



HAL
open science

Amélioration de la prédiction des commandes des pharmacies auprès de la CERP RRM

Jeremy Renaud

► **To cite this version:**

Jeremy Renaud. Amélioration de la prédiction des commandes des pharmacies auprès de la CERP RRM. Intelligence artificielle [cs.AI]. Université Bourgogne Franche-Comté, 2024. Français. NNT : 2024UBFCD010 . tel-04629908

HAL Id: tel-04629908

<https://theses.hal.science/tel-04629908>

Submitted on 1 Jul 2024

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 DE DOCTORAT
DE L'ÉTABLISSEMENT UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ
PRÉPARÉE À L'UNIVERSITÉ DE FRANCHE-COMTÉ

École doctorale n°37
Sciences Pour l'Ingénieur et Microtechniques

Doctorat d'Informatique

par

JEREMY RENAUD

**Amélioration de la prédiction des commandes des pharmacies auprès de la
CERP RRM**
Improvement of prediction of pharmacy orders with CERP RRM

Thèse présentée et soutenue à Belfort, le 17 juin 2024

Composition du Jury :

PROF VERNIER FLAVIEN	Université Savoie Mont-Blanc	Rapporteur
PROF CHRÉTIEN STEPHANE	Université Lumière Lyon 2	Rapporteur
PROF WEMMERT CÉDRIC	Université de Strasbourg	Président du jury / Examineur
PROF COUTURIER RAPHAËL	Université de Franche-Comté	Directeur de thèse
PROF GUYEUX CHRISTOPHE	Université de Franche-Comté	Codirecteur de thèse
COURJAL BENOIT	CERP Rhin Rhône Méditerranée	Invité

RÉSUMÉ

Amélioration de la prédiction des commandes des pharmacies auprès de la CERP RRM

Jeremy RENAUD
Université de Franche Comté, 2024

Encadrants : Raphaël COUTURIER et Christophe GUYEUX

La CERP Rhin Rhone Méditerranée (CERP RRM) est un grossiste répartiteur qui doit assurer l'approvisionnement de pharmacies. Le secteur du médicament, malgré l'évolution récente de la logistique hospitalière, manque considérablement d'outils d'aide à la décision. Le but de la thèse est de mettre en place un système prédictif pour l'ensemble des clients CERP afin de prévoir les commandes avec la meilleure précision possible. Les données sont essentiellement sous forme de séries temporelles.

Dans un premier temps, les travaux de cette thèse ont été de faire un état de l'art des technologies de prédictions de séries temporelles, ainsi que de l'implémentation de systèmes d'IA dans des secteurs industriels liés aux métiers de grossiste répartiteur. La principale contribution de cette thèse a été d'améliorer les prédictions de la CERP RRM à plusieurs niveaux grâce à des techniques de machine learning. Nos résultats montrent une amélioration des prédictions par rapport à la méthode actuelle. La deuxième contribution a été de proposer une nouvelle méthode basée sur l'étude des courbes de ventes pour regrouper les produits entre eux. Cette méthode a été développée pour répondre à une problématique de regroupement des produits de parapharmacie au sein de la CERP RRM. La dernière contribution de cette thèse est une étude comparative de différents modèles de traitement du langage naturel implémentés dans un assistant conversationnel pour le service technique d'un logiciel de gestion d'officine. Cette solution a montré des résultats intéressants en se rapprochant des résultats d'un humain expert.

Mots clés : Réseaux de neurones, Prédiction, Clustering, Répartition pharmaceutique, Traitement du langage naturel

ABSTRACT

Improvement of prediction of pharmacy orders with CERP RRM

Jeremy RENAUD
University of Franche Comté, 2024

Supervisors : Raphaël COUTURIER and Christophe GUYEUX

The CERP Rhin Rhone Méditerranée (CERP RRM) is a wholesale distributor responsible for ensuring pharmacies' supply. Despite recent advancements in hospital logistics, the pharmaceutical sector notably lacks decision support tools. The thesis aims to establish a predictive system for all CERP clients to forecast orders with the highest possible accuracy. The data primarily consists of time series.

Initially, the thesis focused on conducting a state-of-the-art review of time series prediction technologies, as well as implementing AI systems in industrial sectors related to wholesale distribution professions. The main contribution of this thesis was to enhance CERP RRM predictions at multiple levels using machine learning techniques. Our results demonstrate an improvement in predictions compared to the current method. The second contribution was to propose a new method based on sales curve analysis to group products together. This method was developed to address the issue of grouping parapharmacy products within CERP RRM. The final contribution of this thesis is a comparative study of different natural language processing models implemented in a conversational assistant for the technical service of a pharmacy management software. This solution has shown promising results, approaching those of an expert human.

Keywords : Deep Learning, Forecasting, Clustering, Pharmaceutical distribution, Natural language processing

REMERCIEMENTS

J'aimerais tout d'abord remercier mes directeurs de thèse, Raphaël Couturier et Christophe Guyeux, ainsi que mon responsable d'entreprise, Benoit Courjal, pour leur encadrement et leur soutien précieux durant ces trois années de thèse. Leur expertise et leurs conseils ont grandement contribué à la réussite de ce travail.

Je tiens également à exprimer ma gratitude envers l'entreprise CERP RRM, sa filiale Caduciel et leurs collaborateurs, pour m'avoir accueilli et pour avoir partagé avec moi cette expérience enrichissante. Leur soutien et leur collaboration ont été essentiels pour mener à bien ce projet de recherche.

Un grand merci aux membres de l'équipe du département DISC, certains d'entre eux étant mes enseignants depuis 2015. Leurs enseignements et leurs échanges ont nourri ma réflexion et ont été une source d'inspiration tout au long de cette thèse.

Enfin, je souhaite exprimer ma reconnaissance à ma famille et à mes amis pour leur soutien indéfectible depuis le début de ce parcours. Leur encouragement et leur présence ont été d'un grand réconfort et ont contribué à ma motivation tout au long de cette aventure.

SOMMAIRE

I	Introduction, Contexte & État de l'art	1
1	Introduction	3
1.1	Intelligence artificielle appliquée aux données de santé	3
1.2	Système logistique des pharmacies françaises	4
1.3	Études de séries temporelles	5
1.4	Présentations des contributions de la thèse	6
1.5	Plan de la thèse	8
2	Cas d'étude	9
2.1	Glossaire	9
2.2	Le métier de grossiste répartiteur	11
2.2.1	Historique	11
2.2.2	Règles du marché de la répartition pharmaceutique en France	13
2.2.3	Les différents systèmes de répartition	15
2.3	La CERP Rhin-Rhône Méditerranée (CERP RRM)	16
2.3.1	Histoire de la CERP RRM	16
2.3.2	Ses missions	17
2.3.3	Territoire et maillage logistique	18
2.3.4	COALIA : une centrale d'achat pour la pharmacie	19
2.3.5	Caduciel : La filiale numérique au service des pharmaciens	19
2.4	Les difficultés du secteur de la répartition pharmaceutique	20
2.5	Conclusion	22
3	Prédictions pour les séries temporelles	25
3.1	Introduction aux séries temporelles	25

3.1.1	Séries temporelles univariées	26
3.1.2	Séries temporelles multivariées	26
3.2	Modèles classiques pour les séries temporelles	26
3.2.1	Modèles autoregressifs (AR)	27
3.2.2	Modèles à moyenne mobile (MA)	27
3.2.3	Modèles ARIMA (Autoregressive Integrated Moving Average)	27
3.2.4	Modèles exponentiels de lissage (ETS)	28
3.2.5	Apprentissage supervisé	28
3.2.5.1	Fonction de perte	29
3.2.5.2	Descente de gradient (Gradient descent)	29
3.2.5.3	Entraînement des modèles	29
3.2.5.4	Rétropropagation de l'erreur	30
3.2.6	Validation et test	31
3.2.7	Sur-apprentissage et sous apprentissage	31
3.2.8	Optimiseurs en machine learning	32
3.2.8.1	Descente de gradient stochastique (SGD)	32
3.2.8.2	Descente de gradient mini-batch	32
3.2.8.3	Adam (Adaptive moment estimation)	33
3.2.8.4	RMSProp (Root mean square propagation)	33
3.2.8.5	Choix de l'optimiseur	33
3.3	Modèles de machine learning supervisés	34
3.3.1	Régression linéaire	34
3.3.2	Support vector machines (SVM)	34
3.3.3	Arbre de décision	35
3.3.4	Random forests	36
3.3.5	Gradient boosting (XGBoost, LightGBM)	37
3.3.5.1	XGBoost (Extreme gradient boosting)	37
3.3.5.2	LightGBM (Light gradient boosting machine)	38
3.3.6	Ensembles de modèles :	38
3.3.6.1	Bagging (Bootstrap aggregating)	38

3.3.6.2	Stacking	39
3.3.6.3	Modèles de comités (Ensemble Models)	41
3.3.7	Réseaux de neurones : Principes fondamentaux	42
3.3.7.1	Architecture d'un neurone	42
3.3.7.2	Fonctions d'activation	43
3.3.7.3	Calcul de la sortie du réseau de neurones	44
3.3.8	Réseaux de neurones profonds (DNN) pour la régression temporelle	45
3.3.8.1	Architecture des réseaux de neurones profonds	45
3.3.8.2	Fonctionnement des DNN pour la régression temporelle .	45
3.3.8.3	Disparition du gradient	46
3.3.8.4	Causes de la disparition du gradient	46
3.3.8.5	Conséquences de la disparition du gradient	47
3.3.8.6	Solutions à la disparition du gradient	47
3.3.9	Réseaux de neurones récurrents (RNN)	48
3.3.9.1	Architecture des réseaux de neurones récurrents	48
3.3.10	Long short-term memory (LSTM)	49
3.3.10.1	Architecture des LSTM	49
3.3.11	Gated recurrent unit (GRU)	50
3.3.11.1	Architecture des gated recurrent unit (GRU)	50
3.3.12	Réseaux de neurones convolutifs (CNN) pour les séries temporelles	52
3.3.12.1	Architecture des réseaux de neurones convolutifs (CNN) .	52
3.3.13	Modèles Transformers	52
3.3.13.1	Input embeddings	53
3.3.13.2	Positional encoding	53
3.3.13.3	Auto-attention (Self-Attention) / Scaled dot-product attention	54
3.3.13.4	Couches d'attention multi-têtes (Multi-head attention) . . .	55
3.3.13.5	Réseau feed forward	56
3.3.13.6	Modèles Transformers appliqués aux séries temporelles .	56
3.3.14	Validation croisée (Cross-validation)	57
3.3.14.1	Concept de la validation croisée	57

3.3.14.2	Types de validation croisée	58
3.3.14.3	Avantages de la validation croisée	58
3.3.14.4	Limitations de la validation croisée	58
3.3.15	Recherche d'hyperparamètres bayésienne	59
3.3.15.1	Fondements de la recherche d'hyperparamètres bayésienne	59
3.3.15.2	Étapes de la recherche d'hyperparamètres bayésienne . .	60
3.3.15.3	Avantages de la recherche d'hyperparamètres bayésienne	60
3.3.16	Techniques d'optimisation (Grid search, random search)	61
3.3.16.1	Recherche en grille (Grid search)	61
3.3.16.2	Recherche aléatoire (Random search)	61
3.3.16.3	Avantages et inconvénients	62
3.4	Conclusion	63
4	IA pour les secteurs liés aux grossistes-répartiteurs	65
4.1	Introduction	65
4.2	Utilisation de l'IA dans le secteur logistique	66
4.2.1	Prédiction de la demande	66
4.2.2	Gestion et optimisation des Stocks	66
4.2.3	Gestion de la chaîne d'approvisionnement	66
4.2.4	Détection des anomalies	67
4.3	Utilisation de l'IA dans le secteur pharmaceutique	67
4.3.1	Découverte de Médicaments Assistée par l'IA	67
4.3.2	AlphaFold et la prédiction de structure protéique	67
4.3.3	Optimisation des Essais Cliniques	68
4.3.4	Personnalisation des Traitements et de la Médecine de Précision . .	68
4.3.5	Détection des Effets Secondaires et de la Sécurité des Médicaments	68
4.3.6	Minage et analyse de textes cliniques	68
4.4	Utilisation de l'IA pour l'officine de pharmacie	69
4.4.1	Aide à la dispensation de Médicaments	69
4.4.2	Gestion des Stocks et Prévision de la Demande	69

4.4.3	Conseil Pharmaceutique Personnalisé, télé médecine et suivi de santé	69
4.5	Conclusion	70
II	Contributions	71
5	Prédictions pour la CERP RRM	73
5.1	Relations traitements/commandes	73
5.1.1	Choix du prescripteur	73
5.1.2	Choix du patient	74
5.1.3	Choix du pharmacien	75
5.1.4	Conclusion	78
5.2	Échelles d'agrégation	78
5.2.1	Échelle temporelle	78
5.2.2	Échelle produit	80
5.2.3	Échelle de population	81
5.2.4	Conclusion de l'étude préliminaire	82
5.3	Analyse des données d'une officine	83
5.3.1	Analyse de la patientèle	83
5.3.2	Analyse des ventes	85
5.3.2.1	Les différents types de ventes	86
5.3.2.2	Analyse des ventes croisées	88
5.3.3	Amélioration de la prédiction des besoins (achats automatiques sur LEO)	89
5.3.4	Conclusion de l'analyse des données d'officine	90
5.4	Prédiction sur les données de livraisons de la CERP	92
5.4.1	Préparation des données	92
5.4.2	Architecture de machine learning	93
5.4.2.1	Long Short-Term Memory (LSTM)	93
5.4.2.2	LightGBM	93
5.4.2.3	Transformers	94

5.4.3	Prédictions par molécules	94
5.4.3.1	Benchmark de l'erreur quadratique moyenne (RMSE) . . .	94
5.4.3.2	Courbes de prévision	95
5.4.4	Méthode Statistique mise en place par la CERP	101
5.4.4.1	Prédictions par classification ATC5	103
5.4.5	Conclusion des travaux sur les prédictions	104
5.5	Ajout de données externes	104
5.6	Conclusion	106
6	méthode de clustering TSCAPE	109
6.1	Introduction	109
6.2	État de l'art	110
6.2.1	Techniques de comparaison de distances dans l'analyse de séries temporelles	110
6.2.2	Techniques d'incorporation et de réduction de la dimensionnalité . .	111
6.2.3	Algorithmes de regroupement	112
6.3	Description de la méthode TSCAPE	113
6.3.1	Déformation temporelle dynamique (DTW)	114
6.3.2	Projection euclidienne : Positionnement multidimensionnel (MDS) .	114
6.3.2.1	Formulation Mathématique	115
6.3.2.2	Détails opérationnels	115
6.3.3	La réduction multidimensionnelle dans TSCAPE	116
6.3.3.1	Utilisation dans TSCAPE	116
6.3.4	Regroupement (Clustering)	116
6.3.4.1	Formules Mathématiques	117
6.3.4.2	Utilisation dans TSCAPE	117
6.4	Expérimentations	118
6.4.1	Données utilisées	118
6.4.2	Métriques	119
6.4.3	Regroupement par K-means avec une métrique DTW (tslearn) . . .	119

6.4.4	Analyse en Composantes Principales (PCA) pour le regroupement de séries temporelles	120
6.4.5	Expérimentation de TSCAPE	121
6.4.6	Résultats	125
6.5	Conclusion	127
7	Comparaison des LLM	131
7.1	Introduction	131
7.2	Apport de connaissances expertes au LLM	131
7.2.1	création de la base de connaissances	132
7.2.2	recherche de connaissances et réponse aux questions	132
7.2.2.1	Comparaison sémantique des vecteurs	133
7.3	Type de modèles LLM	134
7.3.1	Modèles propriétaires	134
7.3.2	Modèles open source	134
7.4	Agent conversationnel pour le service client d'un LGO	135
7.4.1	Données de test	135
7.4.2	Modèles utilisés	136
7.4.3	Méthode de comparaisons des performances	136
7.4.4	Résultats	137
7.4.5	Perspectives d'améliorations	138
7.4.5.1	Utilisation de modèles plus performants	138
7.4.5.2	LLM : ajout d'un système d'évaluation des sources de connaissances	138
7.4.5.3	Sélection des connaissances pertinentes par rapport à la question	138
7.5	Conclusion	139
III	Conclusion & Perspectives	141
8	Conclusion & Perspectives	143

8.1	Conclusion	143
8.2	Perspectives	145
8.2.1	La donnée de pharmacie : un enjeu pour l'avenir du secteur	145
8.2.2	Un secteur confronté à des problématiques de ruptures	146
8.2.3	Les modèles de langages en assistance	147
8.2.3.1	Les LLM's en assistance aux pharmaciens	147
8.2.3.2	Utilisations des LLM's pour augmenter les données de prédictions	147
8.2.4	Classification et dépendances inter-produits	148

I

INTRODUCTION, CONTEXTE & ÉTAT DE L'ART

INTRODUCTION

Cette thèse est réalisée dans le cadre d'un dispositif CIFRE, entre le Département d'Informatique des Systèmes Complexes (DISC) du laboratoire FEMTO-ST, et l'entreprise Caduciel, filiale numérique de la CERP Rhin Rhone Méditerranée (CERP RRM) : grossiste-répartiteur pour les pharmacies françaises. Les données utilisées sont celles de la CERP RRM, ainsi que celles de certains de leurs clients pharmaciens. Le but principal de cette thèse est d'améliorer les prédictions de commande de la CERP RRM en utilisant des techniques d'intelligence artificielle, notamment d'apprentissage profond (Deep Learning).

1.1/ INTELLIGENCE ARTIFICIELLE APPLIQUÉE AUX DONNÉES DE SANTÉ

L'émergence de l'intelligence artificielle (IA) et des données massives a profondément transformé de nombreux domaines de notre société, notamment celui de la santé. L'intelligence artificielle peut être définie comme la capacité des machines à simuler des processus cognitifs humains, tels que l'apprentissage, le raisonnement et la prise de décision, en s'appuyant sur des algorithmes sophistiqués et des données massives. Ces données massives, également connues sous le nom de "big data", font référence à l'énorme volume, à la variété et à la vitesse des données générées et collectées dans divers contextes.

Cette combinaison de l'IA et des données massives a ouvert de nouvelles perspectives pour la recherche médicale et la prestation des soins de santé. Les données de santé, en particulier, représentent une source d'information précieuse et immense, qui peut être exploitée pour améliorer les diagnostics, les traitements et la gestion des maladies. Les données de santé incluent des informations provenant de diverses sources telles que les dossiers médicaux électroniques, les dispositifs portables de surveillance de la santé, les bases de données génétiques, les forums de discussion en ligne sur la santé, les

publications scientifiques, les données d'imagerie médicale, et bien d'autres encore.

Les enjeux de l'IA et des données massives dans le domaine de la santé sont multiples et complexes. L'énorme volume de données générées dans le domaine de la santé, (dû notamment à la numérisation de nombreux procédés), pose des défis de taille en termes de stockage, de gestion, d'analyse et de sécurité. Les systèmes traditionnels de gestion des données sont souvent incapables de faire face à cette explosion de données, nécessitant ainsi le développement de nouvelles infrastructures et technologies adaptées.

La variété des données de santé provenant de sources hétérogènes rend leur intégration et leur compatibilités difficiles. Les informations médicales sont souvent dispersées dans différents systèmes et formats, ce qui complique leur utilisation pour la recherche et l'analyse. L'IA peut être utilisée dans l'harmonisation et l'intégration de ces données, en facilitant leurs extractions, leurs normalisations et leurs structurations.

Un autre défi majeur réside dans la confidentialité et la sécurité des données de santé. Les informations médicales sont hautement sensibles et personnelles, de ce fait elles tombent donc sous la réglementation de dispositifs comme le règlement général sur la protection des données (RGPD). Ces dispositifs exigent de garantir la protection de la vie privée des individus et de mettre en place des mécanismes de sécurité robustes pour prévenir les violations et les utilisations abusives des données de santé.

En outre, l'exploitation des données massives de santé nécessite des méthodes d'analyse avancées et des algorithmes d'apprentissage automatique adaptés. L'IA offre des possibilités d'identification de schémas, de prédiction et de recommandation, mais il est crucial de développer des modèles d'IA fiables, transparents et interprétables. Les décisions cliniques basées sur l'IA doivent être compréhensibles et justifiables pour les praticiens et les patients

1.2/ SYSTÈME LOGISTIQUE DES PHARMACIES FRANÇAISES

Le système de santé français repose sur un maillon essentiel : les pharmacies. Ces établissements sont responsables de fournir rapidement et efficacement les traitements nécessaires à leurs patients. Pour garantir cette disponibilité, les pharmacies françaises sont approvisionnées par des grossistes-répartiteurs. La principale mission de ces grossistes est de livrer rapidement les commandes des pharmaciens, avec une norme actuelle d'une livraison par jour.

Pour atteindre cette efficacité, les grossistes-répartiteurs doivent mettre en place une chaîne logistique complexe et la plus optimisée possible. Ils doivent stocker une large gamme de produits pharmaceutiques, pour répondre aux besoins de leurs clients pharmaciens. Anticiper les commandes des pharmaciens peut grandement contribuer à

améliorer cette chaîne d'approvisionnement en veillant à ce que chaque entrepôt soit suffisamment approvisionné pour éviter les ruptures de stock.

Dans le domaine de la logistique, les algorithmes d'intelligence artificielle ont prouvé leur utilité depuis de nombreuses années. Que ce soit par le biais de systèmes statistiques et d'intelligence économique ou, plus récemment, grâce à l'utilisation d'algorithmes d'apprentissage automatique, l'IA a permis d'optimiser les opérations logistiques dans divers secteurs.

Cependant, prévoir avec précision les ventes futures parmi les milliers de produits différents référencés sur le marché pharmaceutique français s'avère être un défi complexe, notamment à cause de nombreuses spécificités et règles en vigueur sur ce marché. La diversité des médicaments, des dispositifs médicaux et des produits de santé rend difficile l'établissement de prévisions précises. Par conséquent, une approche consistant à rassembler les produits en groupements cohérents peut aider à simplifier le problème.

Regrouper les produits selon des critères pertinents tels que les molécules actives, les classes thérapeutiques ou encore les variations saisonnières peut permettre de mieux comprendre les besoins et les tendances du marché pharmaceutique. En utilisant des techniques d'intelligence artificielle, nous pouvons développer des modèles prédictifs précis capables d'anticiper les commandes des pharmaciens en fonction de ces regroupements. En optimisant la gestion des stocks dans les entrepôts des grossistes-répartiteurs, il est possible de réduire les coûts liés aux stocks excédentaires ou à l'inverse, d'éviter les ruptures de stock, garantissant ainsi une plus grande disponibilité des produits.

1.3/ ÉTUDES DE SÉRIES TEMPORELLES

À l'ère de la numérisation, l'utilisation de systèmes numériques est devenue une pratique courante dans notre société. Cette utilisation généralisée génère d'énormes quantités de données, qui peuvent être stockées et analysées pour en tirer des informations pertinentes. Parmi ces données, les séries temporelles jouent un rôle crucial. Elles représentent l'évolution d'un système ou d'une variable dans le temps, que ce soit une trajectoire, des courbes de vente ou des relevés de températures. L'étude et l'analyse approfondies de ces séries temporelles offrent de nombreuses possibilités d'exploitation, telles que le regroupement, la prédiction et la détection d'anomalies.

Le regroupement des séries temporelles, également connu sous le nom de clustering, est une technique qui consiste à identifier des sous-ensembles similaires de séries temporelles, basés sur des caractéristiques communes. Cela permet de regrouper les produits

pharmaceutiques en fonction de leur comportement de vente ou de leur évolution temporelle, facilitant ainsi leur analyse et leur compréhension. Par exemple, en regroupant les produits en fonction de leur saisonnalité, il est possible d'identifier les variations saisonnières de la demande et de prendre des mesures appropriées pour gérer les stocks et les approvisionnements.

La prédiction des séries temporelles est une autre tâche importante dans l'analyse des données temporelles. En utilisant des techniques d'intelligence artificielle, notamment d'apprentissage automatique, il est possible de développer des modèles prédictifs capables d'estimer les ventes futures de produits pharmaceutiques. Ces modèles prennent en compte, par exemple, les tendances passées, les motifs saisonniers ou encore les fluctuations à court terme, pour générer des prévisions. Une meilleure précision des prédictions permet aux grossistes-répartiteurs de gérer plus efficacement leurs stocks, d'optimiser les délais de livraison, voir, dans le futur, d'anticiper la demande de leurs clients pharmaciens.

La détection d'anomalies dans les séries temporelles est également un sujet d'application de l'intelligence artificielle important. Dans le cadre du métier de grossiste répartiteur, les anomalies peuvent indiquer des problèmes tels que des ruptures de stock, des erreurs de livraison ou des variations inhabituelles dans les comportements de vente. En utilisant des algorithmes d'IA, il est possible de détecter automatiquement ces anomalies. Cela permet de mettre en évidence des comportements anormaux (clients partant à la concurrence, produit en changement de cycle de vie...), qui n'auraient peut-être pas été détecté.

1.4/ PRÉSENTATIONS DES CONTRIBUTIONS DE LA THÈSE

La principale contribution de cette thèse est l'utilisation de techniques d'intelligence artificielle pour améliorer les prédictions de commande des pharmaciens auprès de la CERP RRM. Pour ce faire, des analyses sur la causalité entre un besoin de traitement et une vente à la CERP RRM, ainsi que sur les possibilités de regroupements des produits, ont été effectuées. Suite à cela, une étude sur les données d'une officine de pharmacie a été réalisée. Cette étude nous a permis de constater que l'application de méthodes de machine learning à la précision d'une pharmacie n'était pas réalisable, mais que des méthodes statistiques plus classiques comme des moyennes roulantes permettaient d'améliorer les prédictions des besoins sur les commandes automatiques du logiciel de gestion d'officine. La suite de ces travaux a consisté à appliquer des techniques de prédictions à base de modèles de réseaux de neurones sur des données historiques

de ventes de CERP RRM. Les premières expérimentations ont été effectuées sur quatre regroupements de produits à base de molécules, et ont donné lieu à une première production scientifique présentée lors d'une conférence scientifique internationale. Dans un deuxième temps, les travaux se sont rapprochés de ceux d'une méthode statistique mise en place par la CERP RRM en parallèle de ceux de cette thèse. La méthode à base de réseaux de neurones a montré une amélioration des résultats par rapport à la méthode statistique actuelle. L'utilisation de réseaux de neurones permet également d'utiliser de nombreuses données en entrée, ce qui a amené à étudier l'intérêt d'ajout de données externes pour améliorer les prédictions.

La deuxième contribution de cette thèse découle d'une problématique apparue durant les travaux réalisés dans la première contribution. Si pour les médicaments il existe des regroupements pertinents sur lesquels faire des prédictions, ce n'est pas le cas pour les produits de parapharmacie. En effet, il n'existe pas de classification préétablie et la CERP RRM a des difficultés pour en trouver une appropriée. De là est venue l'idée de proposer une technique de regroupement basée sur l'étude des courbes de vente des produits. La méthode mise au point dans la deuxième contribution mesure la distance entre les courbes de tous les produits, avant de faire une réduction de dimensions en projetant la matrice distance sur un plan euclidien avant d'appliquer un algorithme de clustering.

