ajustés avec certains hyperparamètres, puis testés avec des données spécifiques. Les résultats obtenus (P\_ML1 à P\_MLn dans la figure 6.2) sont utilisés comme méta-variables, c'est-à-dire les données d'entrée, qui sont ensuite combinées au sein d'un méta-modèle (également appelé méta-régresseur) pour obtenir le modèle hybride définitif.

Dans ce processus, le méta-régresseur, qui est ici un Random Forest, est entraîné en ajustant ses hyperparamètres afin d'améliorer la qualité des prédictions et d'éviter le surajustement. Cette approche est inspirée de la technique de *Stacked Regression*, également connue sous le nom de "*Stacked Generalization*", qui a été proposée par Breiman en 1996 [23]. Elle consiste à combiner plusieurs modèles de base (ou modèles de "niveau 0") pour créer un méta-modèle plus puissant et robuste (ou modèle de "niveau 1"). Les modèles de base sont

généralement entraînés sur les mêmes données d'entrée en utilisant différentes méthodes ou configurations, et leurs prédictions sont ensuite combinées à l'aide du méta-modèle. Ce dernier est entraîné pour optimiser la combinaison des prédictions des modèles de base. L'objectif de la *Stacked Regression* est d'améliorer la précision des prédictions finales en réduisant les biais des modèles de base individuels et en tirant parti de leurs forces complémentaires. Pour plus d'informations, nous vous recommandons de consulter les travaux de [23, 36, 132, 164].

L'analyse de corrélation effectuée dans le chapitre 4 révèle que l'influence des facteurs identifiés varie en fonction du médicament. Cela implique qu'un modèle de prédiction distinct est nécessaire pour chaque médicament. Aussi, le premier modèle hybride comprendra l'ensemble des modèles de base d'apprentissage automatique que nous avons appliqués, tandis que le second inclura les trois meilleurs modèles d'apprentissage automatique pour chaque médicament. Les performances sont relative aux métriques RMSE, MAE et MAPE. Ensuite, nous analyserons l'ensemble de ces modèles et discuterons de leur efficacité pour un nouvel ou un autre hôpital ( $H_0$ ), comme indiqué précédemment dans l'équation (1).

## 6.3 Résultats expérimentaux et analyses

Dans cette section, deux groupes d'expériences numériques sont fournis, et leurs résultats sont analysés. Nous présentons tout d'abord les conditions expérimentales. Ensuite, nous présentons les performances des modèles d'apprentissage automatique décrits dans la section 6.2. Les performances des modèles sont d'abord évaluées sur les données test des trois hôpitaux choisis pour l'entraînement des ces derniers (H1, H2, H3). Les modèles obtenus, tels que décrits dans la démarche générale, sont ensuite utilisés pour prédire la consommation de deux autres hôpitaux (H4 et H5) dont les données n'ont pas fait partie de l'ensemble de données pour la construction des modèles ni lors des tests. Ainsi, leurs performances sont observées, analysées par médicament et par hôpital dans une étape finale. Les principaux modules utilisés pour les différentes applications sont Scikit-learn, PyGRNN, TensorFlow, Keras, mlxtend et scikeras. `StackingCVRegressor` est le module `mlxtend` que nous avons utilisé.

### 6.3.1 Conditions expérimentales

Chaque modèle sélectionné possède un ou plusieurs hyperparamètres qui doivent être ajustés lors de l'étape d'entraînement, à l'exception de MLR et GRNN. Cela est réalisé à l'aide de modules tels que `GridSearchCV` issue de Scikit-Learn. `GridSearchCV` est une méthode permettant d'ajuster les hyperparamètres d'un modèle par validation croisée. Elle essaie de manière exhaustive toutes les combinaisons d'hyperparamètres dans une grille spécifiée, évalue les performances du modèle pour chaque combinaison par validation croisée, puis retourne la combinaison d'hyperparamètres avec le meilleur score de performance. Compte tenu de la capacité limitée de notre ordinateur (décrite dans le chapitre précédent), le nombre d'hyperparamètres est réduit par modèle. C'est la même approche utilisée dans le chapitre 5 lors de la phase de sélection des modèles. Le tableau 6.2 ci-dessous présente les hyperparamètres considérés dans les différents modèles avec leurs réglages.

Modèles	Hyperparamètres
RF	n_estimators : Entier dans l'intervalle [2, 100] max_depth : Entier dans l'intervalle [4, 12]
SVR	C : {0.1, 1, 10, 100} gamma : {1, 0.1, 0.01, 0.001, 0.0001} kernel : {rbf, linear, poly}
MLP	Fonction d'activation : ReLU (Unité Linéaire Rectifiée) Optimiseur : Adam Taux d'apprentissage : {0.01, 0.001, 0.0001} Époques : 300 Fonction de perte : MSE Couches cachées : {(8, 4, 4), (16, 8, 8), (32, 16, 8), (30, 30, 30), (50, 50, 50), (128, 64, 32), (128, 64)}
Hybrid_1 & Hybrid_2	Meta-regresseur : Random Forest n_estimators : {10, 20, 30, 50, 100} max_depth : {5, 6, 7}

TABLE 6.2 – Hyperparamètres utilisés pour chaque modèle

Med	Hybrid_2	Med	Hybrid_2
1	RF-SVR-MLP	24	MLR-SVR-MLP
2	MLR-RF-SVR	25	RF-SVR-GRNN
3	MLR-RF-SVR	26	RF-SVR-MLP
4	MLR-RF-SVR	27	RF-SVR-MLP
5	MLR-RF-SVR	28	MLR-SVR-MLP
6	MLR-RF-SVR	29	MLR-RF-GRNN
7	SVR-MLP-GRNN	30	MLR-RF-SVR
8	MLR-SVR-MLP	31	MLR-SVR-MLP
9	MLR-RF-SVR	32	RF-SVR-MLP
10	RF-SVR-GRNN	33	MLR-RF-SVR-MLP
11	SVR-MLP-GRNN	34	MLR-RF-MLP
12	MLR-RF-SVR	35	SVR-MLP-GRNN
13	MLR-RF-MLP	36	SVR-MLP-GRNN
14	SVR-MLP-GRNN	37	MLR-RF-SVR
15	MLR-RF-SVR	38	MLR-RF-SVR
16	MLR-RF-SVR	39	MLR-RF-MLP
17	MLR-RF-SVR	40	MLR-RF-SVR
18	MLR-SVR-MLP	41	MLR-RF-SVR
19	MLR-RF-SVR	42	MLR-RF-GRNN
20	RF-SVR-MLP	43	RF-SVR-MLP-GRNN
21	MLR-RF-SVR-MLP	44	RF-MLP-GRNN
22	MLR-RF-MLP	45	RF-SVR-GRNN
23	MLR-SVR-GRNN		

TABLE 6.3 – Modèle hybride 2 par médicament

La première version de la méthode hybride (Hybrid\_1) combine toutes les méthodes de base décrites dans la section 6.3. La deuxième version (Hybrid\_2) est construite avec les trois

meilleurs modèles selon les classements des métriques RMSE, MAE et MAPE. Le choix de ces trois métriques est basé sur leur convergence (la convergence vers 0 signifie que le modèle est meilleur). Dans le tableau 6.3, nous présentons les trois meilleurs modèles utilisés dans Hybrid\_2 pour chaque médicament lors de sa construction.

Dans ce tableau, nous remarquons la présence de 4 méthodes sélectionnées au lieu de 3 méthodes pour certains modèles hybrides (Hybrid\_2). Cela est dû à une situation d'égalité lors de la sélection des modèles de base pour construire les modèles Hybrid\_2. StackingCVRRegressor est le module mlxtend que nous avons utilisé pour la construction des modèles hybrides. À première vue, un constat est que les modèles SVR et RF sont les plus couramment utilisés pour les modèles hybrides correspondants.

### 6.3.2 Évaluations de la performance des prévisions après l'entraînement

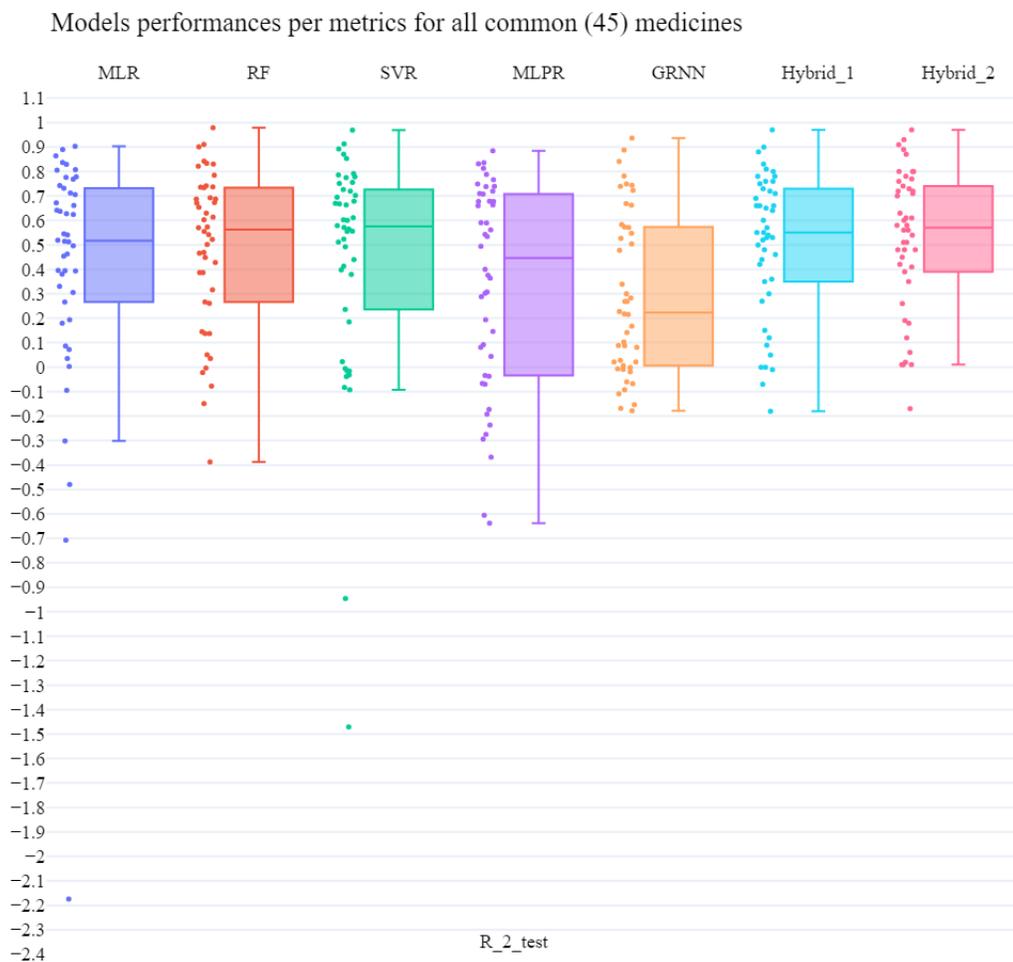
Après avoir construit et entraîné les différents modèles avec 80% des données des hôpitaux sélectionnés (H1, H2 et H3), nous les avons appliqués aux données de test, qui représentent les 20% restants. Nous avons évalué leurs performances en utilisant les métriques telles que  $R^2$ , RMSE, MAE et MAPE pour chaque médicament. Dans cette section, nous présentons les performances des modèles lors de la phase de test, utilisant les données du reste des trois hôpitaux qui ont servi à leur construction. Pour faciliter l'interprétation dans ce chapitre, nous utilisons les métriques  $R^2$  et MAPE.

Les figures 6.3 et 6.4 présentent respectivement en boîte à moustaches les performances de chaque modèle sur les 45 médicaments relativement à la métrique  $R^2$  et MAPE. Les points le long des boîtes à moustaches représentent les scores des métriques de chaque médicament individuel. Les tableaux de la section D.1 de l'annexe D fournissent les détails les scores des métriques des modèles de chaque médicament après leurs tests sur les 20% des données d'entraînement. Ces informations sont résumés dans les tables 6.4 et 6.5.

Modèle	Maximum des prédictions	Q3 des prédictions	Médiane des prédictions	Écart type	Score $\geq$ 70%	Prédiction négative
MLR	0.90	0.725	0.515	0.531	14	5
RF	<b>0.98</b>	0.720	0.560	0.319	12	4
SVR	0.97	0.727	<b>0.575</b>	0.469	14	9
MLP	0.88	0.703	0.445	0.423	12	12
GRNN	0.94	0.570	0.225	0.330	8	10
Hybrid_1	0.97	0.727	0.550	0.292	14	3
Hybrid_2	0.97	<b>0.737</b>	0.570	<b>0.288</b>	<b>17</b>	<b>1</b>

TABLE 6.4 – Résumé des performances des modèles selon la métrique  $R^2$  (avec les données tests)

Le tableau 6.4 résume les performances  $R^2$  pour les sept modèles. À partir de ces résultats, nous pouvons voir, par exemple, que pour le modèle MLR, il y a 14 médicaments dont la

FIGURE 6.3 – Performances des modèles des médicaments en commun (score  $R^2$ )

Modèle	Score minimum (meilleur score) en %	Médiane des scores	Nombre de Score $\leq 20\%$	$20\% < \text{Score}$ $\geq 50\%$	Score $> 50\%$
<b>MLR</b>	15,33	52,87	2	19	25
<b>RF</b>	14,83	<b>43,49</b>	2	24	20
<b>SVR</b>	17,21	46,54	1	24	21
<b>MLP</b>	18,83	56,05	2	17	27
<b>GRNN</b>	22,87	73,11	0	15	31
<b>Hybrid_1</b>	<b>12</b>	47	4	22	20
<b>Hybrid_2</b>	<b>12</b>	44	4	<b>26</b>	<b>19</b>

TABLE 6.5 – Résumé des performances des modèles selon la métrique MAPE (avec les données tests)

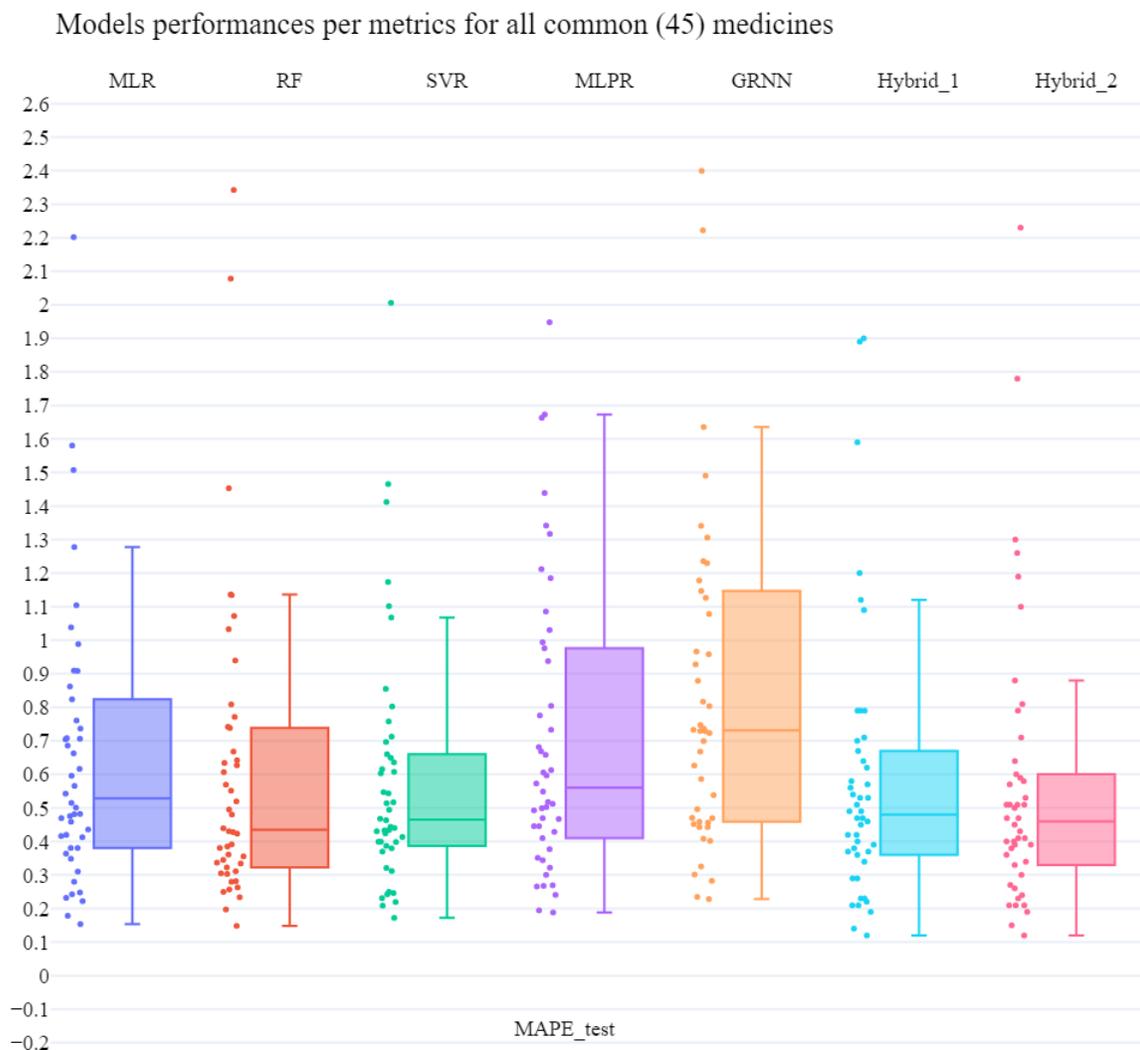


FIGURE 6.4 – Performances des modèles des médicaments en commun (score MAPE)

prédiction est supérieure à 70 % (colonne 6 dans le tableau 6.4) et 5 médicaments avec des scores négatifs (colonne 7). Un score négatif de  $R^2$  signifie que le modèle MLR n'est pas adapté aux données de ces médicaments, c'est-à-dire qu'il ne parvient pas à expliquer la variance dans les données. De plus, à partir des informations du graphique en boîte présenté à la figure 6.3 et de la table 6.4, nous pouvons voir que pour le modèle MLR, la prévision maximale est de 90,26 %, le troisième quartile de prédiction est de 72,5 % et la médiane est de 51,15 %. Une interprétation similaire est donnée pour les autres modèles. Parmi ces sept méthodes, c'est Hybrid\_2 qui présente de meilleur performance avec 17 médicaments prédit avec un score supérieur à 70%. L'écart type des scores de prévision de ce modèle sur les 45 médicaments est le plus faible, soit 0,288.

Selon le MAPE, dont le tableau 6.5 résume également les performances des sept modèles,

on peut constater que le modèle Hybrid\_2 présente les meilleures performances. Ce modèle a pour 4 médicaments pu faire une prévision avec un score inférieur à 20%. 26 médicaments ont également pu avoir au moins un modèle dont le score MAPE est compris entre 20% et 50%. La note négative pour ce modèle est que pour 19 médicaments ont eu des précisions au delà de 50%. La meilleure précision obtenue par ce modèle est de 12%.

### 6.3.2.1 Analyse des performances selon les médicaments

À partir des résultats obtenus, selon le  $R^2$ , nous pouvons regrouper les performances des modèles par médicament en six groupes (comme indiqué dans le tableau 6.6). Dans chaque groupe, sont indiqués les médicaments pour lesquels un modèle parmi les sept testés a obtenu un score dans les plages décrites. Le premier groupe, par exemple, est constitué des médicaments dont la prévision a un score supérieur à 90 %. Dans ce groupe, il y a les médicaments 2, 13 et 25. À l'autre extrémité, le groupe 5 est constitué des médicaments 6, 11, 12, 14, 16, 18, 20, 22, 23, 29, 32, 44 et 45 dont leurs meilleurs modèles n'ont pas excédé 50% selon  $R^2$ .

Groupe	Statut	Intervalle de score	Médicaments dans le groupe
1	Très bon	$\geq 90\%$	2, 13, 25
2	Bon	$[80\%; 90\% [$	1, 7, 10, 30, 31, 41
3	Assez bon	$[70\%; 80\% [$	4, 8, 9, 24, 26, 27, 28, 36, 42
4	Moyen	$[50\%; 70\% [$	3, 5, 15, 17, 19, 21, 33, 34, 35, 37, 38, 39, 40, 43
5	Mauvais	$[0\%; 50\%$	6, 11, 12, 14, 16, 18, 20, 22, 23, 29, 32, 44, 45

TABLE 6.6 – Résumé de la performance des modèles selon la métrique  $R^2$

Il n'y a pas de seuil  $R^2$  universellement accepté qui serait applicable dans tous les cas pour la prévision des besoins en médicament. Le choix du score  $R^2$  acceptable dépend de divers facteurs, y compris la précision requise pour la prévision, les conséquences des erreurs de prévision, la taille de données, etc. Dans le contexte de la prévision des besoins en médicament, un  $R^2$  plus élevé indique généralement un meilleur ajustement du modèle aux données historiques. L'entreprise KLS souhaite une précision supérieure à 95 % pour cette métrique, mais seul le médicament 25 répond à ce critère lors de la phase test. Pour un « *proof of concept* », nous considérons dans notre cas qu'un score  $R^2$  de 70% est acceptable, comme c'est courant dans la littérature [94, 166]. Dans le domaine médical, il est important de noter que le score  $R^2$  ne doit pas être utilisé comme la seule métrique pour évaluer la performance d'un modèle d'apprentissage automatique.

Ainsi, à partir de cet ensemble de résultats, nous pouvons tirer que, pour 18 médicaments sur 45, nous avons pu trouver au moins un modèle parmi les sept qui présente une précision de prédiction acceptable (groupes 1, 2 et 3) selon la métrique  $R^2$ .

La métrique MAPE, qui est la deuxième métrique facilement interprétable, confirme cette conclusion. En effet, les meilleurs modèles de médicaments des trois premiers groupes (groupes

selon  $R^2$ ) affichent des scores MAPE inférieurs à 50% à l'exception de trois médicaments : les médicaments 23, 24 et 27, comme c'est stipulé dans la table 6.7.

Groupe	Statut	Intervalle score MAPE	Médicaments dans le groupe
a	Très Bon	< 10	-
b	Bonne	]10%;20%]	1, 8, 13, 23, 28, 41
c	Raisonnable	]20%;50%]	2, 3, 4, 5, 7, 9, 10, 14, 15, 17, 18, 19, 21, 22, 25, 26, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40, 42, 43, 44
d	Mauvais	> 50%	6, 11, 12, 16, 20, 24, 27, 29, 37, 45

TABLE 6.7 – Répartition des groupes en fonction du score MAPE

Dans la littérature, de manière générale, un score MAPE inférieur à 10 % est excellent, entre 10 % et 20 % est toujours bon, tandis qu'une valeur supérieure à 50 % signifie que le modèle est inexact, car il se trompe plus souvent qu'il ne prédit correctement, comme le souligne respectivement José Montaña Moreno et al. [96] et [141], d'où l'établissement des quatre groupes selon MAPE du tableau 6.7. Cependant, tout dépend du type et de la quantité de données. Il est important de rappeler que le MAPE est sensible aux valeurs aberrantes (outliers), et le fait de combiner des consommations très différentes provenant de trois hôpitaux différents dans le jeu de données de test peut entraîner la présence de consommations dépassant largement ou étant bien en delà des autres. Par conséquent, nous faisons l'hypothèse que le choix d'un score MAPE de 50% comme seuil de prévision est acceptable, même si l'équivalent du score MAPE attendu par l'entreprise KLS devrait être inférieur à 10 %. Ainsi, nous pouvons tirer que les médicaments des trois groupes (a, b et c) ont une prédiction acceptable du point de vue de MAPE.