La dernière contribution de cette thèse est en lien avec les avancées récentes dans le domaine des modèles de traitement du langage naturel, plus précisément sur les "Large Language Models" (LLM). Le but de cette contribution est d'étudier ce que ces nouvelles technologies peuvent apporter à l'environnement CERP RRM et plus largement au secteur de la répartition pharmaceutique ainsi qu'à celui de l'officine de pharmacie. Pour les orienter, ces travaux ont été adossés à une problématique métier. Dans le monde de la pharmacie, il y a de nombreuses informations sous forme non structurée et sous diverses formes, que ce soit de la documentation, des alertes de pharmacovigilance ou encore des documents réglementaires. Les personnes qui doivent prendre connaissance de ces informations passent un certain temps à devoir les assimiler, puis sont ensuite fortement sollicitées en tant que référentes sur ce savoir. Le but de ces travaux est de comparer différents modèles de LLM intégrés dans un agent conversationnel. Cet agent doit pouvoir synthétiser une réponse à une question posée, en s'appuyant sur une base de connaissances dans le but d'assister un expert humain. Le test a été effectué sur les 10 questions les plus posées au service client d'un logiciel de gestion d'officine, en se basant sur la version PDF de sa documentation. Les réponses des modèles ont été mélangées à celles d'un expert humain et ont été évaluées à l'aveugle par des membres du personnel du service client.

1.5/ PLAN DE LA THÈSE

Les travaux de cette thèse sont avant tout exploratoires, c'est pour cela qu'une partie non-négligeable de ces derniers a été de comprendre le contexte, analyser les différents besoins, pour formuler les problématiques et les choix possibles pour y répondre. Pour réaliser ces tâches, il a fallu faire l'acquisition d'une connaissance fine du fonctionnement du secteur métier. Cette analyse a donc été naturellement restituée dans ce mémoire avec ces conclusions. C'est pour cela que cette thèse est divisée en plusieurs parties comme il suit :

- 2 - Cas d'étude. Ce chapitre explique tout le contexte de cette thèse, l'environnement CERP RRM, Caduciel, de l'officine d'une pharmacie, ainsi que les enjeux et concepts liés au monde pharmaceutique utilisés dans cette thèse. Cette partie révélera les problématiques que les contributions de cette thèse tentent de résoudre.
- 3 - État de l'art des méthodes de prédictions pour les séries temporelles. Comme son nom l'indique, ce chapitre introduit l'état de l'art des différentes techniques liées au séries temporelles.
- 4 - Application de l'IA aux problématiques industrielles liées aux grossistes-répartiteurs. Ce chapitre fait un tour d'horizon des travaux liés à l'intelligence artificielle dans les métiers liés aux grossistes répartiteurs. Cela prend en compte la logistique, le secteur pharmaceutique et de la santé, ainsi que le secteur des officines de pharmacies.
- 5 - Prédications de commande des pharmacies auprès de la CERP RRM. Ce chapitre présente la première contribution de la thèse. Il décrit les travaux qui ont été effectués pour améliorer les prédictions de commandes des clients pharmaciens de la CERP RRM grâce à des méthodes de machine learning, ainsi qu'une étude des données d'une officine de pharmacie. Il présente également les réflexions préliminaires aux travaux qui ont été menés pour définir la problématique, ainsi que les choix faits pour y répondre.
- 6 - TSCAPE : Une nouvelle méthode de clustering. Ce chapitre présente la nouvelle méthode de clustering qui a été développée pour répondre à des problèmes de regroupement des produits de pharmacies. Cette méthode s'appuie sur l'étude des formes de courbes des séries temporelles.
- 7 - Comparaison de différents "Large Language Model's" (LLM's) pour l'assistance aux pharmaciens. Ce chapitre montre les travaux réalisés sur les techniques de modèles de langage dans le but de créer des agents conversationnels capables de synthétiser une réponse à partir d'une base de connaissances.
- 8 - Conclusion et perspectives des travaux.

CAS D'ÉTUDE

2.1/ GLOSSAIRE

Avant toute chose, il est important de définir certains termes présents dans ce mémoire.

1. **Forme Galénique** Aussi appelée "forme pharmaceutique" ou "forme médicamenteuse", c'est la forme sous laquelle le médicament se présente. (Exemples : comprimés, gélules, sirops, poudre en sachets, etc.)
2. **LGO** Logiciels de Gestion d'Officine. D'après la définition du Ségur du numérique en santé, sont considérés comme LGO les logiciels qui contiennent toutes les fonctions minimales suivantes : un système de facturation Assurance Maladie validé par le GIE SESAM-Vitale, une compatibilité avec le dossier pharmaceutique validé par le Conseil National de l'Ordre des Pharmaciens (CNOF), posséder un historique des délivrances de médicaments (ordonnancier), un accès à la base de données des médicaments et produits de santé, une gestion des stocks de médicaments et un dossier patient informatisé (informations administratives, observations médicales en saisie libre ou par formulaires type et personnalisables).
3. **Classification ATC** Le système de classification Anatomique, Thérapeutique et Chimique (ATC) [24] a été créé par l'Organisation Mondiale de la Santé (OMS) pour classer les médicaments. Les substances actives sont classées dans une hiérarchie à cinq niveaux différents. Le système comporte quatorze groupes anatomiques/pharmacologiques principaux. Chaque groupe principal est divisé en sous-groupes de niveau 2 qui peuvent être des groupes pharmacologiques ou thérapeutiques. Les 3e et 4e niveaux sont des sous-groupes chimiques, pharmacologiques ou thérapeutiques et le 5e niveau est la substance chimique. Les 2e, 3e et 4e niveaux sont souvent utilisés pour identifier les sous-groupes pharmacologiques lorsqu'ils sont jugés plus appropriés que les sous-groupes thérapeutiques ou chimiques. Le code de classification complet est donc composé comme suit : LC-CLLCC (avec L correspondant à une lettre et C correspondant à un chiffre). Chaque lettre et chaque paire de chiffres ajoutés représente un niveau supplémentaire :

A	Alimentary tract and metabolism (1st level, anatomical main group)
A10	Drugs used in diabetes (2nd level, therapeutic subgroup)
A10B	Blood glucose lowering drugs, excl. insulins (3rd level, pharmacological subgroup)
A10BA	Biguanides (4th level, chemical subgroup)
A10BA02	metformin (5th level, chemical substance)

FIGURE 2.1 – Exemple d'une classification ATC complète

4. **Code CIP** Le Code Identifiant de Présentation (CIP) sert à identifier une présentation unique d'un produit pharmaceutique. Une présentation est définie par : sa dénomination, sa forme galénique, son dosage et son conditionnement ainsi que la contenance de ce dernier. Il existe un code CIP à 7 chiffres qui est maintenant remplacé par un code à 13 chiffres depuis le 1^{er} janvier 2009.
5. **Code EAN** Le Code EAN pour "European Article Numbering" est un numéro d'identification d'un produit utilisé dans le commerce et l'industrie. C'est ce que l'on appelle grossièrement le "code barre" sur nos produits du quotidien. En pharmacie, les produits qui ne possèdent pas de CIP sont identifiés avec leur code EAN. Mais un produit peut également posséder les 2 types de code.
6. **Code UCD** Une troisième façon de codifier les médicaments autre que la classification ATC, ou le code CIP, est l'unité commune de disposition (UCD), elle permet de représenter un médicament sous sa plus petite unité possible (gélule, flacon, comprimé, etc.). Cette codification est le plus souvent utilisée par les services hospitaliers qui délivrent des médicaments à l'UCD, plus qu'à la boîte.
7. **Médicaments OTC** Le terme "OTC" provient de l'expression anglaise "Over The Counter," qui se traduit en français par "par-dessus le comptoir." Dans le domaine pharmaceutique, il désigne l'ensemble des médicaments disponibles sans ordonnance. Il convient de noter que ces médicaments peuvent néanmoins être prescrits par un médecin, bien que cela ne soit pas obligatoire. Les pharmaciens peuvent également conseiller ces médicaments, ce qui s'inscrit dans la pratique de l'automédication de plus en plus répandue. Le marché des médicaments en vente libre est à la fois diversifié et en constante évolution. Cette diversité se manifeste notamment par les variations réglementaires, qui diffèrent d'un pays à l'autre. En effet, certains médicaments sont disponibles en vente libre dans certains pays tandis qu'ils nécessitent une prescription en France. Par ailleurs, ce marché est également caractérisé par son évolutivité, puisque certains médicaments qui étaient auparavant en vente libre en France ont vu leur statut changer, comme c'est le cas pour la Codéine, qui nécessite désormais une prescription. De même, certains médicaments qui nécessitaient autrefois une ordonnance ne l'exigent plus aujourd'hui.

d'hui.

- 8. Groupements de pharmacies** Les groupements de pharmacies représentent des regroupements d'officines de pharmacies visant à optimiser leurs opérations et à renforcer leur positionnement sur le marché. Ces regroupements, qui peuvent revêtir diverses formes telles que les coopératives, les groupes d'achat, les franchises et les réseaux de pharmacies indépendantes, offrent plusieurs avantages aux pharmaciens. Les groupements permettent aux pharmacies de réaliser des achats en volume, ce qui facilite la négociation de prix avantageux avec les fournisseurs et réduit les coûts d'approvisionnement en médicaments et autres produits pharmaceutiques. Ils favorisent également le partage de connaissances et de meilleures pratiques entre les membres, contribuant ainsi à l'amélioration de la qualité des services apportés à leurs patients. Les groupements peuvent aussi proposer des programmes de formation continue, un soutien administratif et juridique, allégeant ainsi la charge de travail des pharmaciens en matière de gestion, de comptabilité, de réglementation et de conformité, etc.

Chaque groupement de pharmacies possède ses propres règles et conditions, ce qui montre l'importance pour un pharmacien de bien les comprendre avant de prendre la décision de rejoindre un groupement.

2.2/ LE MÉTIER DE GROSSISTE RÉPARTITEUR

Il est important de comprendre ce qu'est le métier de grossiste répartiteur, de ses origines à nos jours, ainsi que l'écosystème qui l'entoure. Pour comprendre le travail, les problématiques, les enjeux et les spécificités de ce secteur méconnu, ce chapitre va présenter une petite entrée en matière avec l'historique de ce métier, avant de présenter les spécificités de ce secteur très réglementé. Le chapitre présentera ensuite les différents acteurs et particulièrement l'environnement CERP Rhin Rhône Méditerranée dans lequel cette thèse prend place, avant de finir par mettre en lumière les problématiques que ce secteur rencontre et le besoin d'innovation que les acteurs doivent mettre en place pour se démarquer. Les informations historiques sont issues des travaux d'Arnaud Antoine [54].

2.2.1/ HISTORIQUE

Le métier de répartiteur pharmaceutique trouve ses origines au XII^e siècle, avec l'apparition des importateurs en drogues. Ces ancêtres des répartiteurs étaient des droguistes qui négociaient avec les caravaniers pour obtenir des drogues, épices et plantes exotiques provenant de différentes régions. Ils approvisionnaient ensuite les colporteurs, qui

revendaient ces produits sur les foires et marchés aux apothicaires et médecins. À cette époque, les médecins ont progressivement eu recours à des spécialistes, les apothicaires épiciers, pour exécuter leurs prescriptions en préparant les remèdes dans leurs boutiques aux drogues. Le terme d'epicerie rappelle que les matières premières nécessaires à la préparation des médicaments venaient en partie de contrées lointaines, notamment de l'Orient par la célèbre route des épices.

La profession d'apothicaire s'est différenciée de celle du médecin, comme en témoigne le statut accordé par Louis IX en 1258 pour la préparation et la vente des médicaments. Plus tard, en 1352, Jean le Bon a tenté d'imposer l'antidotaire de Nicolas aux apothicaires épiciers. Cette séparation a été renforcée en 1777, sous le règne de Louis XVI, lorsqu'un décret a remplacé le Jardin des apothicaires par le Collège de pharmacie, consacrant ainsi le nom de pharmaciens pour ces spécialistes et leur accordant l'exclusivité de la préparation des remèdes.

Au XIX^e siècle, la législation française a donné un cadre réglementaire au parcours du médicament, régissant la préparation et la vente des remèdes par les pharmaciens. Certains pharmaciens audacieux ont créé de petites annexes à leurs officines pour fabriquer des spécialités pharmaceutiques en masse. Cela a conduit à la création des premiers laboratoires pharmaceutiques et a nécessité une structure intermédiaire pour approvisionner les officines. Ainsi, en 1851, le premier droguiste pharmaceutique en gros, la Maison Thomas à Agen, est né, considéré comme l'ancêtre des grands répartiteurs modernes.

Dans la seconde moitié du XIX^e siècle, certains marchands d'accessoires médicaux ont développé leur activité pour fournir des médicaments aux pharmaciens. Ce développement a abouti à la création du métier de commissionnaire en spécialités en 1885. La distribution des spécialités pharmaceutiques s'est organisée avec plusieurs circuits, dont les commissionnaires en spécialités, les groupements d'achats de pharmaciens et le circuit direct entre les fabricants et certaines officines importantes.

En 1920, le terme de "répartiteur" a été utilisé pour désigner les grossistes qui achetaient et vendaient en gros et possédaient leur stock. Toutefois, ce métier n'a été officiellement reconnu qu'après la Seconde Guerre mondiale avec la mise en place d'une réglementation visant à assurer la qualité de distribution et l'accès aux médicaments sur tout le territoire, dans un souci de santé publique. Une loi de septembre 1941 a instauré le visa ministériel pour autoriser la mise sur le marché d'une spécialité, permettant de lutter contre les contrefaçons et de normaliser les produits. Cette loi a également permis d'établir les bases réglementaires du métier de "grossiste répartiteur" et de fixer pour la première fois une marge de 12,5 % accordée aux répartiteurs.

Après la Seconde Guerre mondiale, avec l'apparition de la Sécurité Sociale en 1945 et la découverte de nombreux nouveaux produits pharmaceutiques, les répartiteurs ont dû s'adapter. Le législateur a codifié leur exercice en 1962, établissant les bases

réglementaires du métier de "grossiste répartiteur". Une distinction claire a été établie entre les fabricants, les grossistes répartiteurs et les dépositaires de produits pharmaceutiques.

Jusqu'à la fin de la Première Guerre mondiale, seules quelques entreprises familiales exerçaient le métier de commissionnaire en spécialités. Certaines d'entre elles, comme Piot, Lemoine & Royer, Merveau & Cie, et Michelat, Souillard & Cie, ont fusionné en 1924 pour créer l'Office Commercial Pharmaceutique (OCP), aujourd'hui leader sur le marché de la répartition en France. D'autres sociétés capitalistes et coopératives ont également vu le jour à cette époque, donnant naissance aux différentes coopératives de répartition pharmaceutique, dont certaines sont toujours actives aujourd'hui.

Le métier de répartiteur pharmaceutique a évolué au fil des siècles, passant des importateurs en drogues et des apothicaires épiciers du Moyen Âge aux répartiteurs modernes d'aujourd'hui. La législation et la réglementation ont joué un rôle clé dans la reconnaissance et la structuration de cette profession, permettant une distribution efficace des médicaments sur tout le territoire et une protection accrue du consommateur. Les répartiteurs, qu'ils soient capitalistes ou coopératifs, occupent une place essentielle dans le dispositif de distribution du médicament en France.

2.2.2/ RÈGLES DU MARCHÉ DE LA RÉPARTITION PHARMACEUTIQUE EN FRANCE

Les grossistes répartiteurs sont des acteurs indispensables dans le circuit du médicament, résultant de l'imagination des pharmaciens d'officine et de l'essor de l'industrie pharmaceutique. Leur rôle est de fournir en permanence un approvisionnement sûr et rapide à toutes les pharmacies, qu'elles soient situées dans des zones rurales ou urbaines, qu'elles appartiennent à des coopératives ou à des sociétés capitalistes. Ces grossistes répartiteurs sont réglementés et spécialisés, opérant sous le monopole pharmaceutique qui est un fondement de la législation pharmaceutique depuis 1941.

Le terme "grossiste répartiteur pharmaceutique" désigne exclusivement les entreprises spécialisées qui manipulent et détiennent les médicaments en vue de leur distribution en gros. Ces entreprises relèvent de l'ordre professionnel des pharmaciens et doivent obtenir une autorisation administrative spécifique pour exercer leur activité. Un pharmacien responsable est désigné au sein de l'entreprise pour superviser les opérations pharmaceutiques, assurer l'autorité sur les pharmaciens délégués et garantir le respect des règles juridiques et techniques.

La définition officielle du grossiste répartiteur est donnée par l'article R5124-2 du Code de la Santé Publique, qui stipule que c'est une entreprise se livrant à l'achat et au stockage des médicaments en vue de leur distribution en gros et en l'état. Le grossiste répartiteur détient ainsi les stocks de médicaments de la nation et bénéficie d'une exclusivité d'exer-

cice.

Cependant, avec ce monopole pharmaceutique viennent des obligations légales strictes. Le grossiste répartiteur doit disposer d'un assortiment de médicaments comprenant au moins les neuf dixièmes des présentations effectivement exploitées en France. Il doit être capable de satisfaire la demande de sa clientèle habituelle durant au moins deux semaines et de livrer tout médicament faisant partie de son assortiment dans les vingt-quatre heures suivant la commande. Il est tenu de desservir toute pharmacie de son territoire qui en fait la demande, mais peut également livrer exceptionnellement ou en cas d'urgence une officine située en dehors de son secteur.

La création d'un établissement de répartition pharmaceutique est autorisée par l'Agence nationale de sécurité du médicament et des produits de santé (ANSM) selon une procédure fixée à l'avance et soumise à des inspections régulières pour assurer le respect des bonnes pratiques de distribution en gros.

Les grossistes répartiteurs peuvent être classés en deux catégories : les full liners, qui référencent et détiennent la quasi-totalité des produits pharmaceutiques commercialisés dans un pays, et les short liners, qui assurent la distribution en gros d'une gamme plus restreinte de produits pharmaceutiques. Les grossistes répartiteurs short liners sont plus répandus dans certains pays européens où ils ne sont pas soumis aux mêmes obligations de service public qu'en France.

Actuellement, on distingue en France sept acteurs clés, adhérents de la Chambre Syndicale de la Répartition Pharmaceutique (CSRP). Ils représentent 97,14 % du marché de la répartition. Il existe donc également d'autres grossistes-répartiteurs qui ne sont pas adhérents à la Chambre, mais leurs parts de marché sont minimales 2.2.



miro

FIGURE 2.2 – Grossistes-répartiteurs adhérents à la CSRP et leurs parts sur le marché français. Source : "https://www.csrp.fr/le-marche"

2.2.3/ LES DIFFÉRENTS SYSTÈMES DE RÉPARTITION

Depuis près d'un siècle, les entreprises de répartition pharmaceutique, originellement constituées de petites structures familiales ou de groupements de pharmaciens, ont connu une évolution remarquable, prenant des directions très diversifiées. Aujourd'hui, la distinction entre les répartiteurs et leurs concurrents est devenue complexe, car toutes ces sociétés, qu'elles appartiennent à de grands groupes internationaux ou à des coopératives, détiennent un statut de répartiteur pharmaceutique et sont considérées comme un maillon incontournable de la chaîne de distribution du médicament. Deux modèles se sont dégagés pour exercer cette profession : le système capitaliste traditionnel et le système coopératif, chacun adoptant des objectifs et des stratégies qui leur sont propres.

Dans le contexte du système capitaliste, le marché français de la répartition s'est caractérisé par la consolidation des entreprises. Des opérations de rachat et de fusion ont permis un regroupement des acteurs, annonçant leur internationalisation progressive. Actuellement, des acteurs capitalistes majeurs coexistent en France, concurrençant les coopératives de répartition pharmaceutique. On retrouve dans ce genre de système de répartiteurs des représentants de grands groupes comme OCP Répartition, Alliance Healthcare ou encore Phoenix Pharma.

En revanche, le système coopératif, bien qu'étant structuré sous le statut de société anonyme coopérative, a connu une métamorphose importante, se transformant en de véritables entreprises à portée européenne, avec des filiales multiples. Ces coopératives ont désormais la dimension et la puissance des grands groupes capitalistes, bien que leur terminologie traditionnelle évoque un modèle de petite structure locale.

Les coopératives de répartition pharmaceutique fonctionnent sous les mêmes principes que les sociétés capitalistes, tout en présentant quelques particularités régies par la loi sur le statut de la coopération. Leurs objectifs incluent la réduction du coût de revient des produits et services, l'amélioration de la qualité des produits fournis aux membres et aux consommateurs, ainsi que la satisfaction des besoins et la promotion des activités sociales et économiques de leurs membres. Nous reviendrons plus en détail sur ces points avec les activités de la CERP RRM.

Les coopératives de répartition pharmaceutique sont soumises à certaines restrictions concernant la mise à disposition de leurs services aux tiers non associés. Cependant, en cas d'urgence, elles ne peuvent refuser leurs services aux pharmaciens d'officine non associés ou à certains établissements de santé. C'est dans ce système de coopérative que sont situées les différentes CERP, dont la CERP RRM.

2.3/ LA CERP RHIN-RHÔNE MÉDITERRANÉE (CERP RRM)

La CERP Rhin-Rhône Méditerranée, comme précédemment mentionné, figure parmi les grossistes-répartiteurs français qui ont embrassé le modèle coopératif. Cette section sera consacrée à l'exposition détaillée de son historique, de ses missions, de sa structure organisationnelle ainsi que de ses différentes entités et filiales.

2.3.1/ HISTOIRE DE LA CERP RRM

L'histoire de la CERP Rhin-Rhône Méditerranée remonte à 1921, lorsque des pharmaciens se sont réunis à Belfort pour établir "La Pharmacie régionale des Pharmaciens de l'Est" dans le but d'améliorer la qualité de leur approvisionnement en produits pharmaceutiques. Cette initiative confraternelle a ensuite évolué au fil des années, reflétant son développement territorial.

En 1938, l'organisation a été rebaptisée "Les Pharmaciens de l'Est", puis en 1965, elle a pris le nom de "CERP de l'Est" en lien avec l'émergence du mouvement coopératif CERP à travers la France. Cette transformation a été accompagnée par l'ouverture de trois nouveaux établissements entre 1961 et 1965, marquant une phase significative de croissance.

En 1976, la société a fusionné avec la CAPDS pour devenir "CERP Rhin Rhône", comptant alors 12 établissements dans son réseau. L'année 1988 a marqué une étape cruciale avec la fusion entre la CERP Rhin Rhône et la CERP Méditerranée, conduisant à la création définitive de "CERP Rhin Rhône Méditerranée" et intégrant Pharmat, portant ainsi le nombre d'établissements à 17, tout en renforçant sa présence géographique.

Au cours de son histoire, la CERP Rhin Rhône Méditerranée a maintenu une dynamique d'innovation en partenariat avec les pharmaciens. En 2000, l'entreprise a introduit son premier automate de préparation de commandes, marquant une avancée significative dans ses opérations.

Les années suivantes ont été marquées par des développements stratégiques, notamment avec l'acquisition de Caduciel Informatique en 2002, la création de La Centrale en partenariat avec d'autres pharmaciens en 2010, qui deviendra ultérieurement COALIA, et l'inauguration d'un second magasin général en 2020, équipé du système robotique de dernière génération : Scallog.

En 2021, la CERP Rhin Rhône Méditerranée a célébré son centenaire, comptant alors 24 établissements et 3 filiales, illustrant sa longue histoire de croissance et d'innovation dans le secteur pharmaceutique français.

En 2023, la CERP RRM se rapproche de la CERP Rouen, dans le but de fusionner ces

deux entités au cours de l'année 2024.

2.3.2/ SES MISSIONS

La CERP Rhin-Rhône Méditerranée se voit attribuer plusieurs missions essentielles dans le cadre de son engagement envers les pharmaciens. La première et principale mission consiste à exercer en tant que grossiste-répartiteur, impliquant le respect rigoureux des obligations énoncées par le code de la santé publique. Ces obligations englobent la nomination d'un Pharmacien Responsable et d'un Pharmacien Délégué dans chaque établissement, la garantie de livraison dans un délai de 24 heures à toutes les pharmacies du territoire de répartition, à l'exception des périodes de week-ends et de jours fériés. De plus, il est attendu que la CERP RRM référence au moins 90 % de la collection nationale de médicaments, maintienne en permanence un stock équivalent à 15 jours de consommation moyenne des pharmacies, et exécute les commandes sans autorisation de substitution. Enfin, la mise en place d'un système d'astreinte opérationnelle les week-ends et jours fériés vise à assurer des livraisons en situation d'urgence.

La CERP RRM se distingue par son rôle de grossiste-répartiteur, mais elle offre également une gamme complète de services visant à soutenir les pharmaciens dans la gestion quotidienne de leurs officines et à améliorer leur approche envers leurs patients. Pour ce faire, la CERP RRM propose diverses initiatives et accompagnements à destination des pharmaciens.

Tout d'abord, l'entreprise fournit un soutien complet aux futurs acquéreurs d'officines, allant de l'aide à la définition du projet professionnel à l'accompagnement financier grâce à "Passeport Pharma". Cette solution de financement, élaborée en partenariat avec une banque spécialisée dans les professions de santé, offre un accompagnement à l'installation, combinant l'expertise de la CERP RRM et celle de son partenaire bancaire.

La CERP RRM s'engage également à former les jeunes diplômés en pharmacie à travers l'initiative "*École CERP du Titulaire de demain*", une formation de groupe inter-régionale qui couvre divers aspects essentiels de la gestion d'une officine, du financement au management en passant par les responsabilités pharmaceutiques et le droit du travail. Cette formation se caractérise par une approche pratique, favorisant l'apprentissage par la pratique et l'échange d'expérience.

L'entreprise propose également des services de géomarketing, permettant aux pharmaciens d'analyser l'environnement de leur officine, d'évaluer leur zone de chalandise, et de prendre des décisions stratégiques éclairées. Cette analyse peut être complétée par un diagnostic commercial réalisé par la filiale de la CERP RRM, HOLIVIA, spécialisée dans la transaction d'officine, qui semble être un marché prometteur pour les répartiteurs [69]. HOLIVIA offre un accompagnement complet, de la transaction à la gestion du change-

ment en passant par la formation et le réagencement de l'officine.

La CERP RRM propose aussi une formation personnalisée axée sur le management, comprenant six modules visant à doter les pharmaciens des compétences nécessaires pour gérer efficacement leur équipe et leur officine. Ces modules couvrent divers aspects de la communication, du recrutement, de la motivation, de la délégation, et de l'organisation.

Enfin, la CERP RRM met l'accent sur le merchandising en accompagnant les pharmaciens dans la révision de l'aménagement de leur espace de vente pour optimiser leur rentabilité et s'adapter aux besoins changeants de leurs patients. Cette démarche commence par un audit complet de l'officine, suivi de la définition des objectifs et de la mise en œuvre du projet, tout en garantissant un suivi continu pour maintenir les performances au fil du temps.

La CERP Rhin Rhône Méditerranée joue un rôle crucial avec sa mission de grossiste répartiteur mais également en offrant une gamme complète de services visant à renforcer la profession pharmaceutique, à soutenir les nouveaux acquéreurs, à former les jeunes diplômés, à optimiser la gestion des officines et à améliorer l'expérience des patients, tout en s'adaptant constamment aux évolutions du secteur pharmaceutique.

2.3.3/ TERRITOIRE ET MAILLAGE LOGISTIQUE

La CERP RRM se caractérise principalement par son maillage logistique territorial performant, déployé sur l'ensemble de la façade Est de la France. L'entreprise possède son siège social historique à Belfort et exploite un réseau de 24 établissements de répartition stratégiquement implantés au cœur du territoire français. Cette structure organisationnelle permet à l'entreprise d'offrir un service réactif et efficace aux pharmaciens partenaires.

Le maillage territorial est finement conçu, avec en moyenne une agence pour 180 officines, assurant ainsi une proximité maximale, la plupart des officines étant situées à moins de 100 km d'une agence 2.3. Les pharmaciens partenaires bénéficient donc de la possibilité de recevoir leurs commandes deux fois par jour, avec un délai moyen de livraison d'1 heure et 30 minutes. De plus, pour les produits pharmaceutiques d'urgence, un service de livraison express est disponible, et les produits peuvent également être retirés directement au comptoir.

Chaque établissement de répartition gère son propre stock, ce qui permet d'adapter la collection de médicaments au plus près des besoins et des prescriptions locales. En complément de cette structure, la société dispose de deux magasins généraux situés à Dijon et à Montélimar, qui enrichissent la collection en rendant disponibles toutes les



FIGURE 2.3 – Répartitions des agences de la CERP RRM sur son territoire

spécialités pharmaceutiques en moins de 24 heures.

2.3.4/ COALIA : UNE CENTRALE D'ACHAT POUR LA PHARMACIE

COALIA représente une initiative de commerce associé, orchestrée conjointement par la CERP Rhin Rhône Méditerranée et ses partenaires pharmaciens les plus proches, affiliés aux groupements Pharm UPP, Coopérative APM Provence, et APP Sud. Cette collaboration vise à établir une centrale d'achat pharmaceutique, avec pour objectif de mutualiser les achats de près de 1 600 officines indépendantes situées sur la façade Est de la France, partageant un engagement solidaire.

La coopération au sein du réseau COALIA permet aux pharmacies de bénéficier collectivement de remises avantageuses négociées avec les principaux laboratoires spécialisés dans les produits en vente libre (OTC). Ces avantages sont assortis de la qualité de service reconnue de la CERP Rhin Rhône Méditerranée.

Chaque organe décisionnel de COALIA est composé de pharmaciens adhérents et de représentants de la CERP Rhin Rhône Méditerranée, assurant ainsi une co-construction optimale du modèle, conformément aux principes du commerce associé. Le volume de commandes demeure adapté au potentiel de chaque officine au sein du réseau, et des outils d'aide à la vente sont mis à leur disposition pour optimiser leur activité.

2.3.5/ CADUCIEL : LA FILIALE NUMÉRIQUE AU SERVICE DES PHARMACIENS

Alain Malot, un pharmacien qui n'était pas satisfait des logiciels de gestion officinale (LGO) du marché, fonde Caduciel en 1991, où il développe son propre logiciel de ges-

tion d'officine. Comme ce dernier fonctionnait bien, le pharmacien le propose à d'autres confrères. Caduciel continue de se développer, puis est rachetée en 2002 par la CERP RRM pour devenir sa filiale numérique.

En qualité de pionnier dans le domaine des services informatiques et des solutions numériques destinés aux professionnels de la santé, Caduciel a acquis une expérience de 30 ans dans ce secteur. L'entreprise joue un rôle essentiel en accompagnant 300 pharmacies indépendantes et groupements dans leur processus de transformation numérique et d'adaptation de leurs activités aux nouvelles exigences du marché.

Caduciel se positionne aujourd'hui en tant que conseiller et partenaire de confiance pour les pharmacies, leur apportant un soutien quotidien dans le choix et la mise en œuvre de solutions numériques appropriées. L'entreprise vise à optimiser les parcours clients, stimuler les ventes, fidéliser la patientèle, et moderniser les infrastructures des officines.

Originellement reconnue comme éditeur du logiciel de gestion d'officine (LGO), Caduciel a su évoluer en réponse aux nouveaux besoins du secteur pour devenir un acteur numérique polyvalent. Aujourd'hui, elle dispose des compétences nécessaires pour gérer l'intégralité du cycle de vie de ses produits, commençant par l'étude des besoins des utilisateurs, la conception et le développement d'applications, la commercialisation, l'installation, la formation, et la maintenance pour son portefeuille de clients. L'entreprise compte environ trente collaborateurs répartis sur deux sites, à savoir Voray-sur-l'Ognon et Avignon.

En 2024, Caduciel ferme définitivement ses portes, et certains collaborateurs sont repris au sein de la CERP RRM pour former une nouvelle cellule innovation rattachée au siège de l'entreprise.

2.4/ LES DIFFICULTÉS DU SECTEUR DE LA RÉPARTITION PHARMACEUTIQUE

D'après la Chambre Syndicale de la Répartition Pharmaceutique (CSRP) [5][9][3], le secteur des grossistes-répartiteurs fait face à une crise économique importante, résultant de mesures gouvernementales défavorables sur une période de treize ans. Les principaux problèmes incluent un système de rémunération inadapté, des baisses significatives de marges, et une diminution substantielle des prix des médicaments. Cette situation a entraîné une perte d'exploitation de 31 millions d'euros en 2020, malgré un chiffre d'affaires de 17 milliards d'euros, comparé à un gain de 200 millions d'euros pour un chiffre d'affaires de 19 milliards en 2008, ainsi qu'une réduction de 193 millions d'euros de la marge réglementée depuis 2008. On peut noter que la pandémie de COVID-19 a permis aux répartiteurs, sur les années 2020-2021, de reprendre de l'importance dans le parcours du

médicament et auprès des officines. Les pouvoirs publics ont pris conscience du rôle et de la précarité économique du secteur et ont décidé d'une mesure d'urgence en 2020 et d'une revalorisation de la marge en 2021 2.4. Cependant, la profession a été confrontée à des crises majeures, notamment des ruptures d'approvisionnement en médicaments et une inflation croissante. Les coûts énergétiques ont également augmenté, fragilisant davantage le secteur. L'impact de l'inflation sur la profession en 2022 a été de plus de 45 millions d'euros, et il est estimé à 40 millions d'euros pour 2023.

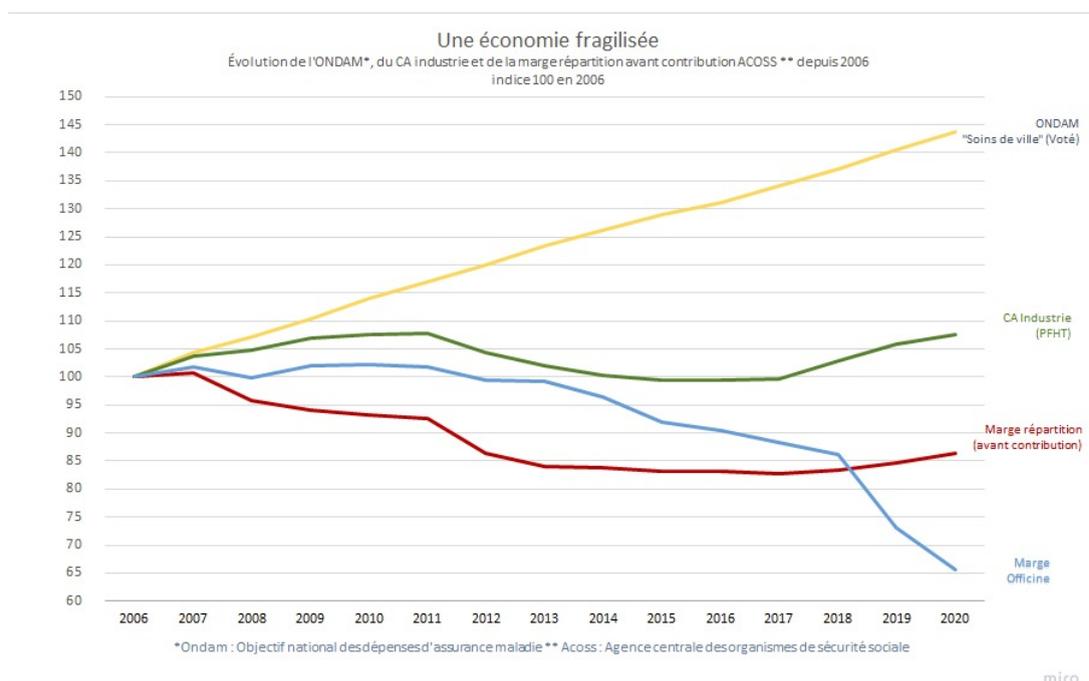


FIGURE 2.4 – Malgré une rehausse sur 2020-2021, on observe une baisse de la marge du secteur de la répartition depuis 2007, alors que l'objectif national des dépenses de l'assurance maladie est en augmentation constante

Le secteur de la répartition pharmaceutique se caractérise par un environnement concurrentiel spécifique, où les prix des médicaments sont uniformément déterminés pour tous les acteurs. En plus d'une concurrence entre les différents grossistes-répartiteurs, les laboratoires proposent également leurs produits directement aux officines et aux hôpitaux, venant marcher sur le secteur de la répartition 2.5. Dans ce contexte, la concurrence s'articule autour de plusieurs paramètres : notamment sur les barèmes d'écart, qui concernent les remises accordées, les conditions d'installation proposées aux pharmaciens, ou encore la qualité des services offerts, y compris la mise à disposition d'outils d'assistance à la pharmacovigilance, les stratégies de merchandising, ainsi que les solutions informatiques spécifiques aux officines.

8 CANAUX DE DISTRIBUTION DU MÉDICAMENT

(en nombre)

Source : Leem, Gers, CSRP, Dress, CNOP

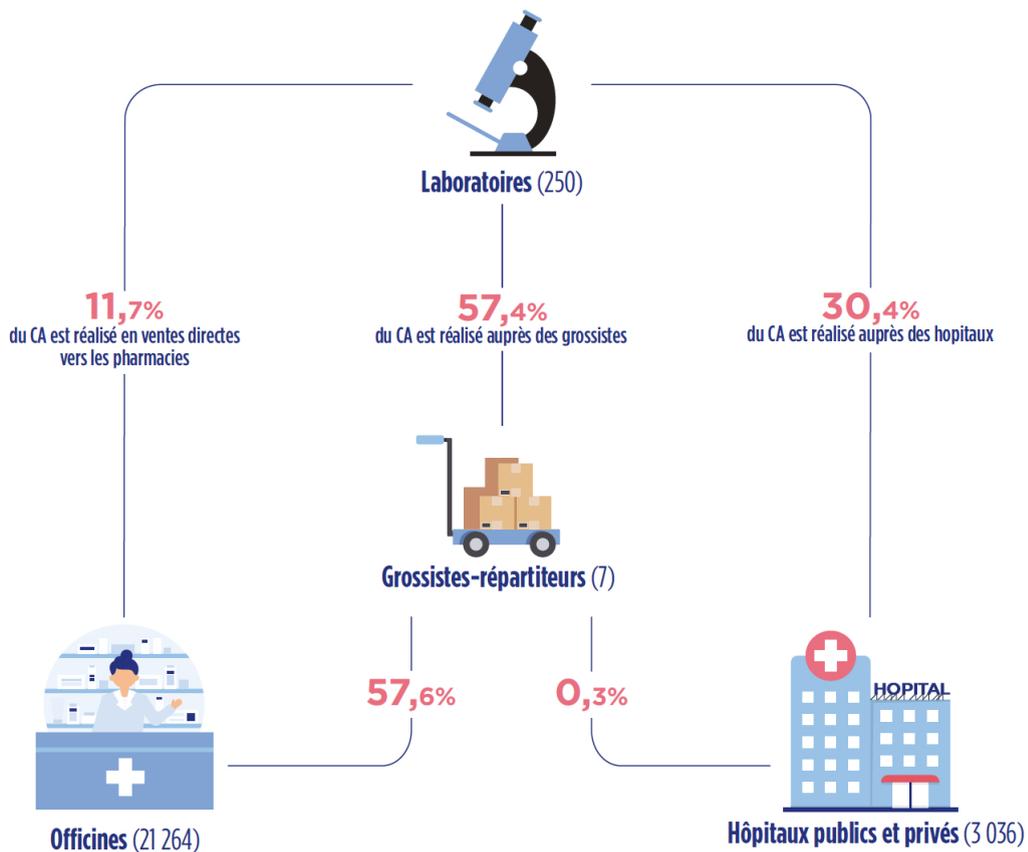


FIGURE 2.5 – Parts de marché de la répartition pharmaceutique dans la distribution des médicaments en 2022. On peut voir que les grossistes-répartiteurs, en plus d'une concurrence entre eux, sont aussi concurrencés par la vente directe des laboratoires. Source : Leem.org [7]

2.5/ CONCLUSION

Ce chapitre a permis d'explorer le secteur de la répartition pharmaceutique, un marché dynamique en constante évolution, caractérisé par des défis et une concurrence accrue. Les grossistes-répartiteurs sont confrontés à la nécessité constante d'innover afin de se démarquer de leurs concurrents, d'offrir de nouveaux services à leur clientèle de pharmaciens et d'acquérir des outils d'aide à la décision pour parfaire leur compréhension du marché pharmaceutique ainsi que leurs processus métier. C'est dans ce contexte que la CERP RRM et sa filiale numérique, Caduciel, ont manifesté un intérêt pour l'intégration de l'intelligence artificielle dans divers outils visant à améliorer leurs processus internes

et les services proposés aux pharmaciens partenaires. Cette réflexion a conduit à la formulation de la problématique centrale de cette thèse, établie en partenariat avec le laboratoire de recherche FEMTO-ST. Le prochain chapitre se consacrera à une revue approfondie de l'état de l'art dans ce domaine spécifique.

PRÉDICTIONS POUR LES SÉRIES TEMPORELLES

3.1/ INTRODUCTION AUX SÉRIES TEMPORELLES

Les séries temporelles sont un type particulier de données que l'on trouve dans de nombreux domaines, notamment en économie, en finance, en météorologie, en sciences sociales, en sciences de la santé, et dans bien d'autres disciplines. Une série temporelle consiste en une séquence chronologique de points de données, où chaque point est associé à un moment spécifique dans le temps. Contrairement à d'autres types de données, comme les données spatiales ou les données croisées, les séries temporelles présentent une dépendance naturelle entre les observations en raison de leur ordre temporel.

Voici quelques caractéristiques clés des séries temporelles :

1. **Ordre temporel** : Chaque observation d'une série temporelle est enregistrée dans un ordre séquentiel, ce qui signifie que la chronologie des données est importante. Par exemple, lorsque l'on analyse les ventes quotidiennes d'un produit, l'ordre des jours est crucial pour comprendre les tendances et les saisons de vente.
2. **Dépendance temporelle** : Les valeurs successives dans une série temporelle sont souvent corrélées ou dépendantes les unes des autres. Cela signifie que la valeur à un instant t peut être influencée par les valeurs passées : $t - 1$, $t - 2$, ... Cette dépendance temporelle peut être utilisée pour modéliser et prévoir les valeurs futures.
3. **Saisonnalité** : De nombreuses séries temporelles présentent des motifs cycliques ou saisonniers. Par exemple, dans notre cas d'étude, on va pouvoir observer une hausse des ventes de produits antihistaminiques (contre les allergies) au printemps de chaque année.
4. **Tendance** : Les séries temporelles peuvent également montrer une tendance à long terme, indiquant une augmentation ou une diminution constante des valeurs

au fil du temps. Dans notre cas d'étude, cela correspondrait par exemple au cycle de vie d'un produit (croissance, stable, en perte de marché).

- 5. Bruit :** Les séries temporelles possèdent également plus ou moins de bruit. Ce bruit est en quelque sorte tout ce qui n'est pas représenté par une tendance ou par la saisonnalité.

3.1.1/ SÉRIES TEMPORELLES UNIVARIÉES

Les séries temporelles univariées consistent en une seule variable temporelle enregistrée à différents instants de temps. Par exemple, la variation quotidienne des prix d'une action est une série temporelle univariée, où la seule variable est le prix de l'action, mesurée à différents jours.

3.1.2/ SÉRIES TEMPORELLES MULTIVARIÉES

Les séries temporelles multivariées impliquent plusieurs variables temporelles enregistrées simultanément à différents instants de temps. Par exemple, pour prédire le trafic routier, vous pourriez avoir des données sur la vitesse du vent, la température, et le volume de trafic enregistrées en même temps à chaque heure de la journée.

Les séries temporelles sont utilisées pour une variété de tâches, notamment la prévision, la détection d'anomalies, l'analyse de tendances, la modélisation de processus, la gestion de la chaîne d'approvisionnement, la surveillance des systèmes et bien plus encore. Pour analyser efficacement les séries temporelles, on utilise diverses techniques statistiques, mathématiques et d'apprentissage automatique.

La suite de ce chapitre propose un état de l'art des techniques d'intelligence artificielle appliquée aux séries temporelles.

3.2/ MODÈLES CLASSIQUES POUR LES SÉRIES TEMPORELLES

Les modèles classiques pour les séries temporelles constituent une catégorie importante de méthodes d'analyse et de prévision des données temporelles. Dans cette section, nous explorerons en détail plusieurs de ces modèles, à commencer par les modèles autoregressifs.

3.2.1/ MODÈLES AUTOREGRESSIFS (AR)

Les modèles autoregressifs sont des modèles statistiques qui reposent sur l'idée que la valeur courante d'une série temporelle est une combinaison linéaire de ses valeurs passées. Plus formellement, un modèle AR(p) est défini comme suit :

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (3.1)$$

où X_t est la valeur à un instant t , c est une constante, $\phi_1, \phi_2, \dots, \phi_p$ sont les coefficients AR, $X_{t-1}, X_{t-2}, \dots, X_{t-p}$ sont les valeurs passées de la série, et ε_t est un terme d'erreur aléatoire.

Les paramètres $\phi_1, \phi_2, \dots, \phi_p$ sont estimés à l'aide de méthodes telles que la méthode des moindres carrés. Les modèles AR sont adaptés aux séries temporelles qui présentent une autocorrélation significative avec leurs valeurs passées.

3.2.2/ MODÈLES À MOYENNE MOBILE (MA)

Les modèles à moyenne mobile sont un autre type de modèle classique pour les séries temporelles. Contrairement aux modèles AR, les modèles MA considèrent que la valeur courante dépend des erreurs passées plutôt que des valeurs passées de la série. Un modèle MA(q) est défini comme suit :

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (3.2)$$

où X_t est la valeur à un instant t , μ est la moyenne de la série, ε_t est le terme d'erreur au temps t , $\theta_1, \theta_2, \dots, \theta_q$ sont les coefficients MA, et $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ sont les erreurs passées.

Comme pour les modèles AR, les coefficients MA sont estimés à l'aide de méthodes d'estimation appropriées [11]. Les modèles MA sont efficaces pour modéliser les séries temporelles avec des erreurs autocorrélées.

3.2.3/ MODÈLES ARIMA (AUTOREGRESSIVE INTEGRATED MOVING AVERAGE)

Les modèles ARIMA [81] combinent les caractéristiques des modèles AR et MA, tout en tenant compte de la différenciation pour rendre la série stationnaire. Un modèle ARIMA(p, d, q) est défini comme suit :

$$\Delta^d X_t = c + \phi_1 \Delta^d X_{t-1} + \phi_2 \Delta^d X_{t-2} + \dots + \phi_p \Delta^d X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (3.3)$$

Ici, Δ^d représente l'opérateur de différenciation, d est le nombre de différenciations nécessaires pour rendre la série stationnaire, $\phi_1, \phi_2, \dots, \phi_p$ sont les coefficients AR, $\theta_1, \theta_2, \dots, \theta_q$ sont les coefficients MA, et les autres termes sont définis de manière similaire aux modèles AR et MA. L'opérateur de différenciation est défini comme il suit :

$$\Delta^d X_t = \sum_{k=0}^d (-1)^k \binom{d}{k} B^k X_t, \quad (3.4)$$

avec $\Delta^d X_t$ représentant la série temporelle différenciée d'ordre d , B est l'opérateur de retard qui décale la série temporelle d'un pas de temps en arrière (défini comme suit : $BX_t = X_{t-1}$), $\binom{d}{k}$ est le coefficient binomial, et X_t est la série temporelle à l'instant t .

Les modèles ARIMA sont largement utilisés pour modéliser des séries temporelles non stationnaires et sont souvent accompagnés d'une analyse des autocorrélations et des autocorrélations partielles pour sélectionner les valeurs optimales de p , d , et q .

3.2.4/ MODÈLES EXPONENTIELS DE LISSAGE (ETS)

Les modèles exponentiels de lissage [61] sont une classe de modèles qui reposent sur l'idée de modéliser la tendance, la saisonnalité, et le bruit dans les séries temporelles à l'aide de composantes exponentielles. Les modèles ETS sont adaptés comme leur nom l'indique aux séries temporelles présentant une saisonnalité et une tendance.

Un modèle ETS est généralement défini par trois composantes principales :

- La composante d'erreur (*Error E*) : elle représente le bruit aléatoire de la série.
- La composante de tendance (*Trend T*) : elle capture la variation à long terme de la série.
- La composante saisonnière (*Seasonal S*) : elle modélise les variations périodiques dans la série.

Les modèles ETS sont généralement notés sous la forme ETS(A, B, C), où A, B et C sont des indicateurs de la présence ou de l'absence des composantes *Error*, *Trend*, et *Seasonal*. Par exemple, ETS(A, A, N) indique un modèle avec une composante d'erreur, une composante de tendance et sans composante saisonnière.

Les paramètres des modèles ETS sont estimés à l'aide de techniques d'optimisation, et les prévisions sont générées en extrapolant les composantes estimées dans le futur.

3.2.5/ APPRENTISSAGE SUPERVISÉ

L'apprentissage supervisé est l'un des principaux paradigmes du machine learning. Dans ce cadre, un modèle est formé à partir d'un ensemble de données d'apprentissage com-

prenant des exemples d'entrées et de sorties souhaitées (étiquettes). L'objectif est d'apprendre une fonction qui peut faire correspondre efficacement les entrées aux sorties.

3.2.5.1/ FONCTION DE PERTE

La fonction de perte mesure l'écart entre les prédictions du modèle et les étiquettes réelles dans l'ensemble de données d'apprentissage. Une bonne fonction de perte doit être conçue de manière à ce que son minimum soit atteint lorsque le modèle produit des prédictions exactes. L'objectif de l'apprentissage supervisé est de minimiser la valeur de la fonction de perte.

Une fonction de perte couramment utilisée pour la régression est l'erreur quadratique moyenne (Mean Squared Error, MSE) définie comme :

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.5)$$

où n est le nombre d'exemples, y_i est la vraie valeur, et \hat{y}_i est la prédiction du modèle.

3.2.5.2/ DESCENTE DE GRADIENT (GRADIENT DESCENT)

La descente de gradient est un algorithme utilisé pour minimiser la fonction de perte d'un modèle en ajustant les paramètres du modèle. L'idée fondamentale est de suivre la direction dans laquelle la pente de la fonction de perte est la plus raide (le gradient est le plus grand négatif) pour converger vers un minimum local ou global.

L'algorithme de descente de gradient met à jour les paramètres du modèle de manière itérative en utilisant la formule suivante :

$$\theta_{t+1} = \theta_t - \alpha \nabla L(\theta_t) \quad (3.6)$$

où θ_t est le vecteur de paramètres à l'itération t , α est le taux d'apprentissage (learning rate), et $\nabla L(\theta_t)$ est le gradient de la fonction de perte par rapport aux paramètres.

3.2.5.3/ ENTRAÎNEMENT DES MODÈLES

L'entraînement des modèles de machine learning consiste à ajuster les paramètres du modèle pour minimiser la fonction de perte sur l'ensemble d'entraînement. Cela se fait généralement à l'aide d'algorithmes d'optimisation tels que la descente de gradient.

Le processus d'entraînement comprend plusieurs étapes, notamment l'initialisation des

paramètres, la propagation avant (forward propagation) pour calculer les prédictions du modèle grâce à la fonction de perte, la rétropropagation (backward propagation) de l'erreur pour calculer les gradients, et la mise à jour des paramètres à l'aide de la descente de gradient.

Une fois le modèle entraîné, il peut être évalué sur un ensemble de test pour estimer sa performance en généralisation.

3.2.5.4/ RÉTROPROPAGATION DE L'ERREUR

La rétropropagation de l'erreur [25] est une technique fondamentale en apprentissage supervisé, particulièrement utilisée dans les réseaux de neurones. Elle permet de calculer efficacement les gradients de la fonction de perte par rapport aux paramètres du modèle, ce qui est essentiel pour ajuster ces paramètres et améliorer la performance du modèle au fil de l'entraînement.

Le processus de rétropropagation peut être divisé en plusieurs étapes clés :

1. Propagation avant (Forward propagation) : Au cours de cette étape, les entrées sont propagées à travers le modèle, de la couche d'entrée jusqu'à la couche de sortie, en calculant les valeurs prédites pour chaque exemple d'entraînement. Cela se fait en appliquant les opérations de transformation définies par les poids et les fonctions d'activation dans chaque couche du modèle. La sortie finale du modèle est comparée aux étiquettes réelles pour calculer la perte.

2. Calcul du gradient local : Une fois que la perte est calculée, l'étape suivante consiste à calculer le gradient local de la perte par rapport aux activations dans la couche de sortie. Cela se fait en utilisant la dérivée de la fonction de perte par rapport à la sortie de la couche de sortie. Pour une fonction de perte telle que l'erreur quadratique moyenne (MSE) pour la régression, le gradient est simplement la différence entre la sortie prédite et l'étiquette réelle, pondérée par la dérivée de la fonction d'activation de la couche de sortie.

Par exemple, pour la MSE, le gradient local est donné par :

$$\nabla L_{\text{output}} = 2 \cdot (\text{output} - \text{label}) \cdot \nabla \sigma(\text{input}) \quad (3.7)$$

où output est la sortie prédite, label est l'étiquette réelle, et $\nabla \sigma(\text{input})$ est la dérivée de la fonction d'activation de la couche de sortie par rapport à son entrée.

3. Rétropropagation du gradient : Le gradient local calculé à l'étape précédente est ensuite rétropropagé à travers le modèle, de la couche de sortie à la couche d'entrée, pour calculer les gradients par rapport aux poids et aux biais de chaque couche. Cela

se fait en utilisant la règle de la chaîne, de sorte que le gradient à une couche donnée dépend des gradients calculés dans les couches supérieures.

Le gradient par rapport aux poids W d'une couche est donné par :

$$\nabla W = \text{input}^T \cdot \nabla L_{\text{output}} \quad (3.8)$$

où input est la sortie de la couche précédente (ou les données d'entrée pour la première couche) et ∇L_{output} est le gradient local calculé à l'étape précédente.

4. Mise à jour des paramètres : Une fois que les gradients par rapport aux poids et aux biais de chaque couche ont été calculés, les paramètres du modèle sont mis à jour à l'aide d'un algorithme d'optimisation tel que la descente de gradient. Les paramètres sont mis à jour de manière à minimiser la fonction de perte en suivant la direction opposée au gradient.

Par exemple, la mise à jour des poids peut être effectuée en utilisant la règle de la descente de gradient :

$$W_{\text{nouveau}} = W_{\text{actuel}} - \alpha \cdot \nabla W \quad (3.9)$$

où W_{actuel} sont les poids actuels, α est le taux d'apprentissage (learning rate), et ∇W est le gradient par rapport aux poids.

Ce processus itératif de rétropropagation de l'erreur, suivi de la mise à jour des paramètres, est répété pour de nombreuses époques (cycles d'entraînement) jusqu'à ce que le modèle atteigne une performance satisfaisante sur l'ensemble d'entraînement.

3.2.6/ VALIDATION ET TEST

Une pratique courante en machine learning est de diviser l'ensemble de données en trois parties : l'ensemble d'entraînement (training set), l'ensemble de validation (validation set), et l'ensemble de test (test set). L'ensemble d'entraînement est utilisé pour former le modèle en réglant les hyperparamètres, l'ensemble de validation est utilisé pour surveiller la performance du modèle pendant l'entraînement, et l'ensemble de test est utilisé pour évaluer la performance finale du modèle.