Les médicaments 24 et 27, lors de la phase de test, avaient des valeurs aberrantes dans leurs jeux de données test, ce qui ont influencé leurs scores MAPE. Cette hypothèse sera vérifiée dans la phase de validation, étant donné que ces derniers ont au moins un modèle performant selon  $R^2$ . Les autres médicaments du "groupe d", le dernier groupe selon MAPE, n'ont pas pu obtenir au moins un modèle de prévision acceptable selon cette métrique. Ce constat est également valable pour le seuil fixé pour le  $R^2$ .

Pour le médicament 23, au regard de la table 6.1, il est clair que les modèles de ce médicament se sont uniquement construits sur les données de l'hôpital H3. Ce qui veut dire que les modèles de ce médicament n'ont pas eu l'occasion de bien identifier les caractéristiques des hôpitaux lors de la phase de construction. Ces 20% de données tests sont constitué principalement de données d'autres hôpitaux. Ce qui a conduit à un mauvais score  $R^2$  et un score biaisé MAPE. Ainsi, ce médicament ne peut être concerné par la phase de validation. Nous retenons les 18 médicaments (y compris 24 et 27) qui ont obtenu au moins un modèle ayant un score  $R^2$  supérieur à 70% et un score MAPE inférieur à 50%. Autrement dit, les trois premiers groupes de la table 6.6 sont retenus pour la suite des analyses.

La capacité de prédiction n'est pas satisfaisante pour les vingt-sept médicaments restants

du point de vue de  $R^2$ , même si certains d'entre eux présentent des prévisions acceptables selon MAPE, comme c'est le cas pour les médicaments 3, 5, 14, 17, 18, etc.

### 6.3.2.2 Analyse des performances des modèles

Les tables 6.8 et 6.9 résument respectivement les performances des modèles par médicament selon les métriques  $R^2$  et MAPE. Dans ces tableaux figurent les modèles qui ont performé relativement aux 18 médicaments des trois groupes (groupe 1, 2 et 3).

Dans la table 6.8, on peut déduire que les modèles hybrides, en particulier la deuxième version, présentent des performances intéressantes selon les intervalles de score  $R^2$  retenus. Cette observation est similaire dans le cas de la métrique MAPE, à l'exception des médicaments 4, 9, 24, 27 et 42.

Cependant, pour certains médicaments, les modèles individuels affichent des précisions du même ordre ou meilleures que celles des modèles hybrides. Le Random Forest qui a été performant dans le chapitre précédent est également intéressant, comme modèle individuel et comme combinaison dans les modèles hybrides (voir table 6.3).

Une recommandation émerge, celle de privilégier le modèle hybride, en particulier Hybrid\_2, pour les prévisions des besoins lorsque l'on prédit dans les hôpitaux dont les données ont servi à construire et à tester le modèle. Cependant, il est important de noter que, dans les cas où les autres modèles ont performé mieux, les modèles hybrides, qui combinent des modèles de base par la technique de "*stacked regression*", sont affectés par les modèles de base moins performants, ce qui les rend moins performants que le meilleur modèle parmi la sélection. Cela a été observé dans la plupart des cas où les modèles hybrides ne sont pas les meilleurs (Voir les tables détaillées dans la section D.1 de l'annexe D).

### 6.3.2.3 Résumé de cette première analyse

Nous pouvons conclure que nos modèles ont réussi à obtenir des prédictions acceptables (avec un  $R^2$  supérieur à 70% et un MAPE inférieur à 50%) pour 16 des 45 médicaments sélectionnés. Nous avons observé que certains modèles sont plus adaptés à certains médicaments qu'à d'autres. De plus, le modèle Hybrid\_2, qui combine les trois meilleurs modèles, a montré des résultats intéressants.

L'approche présentée jusqu'à cette phase d'analyse peut être utilisée pour une prévision des besoins dans l'optique d'une gestion des stocks, mais pas pour le dimensionnement dans d'autres structures. Cependant, notre objectif dans ce travail est de prédire la consommation de médicaments pour d'autres hôpitaux ou de nouveaux établissements de santé. Même si certains de ces modèles ont montré de bonnes performances lors de cette phase de test, ils doivent être évalués et validés lors de la phase suivante. Cela signifie que nous devons les tester sur de nouvelles données et dans d'autres hôpitaux qui n'ont pas été utilisés pour la construction et l'entraînement des modèles (les détails sont expliqués dans la section ci-dessous).

Med	$\geq 90\%$	Entre 90% et 80%	Entre 80% et 70%
1	-	SVR, Hybrid_1, Hybrid_2	MLP
2	MLR, RF, SVR, Hybrid_1, Hybrid_2	MLP, GRNN	-
4	-	MLR	SVR, Hybrid_1, Hybrid_2
7	-	MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2	-
8	-	-	SVR, Hybrid_1
9	-	-	MLR, RF, Hybrid_2
10	-	RF, Hybrid_2	MLR, SVR, MLP, GRNN, Hybrid_1
13	RF, Hybrid_2	MLR, MLP, Hybrid_1	SVR, GRNN
24	-	-	MLR, MLP
25	RF, SVR, GRNN, Hybrid_1, Hybrid_2	MLR, MLP	-
26	-	-	SVR, MLP, Hybrid_1, Hybrid_2
27	-	-	RF, Hybrid_2
28	-	-	MLR, MLP, Hybrid_2
30	-	MLR, RF, SVR	MLP, Hybrid_2
31	-	MLR, Hybrid_2	RF, SVR, MLP, Hybrid_1
36	-	-	MLR, SVR, GRNN, Hybrid_1, Hybrid_2
41	-	RF, Hybrid_1, Hybrid_2	MLR, SVR, MLP, GRNN
42	-	-	MLR, RF, SVR, GRNN, Hybrid_1, Hybrid_2

TABLE 6.8 – Répartition des médicaments en fonction des catégories de score  $R^2$ 

### 6.3.3 Prévision pour d'autres hôpitaux

Comme mentionné précédemment, nos modèles ont été construits à l'aide de données provenant de trois hôpitaux (H1, H2, H3). Dans cette section, nous allons les mettre à l'épreuve en les utilisant pour prédire la consommation de médicaments dans deux autres hôpitaux, H4 et H5. Notre démarche consiste à évaluer les performances de chacun des sept modèles en les appliquant aux 18 médicaments, qui sont subdivisés en trois groupes.

Les prédictions générées par nos modèles seront comparées aux données réelles provenant de H4 et H5 afin d'évaluer leur exactitude. Étant donné que les données de H4 et H5 n'ont pas été utilisées dans les phases précédentes d'apprentissage et de test des modèles, cette étape de validation nous permettra de déterminer si nos modèles sont en mesure de faire de bonnes prévisions lorsqu'ils sont appliqués à d'autres hôpitaux.

Les tableaux 6.10 et 6.11 résument les performances des modèles dans les hôpitaux H4 et H5 pour chacun des 18 médicaments. Ils présentent principalement les scores des métriques  $R^2$  et MAPE, des métriques qui facilitent les analyses et interprétations.

De ces tableaux, on déduit rapidement que les modèles de MLR, les modèles de régression linéaire, ont été très mauvais pour chaque médicament dans chaque hôpital. La plupart des

Med	$\leq$ 10 %	Entre 10% et 20%	Entre 20% et 30%	Entre 30% et 40%	Entre 40% et 50%
1	-	MLP, Hybrid_1, Hybrid_2	-	-	SVR
2	-	-	RF, Hybrid_1, Hybrid_2	-	SVR
4	-	-	-	-	MLR, SVR
7	-	-	Hybrid_2	RF, Hybrid_1	SVR, MLP
8	-	MLR, RF, SVR, Hybrid_1, Hy- brid_2	MLP, GRNN	-	-
9	-	-	-	-	MLR
10	-	-	-	RF	SVR, Hybrid_1, Hy- brid_2
13	-	RF, MLP, Hy- brid_1, Hybrid_2	MLR, SVR, GRNN	-	-
24	-	-	-	-	-
25	-	-	-	SVR	RF, Hybrid_2
26	-	-	-	MLP, Hybrid_1, Hy- brid_2	MLR, RF, SVR, GRNN
27	-	-	-	-	-
28	-	MLR, Hybrid_2	RF, SVR, MLP, Hybrid_1	-	-
30	-	-	-	RF, Hybrid_2	MLR, SVR, Hybrid_1
31	-	-	Hybrid_1, Hy- brid_2	MLR, RF, SVR, MLP	GRNN
36	-	-	-	-	SVR, Hybrid_1
41	-	-	RF, Hybrid_1, Hybrid_2	-	MLR
42	-	-	-	-	RF, MLP, GRNN

TABLE 6.9 – Répartition des médicaments en fonction des catégories de score MAPE

scores  $R^2$  sont négatifs. Pour les 45 médicaments, nous vous renvoyons aux tableaux de la section D.2 de l'annexe D.

Med	MLR		RF		SVR		MLP		GRNN		Hybrid_1		Hybrid_2	
	$R^2$	MAPE	$R^2$	MAPE	$R^2$	MAPE	$R^2$	MAPE	$R^2$	MAPE	$R^2$	MAPE	$R^2$	MAPE
1					<b>0.77</b>	<b>9.37</b>	-1.37	<b>25.15</b>			-1.4	<b>24.67</b>	<b>0.68</b>	<b>9.74</b>
2			0.19	<b>24.58</b>	-27.42	254.97	-60.46	378.13			<b>0.89</b>	<b>17.43</b>	<b>0.71</b>	<b>28.15</b>
4	-206.76	191.28			-18.03	54.98					-15.3	50.4	-19.01	56.13
7	-3.05	<b>35.67</b>	-3.02	<b>34.84</b>	-2.07	<b>30.92</b>	-8.18	65.66	-0.77	<b>27.11</b>	<b>0.73</b>	<b>10.03</b>	<b>0.71</b>	<b>11.56</b>
8	-2.7	<b>22.57</b>	-2.3	<b>19.58</b>	-3.91	<b>21.94</b>	-2.9	<b>22.35</b>	-4.17	<b>26.1</b>	0.15	<b>11.96</b>	-1.29	<b>17.98</b>
9	-5.93	<b>45.39</b>	-0.94	<b>17.81</b>									-0.74	<b>24.29</b>
10	-3074.07	2895.98	<b>0.87</b>	<b>25.9</b>	-6.5	135.62	-6.7	151.86	-0.49	69.75	0.35	<b>45.15</b>	0.38	<b>48.63</b>
13	-5310.2	1976.11	-104.31	279.59	-64.82	221.93	-48.26	192.51	-59.33	213.84	-66.35	214.91	-67.24	224.69
24	-72.67	234.45					-4.09	50.45						
25	-9769.27	1485.74	-11.49	<b>47.98</b>	-4.13	<b>28.57</b>	-20.66	70.36	0.22	<b>12.3</b>	-5.46	<b>32.23</b>	-9.93	<b>43.04</b>
26	-4485.81	481.78	-15.84	<b>24.38</b>	-1.62	<b>10.1</b>	-4.97	<b>15.34</b>	-1.95	<b>11.09</b>	-15.83	<b>28.08</b>	-11.64	<b>24.26</b>
27			<b>0.61</b>	118.59									0.37	181.19
28	-178.08	293.44	-21.1	95.8	-32.06	123.64	-8.52	62.94			-5.37	<b>48.26</b>	-5.52	<b>49.1</b>
30	-1612.19	3636.14	-36.04	565.59	-83.81	817.08	-58.47	691.18			-68.17	780.84	-98.6	928.98
31	-2.23	61.82	-0.59	<b>31.69</b>	-8.67	102.09	-2.05	61.78	-1.89	59.92	0.49	<b>20.25</b>	<b>0.74</b>	<b>13.49</b>
36	-2.54	<b>42.88</b>			-2.52	<b>46.18</b>			-0.83	<b>33.62</b>	-0.69	<b>29.64</b>	0.33	<b>19.0</b>
41	-17223.65	5872.02	-419.42	886.51	-469.72	1003.49	-268.53	736.91	-425.54	949.24	-1096.97	1516.24	-485.61	1009.66
42	-45.44	458.63	-2.41	79.29	-11.67	210.92	-2.89	74.04	-2.72	72.9	-2.95	82.03	-3.04	84.37

TABLE 6.10 – Performances des modèles sur les 18 médicaments dans l'hôpital H4 selon la métrique  $R^2$  & MAPE

Med	MLR		RF		SVR		MLP		GRNN		Hybrid_1		Hybrid_2	
	$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	$R^2$	Mape
1					-18.77	76.86	-2.47	<b>31.41</b>			<b>0.83</b>	<b>11.37</b>	<b>0.85</b>	<b>10.32</b>
2			-11.53	386.15	-142.43	785.66	-446.63	1608.48			-12.69	353.22	-26.32	534.9
4	-1.72e+04	3.38e+03			-4.49	60.96					0.07	<b>21.35</b>	-0.48	<b>23.58</b>
7	-6.19e+02	6.40e+02	-3.85	<b>44.85</b>	-284.44	431.87	-16.81	99.99	-0.8	<b>22.04</b>	-3.33	<b>39.2</b>	0.01	<b>18.97</b>
8	-2.40e+01	1.44e+02	0.33	<b>17.64</b>	-16.93	109.84	0.13	<b>18.9</b>	0.16	<b>22.22</b>	<b>0.68</b>	<b>15.7</b>	0.08	<b>26.14</b>
9	-4.92e+03	2.23e+03	-20.14	141.68									-36.29	193.99
10	-3.25e+04	8.67e+03	-4.7	131.02	-84.2	465.34	0.41	<b>48.83</b>	-0.33	<b>46.45</b>	-3.96	122.28	-0.35	62.9
13	-4.82e+05	2.48e+04	-134.95	414.96	-296.93	631	-160.81	462.72	-252.57	579.29	-110.44	386.39	-81.2	333.23
24	-2.63e+03	1.32e+03					-3.94	<b>45.44</b>						
25	-2.01e+04	3.40e+03	-4.03	<b>41.31</b>	-6.55	54.19	-1.02	<b>29.47</b>	-3.57	<b>40.86</b>	-6.51	53.78	-7.8	59.45
26	-7.69e+03	2.55e+03	-0.5	<b>33.69</b>	-0.03	<b>26.29</b>	-0.18	<b>25.25</b>	0.02	<b>26.73</b>	-0.43	<b>29.56</b>	-0.51	<b>28.04</b>
27			-2.14	<b>41.89</b>									-2.29	<b>49.93</b>
28	-2.68e+03	1.46e+03	-0.92	<b>30.5</b>	-383.23	545.36	-12.06	90.55			-7.65	70.52	-8.68	76.16
30	-5.84e+03	7.91e+03	-4.37	250.23	-206.51	1519.44	-247.15	1649.12			-13.52	441.85	-18.65	507.11
31	-2.70e+05	3.27e+04	-1181.34	2180.92	-	6840.28	-	3614.97	-	2995.48	-809.58	1832.37	-738.81	1777.28
					11698.74		3159.73		2171.22					
36	-7.64e+03	2.56e+03			-0.94	<b>37.09</b>			-0.49	<b>31.08</b>	-0.76	<b>23.87</b>	-1.68	<b>31.09</b>
41	-4.63e+05	2.72e+04	-77.65	362.7	-295.02	699.52	-145.32	489.5	-179.2	547.14	-563.29	954.48	-434.97	846.54
42	-1.46e+05	1.56e+04	0.26	<b>26.2</b>	-	8954.45	-1.53	66.7	-2.86	75.65	0.57	<b>29.39</b>	0.42	<b>29.95</b>
					48243.91									

TABLE 6.11 – Performances des modèles sur les 18 médicaments dans l'hôpital H5 selon la métrique  $R^2$  & MAPE

### 6.3.3.1 Analyse des performance par médicament

Des tableaux 6.10 et 6.11, nous faisons les constats suivants :

- Dans l'hôpital H4 :
  - Dans le groupe 1, constitué des médicaments 2, 13 et 25, le médicament 2 a obtenu de bonnes prédictions selon les deux métriques avec les modèles hybrides (avec 89% de  $R^2$  pour Hybrid\_1). Le médicament 13 n'a obtenu aucun modèle acceptable selon les deux métriques  $R^2$  et MAPE. Pour le médicament 25, seuls les modèles MLR et MLP ont des scores MAPE supérieurs à 50% (mauvaise prévision). Son meilleur modèle est le GRNN selon MAPE.
  - Dans le groupe 2, les médicaments 1, 7, 10 et 31 ont obtenu des modèles avec de prévisions acceptables avec souvent de  $R^2 \geq 70\%$  ou  $\text{MAPE} \leq 50\%$ . Comme c'est le cas pour les modèles RF et SVR pour le médicament 1. Pour cet hôpital, les médicaments 30 et 41 n'ont pas de modèles de prévision acceptable du point de vu de ces deux métriques.
  - Seuls cinq médicaments sur neuf du groupe 3 ont pu obtenir un modèle approximant les quantités de consommation. Il s'agit notamment des médicaments 8, 9, 27, 28 et 36. Pour ces derniers, les meilleurs modèles ont des scores MAPE compris entre 11,96% et 49,74%. Mais le meilleur des modèles n'excède pas 68% selon le  $R^2$ .

Seuls, 11 sur les 18 médicaments ont obtenu des modèles pour une prévision acceptable selon au moins une des deux métriques  $R^2$  ou MAPE. Si nous tenons rigueur pour un MAPE et un  $R^2$  respectant simultanément les seuils fixés pour modèle d'être acceptable, dans cet hôpital, seuls cinq médicaments sont retenus. Comme nous le récapitulons dans le tableau 6.12, il s'agit du médicament 1 avec le modèle SVR, des médicaments 2 et 7 avec les deux modèles hybrides, du médicament 10 avec le Random Forest et le médicament 31 avec l'hybride version 2.

- Dans l'hôpital H5 :
  - Contrairement à ceux de l'hôpital H4, seul le médicament 25 de l'hôpital H5 a obtenu une prévision acceptable avec le modèle MLP, avec un score MAPE de 29,47 %.
  - Dans le groupe 2, les médicaments 1, 7 et 10 ont pu obtenir au moins un modèle pour une prévision acceptable, principalement selon la métrique MAPE. Pour le médicament 1, les meilleurs modèles sont les modèles hybrides. Hybrid\_2 à un score de 85% pour  $R^2$  et 10,32% pour MAPE. Le modèle Hybrid\_2 est également le meilleur modèle du point de vu du score MAPE avec une valeur de 18.97%.
  - Dans le groupe 3, six médicaments ont pu avoir au moins un modèle pouvant approcher au mieux les valeurs réelles dans cet hôpital. Il s'agit des médicaments 4, 8, 26, 27, 28, 36 et 42. Pour ces médicaments, les meilleurs modèles ont des scores MAPE compris entre 15,7% et 29,47%. Et tout comme pour l'hôpital H4 pour ce groupe, le meilleur score selon  $R^2$  des modèles n'excède pas 68%.

Sur l'ensemble des médicaments des trois premiers groupes, pour l'hôpital H5, 12 sur l'effectif de 18 ont obtenu une prévision acceptable selon la métrique MAPE. Seul le médicament 1 respect le critère de performance simultané ( $R^2 \geq 70\%$  et  $MAPE \leq 50\%$ ) avec les deux modèles hybrides.

De ces résultats, on peut tirer que pour ces deux hôpitaux différents, plus de la moitié des médicaments (soient 12 sur 18) ont obtenu au moins un modèle pour une prévision acceptable, pour un MAPE inférieur ou égal à 50% ou un  $R^2$  supérieur ou égal à 70%. Cependant, seuls 5 médicaments ont obtenu au moins un modèle pour une prévision respectant à la fois des deux seuils de score, contre un seul médicament au niveau de l'hôpital H4.

Groupe	Med	Modèles Test	Validation H4		Validation H5		Modèle retenue		
			$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	$R^2$ & Mape
1	2	RF	-	✓	-	-	-	-	-
		MLP	-	-	-	-	-	-	-
		SVR	-	-	-	-	-	-	-
		Hybrid_1	✓	✓	-	-	-	-	-
		Hybrid_2	✓	✓	-	-	-	-	-
	13	MLR	-	-	-	-	-	-	-
		RF	-	-	-	-	-	-	-
		SVR	-	-	-	-	-	-	-
		MLP	-	-	-	-	-	-	-
		GRNN	-	-	-	-	-	-	-
		Hybrid_1	-	-	-	-	-	-	-
	25	MLR	-	-	-	-	-	-	-
		RF	-	✓	-	✓		RF	-
		SVR	-	✓	-	-	-	-	-
		MLP	-	-	-	✓	-	-	-
		GRNN	-	✓	-	✓		GRNN	-
		Hybrid_1	-	✓	-	-	-	-	-
	1	1	SVR	✓	✓	-	-	-	-
MLP			-	✓	-	✓	-	MLP	-
Hybrid_1			-	✓	✓	✓	-	Hybrid_1	-
Hybrid_2			-	✓	✓	✓	-	Hybrid_2	-
7		MLR	-	✓	-	-	-	-	-
		RF	-	✓	-	✓	-	RF	-
		SVR	-	✓	-	-	-	-	-
		MLP	-	-	-	-	-	-	-
		GRNN	-	✓	-	✓	-	GRNN	-
		Hybrid_1	✓	✓	-	✓	-	Hybrid_1	-
10		Hybrid_2	✓	✓	-	✓	-	Hybrid_2	-
		MLR	-	-	-	-	-	-	-
		RF	✓	✓	-	-	-	-	-
	SVR	-	-	-	-	-	-	-	
	MLP	-	-	-	✓	-	-	-	
GRNN	-	-	-	✓	-	-	-		

Continue à la page suivante

Table 6.12 – Suite de la table précédente									
Groupe	Med	Modèles Test	Validation H4		Validation H5		Modèle retenue		
			$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	$R^2$ & Mape
3		Hybrid_1	-	✓	-	-	-	-	-
		Hybrid_2	-	✓	-	-	-	-	-
	30	MLR	-	-	-	-	-	-	-
		RF	-	-	-	-	-	-	-
		SVR	-	-	-	-	-	-	-
		MLP	-	-	-	-	-	-	-
		Hybrid_1	-	-	-	-	-	-	-
		Hybrid_2	-	-	-	-	-	-	-
	31	MLR	-	-	-	-	-	-	-
		RF	-	✓	-	-	-	-	-
		SVR	-	-	-	-	-	-	-
		MLP	-	-	-	-	-	-	-
		GRNN	-	-	-	-	-	-	-
		Hybrid_1	-	✓	-	-	-	-	-
	41	Hybrid_2	✓	✓	-	-	-	-	-
		MLR	-	-	-	-	-	-	-
		RF	-	-	-	-	-	-	-
		SVR	-	-	-	-	-	-	-
		MLP	-	-	-	-	-	-	-
		GRNN	-	-	-	-	-	-	-
		Hybrid_1	-	-	-	-	-	-	-
3	4	Hybrid_2	-	-	-	✓	-	-	-
		MLR	-	-	-	-	-	-	-
		SVR	-	-	-	-	-	-	-
		Hybrid_1	-	-	-	✓	-	-	-
	8	Hybrid_2	-	-	-	✓	-	-	-
		MLR	-	✓	-	-	-	-	-
		RF	-	✓	-	✓	-	RF	-
		SVR	-	✓	-	-	-	-	-
		MLP	-	✓	-	✓	-	MLP	-
		GRNN	-	✓	-	✓	-	GRNN	-
	9	Hybrid_1	-	✓	-	✓	-	Hybrid_1	-
		Hybrid_2	-	✓	-	✓	-	Hybrid_2	-
		MLR	-	✓	-	-	-	-	-
	24	RF	-	✓	-	-	-	-	-
		Hybrid_2	-	✓	-	-	-	-	-
	26	MLR	-	-	-	-	-	-	-
		MLP	-	-	-	✓	-	-	-
		MLR	-	-	-	-	-	-	-
		RF	-	✓	-	✓	-	RF	-
		SVR	-	✓	-	✓	-	SVR	-
		MLP	-	✓	-	✓	-	MLP	-
GRNN		-	✓	-	✓	-	GRNN	-	
27	Hybrid_1	-	✓	-	✓	-	Hybrid_1	-	
	Hybrid_2	-	✓	-	✓	-	Hybrid_2	-	
	RF	-	-	-	✓	-	-	-	

Continue à la page suivante

Groupe	Med	Modèles Test	Validation H4		Validation H5		Modèle retenue		
			$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	$R^2$ & Mape
	28	Hybrid_2	-	-	-	✓	-	-	-
		MLR	-	-	-	-	-	-	-
		RF	-	-	-	✓	-	-	-
		SVR	-	-	-	-	-	-	-
		MLP	-	-	-	-	-	-	-
		Hybrid_1	-	✓	-	✓	-	Hybrid_1	-
	36	Hybrid_2	-	✓	-	-	-	-	-
		MLR	-	✓	-	-	-	-	-
		SVR	-	✓	-	✓	-	SVR	-
		GRNN	-	✓	-	✓	-	GRNN	-
		Hybrid_1	-	✓	-	✓	-	Hybrid_1	-
	42	Hybrid_2	-	✓	-	✓	-	Hybrid_2	-
		MLR	-	-	-	-	-	-	-
		RF	-	-	-	✓	-	-	-
		SVR	-	-	-	-	-	-	-
		MLP	-	-	-	-	-	-	-
		GRNN	-	-	-	-	-	-	-
		Hybrid_1	-	-	-	✓	-	-	-
Hybrid_2	-	-	-	✓	-	-	-		

TABLE 6.12 – Récapitulatif performances des modèles dans les hôpitaux H4 et H5 selon les seuils de  $R^2$  & MAPE fixés

### 6.3.3.2 Analyse des performances selon les modèles

Le tableau 6.12 synthétise les résultats de la validation des modèles des 18 médicaments dans les hôpitaux H4 et H5 figurant respectivement dans les tableaux 6.10 et 6.11. Sur la base des seuils de métriques fixés ( $R^2 \geq 70\%$  et  $MAPE \leq 50\%$ ), ce tableau présente les performances des modèles des 18 médicaments lors de la phase de validation dans les hôpitaux H4 et H5. Les modèles ayant performé dans les deux hôpitaux relativement aux métriques sont présentés dans les deux dernières colonnes.