3.2.7/ SUR-APPRENTISSAGE ET SOUS APPRENTISSAGE

Le sur-apprentissage se produit lorsque le modèle apprend les détails spécifiques de l'ensemble d'entraînement au lieu de généraliser correctement. Cela se traduit par une

excellente performance sur l'ensemble d'entraînement mais une performance médiocre sur l'ensemble de test. Le sous apprentissage, en revanche, se produit lorsque le modèle est trop simple pour capturer les motifs sous-jacents des données et conduit également à une performance médiocre.

Il est essentiel de surveiller ces problèmes et d'ajuster le modèle et les hyperparamètres en conséquence.

3.2.8/ OPTIMISEURS EN MACHINE LEARNING

L'optimisation des modèles de machine learning est une étape cruciale pour entraîner des modèles efficaces. Les optimiseurs sont des algorithmes qui ajustent les paramètres du modèle lors de l'apprentissage pour minimiser une fonction de perte. Cette section présente les principes de quelques exemples d'optimiseurs connus.

3.2.8.1/ DESCENTE DE GRADIENT STOCHASTIQUE (SGD)

La descente de gradient stochastique (Stochastic Gradient Descent, SGD) [18] est l'optimiseur le plus fondamental en machine learning. Il met à jour les paramètres du modèle en se basant sur le gradient de la fonction de perte par rapport à un seul exemple d'entraînement à la fois. L'équation de mise à jour pour un paramètre θ est donnée par :

$$\theta_{t+1} = \theta_t - \alpha \nabla L(\theta_t, x_i, y_i) \quad (3.10)$$

où θ_t est le vecteur de paramètres à l'itération t , α est le taux d'apprentissage (learning rate), $\nabla L(\theta_t, x_i, y_i)$ est le gradient de la fonction de perte L par rapport à θ_t pour l'exemple d'entraînement (x_i, y_i) .

3.2.8.2/ DESCENTE DE GRADIENT MINI-BATCH

La descente de gradient mini-batch [77] est une variante de la SGD qui met à jour les paramètres en se basant sur un petit groupe (mini-batch) d'exemples d'entraînement à la fois, ce qui permet une meilleure utilisation de la parallélisation et une convergence plus rapide. L'équation de mise à jour est similaire à celle de la SGD, mais le gradient est calculé sur le mini-batch.

$$\theta_{t+1} = \theta_t - \alpha \nabla L(\theta_t, \text{Mini-Batch}) \quad (3.11)$$

3.2.8.3/ ADAM (ADAPTIVE MOMENT ESTIMATION)

Adam [76] est un optimiseur qui adapte le taux d'apprentissage pour chaque paramètre en fonction de son historique de gradients. Il utilise deux moments, le moment du gradient (m_t) et le moment du gradient au carré (v_t), pour calculer le taux d'apprentissage adaptatif. L'équation de mise à jour pour un paramètre θ est donnée par :

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla L(\theta_t) \quad (3.12)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) (\nabla L(\theta_t))^2 \quad (3.13)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (3.14)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (3.15)$$

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \quad (3.16)$$

où β_1 et β_2 sont les coefficients d'ajustement des moments, ϵ est une petite valeur pour éviter la division par zéro.

3.2.8.4/ RMSPROP (ROOT MEAN SQUARE PROPAGATION)

RMSProp est un autre optimiseur qui adapte le taux d'apprentissage pour chaque paramètre en utilisant une moyenne mobile pondérée des carrés des gradients. L'équation de mise à jour pour un paramètre θ est donnée par :

$$v_t = \beta v_{t-1} + (1 - \beta) (\nabla L(\theta_t))^2 \quad (3.17)$$

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{v_t} + \epsilon} \nabla L(\theta_t) \quad (3.18)$$

où β est un coefficient d'ajustement des moments, et ϵ est une petite valeur pour éviter la division par zéro.

3.2.8.5/ CHOIX DE L'OPTIMISEUR

Le choix de l'optimiseur dépend souvent de la nature de la tâche et de la nature des données. Il existe d'autres modèles d'optimiseurs [90], et trouver le plus adapté à la problématique est souvent fait de façon empirique.

3.3/ MODÈLES DE MACHINE LEARNING SUPERVISÉS

Les modèles de machine learning supervisés sont des approches couramment utilisées pour la prédiction de séries temporelles. Dans cette section, nous présentons les principales méthodes, y compris la régression linéaire, les machines à vecteurs de support (SVM), les forêts aléatoires, le gradient boosting (notamment XGBoost et LightGBM), ainsi que l'utilisation de réseaux de neurones profonds (RNN) pour la régression temporelle.

3.3.1/ RÉGRESSION LINÉAIRE

La régression linéaire est l'un des modèles les plus simples et les plus couramment utilisés pour la prédiction de séries temporelles. Elle suppose une relation linéaire entre la variable cible et les variables explicatives. Formellement, pour une série temporelle univariée y_t et des prédicteurs x_t , le modèle de régression linéaire s'exprime comme suit :

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t \quad (3.19)$$

Où β_0 et β_1 sont les coefficients de régression et ϵ_t est le terme d'erreur à chaque instant t . La méthode des moindres carrés ordinaires (OLS) [56] est souvent utilisée pour estimer les coefficients.

3.3.2/ SUPPORT VECTOR MACHINES (SVM)

Les machines à vecteurs de support (support vector machines, SVM) sont une classe de modèles de machine learning supervisés largement utilisée pour la classification et la régression. Développées par Vapnik et Cortes (1995) [20], elles se sont avérées efficaces dans un large éventail d'applications, y compris l'analyse de séries temporelles.

Les SVM sont des modèles d'apprentissage supervisé qui cherchent à trouver un hyperplan dans un espace de caractéristiques de manière à maximiser la marge entre les classes d'objets. La marge est la distance entre l'hyperplan et les exemples d'apprentissage les plus proches, appelés vecteurs de support. Les SVM visent à maximiser cette marge tout en minimisant l'erreur de classification.

Le fonctionnement des SVM peut être résumé comme suit :

- 1. Transformation dans un espace de caractéristiques :** Les séries temporelles sont souvent non linéaires, il est donc courant de les projeter dans un espace de caractéristiques de dimension supérieure à l'aide d'une fonction noyau (kernel), telle que le noyau gaussien (RBF) ou le noyau polynomial.

2. **Recherche de l'hyperplan optimal** : Les SVM visent à trouver l'hyperplan qui maximise la marge tout en satisfaisant la contrainte que tous les exemples d'apprentissage sont correctement classés. Cette tâche est formulée comme un problème d'optimisation convexe.
3. **Gestion des données non linéaires** : Les SVM sont capables de gérer des données non linéaires grâce à la transformation dans l'espace de caractéristiques, ce qui leur permet de trouver des frontières de décision complexes.
4. **Régularisation** : Les SVM incluent un paramètre de régularisation C , qui permet de contrôler la tolérance aux erreurs de classification. Un C élevé conduit à une marge plus étroite mais à une meilleure classification des exemples d'apprentissage, tandis qu'un C plus faible permet une marge plus large avec une tolérance aux erreurs plus élevée.

Les SVM offrent plusieurs avantages pour la modélisation des séries temporelles. Leur capacité à gérer des données non linéaires les rend adaptés à des séries temporelles présentant des tendances complexes. De plus, leur capacité à gérer des données à haute dimension en fait un choix approprié lorsque les séries temporelles sont représentées dans un espace de caractéristiques étendu.

Néanmoins, les SVM peuvent être sensibles à la sélection du noyau et aux paramètres de régularisation, ce qui nécessite souvent une validation croisée pour obtenir des performances optimales. De plus, ils ne sont pas intrinsèquement conçus pour la modélisation des séries temporelles multivariées, ce qui peut nécessiter des adaptations ou des extensions spécifiques.

Les SVM demeurent un outil puissant dans l'arsenal des modèles de machine learning pour l'analyse de séries temporelles, en particulier lorsque la non-linéarité et la haute dimensionnalité sont des caractéristiques importantes des données.

3.3.3/ ARBRE DE DÉCISION

Un Arbre de décision [52] est un modèle d'apprentissage supervisé qui peut être utilisé pour la classification ou la régression. Il est particulièrement adapté à la modélisation de séries temporelles lorsque les données présentent des relations complexes et non linéaires. L'Arbre de décision est une représentation graphique sous forme d'arbre, où chaque nœud interne représente une décision basée sur un attribut particulier, chaque branche représente une sortie de cette décision, et chaque feuille représente une classe ou une valeur de sortie.

Le fonctionnement d'un Arbre de décision peut être décomposé en plusieurs étapes :

1. **Sélection de l'attribut de division** : L'algorithme commence par choisir l'attribut qui,

lorsqu'il est utilisé comme critère de division, maximise la pureté des sous-ensembles résultants. La pureté est généralement mesurée à l'aide d'indices tels que l'indice de Gini ou l'entropie.

2. Division des données : Les données sont divisées en sous-ensembles en fonction de la valeur de l'attribut sélectionné. Cette étape divise l'ensemble de données initial en plusieurs sous-ensembles, chacun correspondant à une branche de l'arbre.

3. Récursion : Les étapes 1 et 2 sont répétées de manière récursive pour chaque sous-ensemble jusqu'à ce qu'un critère d'arrêt soit atteint, tel qu'une profondeur maximale de l'arbre, un nombre minimum d'échantillons dans les feuilles, ou une pureté suffisamment élevée.

4. Attribution de classes ou de valeurs : Une fois l'arbre construit, chaque feuille de l'arbre est associée à une classe (dans le cas de la classification) ou à une valeur (dans le cas de la régression). Lors de la prédiction pour de nouvelles données, celles-ci traversent l'arbre en suivant les décisions basées sur les attributs, et la sortie est déterminée en fonction de la feuille atteinte.

Les avantages des Arbres de décision comprennent leur facilité d'interprétation, leur capacité à gérer des données mixtes (numériques et catégoriques), et leur résistance aux valeurs aberrantes. Cependant, ils sont sensibles au sur-apprentissage.

3.3.4/ RANDOM FORESTS

Les "Random forests" [78] sont une technique d'apprentissage supervisé, proposée par L. Breiman [35], qui s'appuie sur un ensemble d'arbres de décision. L'idée fondamentale derrière cette méthode est de créer une multitude d'arbres de décision indépendants, et de les combiner pour obtenir des prédictions plus robustes et précises.

Le fonctionnement des "Random forests" peut être décrit en plusieurs étapes :

1. Sélection aléatoire des échantillons : Pour chaque arbre de décision de la forêt, un échantillon de données est sélectionné de manière aléatoire à partir de l'ensemble de données d'entraînement. Cette sélection aléatoire se fait avec remplacement, ce qui signifie qu'un même exemple peut être choisi plusieurs fois pour un arbre donné.

2. Sélection aléatoire des caractéristiques : À chaque division d'un nœud de l'arbre, un sous-ensemble aléatoire des caractéristiques est choisi parmi l'ensemble complet de caractéristiques. Cela permet de réduire la corrélation entre les arbres et de rendre le modèle plus généralisable.

3. Construction des arbres : Chaque arbre est construit en suivant un processus de division récursif basé sur les caractéristiques sélectionnées. La division est effectuée en

maximisant une métrique d'impureté, généralement l'indice de Gini ou l'entropie, pour séparer les données en classes homogènes.

4. **Vote majoritaire** : Une fois que tous les arbres ont été construits, les prédictions de chaque arbre sont agrégées en effectuant un vote majoritaire (classification) ou une moyenne (régression). Cette agrégation renforce la stabilité et la précision des prédictions.

Les "Random forests" présentent plusieurs avantages :

- Ils sont robustes aux données bruitées et aux valeurs aberrantes grâce à l'agrégation.
- Ils sont adaptés à la classification et à la régression.
- Ils permettent d'estimer l'importance des caractéristiques dans la prise de décision.

Cependant, ils ont aussi quelques inconvénients, notamment la difficulté d'interpréter les résultats en raison de la complexité de l'ensemble d'arbres.

3.3.5/ GRADIENT BOOSTING (XGBOOST, LIGHTGBM)

Le Gradient boosting [37] [42] [30] est une technique d'apprentissage supervisé qui vise à améliorer la performance d'un modèle prédictif en combinant plusieurs modèles plus simples, généralement des arbres de décision. XGBoost et LightGBM sont deux bibliothèques de Gradient boosting largement utilisées dans le domaine des séries temporelles.

3.3.5.1/ XGBOOST (EXTREME GRADIENT BOOSTING)

XGBoost [84] est une bibliothèque de Gradient boosting conçue pour offrir une performance exceptionnelle en matière de prévision et d'analyse de séries temporelles. Son succès réside en grande partie dans les caractéristiques suivantes :

1. **Arbres de décision régularisés** : XGBoost utilise des arbres de décision régularisés qui réduisent le sous-sur-apprentissage et améliorent la généralisation du modèle.
2. **Optimisation de la fonction de perte** : Il utilise une approche d'optimisation itérative pour minimiser la fonction de perte, permettant ainsi une meilleure convergence du modèle.
3. **Élagage des arbres** : Les arbres sont élagués pour éviter la croissance excessive et améliorer l'efficacité du modèle.
4. **Parallélisme** : XGBoost est capable de tirer parti du parallélisme pour accélérer l'apprentissage et les prédictions.

5. **Gestion des valeurs manquantes** : Il gère efficacement les valeurs manquantes, ce qui est courant dans les données temporelles.

3.3.5.2/ LIGHTGBM (LIGHT GRADIENT BOOSTING MACHINE)

LightGBM [89] est une autre bibliothèque de Gradient boosting qui se distingue par sa rapidité d'exécution et sa capacité à gérer de grands ensembles de données. Les caractéristiques clés de LightGBM sont les suivantes :

1. Gradient boosting basé sur l'histogramme : Contrairement à d'autres méthodes, LightGBM utilise un histogramme de valeurs pour la construction des arbres, ce qui réduit la complexité et accélère l'apprentissage.

2. Optimisation de feuille : Il utilise une stratégie d'optimisation de feuille (Leaf-wise) au lieu d'une optimisation en profondeur, ce qui améliore l'efficacité.

3. Parallélisme et mémoire efficace : LightGBM est conçu pour fonctionner de manière efficace en parallèle et utilise moins de mémoire, ce qui le rend adapté à des ensembles de données volumineux.

4. Gestion des valeurs manquantes : Comme XGBoost, LightGBM gère également les valeurs manquantes de manière efficace.

3.3.6/ ENSEMBLES DE MODÈLES :

3.3.6.1/ BAGGING (BOOTSTRAP AGGREGATING)

Le Bagging, ou Bootstrap aggregating [22], est une technique d'ensemble utilisée en apprentissage automatique pour améliorer la robustesse et la performance des modèles prédictifs [95]. Il repose sur le principe de la création de multiples sous-ensembles d'entraînement à partir des données d'origine, permettant ainsi d'atténuer la variance des modèles individuels. Dans cette sous-section, nous détaillons le fonctionnement du Bagging.

Fonctionnement du bagging Le bagging fonctionne selon un ensemble de principes clés :

1. Bootstrapping : Tout d'abord, des échantillons de données sont générés à partir de l'ensemble d'entraînement d'origine en utilisant une méthode appelée bootstrapping. Le bootstrapping consiste à tirer aléatoirement avec remplacement des échantillons de la base de données initiale. En conséquence, chaque échantillon bootstrap est potentiellement différent de l'original et peut contenir des doublons.

2. Modèles individuels : Sur la base de ces ensembles bootstrap, un modèle de machine learning est formé pour chaque ensemble. Il peut s'agir de n'importe quel modèle, tel qu'un arbre de décision, une régression linéaire, ou même un réseau de neurones.

3. Agrégation des prédictions : Une fois que tous les modèles individuels sont formés, leurs prédictions sont agrégées pour obtenir une prédiction finale. La méthode d'agrégation dépend du type de tâche : pour la régression, la prédiction finale est souvent la moyenne des prédictions des modèles, tandis que pour la classification, il peut s'agir du vote majoritaire.

4. Réduction de la variance : L'une des principales motivations du bagging est de réduire la variance des modèles individuels. En tirant aléatoirement des échantillons bootstrap, chaque modèle est exposé à une variation différente des données d'entraînement, ce qui atténue la sensibilité aux fluctuations aléatoires dans les données.

Avantages du bagging Le bagging offre plusieurs avantages significatifs :

- **Amélioration de la performance** : En agrégeant les prédictions de plusieurs modèles, le bagging peut considérablement améliorer la performance par rapport à un modèle unique.
- **Réduction de la variance** : En atténuant la variance, le bagging contribue à rendre les modèles plus stables et à réduire le risque de surajustement.
- **Utilisable avec différents modèles** : Le bagging peut être appliqué à n'importe quel modèle d'apprentissage automatique, ce qui en fait une technique polyvalente.
- **Robustesse aux outliers** : Le fait de tirer des échantillons avec remplacement peut rendre le bagging plus robuste aux valeurs aberrantes (outliers) dans les données.

Limitations du bagging Cependant, le bagging n'est pas une solution universelle et présente également quelques limitations :

- **Pas d'interprétabilité** : L'agrégation des prédictions de plusieurs modèles rend souvent le modèle global moins interprétable que les modèles individuels.
- **Coût de calcul** : Former et agréger plusieurs modèles peut être coûteux en termes de calcul et de mémoire, en particulier avec de grands ensembles de données.

3.3.6.2/ STACKING

Le stacking, également connu sous le nom de stacked generalization [17], est une technique d'ensemble avancée en apprentissage automatique qui vise à améliorer la performance des modèles prédictifs en combinant les prédictions de plusieurs modèles de

base. Contrairement au bagging ou au boosting, qui utilisent un ensemble homogène de modèles de base, le stacking fait appel à une diversité de modèles de base pour tirer parti de leurs forces respectives. Dans cette sous-section, nous explorerons en détail le fonctionnement du stacking.

Fonctionnement du stacking Le stacking fonctionne en plusieurs étapes :

1. Ensemble de Données d'entraînement : L'ensemble de données d'entraînement est divisé en deux parties : une partie est utilisée pour former les modèles de base, et l'autre pour former le modèle de niveau supérieur (ou méta-modèle).

2. Modèles de base : Plusieurs modèles de base hétérogènes sont sélectionnés et formés sur la première partie des données d'entraînement. Ces modèles peuvent être de différents types (par exemple, arbres de décision, réseaux de neurones, régressions linéaires) et sont souvent choisis pour leur diversité.

3. Création des prédictions : Les modèles de base sont utilisés pour générer des prédictions sur la deuxième partie des données d'entraînement qui n'a pas été utilisée pour leur formation. Chaque modèle de base génère une série de prédictions pour chaque exemple de test.

4. Modèle de niveau supérieur (Méta-modèle) : Un modèle de niveau supérieur, appelé méta-modèle, est formé sur les prédictions des modèles de base. Le méta-modèle prend les prédictions des modèles de base comme entrée et apprend à combiner ces prédictions pour produire la prédiction finale. Il est généralement formé à l'aide d'une méthode de régression ou de classification.

5. Prédiction finale : Le méta-modèle est utilisé pour faire des prédictions sur de nouvelles données, c'est-à-dire pour générer la prédiction finale du modèle empilé.

Avantages du stacking Le stacking présente plusieurs avantages significatifs :

- **Amélioration de la performance** : En combinant les prédictions de modèles de base diversifiés, le stacking peut améliorer la performance par rapport à un modèle unique.
- **Flexibilité** : Le stacking est hautement flexible, car il permet d'incorporer différents types de modèles de base, offrant ainsi une grande polyvalence en termes de modélisation.
- **Meilleure gestion de la complexité** : Il peut gérer des problèmes complexes en utilisant des modèles de base spécialisés pour des sous-problèmes spécifiques.

Limitations du stacking Cependant, le stacking n'est pas exempt de limitations :

- **Complexité de mise en oeuvre** : La mise en oeuvre du stacking peut être complexe,

en particulier lorsqu'il s'agit de sélectionner les modèles de base et de définir la structure du méta-modèle.

- **Risque de surajustement** : Comme le méta-modèle est formé sur les prédictions des modèles de base, il existe un risque de surajustement si le nombre de modèles de base est trop élevé ou si la taille de l'ensemble de données est limitée.

3.3.6.3/ MODÈLES DE COMITÉS (ENSEMBLE MODELS)

Les modèles de comités, également appelés ensembles d'apprentissage, sont une famille de techniques d'ensemble en apprentissage automatique visant à améliorer la performance et la robustesse des modèles prédictifs en combinant les prédictions de plusieurs modèles individuels [48, 27, 19, 14]. Contrairement à des approches telles que le bagging ou le stacking, qui se concentrent sur la diversité des modèles, les modèles de comités mettent davantage l'accent sur l'agrégation de modèles de base similaires. Dans cette sous-section, nous explorerons en détail le fonctionnement des modèles de comités.

Fonctionnement des modèles de comités Les modèles de comités fonctionnent en plusieurs étapes :

1. Ensemble de données d'entraînement : L'ensemble de données d'entraînement est utilisé pour former plusieurs modèles individuels, généralement du même type, mais formés sur des sous-ensembles différents des données ou avec des paramètres initiaux différents.

2. Modèles de base : Les modèles individuels, souvent appelés membres du comité, sont formés sur des parties distinctes des données d'entraînement. Par exemple, dans le cas des arbres de décision, chaque membre peut être formé sur un échantillon bootstrap différent.

3. Prédiction des membres : Chaque membre du comité génère des prédictions sur l'ensemble des données d'entraînement.

4. Agrégation des prédictions : Les prédictions de chaque membre sont agrégées pour obtenir une prédiction finale. L'agrégation peut être réalisée par un vote majoritaire dans le cas de la classification ou par une moyenne dans le cas de la régression.

5. Prédiction finale : La prédiction agrégée est utilisée pour faire des prédictions sur de nouvelles données.

Avantages des modèles de comités Les modèles de comités présentent plusieurs avantages importants :

- **Amélioration de la performance** : En combinant les prédictions de plusieurs modèles similaires, les modèles de comités peuvent améliorer la performance par rapport à un modèle unique.
- **Robustesse** : Les modèles de comités sont plus robustes face au surajustement, car ils réduisent la variance des prédictions.
- **Réduction du risque de sélection de modèles** : Étant donné que plusieurs modèles similaires sont utilisés, les modèles de comités sont moins susceptibles de souffrir d'une sélection de modèle inappropriée.
- **Interprétabilité** : Les modèles de comités permettent souvent une meilleure interprétabilité que les ensembles de modèles hétérogènes.

Limitations des modèles de comités Cependant, les modèles de comités ne sont pas sans limitations :

- **Complexité de mise en œuvre** : La mise en œuvre de modèles de comités peut être complexe, en particulier lorsqu'il s'agit de sélectionner les membres du comité et de définir les critères d'agrégation.
- **Coût de calcul** : Comme plusieurs modèles individuels sont formés, les modèles de comités peuvent être coûteux en termes de calcul et de mémoire.
- **Limitation de la diversité** : Les modèles de comités se concentrent sur l'agrégation de modèles similaires, ce qui limite leur capacité à gérer des modèles très diversifiés.

3.3.7/ RÉSEAUX DE NEURONES : PRINCIPES FONDAMENTAUX

Les réseaux de neurones artificiels sont des modèles mathématiques inspirés par le fonctionnement du cerveau humain [10]. Ils sont largement utilisés en apprentissage profond pour résoudre des tâches complexes de classification [41], de régression [15] et d'autres problèmes d'apprentissage automatique. Cette section présente les principes fondamentaux des réseaux de neurones, en se concentrant sur l'architecture d'un neurone, les fonctions d'activation et les équations sous-jacentes.

3.3.7.1/ ARCHITECTURE D'UN NEURONE

Un neurone artificiel est l'unité de base d'un réseau de neurones. Il prend un ensemble de valeurs en entrée, effectue des calculs et produit une valeur en sortie. L'architecture

d'un neurone peut être représentée comme suit :

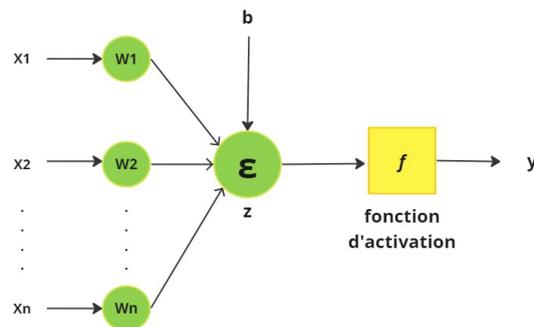


FIGURE 3.1 – Architecture d'un Neurone Artificiel

Un neurone reçoit des entrées x_1, x_2, \dots, x_n , chacune associée à un poids correspondant w_1, w_2, \dots, w_n . Il calcule la somme pondérée des entrées et des poids, à laquelle est ajouté un biais (bias) b :

$$z = \sum_{i=1}^n (x_i \cdot w_i) + b \quad (3.20)$$

La sortie du neurone est ensuite obtenue en appliquant une fonction d'activation f à la valeur z :

$$y = f(z) \quad (3.21)$$

La fonction d'activation joue un rôle crucial dans la capacité du neurone à modéliser des relations complexes.

3.3.7.2/ FONCTIONS D'ACTIVATION

Les fonctions d'activation déterminent la sortie d'un neurone en fonction de son entrée pondérée. Il existe plusieurs types de fonctions d'activation utilisées dans les réseaux de neurones. Voici quelques-unes des plus courantes :

Fonction sigmoïde

La fonction sigmoïde est une fonction d'activation couramment utilisée dans les réseaux de neurones [21]. Elle transforme son entrée en une sortie dans l'intervalle $[0, 1]$. La forme de la fonction sigmoïde ressemble à une courbe en forme de "S". Elle est définie comme suit :

$$f(z) = \frac{1}{1 + e^{-z}} \quad (3.22)$$

où z est la somme pondérée des entrées et des poids.

Unité linéaire rectifiée (ReLU)

La fonction ReLU est une fonction d'activation simple mais puissante [65, 91]. Elle est définie comme suit :

$$f(z) = \max(0, z) \quad (3.23)$$

Elle renvoie la valeur de l'entrée si elle est positive, sinon elle renvoie zéro. La fonction ReLU est particulièrement efficace pour traiter les problèmes de réseaux de neurones profonds en atténuant le problème de la disparition du gradient.

Fonction tangente hyperbolique (Tanh)

La fonction tangente hyperbolique est similaire à la fonction sigmoïde, mais elle renvoie des valeurs dans l'intervalle $[-1, 1]$. Elle est définie comme suit :

$$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (3.24)$$

La fonction tangente hyperbolique est souvent utilisée dans les couches cachées des réseaux de neurones [79, 118].

3.3.7.3/ CALCUL DE LA SORTIE DU RÉSEAU DE NEURONES

Les réseaux de neurones sont généralement construits en empilant plusieurs couches de neurones. L'entrée est propagée à travers le réseau couche par couche, en utilisant les équations décrites précédemment, jusqu'à ce que la sortie soit calculée.

Pour un réseau de neurones à une seule couche cachée, la sortie finale est calculée comme suit :

$$y = f(W \cdot x + b) \quad (3.25)$$

où x est le vecteur d'entrée, W est la matrice des poids de la couche cachée, b est le vecteur de biais, et f est la fonction d'activation.

3.3.8/ RÉSEAUX DE NEURONES PROFONDS (DNN) POUR LA RÉGRESSION TEMPORELLE

Les Réseaux de Neurones Profonds (DNN), également connus sous le nom de réseaux de neurones profonds artificiels, sont une catégorie de modèles de machine learning qui se sont révélés extrêmement puissants pour la régression temporelle. Cette approche repose sur la création de multiples couches de neurones artificiels, ce qui permet de capturer des relations complexes entre les entrées et les sorties. Dans cette sous-section, nous détaillerons le fonctionnement des DNN pour la régression temporelle.

3.3.8.1/ ARCHITECTURE DES RÉSEAUX DE NEURONES PROFONDS

Un réseau de neurones profond est composé de plusieurs couches, chacune contenant un certain nombre de neurones. Les couches sont généralement organisées en trois types principaux :

- 1. Couche d'entrée** : Cette couche reçoit les données d'entrée, généralement sous forme de vecteurs ou de tenseurs. Chaque neurone de cette couche représente une caractéristique d'entrée.
- 2. Couches cachées** : Les couches intermédiaires entre la couche d'entrée et la couche de sortie sont appelées couches cachées. Chaque neurone dans une couche cachée effectue une combinaison linéaire des valeurs provenant de la couche précédente, puis applique une fonction d'activation non linéaire, souvent une fonction sigmoïde, une fonction ReLU (Rectified Linear Unit) ou une fonction tangente hyperbolique.
- 3. Couche de sortie** : La couche de sortie génère les prédictions du modèle. Le nombre de neurones dans cette couche dépend du type de problème. Pour la régression temporelle, il y a généralement un seul neurone en sortie.

3.3.8.2/ FONCTIONNEMENT DES DNN POUR LA RÉGRESSION TEMPORELLE

Le fonctionnement d'un DNN pour la régression temporelle consiste en plusieurs étapes :

- 1. Initialisation** : Les poids et les biais des neurones sont initialisés de manière aléatoire ou avec des valeurs pré-entraînées.
- 2. Propagation avant** : Les données d'entrée sont propagées à travers le réseau en effectuant des opérations de multiplication matricielle et en appliquant les fonctions d'activation à chaque couche cachée. Cela génère une prédiction initiale.
- 3. Calcul de l'erreur** : L'erreur entre la prédiction du modèle et la valeur réelle de la série temporelle est calculée, généralement à l'aide d'une fonction de perte telle que l'erreur

quadratique moyenne (MSE) ou l'erreur absolue moyenne (MAE).

4. Rétropropagation de l'erreur : L'algorithme de rétropropagation ajuste les poids et les biais du réseau pour minimiser l'erreur. Cela se fait en utilisant des techniques d'optimisation telles que la descente de gradient (gradient descent).

5. Entraînement itératif : Les étapes 2 à 4 sont répétées de manière itérative sur un ensemble d'entraînement jusqu'à ce que le modèle atteigne une convergence satisfaisante.

6. Prédiction : Une fois le modèle entraîné, il peut être utilisé pour prédire des valeurs futures de la série temporelle en utilisant la propagation avant.

Ces différentes étapes sont sensiblement similaires pour les modèles présentés par la suite.

Les DNN sont connus pour leur capacité à capturer des modèles complexes dans les séries temporelles, mais ils nécessitent généralement des ensembles de données de grande taille et un réglage minutieux des hyperparamètres pour éviter le surapprentissage. Cependant, leur potentiel de généralisation en fait un outil précieux pour la régression temporelle dans divers domaines d'application.

3.3.8.3/ DISPARITION DU GRADIENT

La disparition du gradient est un problème courant qui peut survenir lors de l'entraînement de réseaux de neurones profonds. Ce problème se manifeste lorsque les gradients des fonctions de perte par rapport aux paramètres du modèle deviennent extrêmement petits à mesure que l'on remonte les couches du réseau, en particulier dans les couches initiales. Cela peut rendre l'apprentissage difficile voire impossible. Les causes et les conséquences de la disparition du gradient sont importantes à comprendre pour concevoir des réseaux de neurones efficaces.

3.3.8.4/ CAUSES DE LA DISPARITION DU GRADIENT

La disparition du gradient est principalement due à deux causes :

1. Fonctions d'activation à gradient faible : Certaines fonctions d'activation, comme la fonction sigmoïde et la fonction tangente hyperbolique, ont des gradients faibles pour des valeurs d'entrée éloignées de zéro. Lorsque ces fonctions sont utilisées dans des couches profondes, les gradients peuvent devenir très petits à mesure que l'information traverse le réseau.

2. Produit de matrices de poids : Lorsque les activations sont propagées à travers des couches profondes, elles sont multipliées par les matrices de poids à chaque couche. Si ces matrices de poids ont des valeurs proches de zéro, cela peut conduire à une

multiplication successive par des valeurs proches de zéro, réduisant ainsi les gradients.

3.3.8.5/ CONSÉQUENCES DE LA DISPARITION DU GRADIENT

La disparition du gradient a plusieurs conséquences :

1. Ralentissement de l'apprentissage : Lorsque les gradients sont très petits, la mise à jour des paramètres du modèle devient négligeable, ce qui ralentit considérablement le processus d'apprentissage.

2. Difficulté à apprendre des représentations profondes : Les réseaux de neurones profonds sont conçus pour apprendre des représentations hiérarchiques des données. La disparition du gradient rend difficile la formation de couches profondes, car les couches initiales ne parviennent pas à transmettre des informations significatives aux couches suivantes.

3. Instabilité du modèle : Les gradients très petits peuvent rendre le modèle instable, car de petites perturbations dans les données d'entrée ou les paramètres du modèle peuvent entraîner des changements importants dans les activations et les gradients.

3.3.8.6/ SOLUTIONS À LA DISPARITION DU GRADIENT

Plusieurs solutions ont été proposées pour atténuer le problème de la disparition du gradient dans les réseaux de neurones profonds :

1. Fonction d'activation ReLU : La fonction d'activation ReLU (Rectified Linear Unit) a un gradient non nul pour les valeurs d'entrée positives, ce qui aide à éviter la disparition du gradient. Elle est couramment utilisée dans les réseaux de neurones profonds.

2. Initialisation des paramètres : L'initialisation appropriée des paramètres du modèle, comme l'initialisation Xavier/Glorot, peut aider à éviter la disparition du gradient en s'assurant que les activations ont une variance constante.

3. Utilisation de couches de normalisation : Les techniques de normalisation, telles que la normalisation par lots (Batch Normalization), permettent de maintenir des activations à une échelle appropriée tout au long de l'entraînement.

4. Réseaux récurrents et architectures spéciales : Pour certaines tâches, l'utilisation de réseaux récurrents ou d'architectures spéciales, comme les réseaux résiduels (ResNets), peut aider à atténuer la disparition du gradient.

3.3.9/ RÉSEAUX DE NEURONES RÉCURRENTS (RNN)

Les Réseaux de Neurones Récurrents (RNN) sont une classe de modèles de réseaux de neurones adaptés au traitement de données séquentielles, ce qui en fait une puissante méthode pour la modélisation et la prédiction de séries temporelles. Dans cette sous-section, nous allons détailler le fonctionnement des RNN dans le contexte de l'analyse des séries temporelles.

3.3.9.1/ ARCHITECTURE DES RÉSEAUX DE NEURONES RÉCURRENTS

La principale caractéristique des RNN réside dans leur capacité à prendre en compte les informations contextuelles dans une séquence de données. La structure des RNN comprend des unités récurrentes (ou cellules) qui maintiennent un état interne. Chaque unité récurrente reçoit une entrée de la séquence actuelle ainsi que l'état interne de la séquence précédente.

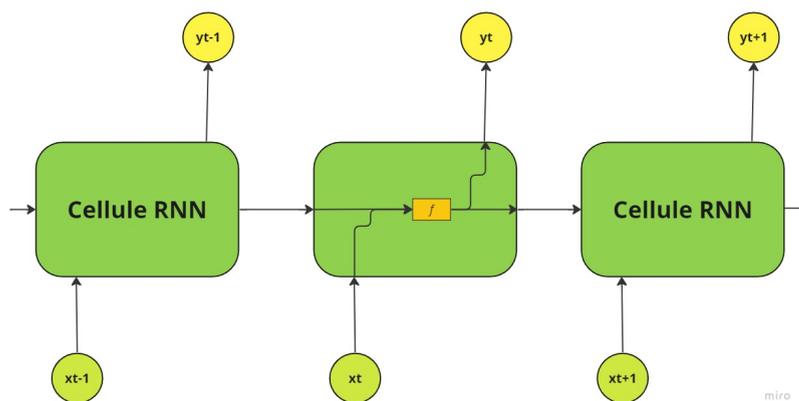


FIGURE 3.2 – Architecture d'une cellule RNN, chaque cellule reçoit en entrée l'état caché du pas de temps précédent en plus de la variable du pas de temps actuel.

Le calcul dans un RNN à l'instant t peut être exprimé comme suit :

$$h_t = f(W_{hh}h_{t-1} + W_{xh}x_t + b_h) \quad (3.26)$$

$$y_t = g(W_{hy}h_t + b_y) \quad (3.27)$$

Où :

- h_t est l'état caché à l'instant t .
- x_t est l'entrée à l'instant t .
- W_{hh} et W_{xh} sont les poids des connexions récurrentes et des entrées, respectivement.
- b_h est le biais de la couche cachée.
- y_t est la sortie à l'instant t .

- W_{hy} est le poids de la connexion entre la couche cachée et la couche de sortie.
- b_y est le biais de la couche de sortie.
- f et g sont des fonctions d'activation telles que la fonction sigmoïde ou la fonction tangente hyperbolique.

3.3.10/ LONG SHORT-TERM MEMORY (LSTM)

Les Long short-term memory (LSTM) sont une variation avancée des réseaux de neurones récurrents (RNN) qui ont été spécialement conçus pour résoudre le problème du "vanishing gradient" dans les séquences temporelles longues. Les LSTM sont particulièrement adaptés pour modéliser des dépendances temporelles à long terme, ce qui les rend extrêmement pertinents dans le contexte de l'analyse des séries temporelles.

3.3.10.1/ ARCHITECTURE DES LSTM

La structure des LSTM est plus complexe que celle des RNN standard. Chaque unité LSTM, ou cellule, maintient un état cellulaire C_t et un état caché h_t . La mise à jour de ces états est contrôlée par des portes, à savoir la porte d'oubli (forget gate), la porte d'entrée (input gate), et la porte de sortie (output gate). Ces portes régulent le flux d'informations à travers la cellule LSTM.

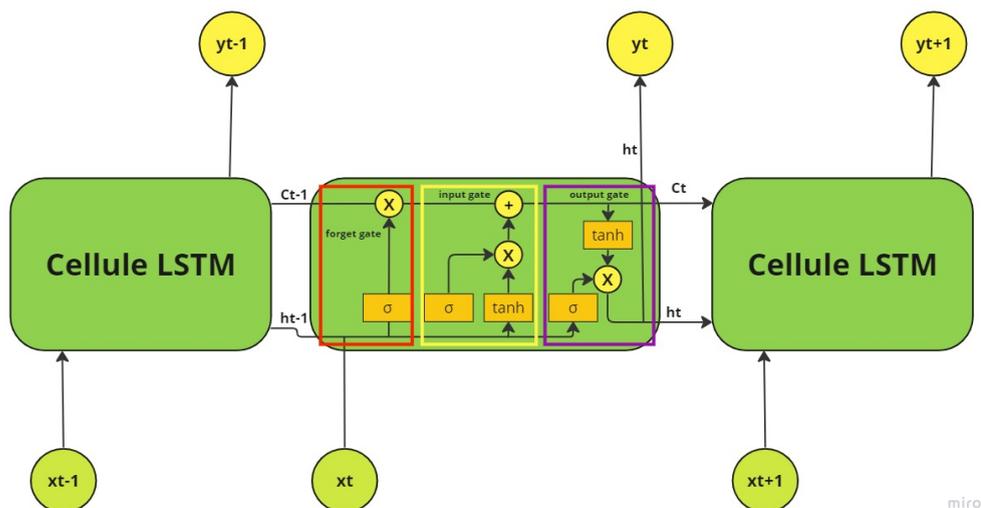


FIGURE 3.3 – Architecture d'une cellule LSTM, chaque cellule reçoit en entrée l'état caché et l'état cellulaire du pas de temps précédent en plus de la variable du pas de temps actuel.

Le calcul de l'état cellulaire et de l'état caché dans une unité LSTM à l'instant t peut être formulé comme suit :

$$\begin{aligned}
 f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) && \text{(Porte d'oubli)} \\
 i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) && \text{(Porte d'entrée)} \\
 \tilde{C}_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) && \text{(Candidat pour l'état cellulaire)} \\
 C_t &= f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t && \text{(Mise à jour de l'état cellulaire)} \\
 o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) && \text{(Porte de sortie)} \\
 h_t &= o_t \cdot \tanh(C_t) && \text{(Mise à jour de l'état caché)}
 \end{aligned}$$

Où :

- f_t , i_t , et o_t sont les vecteurs de la porte d'oubli, de la porte d'entrée, et de la porte de sortie, respectivement.
- \tilde{C}_t est le candidat pour l'état cellulaire.
- σ est la fonction sigmoïde.
- \tanh est la fonction tangente hyperbolique.
- W_f , W_i , W_c , et W_o sont les matrices de poids correspondantes.
- b_f , b_i , b_c , et b_o sont les biais associés.

Les LSTM ont montré une capacité à capturer des dépendances temporelles complexes, ce qui en fait un choix pour de nombreuses applications de prévision de séries temporelles, telles que la finance, la météorologie et la santé.

3.3.11/ GATED RECURRENT UNIT (GRU)

Les Gated recurrent unit (GRU) sont une autre variation avancée des réseaux de neurones récurrents (RNN) qui ont été développés pour résoudre certains des problèmes associés aux RNN classiques, tout en conservant une architecture plus simple que les Long Short-Term Memory (LSTM). Les GRU sont largement utilisés dans l'analyse des séries temporelles en raison de leur capacité à capturer des dépendances à long terme tout en étant plus efficaces en termes de calcul.

3.3.11.1/ ARCHITECTURE DES GATED RECURRENT UNIT (GRU)

Les GRU comportent une seule unité, ou cellule, par pas de temps. Cette unité est dotée de deux portes principales : la porte de réinitialisation (reset gate) et la porte d'actualisation (update gate). Les GRU maintiennent également un état caché qui est mis à jour à chaque instant de temps.

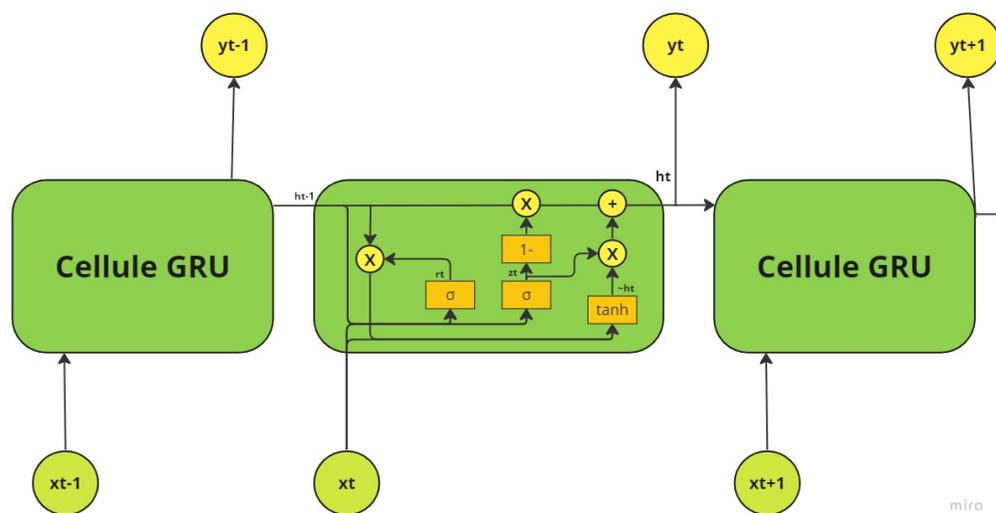


FIGURE 3.4 – Architecture d'une cellule GRU, chaque cellule reçoit en entrée l'état caché du pas de temps précédent en plus de la variable du pas de temps actuel. Sur cette représentation h_t représente le candidat pour l'état caché : \tilde{h}_t .

Le calcul dans une unité GRU à l'instant t peut être formulé comme suit :

$$\begin{aligned}
 r_t &= \sigma(W_r \cdot [h_{t-1}, x_t]) && \text{(Porte de réinitialisation)} \\
 z_t &= \sigma(W_z \cdot [h_{t-1}, x_t]) && \text{(Porte d'actualisation)} \\
 \tilde{h}_t &= \tanh(W \cdot [r_t \odot h_{t-1}, x_t]) && \text{(Candidat pour l'état caché)} \\
 h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t && \text{(Mise à jour de l'état caché)}
 \end{aligned}$$

Où :

- r_t est le vecteur de la porte de réinitialisation.
- z_t est le vecteur de la porte d'actualisation.
- \tilde{h}_t est le candidat pour l'état caché.
- σ est la fonction sigmoïde.
- \tanh est la fonction tangente hyperbolique.
- \odot représente la multiplication élément par élément.
- W_r , W_z , et W sont les matrices de poids correspondantes.

Les GRU ont montré des performances plus modestes par rapport aux LSTM dans la prédiction des séries temporelles, c'est pour cela qu'ils n'ont pas été utilisés par la suite, dans les travaux de cette thèse.

3.3.12/ RÉSEAUX DE NEURONES CONVOLUTIFS (CNN) POUR LES SÉRIES TEMPORELLES

Les Réseaux de neurones convolutifs (CNN), initialement développés pour la vision par ordinateur, ont été adaptés avec succès à l'analyse des séries temporelles. Les CNN se sont révélés efficaces pour extraire des caractéristiques pertinentes des données séquentielles, ce qui les rend pertinents dans le domaine de l'analyse des séries temporelles.

3.3.12.1/ ARCHITECTURE DES RÉSEAUX DE NEURONES CONVOLUTIFS (CNN)

L'architecture des CNN repose sur l'utilisation de filtres de convolution pour extraire des caractéristiques locales des données d'entrée. Dans le contexte des séries temporelles, les filtres de convolution sont appliqués à des fenêtres glissantes de la série pour détecter des motifs temporels. Les couches convolutives sont généralement suivies de couches de sous-échantillonnage (pooling) pour réduire la dimension des caractéristiques extraites.

Le calcul dans une couche de convolution pour les séries temporelles peut être exprimé comme suit :

$$h_i = f(W \cdot x_{i:i+k-1} + b) \quad (3.28)$$

Où :

- h_i est la sortie à l'emplacement i après la convolution.
- f est une fonction d'activation, généralement la fonction ReLU.
- W est le noyau de convolution (filtre).
- $x_{i:i+k-1}$ est la fenêtre de données d'entrée de longueur k à l'emplacement i .
- b est le biais.

Les CNN sont particulièrement efficaces pour capturer des motifs locaux dans les séries temporelles, ce qui les rend utiles pour des tâches telles que la détection d'anomalies ou la classification de motifs. Cependant, ils peuvent nécessiter des ensembles de données de grande taille pour obtenir de bons résultats, en particulier lorsque les motifs temporels sont complexes et variés.

3.3.13/ MODÈLES TRANSFORMERS

Les modèles Transformers ont révolutionné le domaine de l'apprentissage automatique, en particulier pour le traitement du langage naturel et ont eu des résultats intéressants

pour la prédiction des séries temporelles à long terme. Ils ont été introduits pour surmonter les limitations des réseaux de neurones récurrents (RNN) et des réseaux de neurones convolutifs (CNN) en ce qui concerne la capture de dépendances à longue distance dans les données séquentielles. Les principaux concepts des modèles Transformers incluent l'auto-attention, les couches d'attention multi-têtes et les mécanismes d'encodage et de décodage.

3.3.13.1/ INPUT EMBEDDINGS

L'Input embeddings est la première étape dans le traitement des données d'entrée dans un modèle d'IA Transformer. Il est responsable de la conversion des séquences de tokens (mots, caractères, etc.) en vecteurs numériques de dimensions appropriées pour le traitement par le réseau. Cette opération est essentielle car elle permet au modèle de représenter chaque token de manière continue et d'apprendre des relations sémantiques entre les tokens. Pour les séries temporelles composées de valeurs numériques, la même méthode peut être appliquée en substituant les tokens par les valeurs de la séries temporelles (dans le reste de cette partie explicative sur les Transformers, le terme "token" sera utilisé pour désigner indifféremment toute valeur de la séquence, qu'elle soit numérique ou non).

Soit $X = \{x_1, x_2, \dots, x_n\}$ une séquence de n tokens, où x_i est le token à la position i . Les Input Embeddings sont calculés comme suit :

$$E = \text{Embed}(X) \quad (3.29)$$

où Embed est une fonction d'embedding qui associe chaque token x_i à un vecteur d'embedding $e_i \in \mathbb{R}^d$, où d est la dimension de l'embedding. Ainsi, E est une matrice de dimensions (n, d) contenant les embeddings de tous les tokens de la séquence.

3.3.13.2/ POSITIONAL ENCODING

Contrairement aux modèles récurrents, les modèles d'IA Transformers n'ont pas d'invariance à l'ordre des tokens dans une séquence. Pour remédier à cela, un Positional encoding est ajouté aux embeddings des tokens pour incorporer l'information de position. Le Positional Encoding est essentiel pour que le modèle comprenne l'ordre séquentiel des tokens.

Le Positional encoding P est calculé comme suit :

$$P_{i,j} = \begin{cases} \sin\left(\frac{i}{10000^{2j/d}}\right) & \text{si } j \text{ est pair} \\ \cos\left(\frac{i}{10000^{2j/d}}\right) & \text{si } j \text{ est impair} \end{cases} \quad (3.30)$$

où i est la position du token dans la séquence et j est la dimension de l'embedding. Le Positional encoding est ensuite ajouté à la matrice des embeddings E :

$$E' = E + P \quad (3.31)$$

Ainsi, E' contient à la fois les informations sémantiques des tokens et leur position dans la séquence.

3.3.13.3/ AUTO-ATTENTION (SELF-ATTENTION) / SCALED DOT-PRODUCT ATTENTION

Le mécanisme fondamental des Transformers est l'auto-attention, qui permet au modèle de capturer les relations entre les différentes parties de la séquence d'entrée. L'auto-attention est basée sur des poids (ou scores) qui mesurent l'importance de chaque élément de la séquence par rapport à tous les autres éléments.

Le scaled dot-product attention est utilisé pour calculer cette nouvelle représentation. L'attention est calculée en utilisant trois matrices : la matrice des requêtes (Q), la matrice des clés (K) et la matrice des valeurs (V), toutes dérivées des embeddings des tokens.

Pour chaque token à la position i dans la séquence, le scaled dot-product attention calcule un vecteur d'attention A_i en effectuant les étapes suivantes :

1. Calcul des scores d'attention : Les scores d'attention sont calculés en prenant le produit scalaire entre la requête q_i associée au token à la position i et les clés k_j associées à tous les tokens de la séquence :

$$\text{Score}(q_i, k_j) = \frac{q_i \cdot k_j}{\sqrt{d_k}} \quad (3.32)$$

où d_k est la dimension des vecteurs de clés. Le facteur d'échelle $\sqrt{d_k}$ est utilisé pour stabiliser les valeurs des scores.

2. Calcul des poids d'attention : Les scores d'attention sont ensuite transformés en poids d'attention à l'aide de la fonction softmax, ce qui permet de normaliser les scores pour qu'ils soient compris entre 0 et 1 :

$$\text{Attention}(q_i, k_j) = \frac{\exp(\text{Score}(q_i, k_j))}{\sum_{l=1}^n \exp(\text{Score}(q_i, k_l))} \quad (3.33)$$

où n est le nombre de tokens dans la séquence.

3. Calcul de la représentation pondérée : Les poids d'attention sont utilisés pour calculer une nouvelle représentation r_i pour le token à la position i , en combinant les valeurs v_j associées à tous les tokens de la séquence :

$$r_i = \sum_{j=1}^n \text{Attention}(q_i, k_j) \cdot v_j \quad (3.34)$$

Ainsi, la nouvelle représentation r_i est une combinaison pondérée des valeurs v_j , où les poids sont déterminés par l'attention.

La représentation pondérée de chaque élément de la séquence peut se résumer à cette équation :

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3.35)$$

où Q , K , et V sont les matrices des vecteurs de requête, de clé et de valeur, respectivement, et d_k est la dimension des vecteurs de clé.

4. Projection linéaire : Enfin, la représentation pondérée r_i est généralement soumise à une projection linéaire pour produire la sortie finale du scaled dot-product attention :

$$\text{Output}(q_i) = r_i \cdot W_o \quad (3.36)$$

où W_o est une matrice de poids de projection.

3.3.13.4/ COUCHES D'ATTENTION MULTI-TÊTES (MULTI-HEAD ATTENTION)

Pour capturer différentes relations entre les éléments de la séquence, les modèles Transformers utilisent des couches d'attention multi-têtes. Chaque tête d'attention est une projection linéaire des vecteurs de requête, de clé et de valeur. Ces projections permettent au modèle d'apprendre différentes représentations des données d'entrée et d'effectuer des calculs d'attention indépendants.

La sortie des têtes d'attention est ensuite concaténée et transformée linéairement pour produire la sortie finale de la couche multi-têtes :

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^O \quad (3.37)$$

où $\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$ est la sortie de la tête i , W_i^Q , W_i^K , et W_i^V sont les

matrices de projection pour la tête i , et W^O est la matrice de projection finale.

3.3.13.5/ RÉSEAU FEED FORWARD

Le réseau feed forward est une composante clé de chaque couche d'un modèle d'IA Transformer. Il est conçu pour effectuer des transformations non linéaires des embeddings des tokens, en utilisant des couches de neurones entièrement connectées. Le Réseau feed forward est généralement composé de deux couches linéaires séparées par une fonction d'activation, telle que la fonction ReLU.

Soit H la matrice des embeddings de sortie après le positional encoding. Le réseau feed forward est défini comme suit :

$$\text{FFN}(H) = f(H \cdot W_1 + b_1) \cdot W_2 + b_2 \quad (3.38)$$

où W_1 , b_1 , W_2 et b_2 sont les poids et les biais des couches linéaires du réseau feed forward, et f est une fonction d'activation. Cette opération permet d'introduire des transformations non linéaires dans la représentation des tokens.

3.3.13.6/ MODÈLES TRANSFORMERS APPLIQUÉS AUX SÉRIES TEMPORELLES

Encodage des séries temporelles Pour appliquer les modèles Transformers aux séries temporelles, la première étape consiste à représenter les données sous forme de séquences. Chaque séquence temporelle devient une séquence d'entrée pour le modèle. Les valeurs temporelles successives sont traitées comme des éléments consécutifs de la séquence.

Chaque élément de la séquence d'entrée est associé à un vecteur de caractéristiques. Ce vecteur peut inclure des informations sur la valeur temporelle elle-même, les caractéristiques statistiques de la série à un moment donné, ou d'autres informations pertinentes.

Mécanismes d'auto-attention Le mécanisme d'auto-attention des modèles Transformers est essentiel pour la capture des relations entre les différentes parties de la séquence temporelle. L'auto-attention permet au modèle de donner plus de poids aux éléments de la séquence qui sont plus importants pour la prédiction à un instant donné.

Concrètement, chaque élément de la séquence est associé à des vecteurs de requête, de clé et de valeur, et les poids d'auto-attention sont calculés pour mesurer l'importance de chaque élément par rapport à tous les autres éléments. Les poids sont calculés à l'aide de fonctions d'affinité et normalisés pour former une distribution de proba-

bilité. L'auto-attention est appliquée à chaque élément de la séquence pour obtenir des représentations pondérées.