Dans les figures 6.5, 6.6 et 6.7 ci-dessous, il est possible de visualiser les performances des modèles construites avec les données des hôpitaux H1, H2 et H3 et validé dans chacun des hôpitaux H4 et H5. Ces figures sont des exemples de médicament issu de chaque groupes de médicaments établis dans la section précédente dans les hôpitaux H4 et H5. Dans certaines des figures, nous avons exclu le modèle de Régression Linéaire Multiple (MLR) afin d'améliorer la lisibilité des autres graphiques. Les graphiques de chaque modèle corroborent avec les valeurs des métriques présentées dans les tableaux 6.10 et 6.11.

Dans les deux hôpitaux (H4 et H5), les modèles MLR ont montré des performances exceptionnellement faibles en matière de prévision. Même si, lors de la phase de construction des modèles, les modèles MLR semblaient avoir de bonnes performances, lors de la validation leurs résultats sont décevants. Ceci a des implications sur les modèles hybrides qui incluent les MLR. Cette situation explique en partie les performances médiocres de certains médicaments, tels que les médicaments 13, 30 et 41 dont même leur modèles hybrides ont été combiné avec le MLR (Consultez les tableaux 6.3 et 6.8 et 6.9).

En se concentrant sur les scores MAPE de chaque modèle, pour les médicaments présentant un MAPE inférieur à 50%, dans le tableau 6.12, nous observons ce qui suit, dans les deux hôpitaux (H4 et H5) :

- Les modèles RF ont fourni des prévisions acceptables pour 4 médicaments en commun : les médicaments 25, 7, 8 et 26.

- Il y a 2 médicaments ayant des prévisions acceptables selon MAPE avec le SVR : les médicaments 26 et 36.
- Les médicaments 1, 8, et 26 sont 3 médicaments ayant une prévision acceptable avec le MLP.
- Avec le GRNN, on identifie 5 médicaments ayant une prévision acceptable dans les deux hôpitaux, à savoir les médicaments 25, 7, 8, 26, et 36.
- Avec Hybrid\_1, il y a 6 médicaments ayant obtenus des prévisions acceptables selon MAPE dans les deux hôpitaux. Nous énumérons les médicaments 1, 7, 8, 26, 28 et 36 .
- Avec Hybrid\_2, il y a cinq médicaments à savoir 1, 7, 8, 26 et 36.

En résumé, selon MAPE, seuls les médicaments 1, 7, 8, 25, 26, 28 et 36 ont des modèles de prévisions acceptables dans les deux hôpitaux. A l'exception du médicament 25, les modèles hybrides sont des modèles communs acceptable selon le MAPE.

Avec le seuil de score de 70% au minimum selon  $R^2$ , aucun modèle n'a donné satisfaction dans les deux hôpitaux simultanément. Ce qui entraîne qu'aucun modèle n'a de score inférieur à 50% selon MAPE et un score  $R^2$  supérieur à 70% à la fois dans les deux hôpitaux. Cela s'explique en partie par la faiblesse du jeu de donnée disponible pour les deux hôpitaux. Les cas où nous observons des  $R^2$  négatifs et des MAPE inférieurs à 50% résultent de plusieurs facteurs combinés, notamment la taille réduite de l'ensemble de données (données agrégées mensuellement pour les hôpitaux H4 et H5 sur une période maximale de 8 mois), la forte dispersion des données (données déséquilibrées) et l'hétéroscédasticité des données (Voir la table 6.1). Le  $R^2$  est influencé par la capacité du modèle à expliquer la variance globale, tandis que le MAPE mesure l'erreur moyenne en pourcentage. Si le modèle ne parvient pas à expliquer de manière satisfaisante la variance globale, le  $R^2$  peut être négatif. Cependant, lorsque le MAPE est inférieur à 50%, cela suggère que le modèle est capable de généraliser aux données inconnues.

Les figures 6.5, 6.6, 6.7 et 6.8 permet de voir malgré des scores souvent mauvais, que les modèles arrivent à identifier les tendances de consommations ce qui sous-entend qu'ils arrivent à identifier les caractéristiques des hôpitaux qui lui sont soumis.

À ce stade, en ce qui concerne notre objectif principal, nous pouvons conclure que selon le MAPE, certaines méthodes d'apprentissage automatique peuvent aider à prédire les besoins de médicaments dans les hôpitaux en utilisant des données provenant d'autres établissements, mais les résultats varient d'un médicament à l'autre. La conclusion de cette étude se résume dans les colonnes "Med" et "Modèle retenu (MAPE)" du tableau 6.12. Ces résultats permettent aux décideurs de prédire la consommation de certains médicaments à l'ouverture d'un nouvel établissement ou une nouvelle unité (le dimensionnement). Nous recommandons l'utilisation de méthodes hybrides avec des modèles individuels soigneusement choisis comme le Random Forest pour chaque médicament. Une fois que le dimensionnement est fait, les modèles présentés dans ce chapitre 5 permettent d'accompagner les décideurs dans la gestion des quantités à commander chaque mois (ou  $n$  mois).

Les résultats de ce travail de thèse montrent aussi que l'accent devrait être mis sur la qualité et quantité de données en plus de la sélection du modèle. Nous recommandons d'augmenter la quantité de données pour améliorer les performances des modèles. La même démarche de traitement et d'analyse de données proposée dans cette thèse est à appliquer pour trouver des éléments de réponses aux autres médicaments (ou dispositifs médicaux d'une manière générale).



FIGURE 6.5 – Visualisation des performances des modèles pour un médicament dans le groupe 1 (Ex. médicament 2 : H4 vs H5)

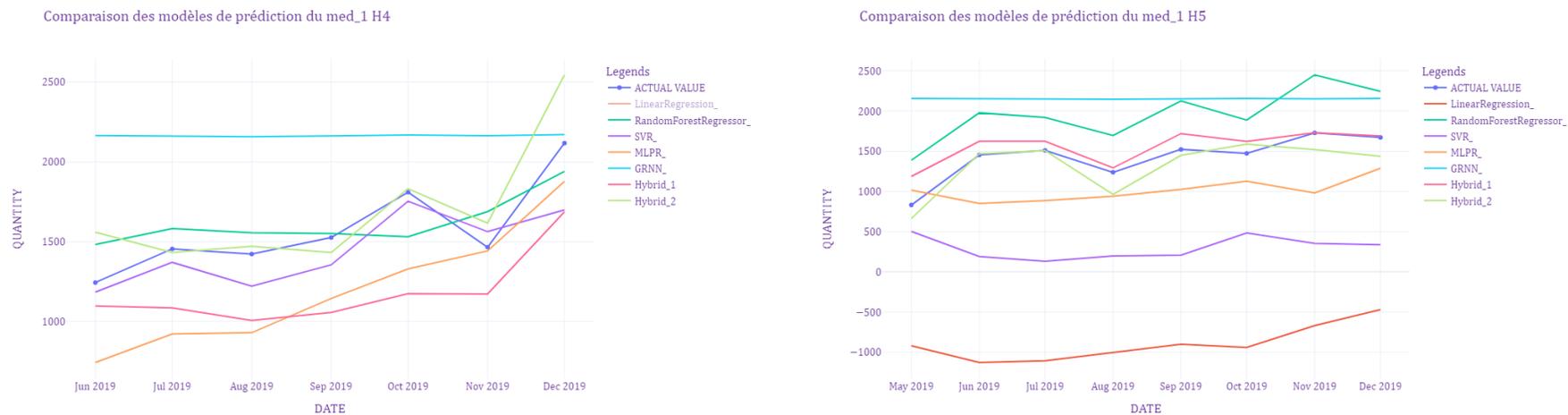


FIGURE 6.6 – Visualisation des performances des modèles pour un médicament dans le groupe 2 (Ex. médicament 1 : H4 vs H5)

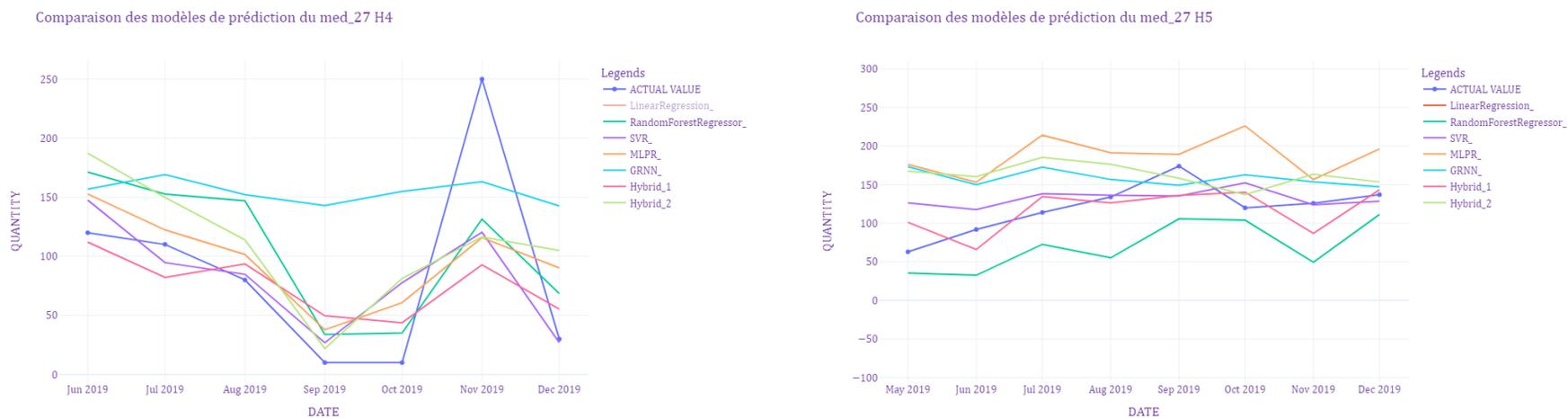


FIGURE 6.7 – Visualisation des performances des modèles pour un médicament dans le groupe 3 (Ex. médicament 27 : H4 vs H5)



FIGURE 6.8 – Visualisation des performances des modèles pour un médicament dans le groupe 5 (Ex. médicament 44 : H4 vs H5)

## 6.4 Conclusions

Le travail de ce dernier chapitre se focalise sur l'utilisation de modèles d'apprentissage automatique pour prédire la demande en médicaments dans des hôpitaux en s'appuyant sur des données de consommation d'autres hôpitaux. Dans l'étude de cas, en fonction des données disponibles pour chacun des hôpitaux, trois hôpitaux ont servi à la construction et aux tests des modèles et deux autres hôpitaux, dont les données n'ont pas concerné la phase d'apprentissage, ont servi à la validation. Sept modèles d'apprentissage automatique, dont 2 modèles hybrides, ont été construits et testés dans une première phase, puis utilisés pour la validation dans une seconde phase. Deux ensembles de résultats se dégagent. Après la phase d'apprentissage et de test, 16 médicaments sur les 45 communs aux hôpitaux ont obtenu au moins un modèle capable de prédire avec une précision supérieure à 70 % selon le  $R^2$  et un MAPE inférieur à 50 %. Pour certains médicaments, nous avons même atteint une précision de 98 % ( $R^2$ ).

Ces premiers résultats confirment les conclusions du chapitre 5, démontrant que l'utilisation conjointe de données provenant de plusieurs hôpitaux, sans les agréger, tout en tenant compte de plusieurs facteurs, permettait aux modèles d'apprentissage automatique, selon une approche causale, de réaliser des prévisions précises. Cette approche peut aider à estimer les besoins en médicaments des hôpitaux et, par conséquent, à améliorer la gestion des stocks.

Le deuxième ensemble de résultats est relatif à la seconde phase de notre étude. Cette seconde phase a consisté à évaluer les modèles construits et entraînés lors de la première phase sur des données de deux hôpitaux (H4 et H5) qui n'avaient pas été utilisées pour l'apprentissage ou les tests. Il en résulte que, sur les 18 médicaments, plus de la moitié d'entre eux ont obtenu au moins un modèle permettant une prévision acceptable selon la métrique MAPE. Une proposition de modèles hybrides d'apprentissage automatique est également effectuée, en mettant l'accent sur la qualité et quantité de données ainsi que le choix préalable des modèles basiques. Également, en respectant des seuils pour les métriques, avec un score supérieur à 70 % pour le  $R^2$  et un MAPE inférieur à 50 %, aucun modèle n'a atteint des performances satisfaisantes simultanément dans les deux hôpitaux. Cela est dû aux limitations de nos ensembles de données dont nous disposons. Les résultats que nous présentons ici n'ont été validés que sur deux hôpitaux. Cela révèle la limitation à laquelle nous avons été confrontés lors de cette étude.

Il est nécessaire de faire fonctionner et de valider les modèles sur d'autres données provenant de plusieurs autres hôpitaux. Cela suggère d'avoir plus de données sur une plus longue période de consommation de plusieurs hôpitaux pour l'expérimentation. Toutefois, la disponibilité et la confidentialité des données peuvent constituer des freins, ceux à quoi nous avons été confrontés. En plus de cette limite, la consommation prise en compte dans ce travail fait référence à la quantité de médicaments livrée à chaque unité fonctionnelle de l'hôpital pour assurer son bon fonctionnement, alors que l'on pourrait utiliser la consommation en relation avec les prescriptions médicales dans les unités fonctionnelles. Mais cela ne remet pas en cause notre approche, la consommation de l'utilisateur final est issue de la consommation des unités fonctionnelles. De même, la méthodologie peut être étendue à d'autres produits de santé, tels que les dispositifs médicaux comme des gants, des masques, des aiguilles, etc. Dans l'objectif de proposer un outil d'aide à la décision aux décideurs du monde hospitalier afin de faciliter le dimensionnement de leurs stocks, il est important qu'une étude comparative soit effectuée pour vérifier les performances de notre approche. Cependant, nous sommes heurtés une nouvelle fois à la disponibilité de données réelles pour en effectuer.

À la lumière de ces premiers résultats, cette étude soulève des questions cruciales et ouvre des pistes pour de futures recherches. Le chapitre suivant, dédié aux conclusions générales et perspectives, explore en détail ces aspects et propose une vision pour l'amélioration de notre approche. Nous discuterons également des possibilités d'extension de cette méthodologie à d'autres produits de santé, et des défis qui se posent en termes de disponibilité et de confidentialité des données, tout en mettant en évidence les limitations que nous avons rencontrées dans cette étude.

# Conclusion générale et perspectives

Dans cette thèse, nous nous sommes intéressés au problème de la prévision de la demande des produits pharmaceutiques dans les hôpitaux pour faciliter le dimensionnement et la gestion des stocks. Un accent particulier est mis sur la prévision des besoins de nouveaux hôpitaux qui sont confrontés à un réel problème de dimensionnement des stocks lors des premières années de fonctionnement. L'objectif principal est de se servir des expériences d'hôpitaux déjà fonctionnels pour aider à prédire les besoins quantitatifs de ces nouveaux hôpitaux dépourvus de données historiques. Pour répondre à ce réel problème posé par l'entreprise KLS-GROUP, nous avons développé une méthodologie et défini les méthodes de résolution permettant de faire des prévisions avec des données réelles.

Il s'agit, bien évidemment, dans un premier temps, de procéder à l'identification des facteurs significatifs qui influent sur la consommation des produits pharmaceutiques et qui constituent les caractéristiques des hôpitaux.

Cette première phase de l'étude se subdivise en trois parties. La première partie a consisté à identifier dans la littérature les différents facteurs impactant la consommation des produits de santé. Nous avons identifié 34 facteurs qui ont été catégorisés en trois groupes qui sont la catégorie socio-démographique, la catégorie des facteurs liés à l'information sanitaire des patients et la catégorie en lien avec les établissements de santé. C'était le lieu aussi d'identifier les méthodes utilisées dans les hôpitaux. Nous avons identifié six méthodes classiques pratiquées depuis les années 80 dont certaines sont recommandées par l'organisation mondiale de la santé.

Les facteurs identifiés ont été soumis à l'avis de certains experts dans les pharmacies hospitalières. En parallèle, nous avons procédé à la collecte de données émanant de deux types de base de données. Il s'agit de données sur des historiques de consommation issues de six hôpitaux français, fournies par l'entreprise partenaire, et de 160 tables de données issues de plusieurs organismes de collecte et de statistiques officiels. Une fois la mise en contexte effectuée, les avis des experts recueillis et en fonction des données à notre disposition, 12 facteurs ont été retenus pour la suite de notre recherche. Ainsi, dans la troisième phase, nous avons pré-traité, organisé et croisé les données pour les rendre exploitables pour les analyses, ce qui est impératif dans notre étude. Nous avons noté que sur les données fournies par l'entreprise partenaire, seul 11,14 % des données relatives aux informations de description des produits de santé sont exploitables, seul 10,73 % des données relatives aux destinations de consommation, les unités fonctionnelles, sont exploitables. Aussi, nous avons constaté seulement 45 médicaments en commun sur 5 674 identifiés dans les six hôpitaux confondus. Au total, la table finale obtenue contient 6 218 lignes et 100 colonnes d'informations. Nous recomman-

---

dons aux décideurs de mettre davantage l'accent sur les informations clés de traçabilité afin de faciliter les utilisations futures. La dernière partie de cette phase a consisté à l'analyse des corrélations entre les facteurs identifiés et la quantité de consommation des médicaments. Ce travail a été réalisé avec l'historique de consommation de 45 médicaments en commun aux six hôpitaux. Les analyses démontrent l'existence d'une corrélation, souvent forte et statistiquement significative, entre la quantité de médicament consommée et certains des facteurs retenus. On observe également que les corrélations sont appréciables par médicament. Ceci implique qu'il faudrait un modèle de prévision pour chaque médicament.

Dans un second temps, nous avons mené des expérimentations pour apporter une réponse au problème posé. L'approche de prévision de la demande adaptée à notre contexte est la méthode causale, une méthode quantitative de prévision de la demande. Cette approche repose sur les méthodes d'apprentissage automatique. L'approche causale est bien indiquée dans notre contexte du fait que nous combinons les historiques de consommation de plusieurs hôpitaux sans les additionner, désirant conserver les patterns et caractéristiques de chacun d'eux.

Dans cette deuxième phase, nous avons proposé en premier lieu une étude expérimentale sur la prévision de la demande mensuelle d'une sélection de 10 médicaments dans les hôpitaux figurant dans notre base de données. L'objectif de ce travail a été double. En plus de l'expérimentation avec notre approche, il a été question de faire une comparaison avec une méthode classique pratiquée dans les hôpitaux, la méthode de consommation, avec des modèles d'apprentissage automatique. Il en résulte que pour la sélection de médicaments, les méthodes d'apprentissage automatique fournissent une meilleure alternative pour les prévisions, notamment le Random Forest avec des précisions dépassant 90% selon  $R^2$  pour certains médicaments. Ces résultats montrent que notre approche peut être utilisée pour prévoir la consommation de médicaments et, par extension, faciliter une gestion efficace des stocks pharmaceutiques dans les hôpitaux. Une confirmation de ces résultats est nécessaire avec plus de médicaments ainsi qu'une vérification des performances de prévisions des modèles dans des hôpitaux dont les données ne serviront pas dans l'apprentissage des modèles, d'où une seconde étude expérimentale lors de cette deuxième phase pour répondre à l'objectif principal de cette thèse.

Le travail dans la seconde étude expérimentale a consisté à construire sept modèles d'apprentissage automatique, dont deux modèles hybrides, pour prédire la demande en médicaments dans deux hôpitaux en se basant sur les données de consommation des autres hôpitaux de notre base de données. La démarche principale se décompose en deux étapes. La première étape consiste à construire les modèles par apprentissage et en effectuant des tests sur trois hôpitaux de référence (H1, H2 et H3). La seconde étape consiste à évaluer les performances des sept modèles sur les données des deux hôpitaux (H4 et H5), il s'agit de la phase de validation.

Nous avons dégagé deux types de résultats. Après la phase d'apprentissage et de test, 16 médicaments sur les 45 communs aux hôpitaux ont obtenu au moins un modèle capable de prédire avec une précision supérieure à 70 % selon le  $R^2$  et un MAPE inférieur à 50 %. Pour certains médicaments, nous avons même atteint une précision de 98 % ( $R^2$ ). Ces résultats

---

qui rejoignent les conclusions de la première étude expérimentale, même si l'expérimentation ne s'est pas faite sur l'ensemble des hôpitaux comme précédemment. Des conclusions qui indiquaient qu'en combinant les données de plusieurs hôpitaux sans les additionner tout en prenant en compte plusieurs facteurs, les modèles d'apprentissage automatique, selon l'approche causale, étaient en mesure de faire de bonne prévision.

Le deuxième ensemble de résultats de l'étude expérimentale, celui relatif aux évaluations des modèles sur les données de deux hôpitaux (H4 et H5), stipule qu'au regard des métriques MAPE et  $R^2$  :

- Dans les deux hôpitaux, les modèles hybrides ont affiché de meilleures performances en termes de MAPE, suivis des modèles de Random Forest. Et seuls 7 médicaments en communs aux deux hôpitaux ont obtenu un au moins un modèle acceptable au regard de leur score MAPE. Une proposition de modèles hybrides d'apprentissage automatique est également effectuée, en mettant l'accent sur la qualité et quantité de données ainsi que le choix préalable des modèles basiques.
- Au regard des seuils de performance " $R^2 \geq 70\%$  et  $MAPE \leq 50\%$ ", aucun modèle n'a donné de prévision acceptable simultanément dans les deux hôpitaux H4 et H5.

Les résultats de cette étude expérimentale montrent que les modèles d'apprentissage automatique, selon l'approche causale, sont capables de faire des prévisions acceptables de la demande en médicaments, même lorsqu'ils sont entraînés sur les données de quelques hôpitaux seulement. Cependant, les seuils de performance fixés sont ambitieux et ont été difficiles à atteindre pour tous les médicaments et dans les deux hôpitaux dus à la limitation des données.

Comme l'utilisation de l'apprentissage automatique se développe dans le domaine des soins de santé, il est essentiel pour les chercheurs et les praticiens de développer une compréhension approfondie du contexte et du rôle des données manquantes. Notre travail sur le traitement et préparation minutieuse des données en collaboration clinique (via la sélection des variables) pour créer des connaissances significatives est une contribution importante au domaine de la prévision de la demande de médicaments dans les hôpitaux. L'identification et l'intégration des facteurs de consommation constituant des caractéristiques des hôpitaux dans des modèles de prévision permettent de capturer certains changements dans l'environnement des hôpitaux. Permettant ainsi de proposer des modèles généralisables à d'autres hôpitaux, une approche applicable dans le secteur industriel.

Une autre contribution principale de nos travaux est la démarche de mise en œuvre des méthodes d'apprentissage automatique pour prédire la demande de médicaments dans les hôpitaux en utilisant des données provenant d'autres hôpitaux sans les additionner et sans se soucier de la chronologie de la consommation, tout en tenant compte des facteurs qui influencent ces consommations de médicaments. Les études de la littérature se concentrent sur des approches de séries temporelles multivariées qui ne sont pas généralisables à d'autres hôpitaux. En effet, ces études se concentrent sur la prévision pour un seul hôpital ou un ensemble d'hôpitaux dont les données sont agrégées avant l'évaluation des modèles. Notre

---

travail, quant à lui, permet de prédire la demande de médicaments dans des hôpitaux différents, ce qui est un avantage important pour aider à un bon dimensionnement.

Cependant, ce travail présente certaines limites. La limitation principale réside dans la disponibilité et dans la qualité des données destinées à ce type d'études. Les hôpitaux sont en train de se conformer aux nouvelles normes de traçabilité et de collecte d'informations, ce qui devrait améliorer la situation à l'avenir.

Nos données s'arrêtent à la consommation des unités fonctionnelles, et non à la consommation du destinataire final (le patient). Ce qui nous a amené à considérer que la "consommation" dans ce travail fait référence à la quantité de médicaments livrés à chaque unité fonctionnelle, et non à celle liée à la prescription. Les données relatives à la consommation des patients nous étaient inaccessibles.

Nous n'avons pas inclus le type d'unité fonctionnelle ou de service où la consommation de médicaments a lieu en tant que variable, car il n'existe pas de nomenclature standard pour les unités fonctionnelles ou les services en France. Nous pensons que la précision des prédictions serait meilleure si cette caractéristique pouvait être incluse dans le modèle, car les types et les quantités de médicaments consommés sont corrélés au type d'unité fonctionnelle ou de service.

Même si nous travaillons avec des données réelles des hôpitaux, nous n'avons pas eu l'opportunité de comparer notre approche à un cas réel lors de l'installation d'une plateforme WMS dans un nouvel hôpital. Cela nous aurait permis de mesurer les marges d'erreur et de prendre des actions correctives pour dimensionner les stocks de médicaments. Dans la mesure où des données seront disponibles pour ce type d'expérimentations, notre approche pourrait aider notre entreprise partenaire à accompagner les décideurs du domaine hospitalier pour le dimensionnement et la gestion des stocks. En outre, l'application de notre approche à plusieurs hôpitaux pourrait contribuer à obtenir un outil robuste d'aide à la décision.

Des analyses de sensibilité des horizons temporels de prévision de la demande et des hyperparamètres pourraient peut-être améliorer les modèles, ce qui constitue une perspective. Au lieu de se focaliser uniquement sur l'horizon temporel des prévisions indiquées par l'entreprise, il pourrait être utile d'analyser l'impact de différentes périodes de prévision sur les prévisions de demande de médicaments si les données sont accessibles. Par exemple, la prévision de la demande de médicaments à court terme (une semaine) peut différer de celle à long terme (trois mois ou plus). L'optimisation des paramètres des modèles d'apprentissage automatique peut être améliorée en ajustant davantage les hyperparamètres pour obtenir des résultats plus précis. Cela peut impliquer une analyse de sensibilité des hyperparamètres et des tests d'autres méthodes d'apprentissage automatique, notamment les réseaux de neurones profonds.

Une piste de recherche consiste à regrouper les médicaments selon les mêmes comportements de consommation ou les mêmes ingrédients actifs afin d'obtenir un modèle de prédiction générique pour un groupe de médicaments.

Une autre piste de recherche consisterait à explorer l'application de techniques de séries chronologiques multivariées et de méthodes causales avec l'apprentissage fédéré (federated learning en anglais). L'apprentissage fédéré est une approche qui permet de construire et

---

de partager des modèles d'apprentissage automatique sans centraliser les données brutes en un seul endroit. En prévision de la demande de médicaments dans les hôpitaux à partir de séries chronologiques, il pourrait agréger les mises à jour de modèle provenant de différentes unités hospitalières, améliorant ainsi la précision des prévisions en considérant des sources d'information variées. Cette approche peut permettre de minimiser la centralisation des données, réduisant les risques de sécurité et de confidentialité tout en permettant une meilleure adaptation locale du modèle aux spécificités de chaque hôpital. En intégrant les données distribuées tout en respectant leur localisation, le federated learning pourrait rendre le modèle global plus robuste et généralisable, améliorant ainsi la qualité des prévisions à l'échelle régionale ou nationale. Ce sont autant de pistes de recherche que nous explorons. Notre travail, ainsi que les perspectives que nous évoquons, peuvent être étendus aux dispositifs médicaux ou aux médicaments rares (spécialisés) pour évaluer leur efficacité et progresser vers une généralisation.

# Bibliographie

- [1] Ahmad Jobran Al-Mahasneh, Sreenatha Anavatti, Matthew Garratt, Mahardhika Pratama, Ahmad Jobran Al-Mahasneh, Sreenatha Anavatti, Matthew Garratt, and Mahardhika Pratama. Applications of general regression neural networks in dynamic systems. *Digital Systems*, 11 2018.
- [2] Ethem Alpaydin. *Introduction to machine learning*. The MIT Press, third edition edition, 2014.
- [3] Yazed Sulaiman AlRuthia, Hadeel Al Kofide, Refaa’H Al Ajmi, Bander Balkhi, Ahmed Alghamdi, Albandari Al Nasser, Areej Alayed, Maali Alshammari, Danah Alsuhaibani, and Asma Alathbah. Drug shortages in large hospitals in riyadh : A cross-sectional study. *Annals of Saudi Medicine*, 37 :375–385, 9 2017.
- [4] Amy Alspaugh, Nikki Lanshaw, Jan Kriebs, and Cheri Van Hoover. Universal health care for the united states : A primer for health care providers. *Journal of Midwifery & Women’s Health*, 66 :441–451, 7 2021.
- [5] Mohsen Sadegh Amalnick, Naser Habibifar, Mahdi Hamid, and Mahdi Bastan. An intelligent algorithm for final product demand forecasting in pharmaceutical units. *International Journal of System Assurance Engineering and Management*, 11 :481–493, 4 2020.
- [6] Brian Archer. Demand forecasting and estimation. *Demand forecasting and estimation*, pages 77–85, 1987.
- [7] Sina Ardabili, Amir Mosavi, and Annamária R. Várkonyi-Kóczy. Advances in machine learning modeling reviewing hybrid and ensemble methods. *Lecture Notes in Networks and Systems*, 101 :215–227, 2020.
- [8] J. Scott Armstrong and Kesten C. Green. Demand forecasting : Evidence-based methods. *SSRN Electronic Journal*, 11 2017.
- [9] Monika Arora and Yogita Gigras. Importance of supply chain management in health-care of third world countries. *International Journal of Supply and Operations Management*, 5 :101–106, 2 2018.
- [10] R. Artusi, P. Verderio, and E. Marubini. Bravais-pearson and spearman correlation coefficients : Meaning, test of hypothesis and confidence interval. *The International Journal of Biological Markers*, 17 :148–151, 4 2002.
- [11] Miftah Asthariq, Wahyuni Nasution, Sri Lestari, and Ramadhani Nasution. Analysis of implementation of drug inventory control using abc-eoq-rop-ss method at arun hospital

- lhokseumawe. *International Journal of Health and Pharmaceutical (IJHP)*, 2 :684–691, 8 2022.
- [12] Gavin Bailey, Tom Cherrett, Ben Waterson, and Robert Long. Can locker box logistics enable more human-centric medical supply chains? *International Journal of Logistics Research and Applications*, 16 :447–460, 2013.
- [13] Rama Baru, Arnab Acharya, Sanghmitra Acharya, Shiva Kumar, and K Nagaraj. Inequities in access to health services in india : Caste, class and region. *Economic and Political Weekly*, 45 :49–58, 2010.
- [14] Sergio Bautista-Arredondo, Edson Serván-Mori, M Arantxa Colchero, Baruch Ramírez-Rodríguez, and Sandra G Sosa-Rubí. Analysis of outpatient healthcare utilization in the context of the universal healthcare coverage reform in mexico. *Salud Pública de México*, 56 :18–31, 2014.
- [15] Martin Beaulieu, Jacques Roy, Sylvain Landry, Martin Michaud, and Clément Roy. La logistique hospitalière au québec : passé, présent et futur. *Gestion*, Vol. 39 :56–62, 10 2014.
- [16] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. *Pearson correlation coefficient*, volume 2, pages 1–4. Springer Science and Business Media B.V., 2009.
- [17] Johan Bernardini-Perinciolo and Marius Bertolucci. Quand la sur-rationalité provoque la sous-rationalité. *Revue internationale de psychosociologie et de gestion des comportements organisationnels*, 23 :91–109, 7 2018.
- [18] Ernst R. Berndt. Pharmaceuticals in u.s. health care : Determinants of quantity and price. *Journal of Economic Perspectives*, 16 :45–66, 2002.
- [19] Steve Beveridge and Cyril Oickle. A comparison of box—jenkins and objective methods for determining the order of a non-seasonal arma model. *Journal of Forecasting*, 13 :419–434, 1994.
- [20] María G. Borgognone, Javier Bussi, and Guillermo Hough. Principal component analysis in sensory analysis : Covariance or correlation matrix? *Food Quality and Preference*, 12 :323–326, 7 2001.
- [21] Carme Borrell, Carles Muntaner, Diana Gil-González, Lucia Artazcoz, Maica Rodríguez-Sanz, Izabella Rohlf, Katherine Pérez, Mar García-Calvente, Rodrigo Villegas, and Carlos Álvarez Dardet. Perceived discrimination and health by gender, social class, and country of birth in a southern european country. *Preventive Medicine*, 50 :86–92, 2010.
- [22] Bertrand Bourgeon, Anne Constantin, Guillaume Karolszyk, Jean-François Marquot, Serena Pedrini, Sylvain Landry, Angel Diaz, and Dominique Estampe. Evaluation des coûts logistiques hospitaliers en france et aux pays-bas. *Logistique & Management*, 9 :81–87, 1 2001.
- [23] Leo Breiman. Stacked regressions. *Machine Learning*, 24 :49–64, 7 1996.
- [24] Charles D. Brennan. Integrating the healthcare supply chain. *Healthcare financial management : journal of the Healthcare Financial Management Association*, 52 :31–34, 1998.

- [25] Santiago-Omar Caballero-Morales and Gladys Bonilla-Enriquez. Six-sigma guidelines to improve inventory management in a bottling company. *International Journal of Entrepreneurial Knowledge*, 10 :20–33, 6 2022.
- [26] Davide Chicco, Matthijs J. Warrens, and Giuseppe Jurman. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, 7 :1–24, 7 2021.
- [27] Il Hwan Chung, Daniel W. Williams, and Myung Rok Do. *Public Performance & Management Review*.
- [28] Roberto R. Corsini, Antonio Costa, Sergio Fichera, and Josè M. Framinan. A new data-driven framework to select the optimal replenishment strategy in complex supply chains. *IFAC-PapersOnLine*, 55 :1423–1428, 1 2022.
- [29] Maria Costin. Logistique hospitalière, un outil du management. *Humanisme et Entreprise*, n° 299 :29–48, 10 2010.
- [30] Gautier Defosse, Sandra Le Guyader-peyrou, Zoé Uhry, Pascale Grosclaude, Marc Colonna, Emmanuelle Dantony, Patricia Delafosse, Florence Molinié, Anne-Sophie Woronoff, Anne marie Bouvier, and Alain Monnereau. Estimation nationale de l'incidence et de la mortalité par cancer en france métropolitaine entre 1990 et 2018. *Étude à partir des registres des cancers du réseau Francim*, 1, 2019.
- [31] Ertha Kusuma Dewi, Maznah Dahlui, Djazuly Chalidyanto, and Thinni Nurul Rochmah. Achieving cost-efficient management of drug supply via economic order quantity and minimum-maximum stock level. *Expert Review of Pharmacoeconomics & Outcomes Research*, 20 :289–294, 5 2019.
- [32] R. Dunford, Q. Su, and E. Tamang. The pareto principle. *The Plymouth Student Scientist*, 07 :140–148, 2014.
- [33] Management Sciences for Health. Managing access to medicines and health technologies. *Arlington, VA : Management Sciences for Health*, 2012.
- [34] Jerome H Friedman. Greedy function approximation : A gradient boosting machine. *The Annals of Statistics*, 29 :1189–1232, 2001.
- [35] Abdrhman Mahmoud Gamil. Pharmaceutical procurement practice aspects. *World Journal of Advanced Research and Reviews*, 8 :337–349, 12 2020.
- [36] M A Ganaie, Minghui Hu, A K Malik, M Tanveer, and P N Suganthan. Ensemble deep learning : A review. 2022.
- [37] Kurubaran Ganasegeran, Sami A.R. Al-Dubai, Ahmad M. Qureshi, Al Abed A.A. Al-Abed, Rizal Am, and Syed M. Aljunid. Social and psychological factors affecting eating habits among university students in a malaysian medical school : A cross-sectional study. *Nutrition Journal*, 11 :1–7, 7 2012.
- [38] K. Ganesh, S. Ganesh, T.T. Narendran, and V.R. Muraleedharan. Drug inventory management at public healthcare institutions : a case study. *International Journal of Logistics Economics and Globalisation*, 3 :155, 2011.

- [39] Mark Goh, Ou Jihong, and Teo Chung-Piaw. Warehouse sizing to minimize inventory and storage costs. *Naval Research Logistics (NRL)*, 48 :299–312, 6 2001.
- [40] Stephen Gorin. Universal health care coverage in the united states :barriers, prospects, and implications. *Health & Social Work*, 22 :223–230, 8 1997.
- [41] Francesco Granata, Fabio Di Nunno, and Giuseppe Modoni. Hybrid machine learning models for soil saturated conductivity prediction. *Water*, 14 :1726, 2022.
- [42] Margaret B. Greenwood-Ericksen, Sabrina J. Poon, Lewis S. Nelson, Scott G. Weiner, and Jeremiah D. Schuur. Best practices for prescription drug monitoring programs in the emergency department setting : Results of an expert panel. *Annals of Emergency Medicine*, 67 :755–764.e4, 6 2016.
- [43] Anna Gu, Albert I Wertheimer, ; Bernard Brown, and Fadia T Shaya. Drug shortages in the us—causes, impact, and strategies. *INNOVATIONS*, 2 :60, 2011.
- [44] Kai Guo, Shanshan Song, Lijuan Qiu, and Xiaohuan Wang. Prediction of red blood cell demand for pediatric patients using a time-series model : A single-center study in china. *Frontiers in Medicine | www.frontiersin.org*, 1 :706284, 2022.
- [45] Hafsa Habebh and Suril Gohel. Machine learning in healthcare. *Current Genomics*, 22 :291, 12 2021.
- [46] Muhtasim Hafiz, Md Sabir Ibna Sazzad, Khalid Ibne Hasan, Jamil Hasnat, and Mahbub C. Mishu. Predicting the demand of prescribed medicines in bangladesh using artificial intelligent (ai) based long short-term memory (lstm) model. *ACM International Conference Proceeding Series*, 1 2020.
- [47] Thouraya Hamdi, Adel Ghith, and Faten Fayala. A principal component analysis (pca) method for predicting the correlation between some fabric parameters and the drape. *Autex Research Journal*, 14 :22–27, 2014.
- [48] Zhongyang Han, Jun Zhao, Henry Leung, King Fai Ma, and Wei Wang. A review of deep learning models for time series prediction, 3 2021.
- [49] Taher Hassan. Logistique hospitalière : organisation de la chaîne logistique pharmaceutique aval et optimisation des flux de consommables et des matériels à usage unique. 2006.
- [50] L. L. Havlicek and N. L. Peterson. Robustness of the pearson correlation against violations of assumptions. *Perceptual and Motor Skills*, 43 :1319–1334, 1976.
- [51] Donald Hebb. *The Organization of Behavior : a neuropsychological theory*. Jon wiley & sons edition, 1949.
- [52] Karin Henricson, Pal Stenberg, Gerhard Rametsteiner, Jonas Ranstam, Bertil S. Hansson, and Arne Melander. Socioeconomic factors, morbidity and drug utilization - an ecological study. *Pharmacoepidemiology and Drug Safety*, 7 :261–267, 1998.
- [53] Jennifer Hill and Elizabeth A. Stuart. Causal inference : Overview. *International Encyclopedia of the Social & Behavioral Sciences : Second Edition*, pages 255–260, 3 2015.

- [54] Charles E. Housley. *Hospital materiel management*. Aspen Systems Corp, 1978.
- [55] Jennifer N. Howard, Ilene Harris, Gavriella Frank, Zippora Kiptanui, Jingjing Qian, and Richard Hansen. Influencers of generic drug utilization : A systematic review. *Research in Social and Administrative Pharmacy*, 14 :619–627, 2018.
- [56] Krisje Hurkens, Wendy van der Valk, and Finn Wynstra. Total cost of ownership in the services sector : A case study. *Journal of Supply Chain Management*, 42 :27–37, 2006.
- [57] Chaitanya Ingle, Dev Bakliwal, Jayesh Jain, Preeyesh Singh, Preeti Kale, and Vaibhav Chhajed. Demand forecasting : Literature review on various methodologies. *2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021*, 2021.
- [58] Mir Javid Iqbal, Mohammad Ishaq Geer, and Parvez Ahamd Dar. Evaluation of medicines forecasting and quantification practices in various public sector hospitals using indicator based assessment tool. *Journal of Applied Pharmaceutical Science*, 7 :072–076, 2017.
- [59] Deepak Janardhanan and Enda Barrett. Cpu workload forecasting of machines in data centers using lstm recurrent neural networks and arima models. *2017 12th International Conference for Internet Technology and Secured Transactions, ICITST 2017*, pages 55–60, 5 2018.
- [60] Christian Janiesch, Patrick Zschech, and Kai Heinrich. Machine learning and deep learning. *Electronic Markets*, 31 :685–695, 9 2021.
- [61] Fouad Jawab, Youness Frichi, and Said Boutahari. Hospital logistics activities. *Proceedings of the International Conference on Industrial Engineering and Operations Management*, pages 3228–3237, 2018.
- [62] Dolores Jiménez-Rubio and Cristina Hernández-Quevedo. Explaining the demand for pharmaceuticals in spain : Are there differences in drug consumption between foreigners and the spanish population ? *Health Policy*, 97 :217–224, 2010.
- [63] Anyona Johnson, Karimi Peter, and Maru Shital. Inventory management practices and supply chain performance of antiretroviral medicines in public hospitals in nyamira county, kenya. *Rwanda Journal of Medicine and Health Sciences*, 4 :257–268, 9 2021.
- [64] Kaluai Julius, Wanja Mwaura Tenambergen, and Job Mapesa. Factors influencing management of pharmaceutical supplies at meru teaching and referral hospital, kenya. *Health Systems Management Journal*, 1 :12–Pages, 2018.
- [65] Prajakta S Kalekar. Time series forecasting using holt-winters exponential smoothing. *Kanwal Rekhi school of information Technology*, 4329008 :1–13, 12 2004.
- [66] Peter Kelle, John Woosley, and Helmut Schneider. Pharmaceutical supply chain specifics and inventory solutions for a hospital case. *Operations Research for Health Care*, 1 :54–63, 2012.
- [67] Prince Waqas Khan, Yung Cheol Byun, Sang Joon Lee, Dong Ho Kang, Jin Young Kang, and Hae Su Park. Machine learning-based approach to predict energy consumption of renewable and nonrenewable power sources. *Energies*, 13 :4870, 9 2020.

- [68] Tomasz Kijewski, Malgorzata Zbawicka, Jakob Strand, Hans Kautsky, Jonne Kotta, Merli Rätsep, and Roman Wenne. Random forest assessment of correlation between environmental factors and genetic differentiation of populations : Case of marine mussels *mytilus*. *Oceanologia*, 61 :131–142, 1 2019.
- [69] Denis Koala, Zakaria Yahouni, Gülgün Alpan, and Yannick Frein. Factors influencing drug consumption and prediction methods. 2021.
- [70] Denis Koala, Zakaria Yahouni, Gülgün Alpan, and Djamal Si Mohand. Correlation analysis of factors impacting health product consumption in french hospitals. *IFAC-PapersOnLine*, 55 :895–900, 1 2022.
- [71] Richard L. Kravitz and Robert A. Bell. Media, messages, and medication : Strategies to reconcile what patients hear, what they want, and what they need from medications. *BMC Medical Informatics and Decision Making*, 13 :1–8, 12 2013.
- [72] Sylvain Landry and Martin Beaulieu. La logistique hospitalière : un remède aux maux du secteur de la santé? *Gestion*, Vol. 26 :34–41, 12 2001.
- [73] Yann Lecun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature 2015* 521 :7553, 521 :436–444, 5 2015.
- [74] M.-K Lee, E A Elsayed, M.-K Leey, and E A Elsayedz. Optimization of warehouse storage capacity under a dedicated storage policy. *International Journal of Production Research*, 43 :1785–1805, 5 2005.
- [75] Juan Alfonso Leonardia and Zambrano Andrew Paul G. Development of a traditional medicine disaster kit. *Southeast Asian journal of tropical medicine and public health*, 40 :96, 2009.
- [76] Li Li, Yuxi Fan, Mike Tse, and Kuo Yi Lin. A review of applications in federated learning. *Computers & Industrial Engineering*, 149 :106854, 11 2020.
- [77] Yanli Liu, Yourong Wang, and Jian Zhang. New machine learning algorithm : Random forest. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7473 LNCS :246–252, 2012.
- [78] Renzhi Lu and Seung Ho Hong. Incentive-based demand response for smart grid with reinforcement learning and deep neural network. *Applied Energy*, 236 :937–949, 2 2019.
- [79] Mehرداد Mahdizadeh and Mahnaz Solhi. Relationship between self-care behaviors and health literacy among elderly women in iran, 2015. *Electronic Physician*, 10 :6462, 3 2018.
- [80] Spyros Makridakis and Micheále Hibon. Arma models and the box-jenkins methodology. *Journal of Forecasting*, 16 :147–163, 1997.
- [81] Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. Statistical and machine learning forecasting methods : Concerns and ways forward. *PLOS ONE*, 13 :e0194889, 3 2018.
- [82] Guillaume Marques, Caroline Thierry, Jacques Lamothe, and Didier Gourc. A review of vendor managed inventory (vmi) : from concept to processes. *Production Planning & Control*, 21 :547–561, 9 2010.