Empilement de couches Comme dans les autres applications des modèles Transformers, les couches d'auto-attention et les couches entièrement connectées sont empilées pour former un modèle profond. Chaque couche capture des dépendances à différentes échelles temporelles, ce qui permet au modèle de modéliser des motifs complexes dans les séries temporelles.

L'empilement de couches permet également d'obtenir des représentations hiérarchiques des séries temporelles, en capturant à la fois des motifs à court terme et des motifs à long terme. Cela rend les modèles Transformers particulièrement adaptés à la modélisation des séries temporelles, qui peuvent avoir des structures temporelles complexes.

3.3.14/ VALIDATION CROISÉE (CROSS-VALIDATION)

La validation croisée [31] est une technique essentielle dans l'évaluation des modèles d'apprentissage automatique, en particulier lors de la sélection des hyperparamètres et de l'estimation de la performance d'un modèle. Elle vise à estimer de manière fiable la capacité d'un modèle à généraliser à de nouvelles données. Dans cette sous-section, nous explorerons en détail le fonctionnement de la validation croisée.

3.3.14.1/ CONCEPT DE LA VALIDATION CROISÉE

La validation croisée est basée sur le principe de diviser l'ensemble de données en plusieurs ensembles d'entraînement et de test, permettant ainsi d'évaluer la performance du modèle sur différentes portions des données. Le processus se déroule en plusieurs étapes :

- 1. Partitionnement des données** : Les données sont divisées en k sous-ensembles mutuellement exclusifs, appelés "plis" (folds). Typiquement, $k = 5$ ou $k = 10$ est choisi.
- 2. Entraînement et évaluation** : Le modèle est entraîné k fois, chaque fois en utilisant $k - 1$ plis comme ensemble d'entraînement et le pli restant comme ensemble de test. La performance est évaluée k fois, une fois pour chaque pli de test.
- 3. Moyenne des performances** : Les mesures de performance (par exemple, l'erreur de validation) obtenues pour chaque pli de test sont moyennées pour obtenir une estimation globale de la performance du modèle.

3.3.14.2/ TYPES DE VALIDATION CROISÉE

Il existe plusieurs variantes de la validation croisée, chacune adaptée à des scénarios particuliers :

1. Validation croisée K-Fold (K-Fold cross-validation) : Comme décrit précédemment, les données sont divisées en k plis, et le modèle est évalué k fois. C'est l'approche de validation croisée la plus couramment utilisée.

2. Validation croisée leave-one-out (LOOCV) : Chaque échantillon est utilisé comme pli de test une seule fois, tandis que les $k - 1$ autres échantillons sont utilisés comme ensemble d'entraînement. Cette méthode est utile lorsque les données sont limitées, mais elle peut être coûteuse en calcul.

3. Validation croisée leave-p-out (LPOCV) : Similaire au LOOCV, mais permet de laisser de côté un nombre spécifié p d'échantillons à chaque itération.

4. Validation croisée stratifiée (Stratified cross-validation) : Utilisée principalement pour les tâches de classification, elle garantit que chaque pli a une distribution similaire de classes que l'ensemble de données global.

5. Validation croisée pour les séries temporelles : Le principe est d'utiliser comme jeu de validation, les données suivant le jeu d'entraînement. De cette façon on conserve l'ordre temporel de la série, contrairement à la sélection aléatoire de données dans les validations croisées "classiques".

3.3.14.3/ AVANTAGES DE LA VALIDATION CROISÉE

La validation croisée présente plusieurs avantages :

- Elle permet une évaluation plus robuste des performances du modèle en utilisant plusieurs partitions des données.
- Elle aide à réduire le biais résultant de l'utilisation d'une seule division fixe des données.
- Elle facilite la sélection des hyperparamètres en fournissant une estimation fiable de la performance.

3.3.14.4/ LIMITATIONS DE LA VALIDATION CROISÉE

Cependant, la validation croisée n'est pas sans limites :

- Elle peut être coûteuse en calcul, en particulier avec un grand nombre de plis.

- Elle peut introduire de la variation dans les mesures de performance, ce qui peut rendre difficile la comparaison entre les modèles.
- Elle ne remplace pas complètement l'importance d'un ensemble de test indépendant pour évaluer la performance finale du modèle sur des données inconnues.

3.3.15/ RECHERCHE D'HYPERPARAMÈTRES BAYÉSIENNE

La recherche d'hyperparamètres est une étape cruciale dans le développement de modèles d'intelligence artificielle, car elle permet de déterminer les valeurs optimales des paramètres qui régissent le comportement du modèle. La recherche d'hyperparamètres bayésienne [106] est une approche systématique et probabiliste visant à optimiser ces paramètres en se basant sur des méthodes d'optimisation bayésienne. Dans cette sous-section, nous explorerons en détail le fonctionnement de la recherche d'hyperparamètres bayésienne.

3.3.15.1/ FONDEMENTS DE LA RECHERCHE D'HYPERPARAMÈTRES BAYÉSIENNE

La recherche d'hyperparamètres bayésienne repose sur deux éléments clés : un modèle probabiliste pour la fonction objectif à optimiser (généralement appelée la "fonction de perte" ou "loss function"), et une méthode pour explorer efficacement l'espace des hyperparamètres. Les éléments fondamentaux de la recherche bayésienne d'hyperparamètres comprennent :

1. Fonction objectif probabiliste : Au lieu de chercher directement les valeurs optimales des hyperparamètres, la recherche bayésienne maintient une distribution probabiliste sur la fonction objectif, généralement modélisée comme une distribution de probabilité gaussienne.

2. Choix de l'explorateur (acquisition function) : L'explorateur est une fonction qui guide la sélection des hyperparamètres à évaluer dans le but de maximiser l'espérance de la fonction de perte. Les choix courants d'explorateurs incluent l'UCB (Upper Confidence Bound), l'EI (Expected Improvement), et le PI (Probability of Improvement).

3. Mise à jour de la distribution postérieure : Après chaque évaluation de la fonction de perte, la distribution probabiliste est mise à jour en utilisant le théorème de Bayes, permettant ainsi d'affiner la compréhension du comportement de la fonction objectif.

3.3.15.2/ ÉTAPES DE LA RECHERCHE D'HYPERPARAMÈTRES BAYÉSIENNE

Le processus de recherche d'hyperparamètres bayésienne peut être divisé en plusieurs étapes :

1. Initialisation : Une première évaluation de la fonction de perte est réalisée pour une configuration d'hyperparamètres initiale.

2. Modélisation probabiliste : Un modèle probabiliste initial est construit pour la fonction de perte en utilisant les résultats de l'évaluation initiale. Cette modélisation probabiliste permet d'estimer la distribution de la fonction de perte.

3. Choix de l'explorateur : L'explorateur est choisi en fonction du modèle probabiliste actuel, et il est utilisé pour déterminer la prochaine configuration d'hyperparamètres à évaluer. L'explorateur vise à équilibrer l'exploration (essayer de nouvelles configurations) et l'exploitation (explorer des configurations prometteuses).

4. Évaluation de la fonction de perte : La configuration d'hyperparamètres suggérée par l'explorateur est évaluée en utilisant la fonction de perte sur les données d'entraînement.

5. Mise à jour de la modélisation probabiliste : Après chaque évaluation, le modèle probabiliste est mis à jour en tenant compte du résultat. Cela permet de raffiner la distribution de la fonction de perte.

6. Répétition : Les étapes 3 à 5 sont répétées un certain nombre de fois ou jusqu'à ce qu'un critère d'arrêt soit atteint, tel que le temps d'exécution maximal ou un nombre d'itérations prédéfini.

3.3.15.3/ AVANTAGES DE LA RECHERCHE D'HYPERPARAMÈTRES BAYÉSIENNE

La recherche d'hyperparamètres bayésienne offre plusieurs avantages :

- Elle est capable de trouver rapidement des configurations d'hyperparamètres prometteuses, ce qui la rend efficace pour les modèles complexes et coûteux en calcul.
- Elle tient compte de l'incertitude dans les évaluations de la fonction de perte, ce qui permet une exploration plus prudente de l'espace des hyperparamètres.
- Elle peut être étendue pour gérer des optimisations multi-objectifs et des contraintes sur les hyperparamètres.

Cependant, la recherche d'hyperparamètres bayésienne n'est pas sans inconvénients, notamment en termes de complexité de mise en œuvre et de coût computationnel. Néanmoins, elle demeure une méthode puissante pour la sélection automatique des hyperparamètres dans le contexte de l'IA et de l'apprentissage automatique.

3.3.16/ TECHNIQUES D'OPTIMISATION (GRID SEARCH, RANDOM SEARCH)

Le processus de recherche des meilleurs hyperparamètres pour un modèle d'apprentissage automatique est essentiel pour obtenir des performances optimales. Deux des approches les plus courantes pour effectuer cette tâche sont la recherche en grille (Grid search) et la recherche aléatoire (random search). Dans cette sous-section, nous explorerons en détail le fonctionnement de ces techniques d'optimisation.

3.3.16.1/ RECHERCHE EN GRILLE (GRID SEARCH)

La recherche en grille est une méthode systématique d'optimisation des hyperparamètres qui explore un ensemble prédéfini de combinaisons d'hyperparamètres [124]. Le processus se déroule en plusieurs étapes :

- 1. Définition de la grille :** Pour chaque hyperparamètre à optimiser, une liste de valeurs candidates est définie. Par exemple, pour un hyperparamètre de taux d'apprentissage, la grille pourrait contenir des valeurs telles que $\{0.1, 0.01, 0.001\}$.
- 2. Combinaisons de grille :** Toutes les combinaisons possibles d'hyperparamètres sont générées en utilisant les valeurs de la grille. Par exemple, si nous avons trois hyperparamètres avec trois valeurs chacun, il y aurait $3^3 = 27$ combinaisons possibles.
- 3. Évaluation des modèles :** Pour chaque combinaison d'hyperparamètres, un modèle est formé et évalué à l'aide d'une métrique de performance prédéfinie, généralement à l'aide de la validation croisée.
- 4. Sélection du meilleur modèle :** La combinaison d'hyperparamètres qui donne la meilleure performance sur la métrique choisie est sélectionnée comme modèle final.

3.3.16.2/ RECHERCHE ALÉATOIRE (RANDOM SEARCH)

La recherche aléatoire [68] est une approche d'optimisation des hyperparamètres qui explore l'espace des hyperparamètres en utilisant des combinaisons aléatoires plutôt qu'une grille prédéfinie. Le processus se déroule comme suit :

- 1. Définition des plages :** Au lieu de spécifier des valeurs précises pour chaque hyperparamètre, des plages ou des distributions de valeurs potentielles sont définies. Par exemple, un taux d'apprentissage pourrait être tiré d'une distribution uniforme entre 0.001 et 0.1.
- 2. Échantillonnage aléatoire :** Des combinaisons aléatoires d'hyperparamètres sont générées en échantillonnant des valeurs dans les plages définies pour chaque hyperparamètre.

3. Évaluation des modèles : Chaque combinaison d'hyperparamètres est utilisée pour former un modèle et évaluer sa performance à l'aide de la validation croisée ou d'un ensemble de validation distinct.

4. Sélection du meilleur modèle : Le modèle avec la meilleure performance sur la métrique choisie est sélectionné comme modèle final.

3.3.16.3/ AVANTAGES ET INCONVÉNIENTS

Chaque technique d'optimisation des hyperparamètres présente ses avantages et ses inconvénients :

Recherche en grille (Grid search) :

Avantages :

- Systématique et exhaustif.
- Garantit que toutes les combinaisons sont explorées.

Inconvénients :

- Coûteux en termes de calcul lorsque l'espace des hyperparamètres est vaste.
- Peut être inefficace si des hyperparamètres sont moins importants que d'autres.

Recherche aléatoire (Random search) :

Avantages :

- Plus efficace en termes de calcul, en particulier pour les grands espaces d'hyperparamètres.
- Peut découvrir des combinaisons performantes de manière inattendue.

Inconvénients :

- Peut ne pas explorer toutes les combinaisons, ce qui peut entraîner une performance sous-optimale dans certains cas.

3.4/ CONCLUSION

Dans ce chapitre, nous avons examiné divers modèles et techniques d'optimisation pertinents pour la prédiction des séries temporelles. Nous avons expliqué ce que sont les séries temporelles dans un premier temps. Puis dans un deuxième temps, nous avons passé en revue quelques modèles classiquement utilisés, comme les modèles autoregressifs, les moyennes mobiles, le modèle ARIMA ou encore les modèles ETS. Par la suite, nous avons expliqué les différentes notions de l'apprentissage supervisé, ainsi que des techniques de validation et de test des modèles. Nous avons également abordé les concepts de sur-apprentissage et de sous-apprentissage, et présenté différents optimiseurs de machine learning. Une présentation de modèles de machine learning a été faite avant de s'intéresser plus en détail aux concepts des modèles à base de réseaux de neurones. Les modèles à base d'attention ont également été présentés. Pour finir, nous avons présenté des techniques d'optimisation des paramètres pour les modèles. Il est important de noter qu'il existe de nombreuses autres approches dans ce domaine, et qui ne seront pas, ou peu, traitées dans cette thèse. On peut citer des techniques de sélection de variables, d'analyse de séries temporelles (détections de points de rupture, détections d'anomalies...) ou des techniques d'IA plus générales comme les algorithmes génétiques ou l'apprentissage par renforcement. Une présentation de certaines méthodes de clustering sera par ailleurs faite dans le chapitre 6 de cette thèse. Cependant, nous allons maintenant étudier les applications potentielles de ces méthodes dans un contexte industriel, en mettant l'accent sur les secteurs associés aux grossistes-répartiteurs pharmaceutiques.

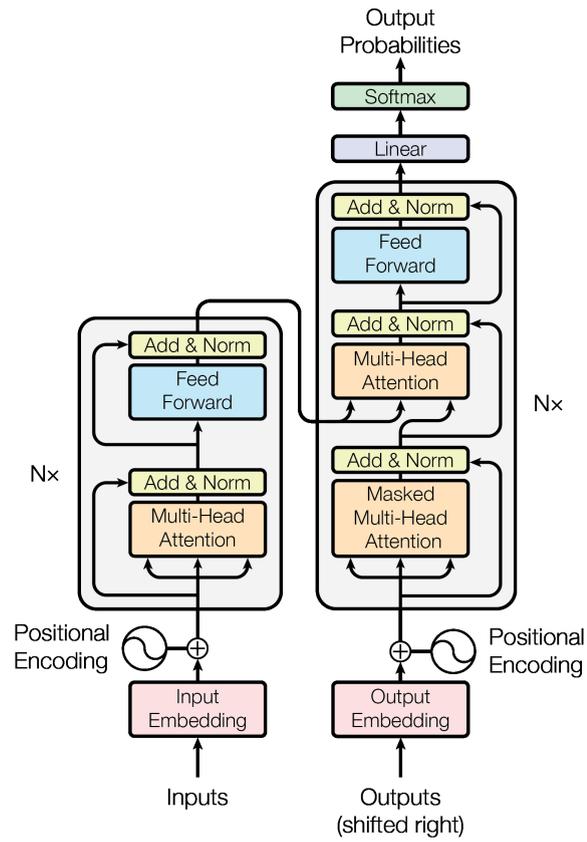


FIGURE 3.5 – Architecture des modèles de transformers comme décrite dans les travaux de Vaswani et al.

TECHNIQUES D'IA POUR LES SECTEURS INDUSTRIELS LIÉES AUX MÉTIERS DE GROSSISTES RÉPARTITEURS

4.1/ INTRODUCTION

L'industrie pharmaceutique est en constante mutation, et les grossistes-répartiteurs jouent un rôle fondamental dans la distribution des médicaments, assurant le lien entre les fabricants et les pharmacies. Dans ce contexte, trois secteurs majeurs interagissent étroitement avec l'activité des grossistes répartiteurs pharmaceutiques : le secteur pharmaceutique, qui englobe la production des médicaments en amont ; le secteur logistique, qui est intimement lié au cœur de métier des grossistes-répartiteurs ; et en aval, le secteur de l'officine de pharmacie, dont les besoins et les exigences exercent une influence directe sur les décisions et les actions des grossistes répartiteurs.

Dans cette perspective, l'intelligence artificielle (IA) a émergé comme un outil particulièrement prometteur pour améliorer divers aspects de l'industrie pharmaceutique. Elle offre des opportunités significatives pour l'optimisation des opérations des grossistes répartiteurs, l'amélioration de la recherche pharmaceutique et la mise à disposition de nouvelles ressources et solutions pour les officines de pharmacie.

Le fait d'utiliser des techniques d'intelligence artificielle pour améliorer le travail et la prise de décisions des utilisateurs peut être également appelé "intelligence augmentée".

Ce chapitre d'état de l'art a pour objectif d'examiner en détail les approches majeures et les travaux de recherche qui font usage de l'IA dans les domaines de la logistique et de la pharmacie. De plus, il vise à mettre en lumière les applications potentielles de ces avancées technologiques pour les officines de pharmacie.

4.2/ UTILISATION DE L'IA DANS LE SECTEUR LOGISTIQUE

Le métier de grossiste-répartiteur est étroitement lié au secteur de la logistique ; de ce fait, les problématiques de ces métiers sont également pertinentes pour notre sujet d'étude. L'analyse des travaux de recherche dans ce domaine nous offre donc des pistes de réflexion et des bases pour nos travaux.

4.2.1/ PRÉDICTION DE LA DEMANDE

La prédiction de la demande (autant celle de la clientèle que celle des officines de pharmacies en amont), est essentielle pour garantir un approvisionnement adéquat et éviter les excédents ou les ruptures de stock. Les techniques d'IA, telles que les réseaux de neurones (comme les LSTM [133] ou les transformers [136] évoqués au chapitre précédent), les forêts aléatoires [113], ou encore des méthodes statistiques plus classiques, comme ARIMA [74], des moyennes mobiles ou encore des lissages exponentiels [72], sont couramment utilisées pour la prédiction de la demande. En effet, on peut considérer la demande au cours du temps comme une série temporelle. L'utilisation de données historiques, de données météorologiques et de données économiques permet d'améliorer la précision des prévisions.

4.2.2/ GESTION ET OPTIMISATION DES STOCKS

La gestion des stocks est un défi complexe pour les entreprises, car elle nécessite de trouver un équilibre entre le maintien de niveaux de stocks suffisants pour répondre à la demande et la minimisation des coûts associés. Les modèles d'IA, tels que les algorithmes de réapprovisionnement basés sur la demande prévue [70], l'apprentissage par renforcement [59], et des algorithmes génétiques [63], sont utilisés pour améliorer la gestion et l'optimisation des stocks.

4.2.3/ GESTION DE LA CHAÎNE D'APPROVISIONNEMENT

L'IA offre la possibilité de surveiller et de gérer la chaîne d'approvisionnement de manière plus transparente et efficace. Les systèmes d'IA sont utilisés pour suivre en temps réel les mouvements des marchandises [123], optimiser les itinéraires de transport [57], et améliorer la visibilité tout au long de la chaîne d'approvisionnement [134].

4.2.4/ DÉTECTION DES ANOMALIES

La détection des anomalies est cruciale pour identifier les problèmes potentiels dans la chaîne d'approvisionnement, comme par exemple les variations imprévues dans les ventes [138]. L'analyse de données à l'aide de techniques d'IA peut permettre de détecter les comportements anormaux, ce qui permet une réponse rapide aux incidents.

4.3/ UTILISATION DE L'IA DANS LE SECTEUR PHARMACEUTIQUE

Nous allons à présent examiner la phase de conception et de fabrication des médicaments, en remontant la chaîne de production pharmaceutique. Il convient de noter que le champ de l'intelligence artificielle offre diverses applications potentielles pour l'amélioration des performances de cette industrie [127]. De telles améliorations pourraient potentiellement avoir un impact sur le secteur des grossistes-répartiteurs, notamment en perturbant le marché lors de la mise sur le marché de nouveaux médicaments.

4.3.1/ DÉCOUVERTE DE MÉDICAMENTS ASSISTÉE PAR L'IA

La découverte de médicaments est un domaine crucial de la recherche pharmaceutique. L'IA peut être utilisée pour aider à concevoir de nouveaux médicaments [96]. Les techniques d'apprentissage automatique, telles que les réseaux de neurones, sont utilisées pour prédire l'activité biologique des composés chimiques [75], pour concevoir des molécules avec des propriétés spécifiques [112], et pour identifier de nouveaux médicaments potentiels [102]. Les modèles de réseaux de neurones ont été notamment utilisés pour simuler le criblage de molécules, une technique longue et coûteuse qui permet de réutiliser une molécule connue dans la création d'un nouveau traitement sur une autre maladie [92].

4.3.2/ ALPHA FOLD ET LA PRÉDICTION DE STRUCTURE PROTÉIQUE

Pouvoir prédire la structure d'une protéine est un challenge intéressant pour la biologie, car cela permet de définir la fonction de cette protéine. La société Deep Mind, filiale de Google a développé un modèle nommé AlphaFold [120] qui prédit le repliement des protéines à partir de leur séquence d'acides aminés. Ce modèle a surclassé ses concurrents dans la compétition "Critical Assessment of Protein Structure Prediction" avec des résultats de 15 à 20% supérieurs [115].

4.3.3/ OPTIMISATION DES ESSAIS CLINIQUES

Les essais cliniques sont coûteux et chronophages. L'IA est déployée pour optimiser la conception des essais cliniques en identifiant des cohortes de patients appropriées [129], en prédisant les résultats [139], en surveillant les données en temps réel [128]. Cela peut conduire à des essais cliniques plus efficaces et à une accélération du développement de médicaments.

4.3.4/ PERSONNALISATION DES TRAITEMENTS ET DE LA MÉDECINE DE PRÉCISION

L'IA peut permettre de faciliter la personnalisation des traitements [119] à chaque patient. Les modèles d'apprentissage automatique analysent les données génomiques, cliniques et de laboratoire pour identifier des traitements personnalisés. Cela ouvre la voie à une médecine de précision, où les médicaments sont adaptés aux besoins individuels des patients.

4.3.5/ DÉTECTION DES EFFETS SECONDAIRES ET DE LA SÉCURITÉ DES MÉDICAMENTS

La sécurité des médicaments est une priorité majeure. L'IA peut être employée pour surveiller les effets secondaires [98] des médicaments en analysant de vastes ensembles de données de patients, y compris des données textuelles non structurées, pour détecter les signaux de sécurité [121]. Cela permet une réponse rapide aux problèmes potentiels de sécurité des médicaments.

4.3.6/ MINAGE ET ANALYSE DE TEXTES CLINIQUES

La récolte et l'analyse de données cliniques sous forme de texte peut être une source d'informations intéressantes pour plusieurs cas d'applications. Cela peut, par exemple, permettre de détecter les cas d'effets indésirables suite à la prise d'un traitement [83], ou améliorer les techniques de pharmacovigilance [109]. Un autre cas d'usage peut être la détection de patients à risques, comme par exemple l'utilisation de technique d'IA pour la prévention de suicide [108], d'un risque d'infection post-opératoire [101] ou encore la prédiction d'un patient violent à partir de données textuelles historiques [94]. L'analyse de données cliniques par des technique d'IA peut également servir à automatiser certaines tâches, comme par exemple l'attribution d'un code CIM10 [116], ce qui permet de classer plus facilement les actes médicaux dans les Services de l'Information Médicale.

4.4/ UTILISATION DE L'IA POUR L'OFFICINE DE PHARMACIE

Le secteur de l'officine de pharmacie est au cœur de la prestation de soins de santé, jouant un rôle essentiel dans la distribution de médicaments, le conseil aux patients et la gestion des stocks. L'intégration de l'intelligence artificielle (IA) dans ce secteur a ouvert de nouvelles perspectives pour l'amélioration des processus, la personnalisation des traitements et la sécurité des patients. Cette section explore l'état de l'art de l'utilisation de l'IA dans l'officine de pharmacie, en se concentrant sur plusieurs domaines clés.

4.4.1/ AIDE À LA DISPENSATION DE MÉDICAMENTS

L'aide à la dispensation de médicaments est un domaine où l'IA a un impact significatif. Les robots pharmaceutiques, ou logiciels d'aide à la dispensation (LAD) peuvent utiliser des algorithmes d'IA pour trier, mesurer et distribuer précisément les médicaments, réduisant ainsi les erreurs humaines et améliorant l'efficacité de la pharmacie. En France, les LAD doivent, pour être certifiés, répondre à certains critères déterminés par la Haute Autorité de santé (HAS) [122]. Les systèmes d'IA peuvent également être utilisés pour vérifier les interactions médicamenteuses potentielles et les allergies des patients [64], aidant les pharmaciens et leurs équipes à fournir un service le plus adapté à leurs patientèles.

4.4.2/ GESTION DES STOCKS ET PRÉVISION DE LA DEMANDE

La gestion des stocks est un défi critique pour les officines de pharmacie. Des recherches ont montré un intérêt de l'utilisation de systèmes d'IA qui analyseraient les données d'inventaire historiques, les tendances de demande et les dates d'expiration pour optimiser les niveaux de stock et réduire les pertes [142]. Ces techniques, déjà présentes dans de nombreux marchés, ne sont pas encore totalement présentes dans l'officine de pharmacie. Mais de plus en plus de logiciels d'officine souhaiteraient incorporer peu à peu ces outils.

4.4.3/ CONSEIL PHARMACEUTIQUE PERSONNALISÉ, TÉLÉ MÉDECINE ET SUIVI DE SANTÉ

Ces dernières années, nous avons vu arriver dans les officines de plus en plus de dispositifs de télé-médecine [117], cela permet d'aider à combler le manque de médecins dans certains secteurs géographiques [67, 93]. Ces outils peuvent utiliser des algorithmes d'IA pour aider le praticien à faire son diagnostic à distance. Dans un autre registre, l'essor

des objets connectés a permis aux pharmaciens d'offrir de nouveaux services à leurs patients, notamment pour le suivi de santé [87, 107]. Ces services coïncident également avec l'arrivée en France du Dossier Médical Partagé (DMP) [104], avec lequel on pourrait imaginer des interactions dans le futur.

4.5/ CONCLUSION

Au cours de ce chapitre, nous avons pu constater que la profession de grossiste-répartiteur se situe à l'interface de multiples secteurs d'activités. Cette nature pluridisciplinaire offre l'opportunité d'explorer diverses applications des techniques d'intelligence artificielle sur une gamme étendue de problématiques. Dans les chapitres à venir, nous examinerons en détail les différentes recherches entreprises pour améliorer les capacités prédictives de la CERP RRM.



CONTRIBUTIONS

5

PRÉDICTIONS DE COMMANDE DES PHARMACIES AUPRÈS DE LA CERP RRM

Comme mentionné précédemment, les travaux de cette thèse étant exploratoires, il a fallu, pour ce chapitre, dans un premier temps, comprendre le contexte de la problématique et définir les choix possibles à étudier. La première section de ce chapitre se consacre à cette tâche. La seconde partie traitera d'une dernière analyse sur des données d'officine de pharmacie, avant de continuer le chapitre sur les travaux de prédictions.

5.1/ RELATION ENTRE LES BESOINS EN PRODUITS PHARMACEUTIQUES ET LES COMMANDES CERP RRM

Afin d'anticiper les futures commandes de produits pharmaceutiques à la CERP RRM, il est légitime d'envisager l'étude de l'évolution des maladies en vue d'en déduire les besoins en traitements. Dans cette optique, nous avons élaboré un scénario visant à conceptualiser le processus décisionnel qui convertit la déclaration d'une maladie en une commande de produit pour la CERP RRM.