- [83] Beaulieu Martin, Bentahar Omar, and Benzidia Smail. The evolution of healthcare logistics : The canadian experience. *Journal of Applied Business & Economics*.
- [84] James W Martin. *Lean Six Sigma for Supply Chain Management : A 10-Step Solution Process*. McGraw-Hill Education, 2014.
- [85] Christine Di Martinelly, Alain Guinet, and Fouad Riane. Chaîne logistique en milieu hospitalier : modélisation des processus de distribution de la pharmacie. pages 1–8, 6 2005.
- [86] Behzad Karami Matin, Saeed Reza Azami, Saeid Mahmoudi, Satar Rezaei, Faramarz Shaahmadi, and Ali Kazemi Karyani. Determinants of pharmaceutical expenditures of urban households : A time series study in kermanshah province (iran). *Electronic Physician*, 7 :1470, 11 2015.
- [87] D Maulud, AM Abdulazeez Journal of Applied Science, Technology, and undefined 2020. A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, 01 :140–147, 2020.
- [88] Franklin E. May, Ronald B. Stewart, and Leighton E. Cluff. Drug use in the hospital : Evaluation of determinants. *Clinical Pharmacology and Therapeutics*, 16 :834–845, 1974.
- [89] Francois Mbonyinshuti, Joseph Nkurunziza, Japhet Niyobuhungiro, and Egide Kayitare. The prediction of essential medicines demand : A machine learning approach using consumption data in rwanda. *Processes*, 10 :26, 2021.
- [90] Marzieh Mehrjoo. Supply chain risk assessment for perishable products applying system dynamics methodology - a case of fast fashion apparel industry. *Electronic Theses and Dissertations*, 1 2014.
- [91] Sajad Ebrahimi Meimand, Hadi Ganjalikhan Hakemi, and Hairul Nizam Ismail. Demand forecasting : A review on qualitative methods and surveyed done by delphi technique. 2012.
- [92] Ahmed U. Mishuk, Jingjing Qian, Jennifer N. Howard, Ilene Harris, Gavriella Frank, Zippora Kiptanui, and Richard Hansen. The association between patient sociodemographic characteristics and generic drug use : A systematic review and meta-analysis. *Journal of Managed Care and Specialty Pharmacy*, 24 :252–264, 2018.
- [93] TM Mitchell. *Machine learning*. McGraw-Hill Science/Engineering/Math, 3 1997.
- [94] David S. Moore, William Notz, and Michael A Flinger. *The basic practice of statistics*. 9 edition, 2021.
- [95] Leo S Morales, Marielena Lara, Raynard S Kington, Robert O Valdez, José J Escarce, and Rand Health. Socioeconomic, cultural, and behavioral factors affecting hispanic health outcomes. *J Health Care Poor Underserved*, 13 :477–503, 2002.
- [96] Juan José Montaña Moreno, Alfonso Palmer Pol, Albert Sesé Abad, and Berta Cajal Blasco. Using the r-mape index as a resistant measure of forecast accuracy. *Psicothema*, 25 :500–506, 11 2013.

- [97] Mohammad Amin Morid, Olivia R. Liu Sheng, and Joseph Dunbar. Time series prediction using deep learning methods in healthcare. *ACM Transactions on Management Information Systems*, 14 :1–26, 8 2023.
- [98] Maryam Motamedi, Na Li, Douglas G. Down, and Nancy M. Heddle. Demand forecasting for platelet usage : from univariate time series to multivariate models. 1 2021.
- [99] Yashuang Mu, Xiaodong Liu, and Lidong Wang. A pearson’s correlation coefficient based decision tree and its parallel implementation. *Information Sciences*, 435 :40–58, 4 2018.
- [100] Joy Muhia, Lynn Waithera, and Rogers Songole. Factors affecting the procurement of pharmaceutical drugs : A case study of narok county referral hospital, kenya. *Medical & Clinical Reviews*, 3, 2017.
- [101] Heather Nachtmann and Edward A. Pohl. The state of healthcare logistics : cost and quality improvement opportunities. *Center for Innovation in Healthcare Logistics, University of Arkansas*, 2009.
- [102] Sushmita A. Narayana, Rupesh Kumar Pati, and Prem Vrat. Managerial research on the pharmaceutical supply chain - a critical review and some insights for future directions. *Journal of Purchasing and Supply Management*, 20 :18–40, 2014.
- [103] H360 – Association nationale des cadres et experts techniques hospitaliers. La logistique hospitalière.
- [104] Angie Nguyen, Robert Pellerin, Samir Lamouri, and Béranger Lekens. Managing demand volatility of pharmaceutical products in times of disruption through news sentiment analysis. *International Journal of Production Research*, 61 :2828–2839, 2022.
- [105] Raheem Abdel Yosef Odeh. Health beliefs, practice, and priorities for health care of arab muslims in the united states implications for nursing care. *Journal of Transcultural Nursing*, 19 :284–291, 4 2008.
- [106] OECD. Health at a glance 2013 : Oecd indicators., 2013. Spending on inpatient care and outpatient care combined accounts for a large proportion of health expenditure across OECD countries – around 62% current health expenditure on average in 2011 (Figure 7.3.1). A further 20% health spending was allocated to medical goods (mainly pharmaceuticals, which accounted for 17% remaining 6% preventions services and administration.
- [107] OECD. Health at a glance 2019 : Oecd indicators, 11 2019.
- [108] OECD. Health at a glance 2021 : Oecd indicators, 2021.
- [109] World Health Organization. *Estimating Drug requirements : A practical manual*. World Health Organisation, 1988.
- [110] World Health Organization. Regional workshop on strengthening quantification and procurement of essential, 6 2014.
- [111] H. Osore. Estimating drug requirements using morbidity data-based method : Cumulative country experience. *Tropical Doctor*, 19 :90–94, 1989.
- [112] Eva Ostertagová, Oskar Ostertag, and Jozef Kováč. Methodology and application of the kruskal-wallis test. *Applied Mechanics and Materials*, 611 :115–120, 2014.

- [113] Ralph G. O'Brien. A general anova method for robust tests of additive models for variances. *Journal of the American Statistical Association*, 74 :877–880, 1979.
- [114] Bijal Pandya and Hemant Thakkar. A review on inventory management control techniques : Abc-xyz analysis. *REST Journal on Emerging trends in Modelling and Manufacturing*, 2 :82–86, 2016.
- [115] E Pappa, N Kontodimopoulos, AA Papadopoulos, Y Tountas, and D Niakas. Pcp15 determinants of pharmaceutical consumption in a general population. *Value in Health*, 11 :A366–A367, 2008.
- [116] Judea Pearl. Causal inference in statistics : An overview. *Statistics Surveys*, 3 :96–146, 1 2009.
- [117] Club Inter Pharmaceutique. Traçabilité du médicament, 2009.
- [118] P.H. Pittman, J.B. Atwater, American Production, and Inventory Control Society.
- [119] Niina Pokela, J. Simon Bell, Katri Lihavainen, Raimo Sulkava, and Sirpa Hartikainen. Analgesic use among community-dwelling people aged 75 years and older : A population-based interview study. *American Journal Geriatric Pharmacotherapy*, 8 :233–244, 2010.
- [120] Etienne Poulin. Benchmarking the hospital logistics process : a potential cure for the ailing health care sector. *CMA MANAGEMENT*, 77 :20–23, 2003.
- [121] Sushil Punia, Konstantinos Nikolopoulos, Surya Prakash Singh, Jitendra K. Madaan, and Konstantia Litsiou. Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. *International Journal of Production Research*, 58 :4964–4979, 8 2020.
- [122] Cao Qingkui and Ruan Junhu. Study on the demand forecasting of hospital stocks based on data mining and bp neural networks. *Proceedings - 2009 International Conference on Electronic Commerce and Business Intelligence, ECBI 2009*, pages 284–289, 2009.
- [123] Ilma Nurul Rachmania and Mursyid Hasan Basri. Pharmaceutical inventory management issues in hospital supply chains. *Management*, 3 :1–5, 2013.
- [124] M. B. Rao, John P. Klein, and Melvin L. Moeschberger. *Survival Analysis Techniques for Censored and Truncated Data*, volume 40. 1998.
- [125] R. Rathipriya, Abdul Aziz Abdul Rahman, S. Dhamodharavadhani, Abdelrhman Meero, and G. Yoganandan. Demand forecasting model for time-series pharmaceutical data using shallow and deep neural network model. *Neural Computing & Applications*, 35 :1945, 1 2023.
- [126] Handanhal Ravinder and Ram Misra. Abc analysis for inventory management : Bridging the gap between research and classroom. *American Journal of Business Education*, 1 2014.
- [127] Payam Refaeilzadeh, Lei Tang, and Huan Liu. *Cross-Validation*, pages 1–7. Springer, New York, NY, 2016.

- [128] Nazaré Rego, João Claro, and Jorge Pinho de Sousa. A hybrid approach for integrated healthcare cooperative purchasing and supply chain configuration. *Health Care Management Science*, 17 :303–320, 2014.
- [129] Laura M Ramirez Restrepo, Aicha Aguezoul, and Sophie Hennequin. Dimensionnement d’entrepôts d’une chaîne logistique durable pilotée par la politique echelon base stock. 2015.
- [130] Díaz Rodríguez Dania Rocío, Hernández Barrera Valentín, Jiménez Trujillo Isabel, and Carrasco Garrido Pilar. Factors associated to medication consumption among the immigrant population residing in spain. *Journal of Immigrant and Minority Health*, 20 :909–919, 2018.
- [131] Marie-Odile Safon. Les réformes hospitalières en france : aspects historiques et réglementaires. *Aspects historiques et réglementaires, Synthèse documentaire, IRDES*, 57, 2017.
- [132] Rosalin Sahoo, Ajit Kumar Pasayat, Bhaskar Bhowmick, Kiran Fernandes, and Manoj Kumar Tiwari. A hybrid ensemble learning-based prediction model to minimise delay in air cargo transport using bagging and stacking. *International Journal of Production Research*, 60 :644–660, 2022.
- [133] Nathalie Sampieri-Teissier. Valoriser la logistique hospitalière au service des patients. *Supply Chain Magazine*, pages 144–145, 2016.
- [134] A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3 :210–229, 7 1959.
- [135] France Assos Santé. Pénurie de vaccins et de médicaments : les inquiétudes de france assos santé confirmées par une enquête exclusive - france assos santé.
- [136] Feri Dian Sanubari, Agusdini Banun Saptaningsih, Rokiah Kusumapradja, and Arrozi Arrozi. Differences month of stock chemotherapy based on consumption and morbidity methods at mrccc hospital. *Journal Health Sains*, 4 :35–39, 3 2023.
- [137] Seddigheh Edalat Sarvestani, Nahid Hatam, Mozghan Seif, Leila Kasraian, Fazilat Sharifi Lari, and Mohsen Bayati. Forecasting blood demand for different blood groups in shiraz using auto regressive integrated moving average (arima) and artificial neural network (ann) and a hybrid approaches. *Scientific Reports*, 12, 12 2022.
- [138] Muhammad Rehan Sarwar, Muhammad Atif, Shane Scahill, Anum Saqib, Muhammad Qamar uz Zaman, and Zaheer Babar. Drug utilization patterns among elderly hospitalized patients on poly-pharmacy in punjab, pakistan. *Journal of Pharmaceutical Policy and Practice*, 10 :1–16, 2017.
- [139] Maximilian Schilling, Lennart Rickmann, Gabriele Hutschenreuter, and Cord Spreckelsen. Reduction of platelet outdating and shortage by forecasting demand with statistical learning and deep neural networks : Modeling study. *JMIR Medical Informatics*, 10, 2 2022.
- [140] Marie-Paule Serre and Déborah Wallet-Wodka. *Marketing des produits de santé*. Dunod, 2014.

- [141] I. Nyoman Setiawan, Robert Kurniawan, Budi Yuniarto, Rezzy Eko Caraka, and Bens Pardamean. Parameter optimization of support vector regression using harris hawks optimization. *Procedia Computer Science*, 179 :17–24, 2021.
- [142] Ye Shi, Xiaolong Guo, and Yugang Yu. Dynamic warehouse size planning with demand forecast and contract flexibility. *International Journal of Production Research*, 56 :1313–1325, 2 2017.
- [143] Murtada Shubbar and Jared Smith. Demand forecasting : An open-source approach. *SMU Data Science Review*, 2, 2019.
- [144] Sundus Shukar, Fatima Zahoor, Khezar Hayat, Amna Saeed, Ali Hassan Gillani, Sumaira Omer, Shuchen Hu, Zaheer Ud Din Babar, Yu Fang, and Caijun Yang. Drug shortage : Causes, impact, and mitigation strategies. *Frontiers in Pharmacology*, 12, 7 2021.
- [145] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dhharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, and Demis Hassabis. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362 :1140–1144, 12 2018.
- [146] Edward A. Silver. Operations research in inventory management : A review and critique. *Operations Research*, 29 :628–645, 8 1981.
- [147] Edward A. Silver, David F. Pyke, and Douglas J. Thomas. *Inventory and Production Management in Supply Chains*. CRC Press, 4th edition edition, 12 2016.
- [148] Monica C. Skewes, Vivian M. Gonzalez, Julie A. Gameon, Paula FireMoon, Emily Salois, Stacy M. Rasmus, Jordan P. Lewis, Scott A. Gardner, Adriann Ricker, and Martel Reum. Health disparities research with american indian communities : The importance of trust and transparency. *American Journal of Community Psychology*, 66 :302–313, 12 2020.
- [149] Robert Soeters and Wilbert Bannenberg. Computerized calculation of essential drug requirements. *Social Science and Medicine*, 27 :955–970, 1988.
- [150] Donald F. Specht. A general regression neural network. *IEEE Transactions on Neural Networks*, 2 :568–576, 1991.
- [151] Henri Sterdyniak, Pierre Villa, and Département analyse et prévision de l’OFCE. Pour une réforme du financement de la sécurité sociale. *Revue de l’OFCE*, 67 :155–205, 1998.
- [152] Milan Stojanović and Dušan Regodić. The significance of the integrated multicriteria abc-xyz method for the inventory management process. *Acta Polytechnica Hungarica*, 14 :29–48, 2017.
- [153] M Supriya and A J Deepa. Machine learning approach on healthcare big data : a review. 2020.
- [154] Ajaya Kumar Swain and Ray Qing Cao. Using sentiment analysis to improve supply chain intelligence. *Information Systems Frontiers*, 21 :469–484, 4 2019.

- [155] Fisaha Tesfay, Sara Javanparast, Lillian Mwanri, and Anna Ziersch. Stigma and discrimination : Barriers to the utilisation of a nutritional program in hiv care services in the tigray region, ethiopia. *BMC Public Health*, 20 :1–13, 6 2020.
- [156] JW Toomey. *Inventory management : principles, concepts and techniques*, volume 12. Springer Science & Business Media, 2000e edition, 2000.
- [157] Vijith Varghese, Manuel Rossetti, Edward Pohl, Server Apras, and Douglas Marek. Applying actual usage inventory management best practice in a health care supply chain. *Int. J Sup. Chain. Mgt*, 1, 2012.
- [158] Aayushi Verma, Aishwarya Karan, Aaroh Mathur, and S. Chethan. Analysis of time-series method for demand forecasting. *Journal of Engineering and Applied Sciences*, 12 :3102–3107, 1 2017.
- [159] Anita R. Vila-Parrish and Julie Simmons Ivy. Managing supply critical to patient care : An introduction to hospital inventory management for pharmaceuticals. *International Series in Operations Research and Management Science*, 184 :447–463, 2013.
- [160] Jonas Volland, Andreas Fügener, Jan Schoenfelder, and Jens O. Brunner. Material logistics in hospitals : A literature review, 2017.
- [161] Shouyi Wang and Wanpracha Art Chaovalitwongse. *Evaluating and Comparing Forecasting Models*. John Wiley & Sons, Inc., 2 2011.
- [162] Alessandro Zananini. Optimal stock sizing in a cutting stock problem with stochastic demands. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10335 LNCS :293–301, 2017.
- [163] Cha Zhang and Yunquian Ma. *Ensemble machine learning : methods and applications*, volume 1. Springer, 2012.
- [164] Yuzhen Zhang, Jingjing Liu, and Wenjuan Shen. A review of ensemble learning algorithms used in remote sensing applications. *Applied Sciences 2022, Vol. 12, Page 8654*, 12 :8654, 8 2022.
- [165] Bo Zhao. Web scraping. *Encyclopedia of Big Data*, pages 1–3, 2017.
- [166] William G Zikmund, Barry J Babin, Jon C Carr, and Mitch Griffin. *Business Research Methods*. Cengage learning, 9 edition, 3 2013.
- [167] Ömer Alkan and Kübra Elmali. The determination of the factors affecting pharmaceutical consumption in respect to provinces in turkey. *Mediterranean Journal of Social Sciences*, 11 2015.

# Troisième partie

## Les annexes

# Annexe A

## Annexe du chapitre 3

	Colonne	Non-Null Count	Dtype
0	ID_SITE_RATTACHE	6218	object
1	ID_REF	6218	object
2	DESIG	6218	object
3	CODE_ATC	6218	object
4	ANNEE	6218	object
5	HOSPI_CODE_UCD	6218	object
6	CODE_CIP	5503	object
7	MOIS	6218	int32
8	CODE_DEPT	6218	object
9	N_ETB	6218	int32
10	MORTALITE	6218	int32
11	POPULATION	6218	int32
12	N_UFS	6218	int32
13	QUANTITE	6218	float64
14	ANA_REA	6218	int32
15	HEMATO	6218	int32
16	CHIR_VD_T_HEP	6218	int32
17	GENY_OBS	6218	int32
18	PHARMA	6218	int32
19	CHIR_GLE_POLYV	6218	int32
20	MED_URG	6218	int32
21	CARDIO	6218	int32
22	ONCO_MED	6218	int32
23	GASTO_ENTERO_HEP	6218	int32
24	MED PHY READA NEURO	6218	int32
25	NEPHR	6218	int32
26	CHIR_ORTHO_TRAUMA	6218	int32
27	OPHTA	6218	int32
28	MED_PEDIA	6218	int32
29	PNEU_ADDIC_ALLER	6218	int32
30	GERIATRIE	6218	int32
31	PATHO_INF_TROP	6218	int32
32	NEURO_CHIR	6218	int32

Continue à la page suivante

Table A.1 – Suite de la table précédente			
	Colonne	Non-Null Count	Dtype
33	LAB_PREL	6218	int32
34	MED_INT_MED_POLY	6218	int32
35	ENDOCRINO	6218	int32
36	S_PALL	6218	int32
37	RHUMATO	6218	int32
38	URO_ANDRO	6218	int32
39	ORL	6218	int32
40	RECHER	6218	int32
41	DERMA_MST	6218	int32
42	CHIR_CTV	6218	int32
43	MED_NUC_IMG	6218	int32
44	CMF_STOMA	6218	int32
45	AUTRES	6218	int32
46	MED_LEGALE_VICT	6218	int32
47	BIOLOGIE_PATHO	6218	int32
48	MED_TRV	6218	int32
49	PSY	6218	int32
50	RADIOTHERAPIE	6218	int32
51	LIT_HC	6218	int32
52	LIT_HP	6218	int32
53	SEJ_MCO	6218	int32
54	SEJ_HAD	6218	int32
55	SEJ_PSY	6218	int32
56	SEJ_SSR	6218	int32
57	SEJ_SLD	6218	int32
58	PNM_ADM	6218	int32
59	PNM_SERV_SOINS	6218	int32
60	PNM_EDU_SOCIO	6218	int32
61	PNM_MED_TEC	6218	int32
62	PNM_TEC_OUVR	6218	int32
63	PERS_ANA_REA	6218	int32
64	PERS BIOLOGIE PATHO	6218	int32
65	PERS_CARDIO	6218	int32
66	PERS_CHIR_CTV	6218	int32
67	PERS CHIR ORTHO TRAUMA	6218	int32
68	PERS_CHIR_VD_T_HEP	6218	int32
69	PERS_CMF_STOMA	6218	int32
70	PERS_DERMA_MST	6218	int32
71	PERS_ENDOCRINO	6218	int32
72	PERS GASTO ENTERO HEP	6218	int32
73	PERS_GENY_OBS	6218	int32
74	PERS_GERIATRIE	6218	int32
75	PERS_HEMATO	6218	int32
76	PERS_LAB_PREL	6218	int32
77	PERS MED INT MED POLY	6218	int32
78	PERS_MED_NUC_IMG	6218	int32
79	PERS_MED_PEDIA	6218	int32

Continue à la page suivante

Table A.1 – Suite de la table précédente			
	Colonne	Non-Null Count	Dtype
80	PERS MED PHY READA NEURO	6218	int32
81	PERS_MED_TRV	6218	int32
82	PERS_MED_URG	6218	int32
83	PERS_NEPHR	6218	int32
84	PERS_NEURO_CHIR	6218	int32
85	PERS_ONCO_MED	6218	int32
86	PERS_OPHTA	6218	int32
87	PERS_ORL	6218	int32
88	PERS_PHARMA	6218	int32
89	PERS_PNEU_ADDIC_ALLER	6218	int32
90	PERS_PSY	6218	int32
91	PERS_RADIOThERAPIE	6218	int32
92	PERS_RHUMATO	6218	int32
93	PERS_URO_ANDRO	6218	int32
94	PERS_CHIR_GLE	6218	int32
95	PERS_MED_GLE	6218	int32
96	PERS_SANT_PUB	6218	int32
97	PN_MEDICAL	6218	int32
98	P_MEDICAL	6218	int32
99	PERIODE	6218	datetime64[ns]

TABLE A.1 – Description des variables de la base de données structurée et exploitable

## Annexe B

### Étude de Terrain : Questionnaire

# Questionnaire

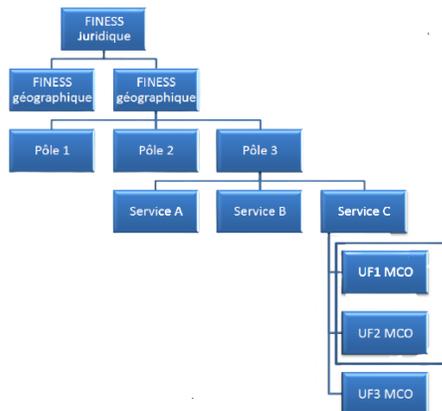
Les objectifs de ce questionnaire tournent autour de trois grands points :

- Le premier point est la structuration de l'hôpital. L'idée est de déterminer au plus fin et au plus précis l'unité qui permet de mieux juger de la consommation des médicaments et produits pharmaceutique en général, c'est-à-dire de déterminer le niveau d'agrégation de la donnée de consommation. Aussi s'associe à cette section la recherche d'informations sur la nomenclature standard des Unités Fonctionnelles (UF) et des services de soins hospitaliers. On désire également avoir l'opinion du spécialiste relativement au regroupement de services selon les types de médicaments consommés ou prises en charges
- La détermination des variables influençant la consommation des médicaments et produits pharmaceutiques en général est recherchée dans le deuxième point. Il s'agit dans cette section d'enquêter sur les critères privilégiés, en pratique, au sein des structures sanitaires et d'obtenir un avis sur les facteurs que nous avons identifié dans la littérature. Ça sera le lieu aussi de comprendre la relation conditionnement et consommation.
- La troisième partie est pour enquêter sur la disponibilité des données comme le nombre de lits, le nombre de personnel par service/UF ... au sein de l'hôpital.

**NB : N'hésitez pas à faire des commentaires.**

## I. Question relative à la structuration de l'hôpital /choix de l'unité décisionnelle

**Q.1 -Confirmez- vous la hiérarchisation des centres hospitaliers ci-dessous ?**



1. Entité Juridique (FINESS Juridique) (non-physique) regroupant,
2. Etablissements (FINESS Géographique) regroupant,
3. Les pôles regroupant,
4. Les services et activités regroupant,
5. Les unités médicales fonctionnelles.

OUI       NON

Si non, veuillez proposer un ordre .....

.....

**Q.2 -Parmi les découpages ci-dessous, selon vous, est-il pertinent d'étudier les besoins en médicaments et dispositifs médicaux selon :**

	OUI	NON
Entité Juridique (FINESS Juridique) .....	<input type="checkbox"/>	<input type="checkbox"/>
Etablissements (FINESS Géographique) .....	<input type="checkbox"/>	<input type="checkbox"/>
Les pôle .....	<input type="checkbox"/>	<input type="checkbox"/>
Les services .....	<input type="checkbox"/>	<input type="checkbox"/>
Les unités médicales fonctionnelles .....	<input type="checkbox"/>	<input type="checkbox"/>
Les groupes d'activités (MCO, SSR, SLD, PSY, HAD).....	<input type="checkbox"/>	<input type="checkbox"/>

---

Les activités .....    
Les spécialités médicales .....

Donnez votre raison : .....  
.....

**Q.3 - Dans nos recherches, nous avons constaté que la nomenclature des services diffère très souvent d'un hôpital à un autre, n'existe-t-il pas une nomenclature standard ?**

OUI  NON

Si oui, comment expliquer cette différence de désignation ?

.....  
.....

Si non, sur quel critère repose la nomenclature des services de soins au sein des hôpitaux ?

.....  
.....  
.....

**Q.4 - Existe-t-il une nomenclature standard ou une référence nationale des unités médicales fonctionnelles hormis celle proposée par l'Agence Technique de l'Information sur l'Hospitalisation (ATIH) dans le cadre du PMSI ?**

OUI  NON

Si oui laquelle ? .....  
.....  
.....

Comment sont définies les unités fonctionnelles ? .....  
.....  
.....

Si non, disposez-vous de base de données des unités fonctionnelle ayant un lien avec la nomenclature SAE de l'ATIH ? .....  
.....  
.....

**Q.5 - En tenant compte de la consommation des médicaments et dispositifs médicaux, est-il pertinent de regrouper les services et activités comme suit ?**

Objectif de cette question est d'avoir l'avis de l'expert sur l'hypothèse de regroupement des services de soins (au cas où il n'y a pas de nomenclature standard)

N°	Services groupés	OUI	NON
1	<b>Anesthésie réanimation chirurgicale</b>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>Anesthésie réanimation</b>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>Anesthésie</b>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>Médecine intensive et réanimation</b>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>Réanimation</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....		
2	<b>Cardiologie</b>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>Maladie vasculaire</b>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>Unité de de soins intensifs de cardiologie</b>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>Rythmologie et simulation cardiaque</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....		
	.....		
3	<b>Chirurgie cardiaque</b>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>Chirurgie thoracique</b>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>Chirurgie vasculaire</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....		
	.....		
	4	<b>Chirurgie digestive</b>	<input type="checkbox"/>
<b>Chirurgie viscérale</b>		<input type="checkbox"/>	<input type="checkbox"/>
	<b>Transplantation hépatique</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....		
	.....		
	5	<b>Chirurgie générale</b>	<input type="checkbox"/>
<b>Chirurgie polyvalente</b>		<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....		
	.....		
6	<b>Chirurgie orthopédique</b>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>Traumatologique</b>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>Chirurgie osseuse adulte</b>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>Chirurgie plastique reconstructrice et esthétique - Brûlé</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non .....		
	.....		
7	<b>Dermatologie, maladies sexuellement transmissibles</b>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>Allergologie et explorations cutanées</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non .....		
	.....		
8	<b>Gériatrie</b>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>Gérontologie</b>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>Oncogériatrie</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....		
	.....		
	9	<b>Médecine du travail</b>	<input type="checkbox"/>
	<b>Pathologies professionnelles et santé au travail</b>		

	<b>Service d'explorations fonctionnelles respiratoires-Médecine du sport et de l'activité physique</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....	<input type="checkbox"/>	<input type="checkbox"/>
10	<b>Rhumatologie</b> <b>Médecine du sport</b> <b>Kinésithérapie</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....	<input type="checkbox"/>	<input type="checkbox"/>
11	<b>Hématologie</b> <b>Hématologie clinique</b> <b>Hématologie biologique</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....	<input type="checkbox"/>	<input type="checkbox"/>
12	<b>Médecine interne et maladies systémiques</b> <b>Médecine polyvalente</b> <b>HAD polyvalente</b> <b>Polyclinique médicale</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....	<input type="checkbox"/>	<input type="checkbox"/>
13	<b>Analyses de laboratoire</b> <b>Laboratoires</b> <b>Centre de prélèvements</b>  <b>Laboratoire de biochimie et biologie moléculaire</b> <b>Microbiologie moléculaire</b> <b>Bactériologie et Hygiène hospitalière</b>  <b>Biologie médicale</b> <b>Bio-banques et Etablissement de sang</b> <b>Hématologie biologique</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....	<input type="checkbox"/>	<input type="checkbox"/>
14	<b>Médecine physique et réadaptation (MPR) neurologique</b> <b>Unité de soins intensifs neuro-vasculaires</b> <b>Neurologie</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....	<input type="checkbox"/>	<input type="checkbox"/>
15	<b>Neurochirurgie et chirurgie de la douleur et du rachis</b> <b>Neurochirurgie</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non .....	<input type="checkbox"/>	<input type="checkbox"/>
16	<b>Oncologie médicale</b> <b>Gynécologie-obstétrique</b> <b>Oncologie médicale polydisciplinaire</b> <b>Orthogénie</b> <b>Maternité</b> <b>Planification familiale – ivg</b>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
	Justification si non : .....	<input type="checkbox"/>	<input type="checkbox"/>

	Justification si non : .....	<input type="checkbox"/>	<input type="checkbox"/>
17	<b>ORL, chirurgie cervico-faciale</b> <b>Chirurgie Maxillo-faciale et stomatologie</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....	<input type="checkbox"/>	<input type="checkbox"/>
18	<b>Radiologie</b> <b>Imagerie</b> <b>Neuroradiologie</b> <b>Médecine nucléaire et traitement de l'image</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....	<input type="checkbox"/>	<input type="checkbox"/>
19	<b>Médecine pédiatrique</b> <b>Pédiatrie néonatale et réanimations</b> <b>Soins intensifs de néonatalogie et médecine néonatale / réanimation et surveillance continue de pédiatrie</b> <b>Chirurgie infantile</b> <b>Puériculture</b> <b>Néonatalogie</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....	<input type="checkbox"/>	<input type="checkbox"/>
20	<b>Pneumologie et réanimation respiratoire</b> <b>Pneumologie et addictologie</b> <b>Pneumologie, oncologie thoracique et allergologie respiratoire</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....	<input type="checkbox"/>	<input type="checkbox"/>
21	<b>Psychiatrie adulte</b> <b>Autisme</b> <b>Addiction</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....	<input type="checkbox"/>	<input type="checkbox"/>
22	<b>Anatomie et cytologie pathologiques</b> <b>Biochimie et d'Hormonologie</b> <b>Biologie de la reproduction</b> <b>Biologie pathologie</b> <b>Médecine et biologie de la reproduction et cytogénétique</b> <b>Génétique</b> <b>Génétique Médicale, Maladies rares et médecine personnalisée</b> <b>Parasitologie - Mycologie</b> <b>Pathologie cellulaire et tissulaire</b>	<input type="checkbox"/>	<input type="checkbox"/>
	Justification si non : .....	<input type="checkbox"/>	<input type="checkbox"/>

**Q.6 - Dans la chaîne logistique hospitalière, quand pouvons-nous dire qu'un produit est consommé ?**

Disponibilisé dans le service ou l'unité fonctionnelle .....

Consommation réelle par le patient .....

Autre : .....

## II. Quels sont les facteurs qui rentrent en compte dans la quantification des besoins de l'hôpital ?

**Q.7 - Parmi les critères ci-dessous, quels sont ceux que vous jugez pertinents pour le choix et la prédiction des médicaments :**

	OUI	NON
<b>Critère relatif à l'établissement :</b>		
1. Position géographique de l'hôpital .....	<input type="checkbox"/>	<input type="checkbox"/>
2. Le type d'établissement .....	<input type="checkbox"/>	<input type="checkbox"/>
3. Le nombre de personnel médical (médecins, pharmaciens ...)	<input type="checkbox"/>	<input type="checkbox"/>
4. Le nombre de personnel non-médical (infirmiers, ...)	<input type="checkbox"/>	<input type="checkbox"/>
5. Nombre de séjours de patients/ou nombre de patients.....	<input type="checkbox"/>	<input type="checkbox"/>
6. Le nombre et types de services et activités hospitalières (ou unités médicales fonctionnelles) .....	<input type="checkbox"/>	<input type="checkbox"/>
7. Le nombre de lits et places de l'hôpital/UF .....	<input type="checkbox"/>	<input type="checkbox"/>
8. Délais de livraison et de péremption des produits.....	<input type="checkbox"/>	<input type="checkbox"/>
9. La saisonnalité .....	<input type="checkbox"/>	<input type="checkbox"/>
10. Capacité de stockage .....	<input type="checkbox"/>	<input type="checkbox"/>
<b>Critère socio-démographique / socio-économique :</b>		
11. Nombre de Population locale (département/région) .....	<input type="checkbox"/>	<input type="checkbox"/>
12. Mortalité de la localité /hôpital .....	<input type="checkbox"/>	<input type="checkbox"/>
13. Genre (sexe) .....	<input type="checkbox"/>	<input type="checkbox"/>
14. Age ou Catégorie d'âge .....	<input type="checkbox"/>	<input type="checkbox"/>
<b>Autres (Ajouter si vous avez d'autres critères) :</b> .....		
.....		
.....		
.....		

**Q.8 – Dans vos pratiques actuelles, quelles sont les variables (indicateurs) prises en compte pour appréhender les besoins pharmaceutiques ?**

.....  
.....  
.....  
.....

**Q.9 - Pour l'identification de la molécule en vue de la prise en charge des patients, dans la pratique, quelle est la sérialisation la plus déterminante ?**

	OUI	NON
Code CIP .....	<input type="checkbox"/>	<input type="checkbox"/>

Code UCD .....

Code ATC .....

**III. Question relative à la disponibilité des données au sein de l'hôpital/ au niveau national**

**Q.10- Le personnel médical et non-médical, exercent-ils dans plusieurs unités fonctionnelles au sein de la même structure ?**

OUI  NON

Si non, disposez-vous de données mettant en lien nombre de personnel et les UFs /services

OUI  NON

Si oui, pouvons-nous disposer de ces données ? .....  
.....  
.....

Ces données, sont-elles open source pour tous les hôpitaux publics de France et comment en disposer ?

.....  
.....  
.....

**Q.11- Le séjour des patients étant un indicateur de consommation des produits pharmaceutiques, disposez-vous de données liant le nombre de séjours par unité fonctionnelle ou service ?**

OUI  NON

Si oui, pouvons-nous disposer de ces données ?

.....  
.....

Ces Données, sont-elles open source pour tous les hôpitaux publics de France et comment en disposer ?

.....  
.....

**Q.12 – Dans votre système de gestion est-il possible de faire le lien entre produits consommés et nombres de patients bénéficiaires par unité fonctionnelle ou service ?**

OUI  NON

Si oui, pouvons-nous disposer de ces données ?

.....  
.....

Ces Données, sont-elles open source pour tous les hôpitaux publics de France et comment en disposer ? .....

---

.....  
.....

**Q.13 - Les unités fonctionnelles étant un regroupement d'équipement médico-technique dont les lits, disposez-vous de données liant le nombre de lits par unité fonctionnelle ou service ?**

OUI       NON

Si oui, pouvons-nous disposer de ces données ? .....

.....  
.....

Ces données, sont-elles "open source" pour tous les hôpitaux publics de France ? .....

.....  
.....

Si oui, comment en disposer et sur quelle plateforme ? .....

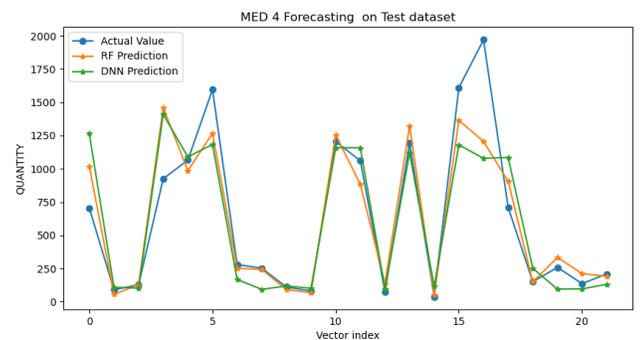
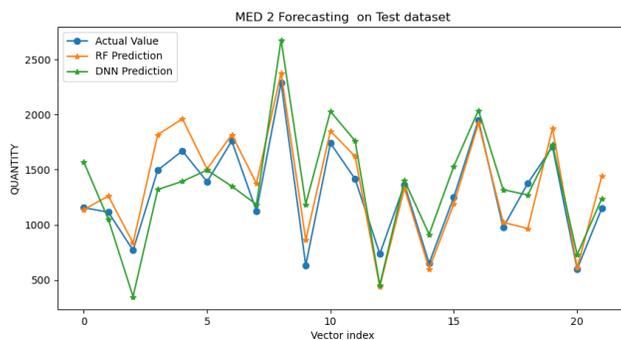
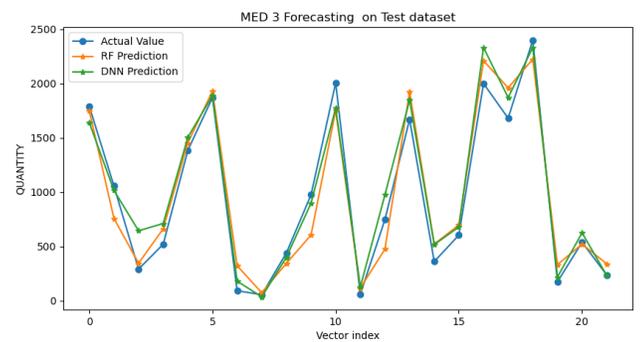
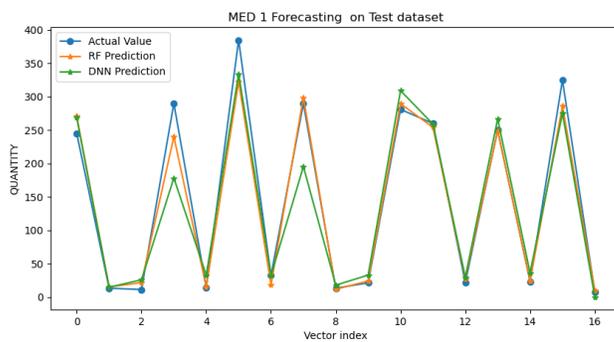
.....

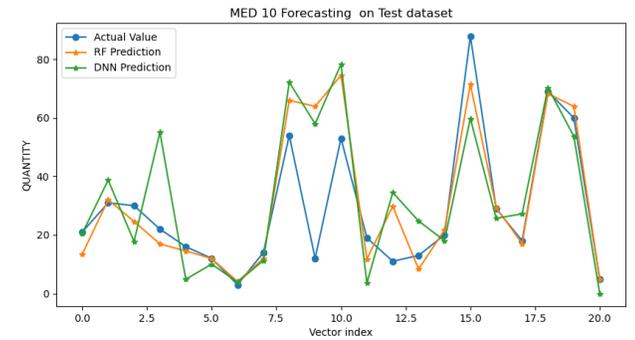
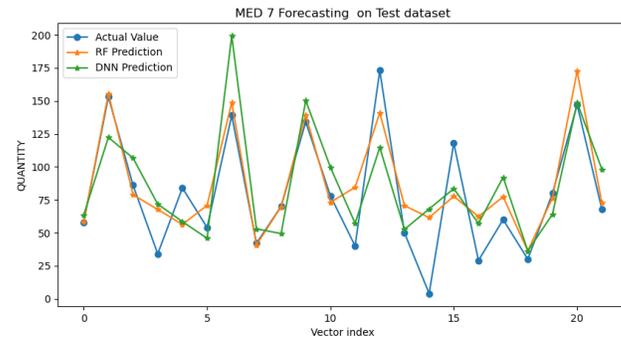
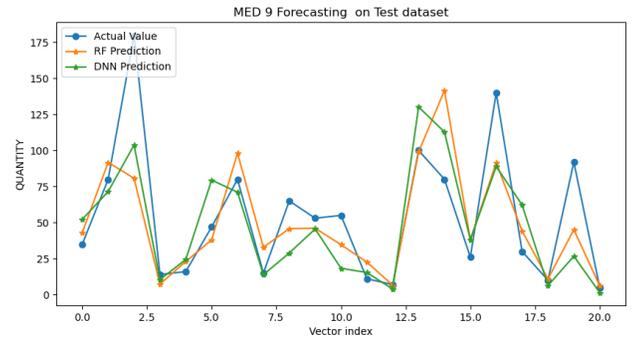
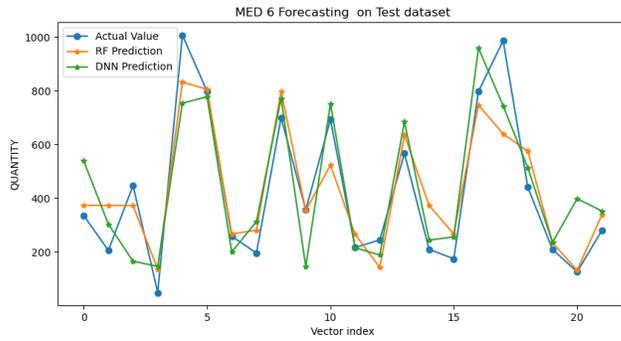
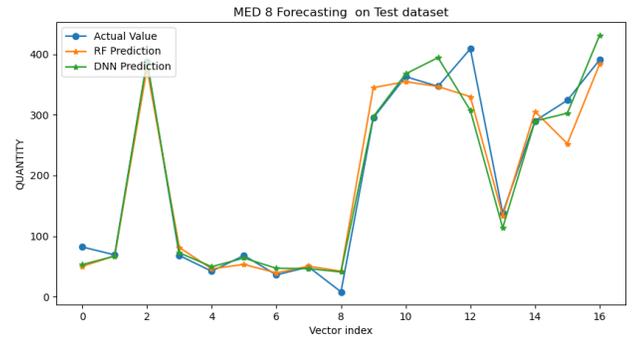
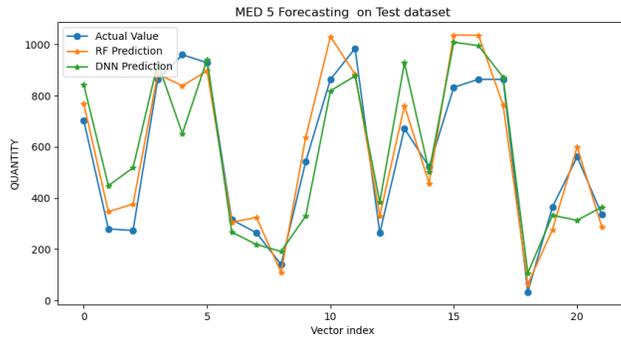
**Fin merci pour votre participation.**

# Annexe C

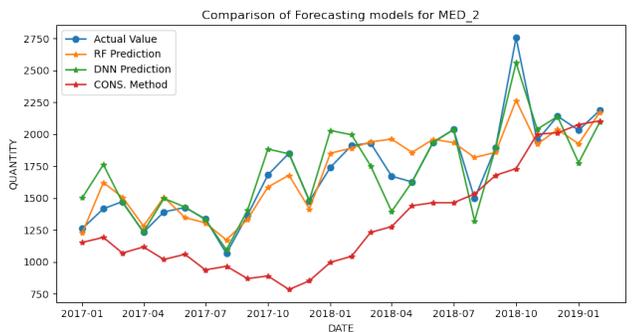
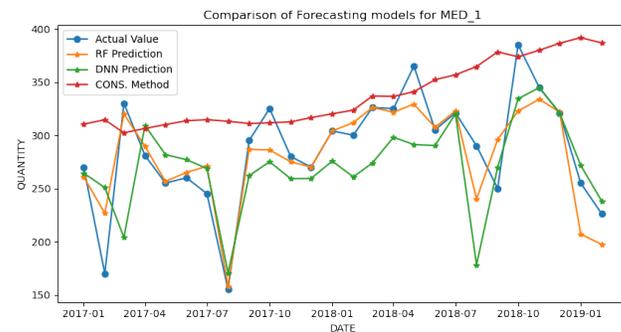
## Annexe du chapitre 5