5.1.1/ CHOIX DU PRESCRIPTEUR

Le scénario débute par le diagnostic d'une maladie effectué par un professionnel de santé. À la suite de ce diagnostic, le professionnel peut envisager différentes alternatives, notamment la proposition d'une alternative non pharmaceutique, telle qu'une consultation chez un spécialiste, ce qui met fin au scénario pour cette situation particulière. Le professionnel de santé a également la possibilité de prescrire un traitement, que ce soit en

complément de la première option ou de manière indépendante, ouvrant ainsi la voie à la poursuite du scénario.

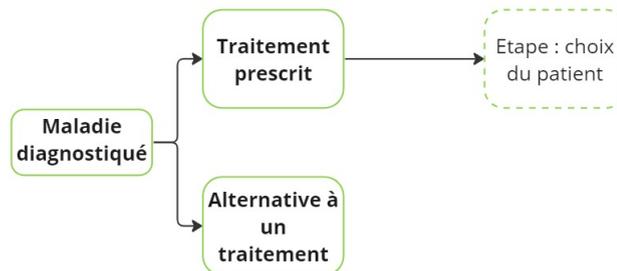


FIGURE 5.1 – Schéma décisionnel au niveau du prescripteur

5.1.2/ CHOIX DU PATIENT

Une fois qu'un traitement pharmaceutique a été prescrit, le patient peut soit déjà posséder le traitement, soit opter pour une pharmacie où il peut se le faire délivrer. À ce stade, l'étape de la dispensation en pharmacie peut également impliquer, en complément ou non, un patient qui vient réapprovisionner son stock de médicaments sans ordonnance (OTC) ou acheter des produits de parapharmacie. Ces trois situations conduisent à une dispensation de produit.

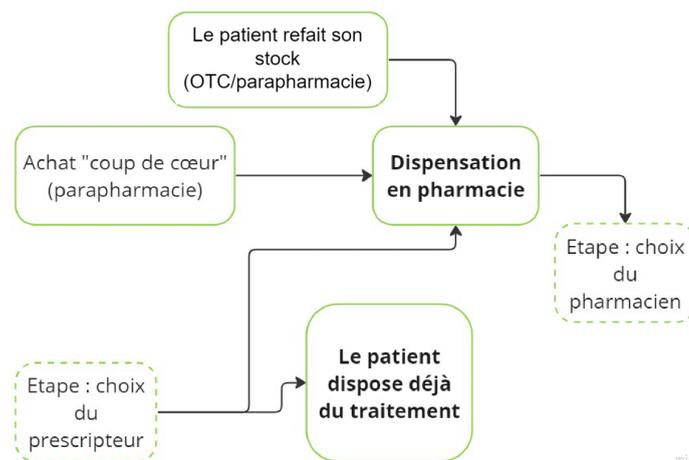


FIGURE 5.2 – Schéma décisionnel au niveau du patient

5.1.3/ CHOIX DU PHARMACIEN

Lorsqu'une dispensation de produit est nécessaire, plusieurs scénarios se présentent. Le pharmacien peut soit disposer du produit requis ou de l'un de ses génériques, ou dans certains cas, il peut proposer un produit équivalent. Ces trois situations aboutissent à une vente de produit. Le pharmacien a également la possibilité d'enregistrer une reconduction de traitement, anticipant ainsi le stockage de quantités du produit requis pour des dispensations futures. Dans tous ces cas, on peut considérer une action de réapprovisionnement du stock de la pharmacie, impliquant par conséquent une commande de produit.

Une autre éventualité est que le pharmacien enregistre une commande en attente pour un produit qu'il n'a pas actuellement en stock (un promis), ce qui entraîne également une commande de produit. La dernière phase consiste en la décision du pharmacien concernant la source de sa commande. Il peut choisir de commander directement auprès du laboratoire pharmaceutique, ou opter pour l'un des grossistes-répartiteurs en France, parmi lesquels figure la CERP RRM. Ce choix peut être influencé par diverses variables à prendre en compte.

Les facteurs qui peuvent influencer la décision du pharmacien ne sont pas nécessairement classés par ordre d'importance, car cela peut varier d'un pharmacien à un autre :

- 1. Les habitudes de commande / la configuration du LGO :** Le pharmacien a tendance à commander ses produits chez son distributeur de prédilection, en particulier lorsque des commandes automatiques sont gérées par le LGO, qui est généralement configuré pour commander auprès d'une entité spécifique.
- 2. La disponibilité du produit :** La disponibilité du produit chez la source d'approvisionnement choisie est un élément crucial. Dans des situations urgentes, le pharmacien peut être amené à changer de source si les délais de disponibilité sont excessivement longs.
- 3. Les délais de livraison :** De manière similaire au point précédent, si les délais de livraison ne correspondent pas aux attentes du pharmacien, il peut opter pour un changement de source d'approvisionnement pour cette commande.
- 4. Les directives des groupements :** Certains pharmaciens sont membres de groupements, lesquels peuvent imposer des directives de commandes que leurs adhérents sont parfois tenus de respecter. Cela peut inclure des partenariats commerciaux avec des fabricants encourageant la commande directe au laboratoire en échange de tarifs préférentiels.
- 5. La quantité de produits commandés :** En fonction de la quantité de produits commandés, certains fournisseurs peuvent offrir un meilleur rapport qualité-prix, ainsi

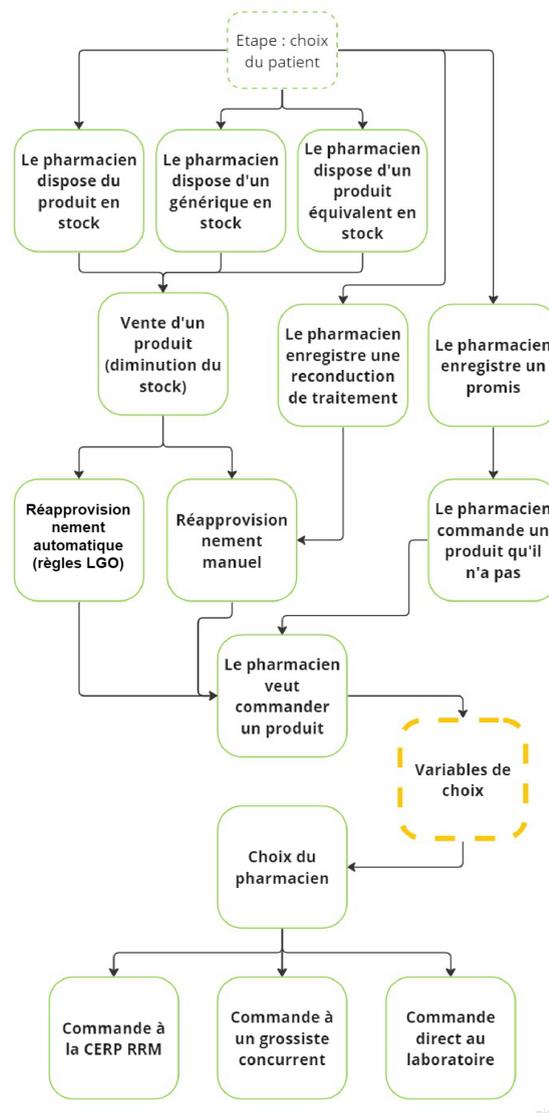


FIGURE 5.3 – Schéma décisionnel au niveau du pharmacien

que la capacité à répondre aux quantités demandées.

6. **Les offres commerciales** : Certains fournisseurs peuvent proposer des réductions ou des remises sur les produits, ces offres étant parfois soumises à certaines conditions (limites de temps, quantités commandées, fidélité du pharmacien, etc.). Si les conditions sont réunies, le pharmacien peut être incité à modifier ses habitudes de commande pour réaliser des économies sur sa commande actuelle ou sur les prochaines.
7. **Les opportunités de commandes groupées** : Les commandes groupées sont fréquemment utilisées par les groupements de pharmacies, mais il peut également arriver que des pharmaciens s'associent pour réaliser des commandes groupées, ce qui leur permet de commander des quantités plus importantes et de négocier des prix avantageux. Une opportunité de ce type peut inciter le pharmacien à changer

de source d'approvisionnement.

8. **Les contraintes de conditionnement** : Certains fournisseurs ne proposent certains produits qu'en conditionnements spécifiques (par exemple, par lot de colisage, tel qu'un colisage de 10 boîtes). Ces contraintes peuvent parfois être perçues comme gênantes par le pharmacien, qui préférera donc se tourner vers une autre source d'approvisionnement.
9. **Le prix du produit** : Cette variable a déjà été mentionnée à travers d'autres facteurs évoqués précédemment. Toutefois, comme dans tout marché, le prix des produits peut influencer le choix du fournisseur. Il convient de noter que, dans le cas des médicaments en France, les prix sont réglementés et les produits sont remboursés dans certains cas, ce qui limite l'influence de cette variable, contrairement à la parapharmacie.
10. **Les contraintes logistiques** : Selon le fournisseur, les contraintes logistiques pour la livraison des commandes peuvent varier en complexité. Certains fournisseurs peuvent offrir un service de livraison quotidien en camionnette pour des commandes de faible volume, tandis que d'autres livreront des commandes hebdomadaires, généralement de grandes quantités sur palettes. Ces contraintes s'entremêlent avec celles de conditionnement et de délais de livraison, mais ajoutent également des problématiques organisationnelles pour la pharmacie, telles que la nécessité de prévoir un employé pour décharger et ranger ces commandes volumineuses, ou les difficultés de stationnement pour les pharmacies en zone urbaine.
11. **La durée d'écoulement** : La durée d'écoulement d'un produit représente le temps moyen nécessaire pour écouler une quantité donnée de ce produit. Bien que cette variable ne soit pas toujours déterminante dans le choix du fournisseur, elle influence généralement sur le volume de la commande passée par le pharmacien à ce dernier. Un produit présentant une durée d'écoulement relativement courte pour un pharmacien donné est susceptible de l'inciter à effectuer une commande de plus grande envergure, afin d'assurer un stockage à plus longue durée.
12. **Apparition d'une nouveauté dans le catalogue** : Lors de l'introduction d'un nouveau produit, le pharmacien peut être incité, par exemple par le biais d'un commercial, à commander cette nouvelle référence en remplacement d'une ancienne. Cependant, la disponibilité d'un nouveau produit n'est pas uniforme chez tous les fournisseurs. Ainsi, le choix du pharmacien d'acquiescer cette nouveauté, outre sa capacité à modifier les ventes des anciennes références, peut également l'amener à changer de fournisseur en fonction de la disponibilité de ce nouveau produit.

5.1.4/ CONCLUSION

Dans cette section, nous avons vu qu'il n'y a pas de relation forte entre la prescription d'un traitement et la commande de son médicament d'une officine à la CERP RRM. La CERP RRM n'est bien évidemment pas le seul acteur sur le marché et les besoins en médicaments sont répartis entre les différents concurrents du secteur. Les capacités de stockages des différents acteurs ont aussi tendance à diluer les besoins.

5.2/ ÉCHELLES ET AGRÉGATIONS DES DONNÉES DANS LE DOMAINE DE LA RÉPARTITION PHARMACEUTIQUE

La multitude de produits distincts, ainsi que la variabilité temporelle de leurs ventes, nous incite à considérer la pertinence de regrouper les données avant de les soumettre en entrée de nos modèles prédictifs. Notre analyse a permis d'identifier trois axes d'agrégation distincts, sur lesquels nous pouvons travailler avec des degrés de précision variables en fonction des besoins.

5.2.1/ ÉCHELLE TEMPORELLE

- 1. À l'année :** L'agrégation des données de vente à l'échelle annuelle présente des avantages tels que la réduction de l'aléatoire et du bruit grâce à la période prolongée considérée, ainsi que des volumes de ventes importants permettant une vue d'ensemble du marché. Cependant, cette approche souffre d'un manque de précision, car les données sont regroupées sur une période étendue, ce qui limite son intérêt pour des analyses fines. De plus, le jeu de données disponible pour l'apprentissage est réduit, ce qui peut affecter la performance des modèles prédictifs. En outre, l'agrégation annuelle ne permet pas de prendre en compte la saisonnalité, ce qui limite la capacité à détecter les variations périodiques de la demande pharmaceutique.
- 2. Au mois :** Le regroupement des données de vente au niveau mensuel offre certains avantages, tels que la disponibilité d'un volume de données adéquat pour l'apprentissage, permettant des analyses plus détaillées. De plus, les variations de la demande au cours des mois peuvent être modérées, ce qui permet une meilleure compréhension des tendances mensuelles. Par ailleurs, cette approche prend en compte la saisonnalité, offrant ainsi des informations sur les variations périodiques. Cependant, le niveau actuel de prédiction avec les outils statistiques existants peut nécessiter des améliorations pour obtenir des prédictions plus précises. Enfin, la catégorisation au niveau mensuel peut encore manquer de granularité pour des

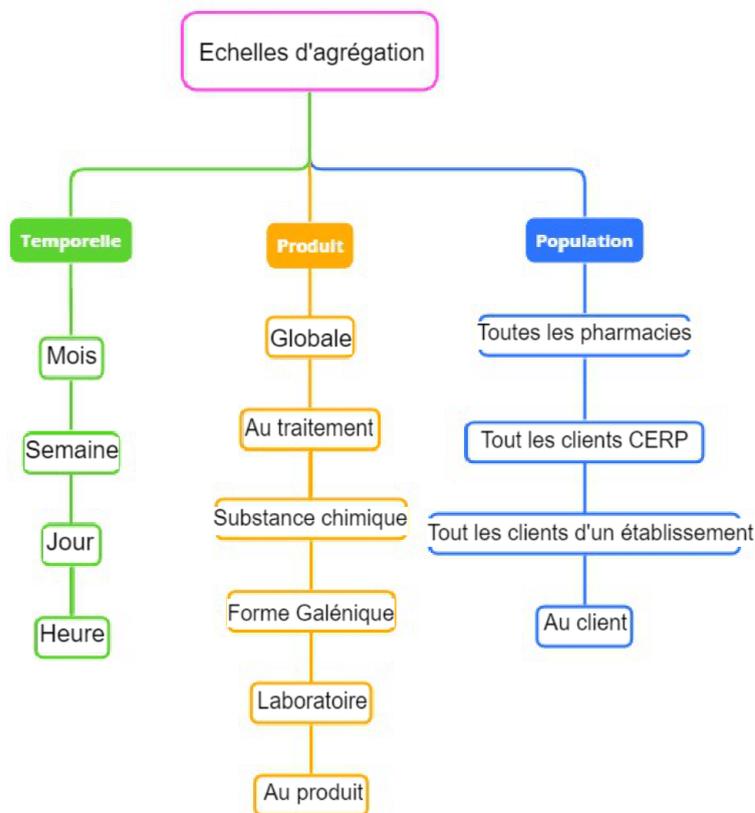


FIGURE 5.4 – MindMap des différentes échelles d'agrégations possibles pour les données de ventes

analyses approfondies.

3. **À la semaine** : L'agrégation des données de vente à l'échelle hebdomadaire présente l'avantage d'une précision supérieure par rapport au niveau actuel de regroupement. De plus, cette approche permet de prendre en compte la saisonnalité, offrant ainsi des informations sur les variations hebdomadaires de la demande pharmaceutique. Cependant, il convient de noter que l'introduction de comportements aléatoires, ou du bruitage des courbes, est un défi associé à cette échelle. Enfin, il peut y avoir une perte d'informations en raison de l'agrégation, ce qui nécessite une analyse minutieuse des résultats obtenus.
4. **À la journée** : L'agrégation des données de vente à l'échelle journalière permet une précision très fine, ce qui permet de capturer les variations quotidiennes de la demande pharmaceutique. De plus, il existe un grand jeu de données disponible pour l'apprentissage, facilitant ainsi la modélisation prédictive. Cependant, cette approche présente également des problématiques, telles que l'introduction de comportements aléatoires élevés, ou du bruitage des courbes, en raison de l'agrégation

à une échelle très fine. De plus, il peut y avoir une perte d'informations potentielle en raison de l'agrégation, notamment pour les produits à faible volume de ventes. Par ailleurs, l'agrégation journalière peut entraîner une diminution de la saisonnalité, limitant ainsi la détection des tendances périodiques.

- 5. À l'heure :** L'agrégation des données de vente à l'échelle horaire offre une précision maximale en capturant les variations horaires de la demande pharmaceutique. De plus, un grand jeu de données est disponible pour l'apprentissage, ce qui facilite la modélisation prédictive. Cependant, cette approche présente des problématiques majeures, notamment l'introduction de comportements aléatoires très élevés, ou du bruitage des courbes, en raison de l'agrégation à une échelle extrêmement fine. De plus, il peut y avoir une perte importante d'informations en raison de l'agrégation, en particulier pour les produits à faible volume de ventes. En conséquence, le jeu de données obtenu à cette échelle peut être presque inexploitable pour l'apprentissage en raison de la complexité induite par le bruit et la faible saisonnalité restante.

5.2.2/ ÉCHELLE PRODUIT

- 1. Globale :** L'agrégation des données de vente à l'échelle globale présente l'avantage d'un volume de ventes important, ce qui permet d'obtenir une vision d'ensemble du marché pharmaceutique. Cependant, cette approche présente une problématique majeure, à savoir le manque d'intérêt en raison de l'absence d'informations sur les produits vendus. En ne distinguant pas les produits spécifiques, il devient difficile de tirer des insights significatifs pour des analyses approfondies ou des prédictions précises.
- 2. Au traitement (niveau 2 de la classification ATC) :** Le regroupement des données de vente selon le niveau du traitement, défini par la classification ATC (Anatomical Therapeutic Chemical), offre l'avantage d'un volume de ventes important et d'une précision accrue sur les différents traitements. Cela permet une meilleure compréhension des variations de la demande selon les classes thérapeutiques. Cependant, cette approche présente des problématiques, notamment une précision encore relativement faible pour obtenir des prédictions correctes. De plus, elle ne fonctionne pas pour la parapharmacie, qui comprend des produits non médicamenteux.
- 3. Substance chimique (Niveau 5 de la classification ATC) :** L'agrégation des données de vente au niveau de la substance chimique, correspondant au niveau le plus fin de la classification ATC, présente l'avantage d'une précision actuelle sur l'outil statistique de la CERP ainsi qu'une précision élevée grâce à la granularité des données. De plus, les volumes de ventes sont généralement importants

et le nombre de regroupements de produits est modéré. Cependant, cette approche présente une problématique liée à une précision améliorable, qui nécessite des efforts supplémentaires pour des prédictions plus précises. De plus, elle ne convient pas à la parapharmacie.

4. **Forme galénique** : L'agrégation des données de vente selon la forme galénique des produits présente l'avantage d'améliorer la précision en différenciant les différentes formes galéniques disponibles sur le marché pharmaceutique. Cela permet une meilleure compréhension des variations de la demande selon les formulations des produits. Cependant, cette approche présente des problématiques, telles qu'une augmentation du nombre de regroupements, une baisse du volume des ventes par regroupements et l'ajout d'un niveau de complexité lié aux concurrences entre les formes galéniques. Cela peut rendre l'analyse et les prédictions plus complexes et nécessiter une gestion minutieuse des données.
5. **Au laboratoire** : Le regroupement des données de vente selon le laboratoire producteur présente l'avantage d'améliorer la précision en différenciant les différents laboratoires. Cela permet de mieux comprendre les variations de la demande en fonction des fabricants. Cependant, cette approche présente des problématiques similaires à celle de la forme galénique, notamment une augmentation du nombre de regroupement, une baisse du volume des ventes par regroupements et l'ajout d'un niveau de complexité dû aux concurrences entre les laboratoires. Il est donc essentiel de prendre en compte ces facteurs lors de l'analyse des données et des prédictions.
6. **Au produit** : L'agrégation des données de vente au niveau du produit offre une précision très élevée en permettant de distinguer chaque produit individuellement. Cela permet une compréhension approfondie des variations de la demande pour chaque produit spécifique. Cependant, cette approche présente des problématiques, notamment une augmentation significative du nombre de regroupements, une baisse du volume des ventes par regroupement et l'ajout d'un niveau de complexité lié aux concurrences entre les produits. Cela peut rendre l'analyse et les prédictions plus complexes et nécessiter une gestion minutieuse des données pour obtenir des informations significatives.

5.2.3/ ÉCHELLE DE POPULATION

1. **Toutes les pharmacies** : L'agrégation des données de vente en regroupant toutes les pharmacies présente l'avantage de permettre la prédiction du besoin global au niveau national. Cela offre une vue d'ensemble de la demande pharmaceutique à l'échelle du pays. Cependant, cette approche comporte des problématiques, notamment des données qui ne sont pas facilement accessibles, ce qui peut rendre

la collecte et l'analyse de ces informations plus difficiles. De plus, il peut y avoir un manque de précision géographique, ce qui limite la capacité à identifier les variations de la demande selon les régions ou les localités spécifiques.

2. **Tous les clients CERP RRM** : Le regroupement des données de vente en incluant tous les clients de la CERP RRM (Répartition, Répartition Médicale) offre l'avantage de gérer les besoins des clients de cette organisation spécifique. De plus, la précision actuelle dans le logiciel statistique de la CERP facilite l'analyse et les prédictions. Cependant, il existe une problématique liée à la précision géographique, qui nécessite encore des améliorations pour obtenir une localisation plus précise des données. Cela peut limiter la capacité à identifier les variations de la demande selon les régions spécifiques ou les localités.
3. **Tous les clients d'un établissement (département)** : L'agrégation des données de vente en regroupant tous les clients d'un établissement, tel qu'un département, présente l'avantage de permettre la segmentation du besoin spécifique à cet établissement. Cela permet d'allouer directement les produits en fonction de ces besoins particuliers. Cependant, cette approche comporte des problématiques, notamment une baisse du volume de vente global en raison de l'agrégation spécifique à un établissement. De plus, il peut y avoir une augmentation du nombre de prédictions à effectuer, ce qui nécessite une gestion plus complexe des données et des prévisions.
4. **Au client (pharmacie)** : L'agrégation des données de vente au niveau du client, en l'occurrence la pharmacie, offre le niveau de précision le plus élevé. Cela permet de capturer les variations spécifiques à chaque pharmacie individuellement. Cependant, cette approche comporte des problématiques, notamment une baisse du volume de vente global en raison de l'agrégation à un niveau très spécifique. De plus, il peut y avoir une augmentation du bruit dans les données en raison des nombreuses variations induites par le comportement individuel de chaque pharmacie. Cela rend pratiquement impossible l'exploitation de ce niveau de granularité pour des analyses ou des prédictions significatives.

5.2.4/ CONCLUSION DE L'ÉTUDE PRÉLIMINAIRE

Dans cette section, nous avons examiné les avantages et les contraintes des divers niveaux de granularité pour l'agrégation des ventes de produits. Bien que les niveaux les plus détaillés puissent sembler complexes à exploiter, il demeure intéressant d'analyser les données de pharmacie, car elles fournissent des informations complémentaires à celles obtenues à partir des données de ventes du grossiste-répartiteur. La section suivante présentera ainsi les travaux réalisés sur les données provenant d'une officine de pharmacie.

5.3/ ANALYSE DES DONNÉES D'UNE OFFICINE

Les officines de pharmacies sont le dernier maillon (avec les services hospitaliers) de la chaîne de distribution des produits pharmaceutiques. Elles délivrent aux patients, peuvent être connectées à tous les fournisseurs (laboratoires et différents grossistes-répartiteurs). Par conséquent, leurs données de ventes correspondent à la demande réelle de produit pharmaceutique d'un territoire, contrairement aux données de ventes de la CERP RRM qui correspondent uniquement à une partie de la demande. De plus, les officines de pharmacies possèdent également des données sur les patients, qui une fois anonymisées, offrent une source d'informations potentielles à exploiter. Dans cette section, nous allons présenter les travaux réalisés sur les données extraites du LGO d'une pharmacie située dans la ville de Gap.

5.3.1/ ANALYSE DE LA PATIENTÈLE

Les données de santé des patients sont considérées comme des données sensibles et doivent donc être anonymisées pour être utilisées. Par conséquent, deux paramètres restent exploitables sans risquer de compromettre le travail d'anonymisation : l'âge et le sexe. Dans cette analyse, nous nous sommes donc concentrés uniquement sur ces deux variables.

Dans un premier temps, nous avons comparé la population de la ville de Gap où est située la pharmacie étudiée, avec la patientèle de cette dernière (figure 5.5). Les données de recensement de la population les plus récentes et correspondant aux années d'exploitation de la pharmacie que nous possédons sont celles de l'année 2020. Comme les pharmacies sont des commerces régulés par la loi dans leurs implantations, on devrait théoriquement voir une homogénéité dans la distribution de la population et de celle de la patientèle.

En étudiant les pourcentages de répartition de la population et de la patientèle à travers les tranches d'âges et le sexe de l'individu (figure 5.6 & 5.7), on peut observer quelques différences dans les distributions des deux populations observées :

1. La tranche d'âge des 0-14 ans est sous-représentée dans la patientèle : Cela peut s'expliquer par le fait qu'en France la dispensation de médicaments aux mineurs est très encadrée par la loi [100]. De plus, même s'il est possible d'avoir sa propre carte vitale dès l'âge de 12 ans, la plupart des personnes sont souvent rattachées à la carte vitale de leurs parents, en attendant la création automatique de leur dossier à l'assurance maladie à l'âge de 16 ans.
2. Dans la tranche des 15-29 ans, les femmes sont sur-représentées dans la patientèle. Deux hypothèses peuvent l'expliquer, la prescription de solutions contra-

ceptives pour les jeunes femmes d'une part et d'autre part, on commence à retrouver également dans cette tranche d'âge, les femmes qui ont leur premier enfant [147], et qui comme on l'a vu précédemment, sont très souvent rattachées aux comptes sécurité sociale des parents.

3. Les points soulevés précédemment se retrouvent encore dans la tranche 30-44 ans où les femmes sont encore sur-représentées dans la distribution. Cet effet s'estompe peu à peu dans les tranches d'âge supérieures, ce qui correspond bien au passage sous leurs propres carte vitale des enfants.
4. Comme dit précédemment, il me semble que les femmes ont d'une manière générale un plus grand suivi médical et une plus grande consommation de soins, pour diverses raisons (gynécologique, grossesses, etc.) ce qui expliquera en partie leurs sur-représentation dans la patientèle
5. Un dernier point de différence entre les deux populations, c'est que l'âge global de la patientèle est plus élevé que celui de la population de Gap, dû comme on a pu le voir à une fréquentation moins élevée des moins de 30 ans, notamment chez les hommes.