### C.1 Les performances des modèles RF et DNN avec l'ensemble de données de test

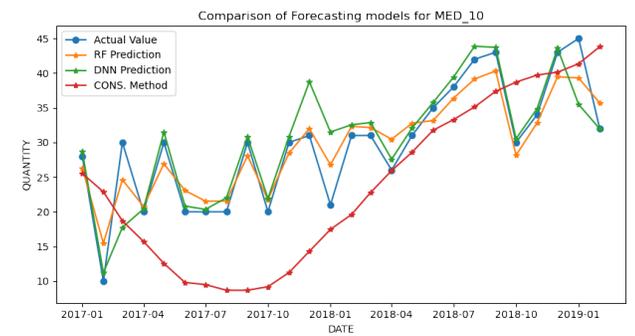
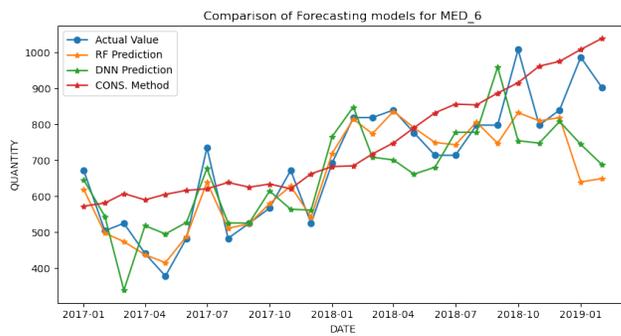
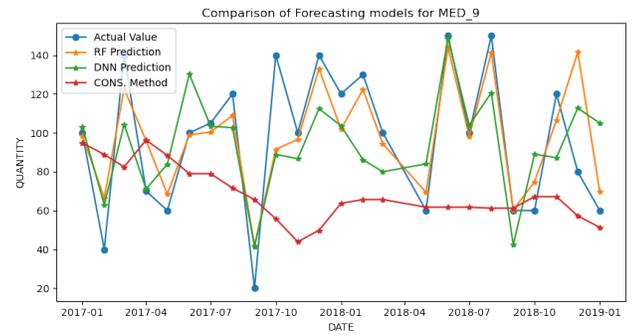
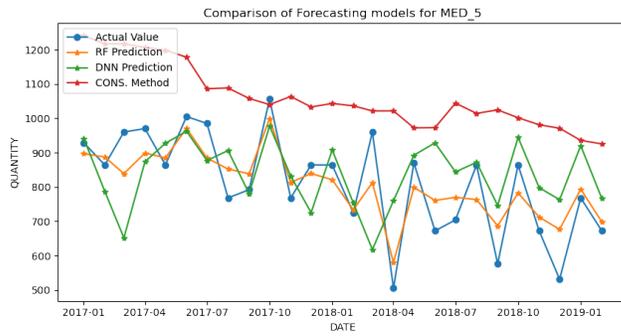
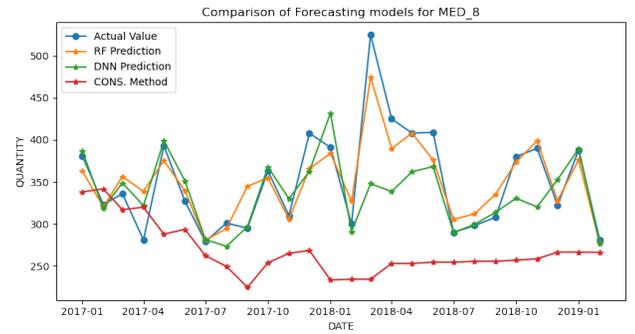
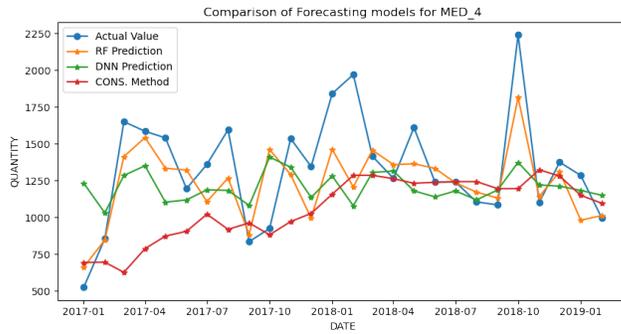
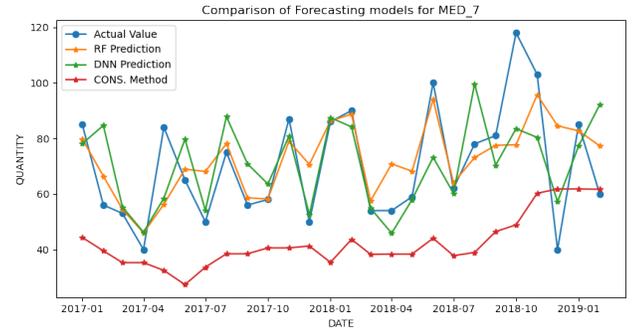
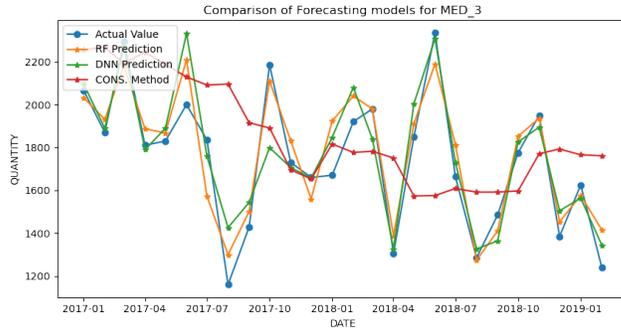




## C.2 Comparaison des modèles de Random Forest, du DNN et de la méthode de consommation : Exemple avec le jeu de donnée de l'hôpital H3



## C.2. COMPARAISON DES MODÈLES DE RANDOM FOREST, DU DNN ET DE LA MÉTHODE DE CONSOMMATION : EXEMPLE AVEC LE JEU DE DONNÉE DE L'HÔPITAL H3



# Annexe D

## Annexe du chapitre 6

### D.1 Résultats suite aux tests des modèles construits

Med	MLR	RF	SVR	MLP	GRNN	Hybrid_V1	Hybrid_V2
1	0.62	0.65	0.85	0.79	-0.07	0.80	<b>0.89</b>
2	0.90	<b>0.91</b>	<b>0.91</b>	0.88	0.89	0.90	<b>0.91</b>
3	0.52	0.43	<b>0.60</b>	-0.04	0.09	0.53	0.51
4	<b>0.81</b>	0.69	0.72	0.66	0.28	0.78	0.73
5	<b>0.64</b>	0.60	0.57	0.56	0.50	0.53	0.48
6	0.04	0.14	-0.02	-0.29	-0.15	0.15	<b>0.18</b>
7	0.84	0.83	<b>0.89</b>	0.84	0.84	0.83	0.87
8	0.64	0.45	<b>0.73</b>	0.68	0.58	<b>0.73</b>	0.63
9	<b>0.77</b>	0.74	0.69	0.59	0.01	0.69	0.74
10	0.73	<b>0.84</b>	0.78	0.71	0.74	0.78	0.80
11	-0.48	-0.39	-0.02	-0.17	-0.02	<b>0.09</b>	0.02
12	0.00	0.26	-0.09	-0.19	-0.01	<b>0.36</b>	0.35
13	0.86	0.90	0.79	0.81	0.78	0.88	<b>0.93</b>
14	0.07	-0.00	<b>0.38</b>	0.31	0.30	0.00	0.26
15	-0.30	0.32	-0.01	-0.61	-0.11	0.35	<b>0.54</b>
16	<b>0.27</b>	0.14	-0.04	-0.07	-0.09	0.12	0.01
17	0.51	0.50	0.41	-0.64	0.03	0.55	<b>0.61</b>
18	0.33	0.27	0.40	0.30	-0.18	<b>0.48</b>	0.42
19	0.18	0.57	<b>0.67</b>	-0.03	-0.17	0.42	0.48
20	0.19	0.47	<b>0.49</b>	0.40	0.23	0.27	0.41
21	0.46	<b>0.57</b>	0.55	0.38	0.34	0.44	0.48
22	<b>0.38</b>	0.03	-0.03	0.15	-0.01	-0.01	0.19
23	-0.09	-0.15	-0.08	-0.27	0.02	<b>0.30</b>	0.06
24	0.71	0.52	0.66	<b>0.77</b>	-0.06	0.55	0.58
25	0.89	<b>0.98</b>	0.97	0.83	0.94	0.97	0.97
26	0.67	0.67	0.70	0.74	0.66	0.76	<b>0.77</b>
27	0.50	<b>0.74</b>	0.56	0.59	0.17	0.66	0.72
28	<b>0.78</b>	0.14	0.68	0.75	0.02	0.66	0.72
29	<b>0.31</b>	0.05	0.02	0.04	0.09	-0.07	-0.17
30	0.81	0.82	<b>0.87</b>	0.74	0.55	0.66	0.71

Continue à la page suivante

Table D.1 – Suite de la table précédente							
Med	MLR	RF	SVR	MLP	GRNN	Hybrid_V1	Hybrid_V2
31	<b>0.83</b>	0.73	0.78	0.72	0.67	0.76	0.80
32	0.09	<b>0.39</b>	0.24	0.19	0.22	-0.00	0.12
33	0.51	0.47	0.44	0.53	0.48	<b>0.57</b>	0.39
34	<b>0.66</b>	0.61	0.51	<b>0.66</b>	0.27	0.64	0.58
35	0.40	0.54	<b>0.60</b>	0.54	0.57	<b>0.60</b>	0.56
36	0.74	0.67	0.75	0.68	0.72	0.72	<b>0.76</b>
37	0.39	0.55	0.52	-0.07	0.22	<b>0.65</b>	0.51
38	<b>0.63</b>	<b>0.63</b>	0.61	0.36	0.08	0.52	0.56
39	0.54	<b>0.69</b>	-1.47	0.09	-0.00	0.46	0.60
40	0.39	<b>0.69</b>	0.56	0.08	0.10	0.50	0.45
41	0.71	<b>0.83</b>	0.79	0.71	0.74	0.81	0.80
42	<b>0.78</b>	<b>0.78</b>	0.76	-0.37	0.75	0.75	<b>0.78</b>
43	0.54	0.39	0.58	0.49	0.53	0.54	<b>0.61</b>
44	-2.17	-0.08	-0.95	<b>0.29</b>	0.27	-0.18	0.01
45	-0.71	-0.02	<b>0.19</b>	-0.24	0.14	0.05	0.01

TABLE D.1 – Scores selon  $R^2$  (Données test)

Med	MLR	RF	SVR	MLP	GRNN	Hybrid_V1	Hybrid_V2
1	439.16	421.67	274.12	330.08	740.09	323.22	<b>237.32</b>
2	43.74	42.03	<b>41.50</b>	47.62	47.05	43.26	42.69
3	578.54	631.01	<b>527.62</b>	849.84	797.07	571.00	584.05
4	<b>264.05</b>	334.40	316.54	346.99	506.61	283.17	310.50
5	<b>520.99</b>	545.44	566.68	573.13	609.80	596.11	624.74
6	471.47	445.77	483.76	546.14	515.38	441.55	<b>435.56</b>
7	275.42	277.69	<b>223.40</b>	276.44	271.12	284.25	248.23
8	186.90	231.77	163.13	176.98	201.51	<b>162.87</b>	189.42
9	<b>125.40</b>	133.84	143.95	166.84	259.51	145.60	131.97
10	268.96	<b>206.29</b>	244.76	279.83	262.68	242.18	232.57
11	320.12	310.00	265.23	284.97	265.58	<b>251.23</b>	260.50
12	371.03	319.43	388.47	405.76	373.25	<b>297.64</b>	299.07
13	108.42	92.39	134.28	127.20	137.48	100.65	<b>77.02</b>
14	197.82	205.73	<b>161.77</b>	170.88	171.78	205.05	176.77
15	134.10	97.19	117.85	148.93	123.76	94.86	<b>79.82</b>
16	<b>152.16</b>	164.97	181.05	183.47	185.72	166.95	176.37
17	85.61	86.71	94.10	157.24	121.12	81.97	<b>76.74</b>
18	103.25	108.12	98.00	105.40	136.99	<b>90.93</b>	96.37
19	85.26	61.49	<b>54.30</b>	95.73	101.78	71.95	67.65
20	46.92	38.20	<b>37.24</b>	40.48	45.93	44.79	40.28
21	108.96	<b>97.46</b>	99.17	117.42	120.86	110.77	107.62
22	<b>71.99</b>	89.86	92.93	84.54	91.78	91.74	82.33
23	120.97	123.90	120.29	130.52	114.39	<b>96.73</b>	111.92
24	48.22	61.39	51.54	<b>42.91</b>	91.43	59.35	57.51
25	50.59	<b>22.20</b>	27.04	62.77	38.55	24.33	24.38

Continue à la page suivante

Table D.2 – Suite de la table précédente							
Med	MLR	RF	SVR	MLP	GRNN	Hybrid_V1	Hybrid_V2
26	81.07	80.97	77.24	72.49	82.23	69.98	<b>68.58</b>
27	91.08	<b>65.28</b>	85.52	82.09	117.04	74.48	68.30
28	<b>61.74</b>	121.40	74.48	65.83	129.85	76.03	68.98
29	<b>52.13</b>	60.95	61.84	61.17	59.70	64.84	67.57
30	62.83	60.34	<b>51.55</b>	73.13	96.31	83.07	76.61
31	<b>39.41</b>	49.14	45.15	50.42	54.84	46.47	42.75
32	78.22	<b>64.09</b>	71.50	73.47	72.36	81.83	76.74
33	28.69	29.96	30.76	28.06	29.70	<b>26.95</b>	32.13
34	<b>43.99</b>	46.99	52.87	44.09	64.70	45.49	48.72
35	76.52	66.62	<b>62.05</b>	66.54	64.49	62.62	65.32
36	52.09	58.98	51.25	58.32	54.08	54.06	<b>50.39</b>
37	24.32	20.83	21.54	32.30	27.66	<b>18.57</b>	21.86
38	15.09	<b>15.03</b>	15.40	19.71	23.69	17.08	16.31
39	13.68	<b>11.31</b>	31.75	19.25	20.21	14.89	12.75
40	19.86	<b>14.16</b>	16.84	24.48	24.19	18.00	18.95
41	8.88	<b>6.81</b>	7.66	8.94	8.45	7.27	7.44
42	4.87	<b>4.78</b>	5.08	12.04	5.16	5.11	<b>4.78</b>
43	3.58	4.15	3.44	3.77	3.65	3.59	<b>3.31</b>
44	3.00	1.75	2.35	<b>1.42</b>	1.44	1.83	1.68
45	8.60	6.65	5.94	7.31	<b>6.09</b>	6.40	6.55

TABLE D.2 – Scores selon RMSE (donnée test)

Med	MLR	RF	SVR	MLP	GRNN	Hybrid_V1	Hybrid_V2
1	316.67	332.07	212.13	227.68	584.72	263.30	<b>192.31</b>
2	34.50	<b>24.25</b>	30.46	39.69	35.49	25.94	27.28
3	506.49	474.25	<b>420.96</b>	569.49	541.18	461.05	510.39
4	<b>203.58</b>	239.98	245.95	278.10	426.15	206.84	242.28
5	277.08	<b>252.94</b>	275.02	281.90	283.69	267.27	272.28
6	387.14	<b>366.91</b>	385.30	443.73	395.11	391.47	388.59
7	220.41	222.53	<b>168.39</b>	216.37	220.22	234.75	187.31
8	119.47	175.90	114.98	135.13	159.63	<b>112.20</b>	134.34
9	97.18	97.43	115.44	144.24	210.71	116.81	<b>90.28</b>
10	198.85	<b>131.29</b>	160.63	225.71	196.21	153.39	154.15
11	250.14	180.58	185.26	199.91	202.36	<b>169.11</b>	184.53
12	248.73	253.53	236.27	260.26	244.86	<b>213.51</b>	220.51
13	88.59	70.37	108.33	101.62	116.82	72.56	<b>59.01</b>
14	137.07	116.26	120.23	130.08	121.82	129.56	<b>109.43</b>
15	103.52	74.00	98.38	133.34	105.10	75.08	<b>66.18</b>
16	<b>123.40</b>	143.58	150.42	153.10	152.54	134.68	144.71
17	66.50	71.23	72.36	116.87	82.47	67.84	<b>64.18</b>
18	79.29	88.73	<b>71.89</b>	86.26	98.84	72.49	74.55
19	68.34	49.74	43.72	72.09	76.42	61.31	<b>53.60</b>
20	32.95	32.30	<b>27.23</b>	35.34	35.85	36.93	33.56

Continue à la page suivante

Table D.3 – Suite de la table précédente							
Med	MLR	RF	SVR	MLP	GRNN	Hybrid_V1	Hybrid_V2
21	87.13	83.20	<b>78.78</b>	88.37	89.98	92.20	89.03
22	<b>60.68</b>	69.75	75.91	67.70	76.15	70.03	67.11
23	94.44	96.72	95.79	107.89	90.42	<b>73.49</b>	81.91
24	40.50	45.25	41.67	<b>37.60</b>	75.88	43.66	46.37
25	37.28	<b>16.87</b>	17.82	58.67	30.88	19.38	18.79
26	67.04	62.93	59.16	55.68	60.76	54.13	<b>51.40</b>
27	72.71	<b>53.73</b>	65.15	64.47	94.84	58.57	55.92
28	<b>46.19</b>	76.72	56.53	51.89	107.44	49.04	49.18
29	<b>44.94</b>	52.40	53.63	51.84	48.96	49.97	54.70
30	52.61	48.13	<b>41.90</b>	59.67	75.79	67.20	59.65
31	31.50	36.23	36.03	35.55	41.96	34.35	<b>31.26</b>
32	63.48	<b>50.60</b>	55.70	58.97	60.64	65.18	59.42
33	22.60	21.26	21.97	21.70	23.10	<b>20.06</b>	22.29
34	<b>32.76</b>	40.63	42.42	35.21	59.98	33.12	38.54
35	54.70	42.88	41.47	40.88	<b>35.79</b>	37.95	38.02
36	39.38	41.52	<b>34.83</b>	36.34	36.61	36.82	35.59
37	16.98	15.97	14.29	26.03	22.79	<b>12.86</b>	15.64
38	10.64	<b>9.83</b>	11.60	13.17	20.19	10.90	10.11
39	9.70	<b>6.45</b>	17.83	13.09	15.19	10.78	8.32
40	15.66	<b>11.18</b>	12.86	19.39	21.23	14.73	14.93
41	6.71	4.35	5.73	7.84	7.11	5.00	<b>4.75</b>
42	3.91	<b>3.52</b>	4.32	6.93	3.76	3.93	3.63
43	2.91	2.85	2.52	2.56	2.52	2.56	<b>2.42</b>
44	2.25	1.39	1.78	<b>1.21</b>	1.29	1.53	1.29
45	7.41	5.10	<b>4.00</b>	5.17	4.40	4.97	4.80

TABLE D.3 – Scores selon MAE (donnée test)

Med	MLR	RF	SVR	MLP	GRNN	Hybrid_V1	Hybrid_V2
1	380.30	234.25	49.42	<b>18.83</b>	584.62	19.5	19.0
2	98.85	28.01	46.32	103.04	66.84	<b>21.0</b>	<b>21.0</b>
3	<b>31.00</b>	60.67	51.38	77.62	73.65	57.0	53.0
4	41.64	64.24	<b>41.39</b>	54.85	117.84	62.0	60.0
5	<b>23.20</b>	51.92	69.64	61.33	72.30	58.0	45.0
6	73.62	77.11	117.41	131.70	123.01	79.0	<b>71.0</b>
7	51.48	39.14	42.39	44.55	95.79	38.0	<b>30.0</b>
8	15.33	19.68	17.22	24.08	28.22	<b>12.0</b>	15.0
9	<b>41.29</b>	113.43	71.27	143.85	414.46	70.0	81.0
10	110.45	<b>32.28</b>	43.43	194.78	130.66	47.0	47.0
11	90.92	93.99	<b>54.72</b>	59.70	62.64	64.0	64.0
12	61.60	62.74	61.54	73.29	74.68	<b>53.0</b>	58.0
13	22.21	14.83	20.90	19.44	23.45	14.0	<b>12.0</b>
14	70.49	66.79	51.65	<b>49.27</b>	53.84	67.0	57.0
15	38.10	25.71	43.07	57.24	49.66	29.0	<b>24.0</b>

Continue à la page suivante

Table D.4 – Suite de la table précédente							
Med	MLR	RF	SVR	MLP	GRNN	Hybrid_V1	Hybrid_V2
16	<b>76.04</b>	103.27	141.21	108.59	149.07	120.0	126.0
17	24.81	36.10	24.63	118.54	96.63	<b>23.0</b>	27.0
18	28.03	28.20	21.97	26.89	30.12	<b>22.0</b>	23.0
19	45.92	34.46	<b>23.09</b>	35.14	80.31	37.0	34.0
20	220.14	<b>145.32</b>	146.54	166.33	239.93	159.0	178.0
21	36.32	38.10	<b>32.09</b>	32.19	32.55	40.0	39.0
22	<b>34.82</b>	43.94	44.03	40.95	47.00	46.0	41.0
23	24.25	24.96	24.97	26.78	22.87	<b>19.0</b>	21.0
24	158.02	207.82	200.55	<b>39.37</b>	222.25	189.0	223.0
25	70.78	47.99	<b>39.91</b>	167.27	73.32	54.0	50.0
26	48.07	42.34	46.76	37.74	45.88	36.0	<b>33.0</b>
27	90.86	<b>44.20</b>	65.98	80.45	163.59	79.0	79.0
28	<b>17.82</b>	26.29	24.29	26.56	72.91	21.0	19.0
29	86.17	107.26	<b>64.93</b>	121.22	114.70	112.0	110.0
30	47.55	<b>33.71</b>	44.40	51.76	87.90	42.0	39.0
31	38.09	31.12	36.96	30.04	40.85	29.0	<b>26.0</b>
32	43.59	<b>35.59</b>	38.70	42.90	44.33	45.0	40.0
33	41.96	38.50	<b>31.18</b>	44.59	47.06	39.0	38.0
34	56.61	73.86	63.60	51.27	134.14	<b>49.0</b>	59.0
35	127.75	63.42	80.26	60.58	<b>45.59</b>	49.0	47.0
36	59.63	55.10	43.29	50.29	58.59	47.0	<b>41.0</b>
37	103.79	80.89	85.51	93.81	107.86	<b>79.0</b>	88.0
38	54.25	<b>49.57</b>	60.31	65.83	92.78	53.0	51.0
39	70.60	<b>33.43</b>	106.74	97.59	123.63	71.0	51.0
40	47.00	<b>30.51</b>	37.94	66.92	81.70	42.0	43.0
41	48.23	23.39	54.37	68.11	72.88	23.0	<b>21.0</b>
42	68.59	<b>42.76</b>	75.79	46.72	44.33	51.0	51.0
43	46.26	43.04	39.88	46.95	45.17	<b>37.0</b>	40.0
44	82.40	56.96	60.77	49.87	69.95	56.0	<b>41.0</b>
45	150.73	113.64	110.18	134.16	112.68	<b>109.0</b>	119.0

TABLE D.4 – Scores selon MAPE (donnée test)

Med	$\geq 90\%$	Entre 90% et 80%	Entre 80% et 70%	Entre 70% et 60%	Entre 60% et 50%	$\leq 50\%$
1		SVR, Hybrid_1, Hybrid_2	MLP	MLR, RF		GRNN
2	MLR, RF, SVR, Hybrid_1, Hybrid_2	MLP, GRNN				

Continue à la page suivante

Table D.5 – Suite de la table précédente						
Med	≥ 90%	Entre 90% et 80%	Entre 80% et 70%	Entre 70% et 60%	Entre 60% et 50%	≤ 50%
3				SVR	MLR, Hybrid_1, Hybrid_2	RF, MLP, GRNN
4		MLR	SVR, Hybrid_1, Hybrid_2	RF, MLP		GRNN
5				MLR, RF	SVR, MLP, GRNN, Hybrid_1	Hybrid_2
6						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2
7		MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2				
8			SVR, Hybrid_1	MLR, MLP, Hybrid_2	GRNN	RF
9			MLR, RF, Hybrid_2	SVR, Hybrid_1	MLP	GRNN
10		RF, Hybrid_2	MLR, SVR, MLP, GRNN, Hybrid_1			
11						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2
12						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2
13	RF, Hybrid_2	MLR, MLP, Hybrid_1	SVR, GRNN			
14						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2
15					Hybrid_2	MLR, RF, SVR, MLP, GRNN, Hybrid_1
16						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2

Continue à la page suivante

Table D.5 – Suite de la table précédente						
Med	≥ 90%	Entre 90% et 80%	Entre 80% et 70%	Entre 70% et 60%	Entre 60% et 50%	≤ 50%
17				Hybrid_2	MLR, RF, Hybrid_1	SVR, MLP, GRNN
18						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hy- brid_2
19				SVR	RF	MLR, MLP, GRNN, Hy- brid_1, Hy- brid_2
20						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hy- brid_2
21					RF, SVR	MLR, MLP, GRNN, Hy- brid_1, Hy- brid_2
22						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hy- brid_2
23						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hy- brid_2
24			MLR, MLP	SVR	RF, Hy- brid_1, Hybrid_2	GRNN
25	RF, SVR, GRNN, Hy- brid_1, Hybrid_2	MLR, MLP				
26			SVR, MLP, Hybrid_1, Hybrid_2	MLR, RF, GRNN		
27			RF, Hy- brid_2	Hybrid_1	MLR, SVR, MLP	GRNN
28			MLR, MLP, Hybrid_2	SVR, Hy- brid_1		RF, GRNN
29						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hy- brid_2
30		MLR, RF, SVR	MLP, Hy- brid_2	Hybrid_1	GRNN	

Continue à la page suivante

Table D.5 – Suite de la table précédente						
Med	$\geq 90\%$	Entre 90% et 80%	Entre 80% et 70%	Entre 70% et 60%	Entre 60% et 50%	$\leq 50\%$
31		MLR, <b>Hybrid_2</b>	RF, SVR, MLP, Hybrid_1	GRNN		
32						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2
33					MLR, MLP, Hybrid_1	RF, SVR, GRNN, Hybrid_2
34				MLR, RF, MLP, Hybrid_1	SVR, Hybrid_2	GRNN
35				SVR, Hybrid_1	RF, MLP, GRNN, Hybrid_2	MLR
36			MLR, SVR, GRNN, <b>Hybrid_1</b> , <b>Hybrid_2</b>	RF, MLP		
37				Hybrid_1	RF, SVR, Hybrid_2	MLR, MLP, GRNN
38				MLR, RF, SVR	Hybrid_1, Hybrid_2	MLP, GRNN
39				RF, Hybrid_2	MLR	SVR, MLP, GRNN, Hybrid_1
40				RF	SVR, Hybrid_1	MLR, MLP, GRNN, Hybrid_2
41		RF, <b>Hybrid_1</b> , <b>Hybrid_2</b>	MLR, SVR, MLP, GRNN			
42			MLR, RF, SVR, GRNN, <b>Hybrid_1</b> , <b>Hybrid_2</b>			MLP
43				Hybrid_2	MLR, SVR, GRNN, Hybrid_1	RF, MLP
44						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2

Continue à la page suivante

Table D.5 – Suite de la table précédente						
Med	$\geq 90\%$	Entre 90% et 80%	Entre 80% et 70%	Entre 70% et 60%	Entre 60% et 50%	$\leq 50\%$
45						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hy- brid_2

TABLE D.5 – Performance des modèles par médicament selon la métrique  $R^2$ 

Med	$\leq 10$	Entre 10 et 20	Entre 20 et 30	Entre 30 et 40	Entre 40 et 50	$\geq 50$
1		MLP, Hy- brid_1, Hybrid_2			SVR	MLR, RF, GRNN,
2			RF, Hybrid_1, Hybrid_2		SVR	MLR, MLP, GRNN
3				MLR		RF, SVR, MLP, GRNN, Hybrid_1, Hy- brid_2
4					MLR, SVR	RF, MLP, GRNN, Hy- brid_1, Hy- brid_2
5			MLR		Hybrid_2	RF, SVR, MLP, GRNN, Hy- brid_1
6						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hy- brid_2
7			Hybrid_2	RF, Hybrid_1	SVR, MLP	MLR, GRNN
8		MLR, RF, SVR, Hy- brid_1, Hybrid_2	MLP, GRNN			
9					MLR	RF, SVR, MLP, GRNN, Hybrid_1, Hy- brid_2
10				RF	SVR, Hy- brid_1, Hy- brid_2	MLR, MLP, GRNN
11						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hy- brid_2
12						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hy- brid_2
13		RF, MLP, Hybrid_1, Hybrid_2	MLR, SVR, GRNN			

Continue à la page suivante

Table D.6 – Suite de la table précédente						
Med	≤ 10	Entre 10 et 20	Entre 20 et 30	Entre 30 et 40	Entre 40 et 50	≥ 50
14					MLP	MLR, RF, SVR, GRNN, Hybrid_1, Hybrid_2
15			RF, Hybrid_1, Hybrid_2	MLR	SVR, GRNN	MLP
16						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2
17			MLR, SVR, Hybrid_1, Hybrid_2	RF		MLP, GRNN
18			MLR, RF, SVR, MLP, Hybrid_1, Hybrid_2	GRNN		
19			SVR	RF, MLP, Hybrid_1, Hybrid_2	MLR	GRNN
20						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2
21				MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2		
22				MLR	RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2	
23		Hybrid_1	MLR, RF, SVR, MLP, GRNN, Hybrid_2			
24						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2
25				SVR	RF, Hybrid_2	MLR, MLP, GRNN, Hybrid_1
26				MLP, Hybrid_1, Hybrid_2	MLR, RF, SVR, GRNN	
27					RF	MLR, SVR, MLP, GRNN, Hybrid_1, Hybrid_2
28		MLR, Hybrid_2	RF, SVR, MLP, Hybrid_1			GRNN
29						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2
30				RF, Hybrid_2	MLR, SVR, Hybrid_1	MLP, GRNN
31			Hybrid_1, Hybrid_2	MLR, RF, SVR, MLP	GRNN	

Continue à la page suivante

Table D.6 – Suite de la table précédente						
Med	$\leq 10$	Entre 10 et 20	Entre 20 et 30	Entre 30 et 40	Entre 40 et 50	$\geq 50$
32				RF, SVR, Hybrid_2	MLR, MLP, GRNN, Hybrid_1	
33				RF, SVR, Hybrid_1, Hybrid_2	MLR, MLP, GRNN	
34					Hybrid_1	MLR, RF, SVR, MLP, GRNN, Hybrid_2
35					GRNN, Hybrid_1, Hybrid_2	MLR, RF, SVR, MLP
36					SVR, Hybrid_1	MLR, RF, MLP, GRNN, Hybrid_2
37						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2
38					RF	MLR, SVR, MLP, GRNN, Hybrid_1, Hybrid_2
39				RF		MLR, SVR, MLP, GRNN, Hybrid_1, Hybrid_2
40				RF, SVR	MLR, Hybrid_1, Hybrid_2	MLP, GRNN
41			RF, Hybrid_1, Hybrid_2		MLR	SVR, MLP, GRNN
42					RF, MLP, GRNN	MLR, SVR, Hybrid_1, Hybrid_2
43				SVR, Hybrid_1, Hybrid_2	RF, MLP, GRNN	MLR
44					MLP, Hybrid_2	MLR, RF, SVR, GRNN, Hybrid_1
45						MLR, RF, SVR, MLP, GRNN, Hybrid_1, Hybrid_2

TABLE D.6 – Performance des modèles par médicament selon la métrique MAPE

## D.2 Résultats suite aux validations des modèles dans les deux hôpitaux (H4 & H5)

### D.2.1 Pour l'hôpital H4

D.2. RÉSULTATS SUITE AUX VALIDATIONS DES MODÈLES DANS LES DEUX HÔPITAUX (H4 & H5)

Med	MLR		RF		SVR		MLP		GRNN		Hybrid 1		Hybrid 2		
	$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	
1	-156.9	221.62	<b>0.81</b>	11.15	0.77	<b>9.37</b>	-1.37	25.15	-4.69	40.87	-1.4	24.67	0.68	9.74	
2	-	9587.24	0.19	24.58	-27.42	254.97	-60.46	378.13	-7.23	135.59	<b>0.89</b>	<b>17.43</b>	0.71	28.15	
3	-	33265.16	-23.91	76.15	-3.91	31.94	-13.91	59.15	-18.7	68.14	-29.02	83.53	-6.92	42.28	
4	-	2623.86	-22.01	60.89	-18.03	54.98	-18.98	56.97	-15.18	<b>49.74</b>	<b>-15.3</b>	50.4	-19.01	56.13	
5	-	206.76	-6.77	39.07	-6.8	39.1	-5.09	34.36	-6.64	38.5	-6.19	37.44	-10.88	50.36	
6	-	369.99	-0.04	22.85	-0.71	25.83	-0.38	30.06	-0.18	18.71	0.01	17.36	-0.08	21.63	
7	-3.05	35.67	-3.02	34.84	-2.07	30.92	-8.18	65.66	-0.77	27.11	<b>0.73</b>	<b>10.03</b>	0.71	11.56	
8	-2.7	22.57	-2.3	19.58	-3.91	21.94	-2.9	22.35	-4.17	26.1	0.15	<b>11.96</b>	-1.29	17.98	
9	-5.93	45.39	-0.94	17.81	<b>-0.48</b>	15.44	-0.92	<b>14.19</b>	-0.77	18.95	-1.16	25.61	-0.74	24.29	
10	-	3074.07	<b>0.87</b>	<b>25.9</b>	-6.5	135.62	-6.7	151.86	-0.49	69.75	0.35	45.15	0.38	48.63	
11	-	537.66	-4.0	38.18	-4.89	43.19	-4.97	40.29	-3.73	37.21	-4.14	39.34	-4.84	44.33	
12	-	161.95	-1.94	80.7	-0.71	69.38	-1.23	71.48	-0.36	78.31	-1.54	94.63	-1.62	94.1	
13	-	5310.2	-	279.59	-64.82	221.93	-48.26	192.51	-59.33	213.84	-66.35	214.91	-67.24	224.69	
14	-	5143.46	-144.6	68.67	-	61.26	-	70.86	-134.5	66.27	-	58.83	-	69.22	
15	-	4079.29	-8.92	32.8	-2.69	18.68	-8.61	34.24	-0.98	11.62	-2.23	15.32	-2.84	18.89	
16	-	122.39	-0.99	125.5	-1.48	160.46	0.13	75.2	-1.93	174.0	-2.08	175.33	-1.48	156.99	
17	-9.36	87.14	0.3	17.03	0.2	21.98	-0.51	19.67	-0.13	15.92	-0.7	35.0	-1.33	40.66	
18	-78.71	396.96	0.33	30.43	-0.21	43.87	-0.04	29.6	0.04	38.22	-9.18	148.02	-9.72	151.79	
19	-	221.11	<b>0.84</b>	<b>22.6</b>	0.66	24.59	-1.24	44.13	-1.89	84.67	-0.06	51.25	0.03	48.52	
20	-	406.77	-7.66	51.86	-0.71	32.35	-0.03	22.02	-2.66	53.43	<b>0.53</b>	<b>19.67</b>	-0.92	28.97	
21	-	346.16	-6.7	309.57	-1.2	185.29	-30.14	793.61	-56.76	1109.21	-11.58	439.82	-12.1	461.77	
22	-	1725.86	-0.25	26.91	-0.03	24.34	-1.51	40.54	-0.18	24.89	-0.8	33.57	-0.58	31.08	
23	-	1.1e+28	5.7e+15	-	1152.25	-	1339.16	-30.62	293.52	-	1619.99	-	1190.47	-	1125.94
24	-72.67	234.45	-1.24	29.91	-4.23	52.87	-4.09	50.45	-1.68	30.91	-2.44	40.15	-2.17	35.79	
25	-	9769.27	-11.49	47.98	-4.13	28.57	-20.66	70.36	<b>0.22</b>	<b>12.3</b>	-5.46	32.23	-9.93	43.04	
26	-	4485.81	-15.84	24.38	-1.62	10.1	-4.97	15.34	-1.95	11.09	-15.83	28.08	-11.64	24.26	
27	-132.2	3365.01	0.61	118.59	<b>0.68</b>	135.76	0.58	157.82	-0.59	480.49	0.63	132.82	0.37	181.19	
28	-	178.08	-21.1	95.8	-32.06	123.64	-8.52	62.94	-3.45	<b>45.86</b>	-5.37	48.26	-5.52	49.1	
29	-	2968.3	-79.64	946.79	-45.1	713.84	-46.51	753.87	-40.39	727.86	-46.89	616.94	-61.64	885.62	
30	-	1612.19	-36.04	565.59	-83.81	817.08	-58.47	691.18	-	1500.67	-68.17	780.84	-98.6	928.98	
31	-2.23	61.82	-0.59	31.69	-8.67	102.09	-2.05	61.78	-1.89	59.92	0.49	20.25	<b>0.74</b>	<b>13.49</b>	
32	-	156.54	-48.57	470.01	-6.03	100.14	-26.56	358.64	-51.69	479.51	-49.81	517.3	-2.33	134.45	
33	-	137.65	-17.46	57.55	-15.28	54.01	-12.92	49.08	-18.98	60.08	-18.64	59.86	-17.4	57.47	
34	-	402.49	-0.37	125.48	<b>0.73</b>	<b>25.52</b>	0.39	63.73	-4.64	255.78	0.70	53.66	-1.85	137.66	
35		3102.83		22.0		386.03		430.68		164.88		470.3		30.91	
36	-2.54	42.88	<b>0.41</b>	<b>17.79</b>	-2.52	46.18	-1.93	41.65	-0.83	33.62	-0.69	29.64	0.33	19.0	
37	-	1647.56	0.11	10.92	-	170.79	-7.67	37.76	-0.14	11.47	-1.7	22.72	0.1	10.82	
38	-	262.56	-0.58	27.29	-	484.96	-0.17	<b>18.24</b>	-81.08	250.36	-0.25	21.06	-0.33	19.48	
39	-	1.28e+5	-13.28	54.66	-11.45	39.36	-6.6	33.21	-3.41	27.91	-7.35	38.5	-7.33	38.17	
40	-45.76	186.42	-1.05	32.39	-7.76	73.68	-0.31	22.64	-0.34	<b>19.17</b>	-2.86	40.27	-4.15	48.82	
41	-	1.72e+5	-	886.51	-	1003.49	-	736.91	-	949.24	-	1516.24	-	1009.66	
42	-45.44	458.63	-2.41	79.29	-11.67	469.72	-2.89	74.04	-2.72	72.9	-2.95	82.03	-3.04	84.37	
43	-	125.85	-6.26	56.05	-5.82	49.81	-4.42	46.54	-5.7	54.48	-1.23	22.16	-7.24	60.79	
44	-53.27	430.64	-0.32	48.26	-0.05	35.5	-0.43	28.95	-0.18	32.18	-0.79	53.89	-0.84	62.07	
45	-	244.95	-0.33	71.72	-0.22	85.52	-1.24	112.04	-0.38	93.29	-0.58	91.28	-0.57	89.47	

TABLE D.7 – Performance des modèles par médicament selon la métrique  $R^2$  & MAPE dans H4

### D.2.2 Pour l'hôpital H5

Med	MLR		RF		SVR		MLP		GRNN		Hybrid 1		Hybrid 2				
	R <sup>2</sup>	Mape	R <sup>2</sup>	Mape	R <sup>2</sup>	Mape	R <sup>2</sup>	Mape	R <sup>2</sup>	Mape	R <sup>2</sup>	Mape	R <sup>2</sup>	Mape			
1	-76.94	166.6	-3.19	38.88	-18.77	76.86	-2.47	31.41	-7.51	58.14	0.83	11.37	<b>0.85</b>	<b>10.32</b>			
2	-	1.6e+5	-11.53	386.15	-	785.66	-	1608.48	-50.64	511.78	-12.69	353.22	-26.32	534.9			
3	-	3.01e+6	-	798.2	-142.43	889.87	-	446.63	-	620.88	-	377.8	-36.16	345.56			
4	-	458763.78	-	203.68	-289.0	121.57	-	144.41	-	669.65	-43.57	377.8	-36.16	345.56			
5	-	17200.34	-1.84	32.01	-4.49	60.96	<b>0.62</b>	<b>17.89</b>	0.03	22.47	0.07	21.35	-0.48	23.58			
6	-	18641.21	-13.1	129.31	-16.93	163.77	-22.7	186.62	-17.87	167.38	-18.94	165.38	-11.24	130.98			
7	-	165091.75	-2.19	<b>30.27</b>	-	235.19	-7.89	62.13	-15.66	93.43	-2.34	40.52	-27.87	100.68			
8	-	619.23	-3.85	44.85	-	108.01	-	284.44	-	431.87	-16.81	99.99	-0.8	22.04			
9	-23.97	143.55	0.33	17.64	-16.93	109.84	0.13	18.9	0.16	22.22	<b>0.68</b>	<b>15.7</b>	0.08	26.14			
10	-	4920.99	-20.14	141.68	-	388.76	-	625.74	-96.25	310.26	-23.39	168.21	-62.67	268.54			
11	-	32537.41	-4.7	131.02	-84.2	465.34	<b>0.41</b>	<b>48.83</b>	-0.33	46.45	-3.96	122.28	-0.35	62.9			
12	-	17408.14	-7.77	204.52	-0.54	88.37	-1.1	102.92	-1.14	105.29	-1.5	112.55	-1.88	122.07			
13	-	107013.94	-	8864.8	<b>-0.1</b>	<b>28.47</b>	-0.83	37.01	-1.07	34.95	-32.9	164.6	-17.78	123.48			
14	-	482440.05	-	24765.2	-	134.95	414.96	-	296.93	631.0	-	160.81	462.72	-	252.57		
15	-	39485.95	-12.81	250.81	-2.42	128.34	0.14	59.0	-1.11	105.61	-45.62	401.56	-7.82	206.19			
16	-	30239.4	<b>0.54</b>	<b>20.67</b>	-2.98	42.39	-6.8	54.57	-2.1	39.13	-0.18	26.6	-0.04	24.11			
17	-	7094.96	-1.12	158.26	-1.74	186.45	-0.02	101.08	-2.15	199.86	-2.58	211.43	-1.53	179.54			
18	-	5480.95	-6.45	107.55	-	126.79	456.67	0.04	39.95	-0.36	50.37	-30.47	235.28	-16.99	179.02		
19	-	63115.12	-1.63	34.05	-24.69	161.88	-7.55	78.79	-10.23	116.21	-	132.75	380.81	-	120.31		
20	-	890.27	-16.87	676.58	-1.57	85.26	-1.8	82.24	-29.91	894.8	-7.83	442.47	-9.28	479.9			
21	-	7911.36	-4.02	88.47	-1.11	50.25	-97.0	398.74	-32.3	245.01	-5.35	98.45	-13.07	159.65			
22	-	147246.63	<b>0.06</b>	<b>47.21</b>	-	104.96	485.67	-0.93	72.32	-7.02	145.87	-0.92	69.98	-0.44	57.56		
23	-	5.98e+29	-59.9	234.41	-18.04	134.48	-32.03	174.89	-22.62	149.5	-66.13	245.46	-40.72	196.47			
24	-	2629.69	-	1885.81	-	2336.19	-	1409.6	-	2625.45	-	1930.8	-	1889.07			
25	-	20077.01	-0.27	20.52	-8.27	79.01	-3.94	45.44	-0.74	31.99	0.14	<b>20.11</b>	<b>0.26</b>	20.43			
26	-	7693.63	-4.03	41.31	-6.55	54.19	-1.02	<b>29.47</b>	-3.57	40.86	-6.51	53.78	-7.8	59.45			
27	-	16804.66	-0.5	33.69	-0.03	26.29	-0.18	<b>25.25</b>	<b>0.02</b>	26.73	-0.43	29.56	-0.51	28.04			
28	-	2680.18	-2.14	41.89	-0.05	26.03	-5.11	67.75	-2.04	48.21	<b>0.2</b>	<b>23.39</b>	-2.29	49.93			
29	-	46371.21	-0.92	30.5	-	383.23	545.36	-12.06	90.55	<b>-0.28</b>	<b>21.84</b>	-7.65	70.52	-8.68	76.16		
30	-	5838.33	-10.33	364.71	-86.48	1041.53	-16.25	483.25	-16.4	487.6	-17.64	539.25	-16.02	444.38			
31	-	270420.17	-4.37	250.23	-	206.51	1519.44	-	1649.12	-10.09	391.89	-13.52	441.85	-18.65	507.11		
32	-	44377.87	-	2180.92	-	11698.74	6840.28	-	3159.73	-	2171.22	2995.48	-	1832.37	-	1777.28	
33	-	23628.54	-14.5	808.4	-	413.89	4084.49	-87.34	1755.72	-	136.95	2338.12	-26.46	1126.34	-11.16	772.13	
34	-	6724.24	-38.64	976.49	-3.55	362.32	-11.73	592.44	-10.66	567.14	-0.29	200.2	-5.36	462.42			
35	-	311.26	-0.18	42.83	-3.25	103.63	-0.94	48.71	-0.04	50.94	-1.1	50.31	-1.6	55.33			
36	-	7643.75	<b>37.86</b>	<b>37.86</b>	-	112.18	-	71.67	-	82.35	-	61.58	-	85.72			
37	-	30585.88	-2.3	39.43	-0.94	37.09	-1.06	41.26	-0.49	31.08	-0.76	23.87	-1.68	31.09			
38	-	240377.82	-8.48	119.88	-	5277.81	2542.83	-13.96	113.45	-27.9	205.23	-12.82	147.14	-13.94	148.97		
39	-	10558.84	-	1622.81	-	101192.47	10525.84	-11.98	121.78	-	807.84	947.3	-735.5	905.39	-	928.32	
40	-	-32.12	-6.87	118.18	-1.31	68.27	<b>-0.27</b>	<b>39.82</b>	-1.46	64.39	-5.49	114.03	-6.65	120.22			
	-	405.47	-	1047.51	-	3643.8	5580.75	-	603.23	2347.96	-	326.98	1714.31	-21.43	467.74	-9.65	303.18
			-134.26														

Continue à la page suivante

D.2. RÉSULTATS SUITE AUX VALIDATIONS DES MODÈLES DANS LES DEUX HÔPITAUX (H4 & H5)

Table D.8 – Suite de la table précédente														
Med	MLR		RF		SVR		MLP		GRNN		Hybrid 1		Hybrid 2	
	$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	$R^2$	Mape	$R^2$	Mape
41	-	27183.15	-77.65	362.7	-	699.52	-	489.5	-179.2	547.14	-	954.48	-	846.54
	462639.19				295.02		145.32				563.29	434.97		
42	-	15595.25	0.26	<b>26.2</b>	-	8954.45	-1.53	66.7	-2.86	75.65	<b>0.57</b>	29.39	0.42	29.95
	146375.36				48243.91									
43	-	18146.68	-21.49	186.46	-	2093.2	-23.27	192.67	-2.99	77.48	-	496.43	-418.5	714.62
	233690.61				3077.95					166.92				
44	-	11096.63	-18.86	411.94	<b>0.55</b>	55.84	-1.32	148.77	-0.44	117.81	-15.54	362.16	-20.3	421.79
	14966.48													
45	-	4791.06	<b>-0.1</b>	34.89	-0.33	<b>20.82</b>	-0.38	36.12	-0.73	32.13	-0.21	28.93	-0.35	25.92
	14906.0													

TABLE D.8 – Performance des modèles par médicament selon la métrique  $R^2$  & MAPE dans H5