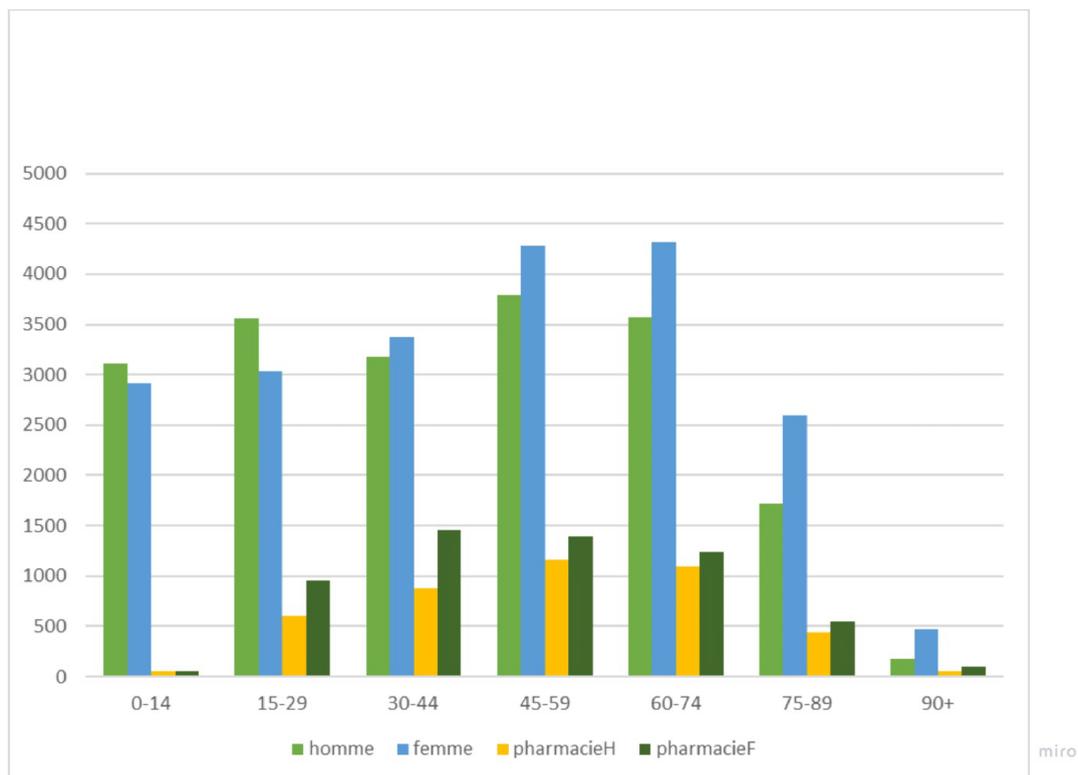


FIGURE 5.5 – Population de la ville de Gap comparée à la patientèle d'une pharmacie de Gap, regroupées par tranche d'âge et sexe, en nombre de personnes, pour l'année 2020

Nous allons maintenant étudier la relation entre les ventes par rapport à cette patientèle, pour essayer de découvrir des informations utiles pour le pharmacien.

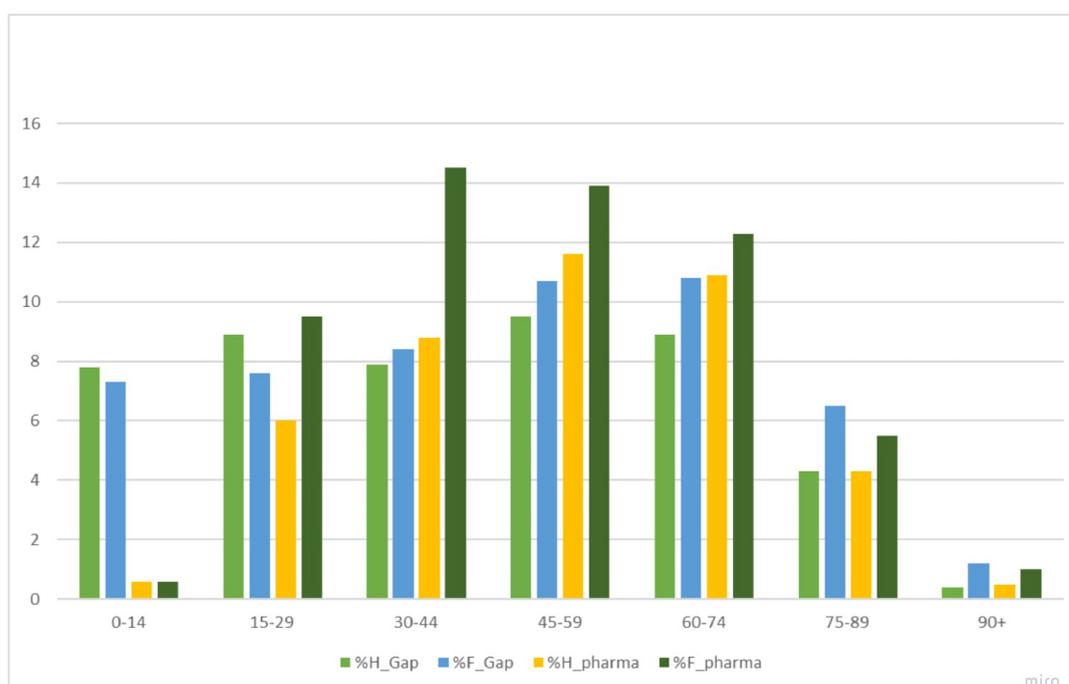


FIGURE 5.6 – Pourcentage de la population de la ville de Gap comparé au pourcentage de la patientèle d'une pharmacie de Gap, regroupées par tranche d'âge et sexe, pour l'année 2020

5.3.2/ ANALYSE DES VENTES

En se concentrant sur les données de la pharmacie, nous pouvons entrer plus en détails et comparer le pourcentage de patientèle par âge avec les pourcentages de vente, de chiffre d'affaires (CA) et de marge réalisée par la pharmacie sur les ventes. L'étude est réalisée sur l'agrégation de toutes les données de ventes disponibles, soit sur la période 2008-2021 (figure 5.8).

La première observation est que le chiffre d'affaires et la marge suivent bien le nombre de ventes à l'exception d'un pic pour les personnes âgées de 41 ans. Après une recherche plus poussée dans les données, on observe que ce pic est créé par un patient qui a eu des soins particulièrement coûteux.

La deuxième observation est que le pourcentage du nombre de ventes n'est pas corrélé à celui de la patientèle : la population de plus de 50 ans représente une grande partie des ventes. En effet, même si elle ne représente que 38% de la patientèle, elle cumule 67,35% des ventes de l'officine (figure 5.9).

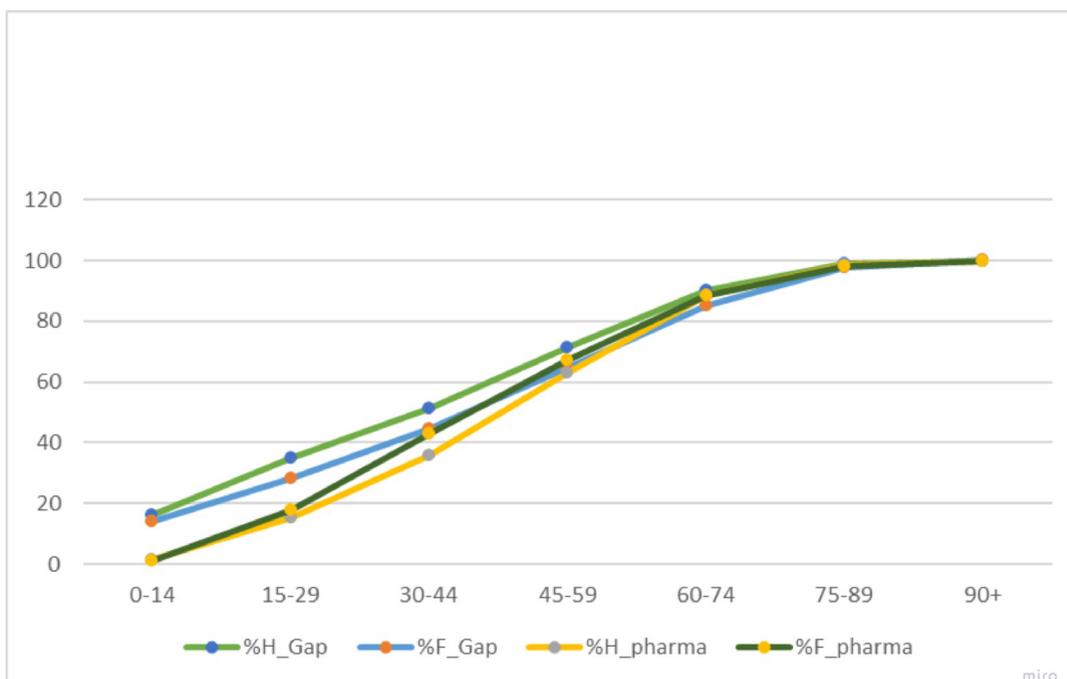


FIGURE 5.7 – Pourcentages cumulés de la population de la ville de Gap comparés au pourcentage de la patientèle d'une pharmacie de Gap, regroupés par tranche d'âge et sexe, pour l'année 2020. On peut observer que la répartition de la patientèle est plus faible que celle de la population sur les tranches d'âge. Cet écart se comble plus rapidement chez les femmes que chez les hommes

5.3.2.1/ LES DIFFÉRENTS TYPES DE VENTES

Il est temps de définir plus en détail ce que l'on appelle une vente dans le cadre des données d'officine de pharmacie : il s'agit de l'ensemble des produits vendus en une seule fois pour le même patient. Une vente est constituée de lignes, chaque ligne correspondant à un produit et possédant un nombre d'unités.

Dans l'univers de l'officine de pharmacie, on distingue trois types de ventes (figure 5.10) :

1. **Les ventes remboursées**, qui sont, comme leur nom l'indique, des ventes constituées uniquement de lignes correspondant à des produits remboursés, ce sont les ventes issues d'une ordonnance. Le pharmacien n'a que peu d'influence sur ce type de vente, car le patient vient suite à une prescription d'un autre professionnel de santé et les prix des médicaments remboursés sont imposés. Dans la pharmacie étudiée, ces ventes représentent **53,76%** des ventes totales.
2. **Les ventes non-remboursées**, qui sont des ventes uniquement composées de produits sans ordonnance (OTC, parapharmacie...). C'est sur ce type de vente que le pharmacien a le plus d'influence, car le patient ne vient pas avec une prescription

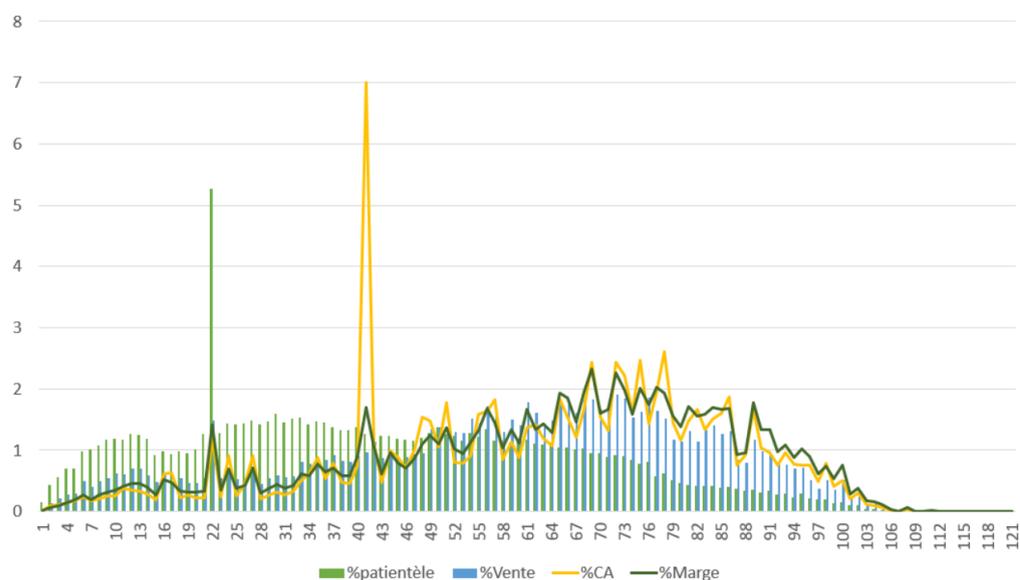


FIGURE 5.8 – Comparaison entre les pourcentages de la clientèle, du chiffre d'affaires (CA), de la marge et des ventes d'une pharmacie de Gap regroupés par âge du patient sur la période 2008-2021. On peut observer que les moins de 40 ans ont une part moins importante dans les ventes/CA/Marge que leur représentativité dans le pourcentage de la clientèle.

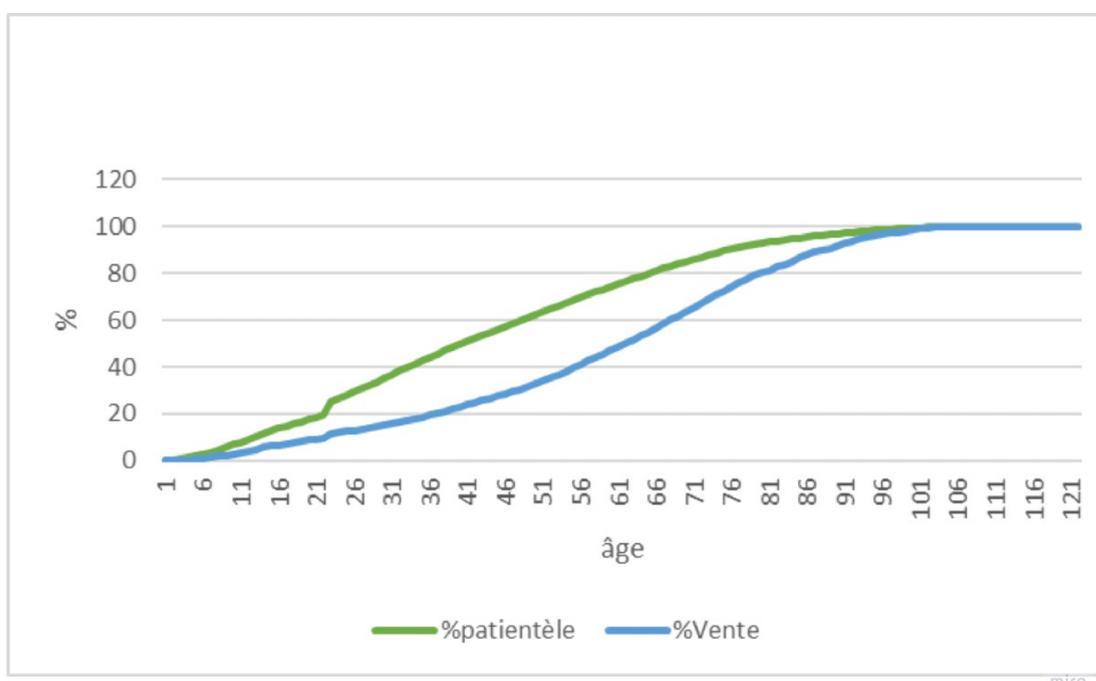


FIGURE 5.9 – Comparaison des pourcentages cumulés de la clientèle et des ventes de la pharmacie, regroupés par âge.

et les prix des produits non remboursés sont libres. Ainsi, le pharmacien doit attirer le patient et le conseiller pour finalement pouvoir conclure la vente. De plus, il peut

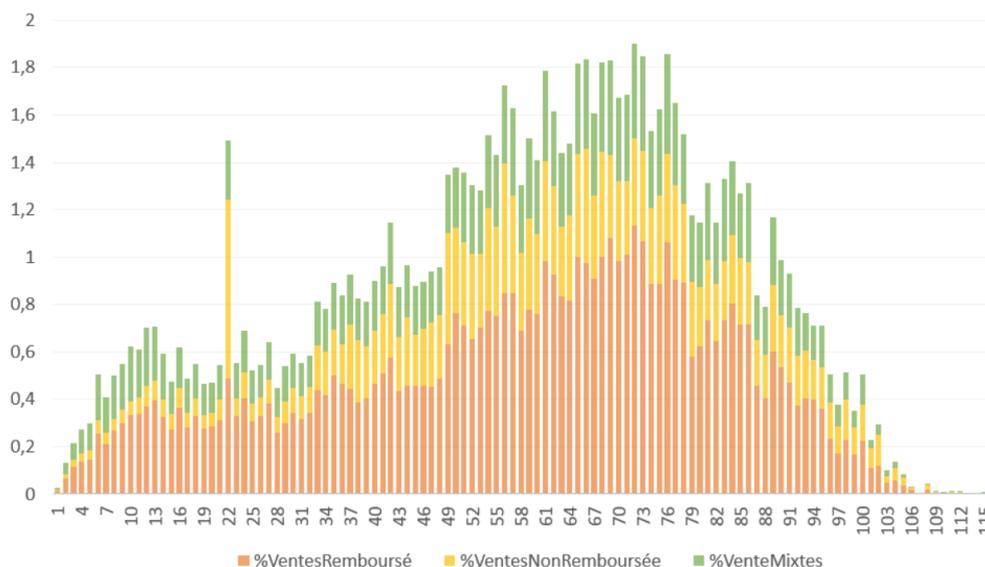


FIGURE 5.10 – Pourcentages des ventes d’une pharmacie de Gap, regroupés par type de ventes (remboursées, non-remboursées, mixtes) et par âge du patient.

jouer sur les prix dans sa pharmacie pour dégager plus ou moins de marge, ou être compétitif et attirer plus de patients. Dans la pharmacie étudiée, ces ventes représentent **23,12%** des ventes totales.

- 3. Les ventes mixtes** sont entre les deux autres types de ventes présentés précédemment. Elles sont constituées à la fois de produits remboursés et de produits non remboursés. Ce type de vente est intéressant car c’est un cas où un patient est venu pour des produits prescrits sur une ordonnance mais auquel le pharmacien a pu vendre un ou plusieurs produits non prescrits en supplément, par exemple pour améliorer le traitement et desquels il peut retirer une marge plus importante. Dans la pharmacie étudiée, ces ventes représentent **23,12%** des ventes totales.

Dans un objectif d’amélioration de la marge de la pharmacie, il est intéressant d’étudier les moyens de maximiser la transformation de ventes remboursées en ventes mixtes. Pour ce faire, apporter de l’information au pharmacien sur quelles paires de produits remboursés/non-remboursés se vendent le mieux ensemble pourrait aider ce dernier à optimiser son conseil et sa vente de compléments de traitement.

5.3.2.2/ ANALYSE DES VENTES CROISÉES

Suite à l’idée d’optimisation des ventes croisées des officines de pharmacie, une analyse approfondie des données de vente de produits remboursés et non remboursés (ventes croisées) s’est révélée pertinente. L’objectif était de déterminer quelles paires de produits, combinant des médicaments remboursés avec des produits non remboursés,

présentaient les meilleures performances en termes de transformation des ventes. Cela permettrait de proposer un outil de visualisation des produits compatibles entre eux basé sur l'empirisme des ventes. Dans le cas où des paires de produits seraient très liées, la CERP RRM aurait aussi un moyen de prédire les commandes d'un de ces produits en fonction des ventes du deuxième.

Malheureusement, les volumes de vente d'une seule pharmacie ne sont pas assez conséquents pour pouvoir réaliser ce genre de travaux de manière pertinente. Il serait intéressant de pouvoir regrouper les données de ventes d'un nombre plus conséquent d'officines et d'effectuer l'analyse sur ces données. Cela permettrait de consolider les résultats et de permettre aux pharmaciens d'avoir une vision globale de la stratégie de leurs confrères.

5.3.3/ AMÉLIORATION DE LA PRÉDICTION DES BESOINS (ACHATS AUTOMATIQUES SUR LEO)

La problématique des volumes de ventes décrite précédemment montre que déployer des techniques "lourdes" comme du machine learning nécessitant des ressources (données, temps, et puissance de calculs) assez importantes peut être difficile à mettre en place sur les données d'une officine de pharmacie seule.

Si certains pharmaciens commandent encore manuellement leur stock, d'autres ont opté pour une gestion de stock "automatique" en configurant leur LGO pour qu'il commande des produits si la quantité en stock descend sous un certain seuil.

La pharmacie étudiée utilisait également une fonction de commande automatique tous les mois. La prévision du besoin de cette fonction est une simple copie du passé : le logiciel commande le nombre de ventes qu'il y a eu au même mois de l'année précédente.

Nous avons proposé de remplacer cette technique par une moyenne roulante pour voir si cela permettait d'améliorer la prédiction du besoin. L'étude a été réalisée sur des produits fournis et considérés comme pertinents par le pharmacien, dont 2 produits saisonniers.

Le tableau 5.1 montre que pour ces dix produits, une moyenne roulante sur 6 mois a réduit la MAE de **59.05%** en moyenne, tandis qu'une moyenne roulante sur trois mois a réduit l'erreur de **61.08%**. Cela est possible car ces deux techniques permettent de corriger certains défauts d'une simple copie de l'année précédente :

1. Pour les produits qui ont des ventes "fluctuantes" et qui sont en rythme de croisière (ni morts, ni en croissance), les moyennes roulantes permettent de réduire les écarts dus à une fluctuation d'année en année (voir figure 5.11).
2. Pour les produits "en croissance", se baser sur l'année précédente cause souvent une commande plus basse que les besoins de l'instant T. Les moyennes roulantes

Produit	MAE RM 6	MAE RM 3	MAE Année N-1 (Baseline)
CHLORHEX CHL EG 0.5ML 0.5G P 100ML FL 90ML	24.71	30.72	31.08
CHOLECALCIFEROL EG 100 000 UI SOLUTION BUVABLE	39.61	35.64	85.33
KARDEGIC 75MG PDR SBUB SACH DOS 30SACH	23.18	20.11	95.92
LIDOC PRILO BGR 5% PANS ADH CUTANE B	29.18	28.69	98.25
PANTOPRAZ EG 20MG CPR GAST PLQ 28	15.25	16.75	18.5
PAROXET EG 20MG CPR PESE PLQ 14	21.78	18.86	34.25
PREDNISOL TEVA 20MG CPR OROD PLQ 20	5.36	3	31.82
ZYMAD 50 000 UI (CHOLECALCIFEROL) SOLUTION BUVABLE	22.18	19.56	60.42
LEVOCETIR EG 5MG CPR PELL PLQ 28 CPR PELL PLQ 28	8.28	6.56	7.83
LEVOCETIR EG 5MG CPR PELL PLQ 14 CPR PELL PLQ 28	0.82	1	1.42
Moyenne	19.04	18.09	46.48
Réduction de l'erreur (%)	59.05	61.08	

TABLE 5.1 – Comparatif de l'erreur absolue moyenne (MAE) des prédictions mensuelles pour les techniques de moyenne roulante sur 3 mois et 6 mois, par rapport aux ventes réelles pour l'année 2020. (meilleurs résultats en gras)

permettent de prendre en compte la tendance actuelle des derniers mois (voir figure 5.12).

3. Dans le cas contraire, pour un produit "mort" ou "en chute", commander les ventes de l'année précédente cause un sur-stock qui, selon les cas, peut être difficile à écouler pour le pharmacien (voir figure 5.13).

5.3.4/ CONCLUSION DE L'ANALYSE DES DONNÉES D'OFFICINE

Les expériences réalisées sur les données de pharmacie ont montré qu'il est possible d'améliorer la prédiction du besoin pour des produits vendus régulièrement grâce à des techniques comme une moyenne roulante. Elles ont également montré qu'il n'est pas utile pour la CERP RRM de faire des analyses à l'échelle de la pharmacie, car le volume de données est trop faible pour en tirer des analyses pertinentes. Nous allons donc remonter d'un niveau et étudier les données de commande de la CERP RRM.

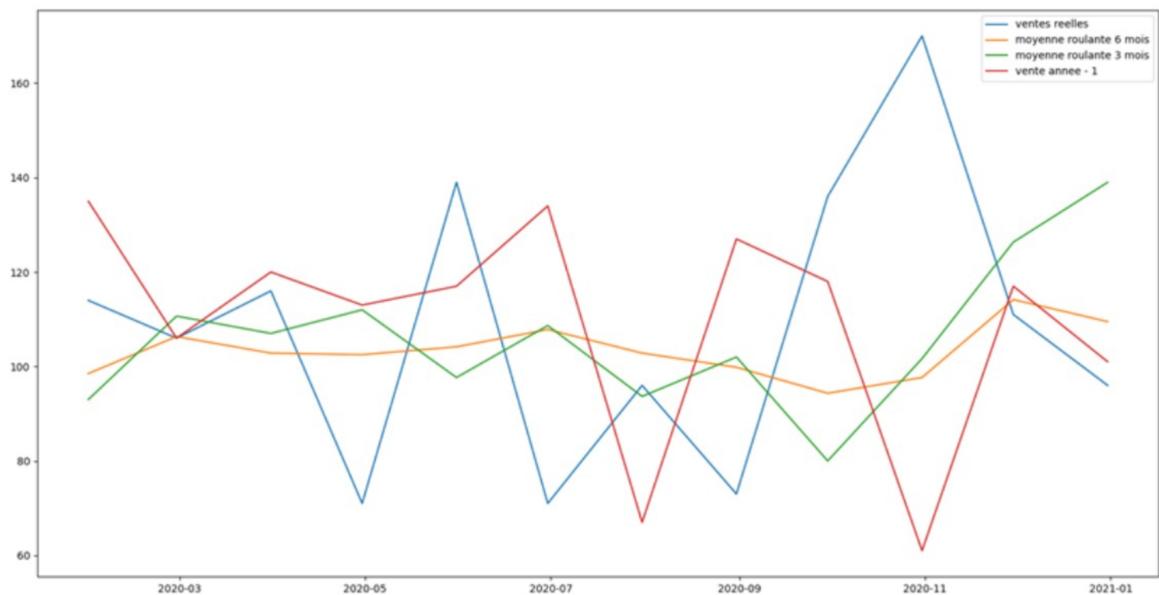


FIGURE 5.11 – Comparaison de la copie de l'année précédente, d'une moyenne roulante sur 6 mois, d'une moyenne roulante sur 3 mois et des ventes réelles pour un produit stable.

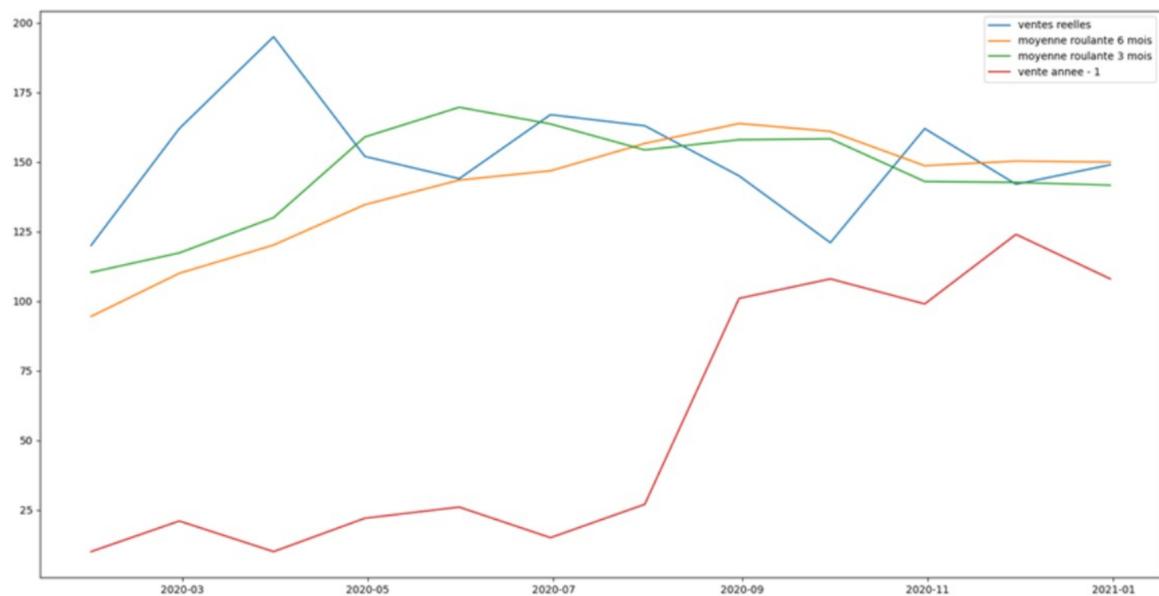


FIGURE 5.12 – Comparaison de la copie de l'année précédente, d'une moyenne roulante sur 6 mois, d'une moyenne roulante sur 3 mois et des ventes réelles pour un produit en croissance.

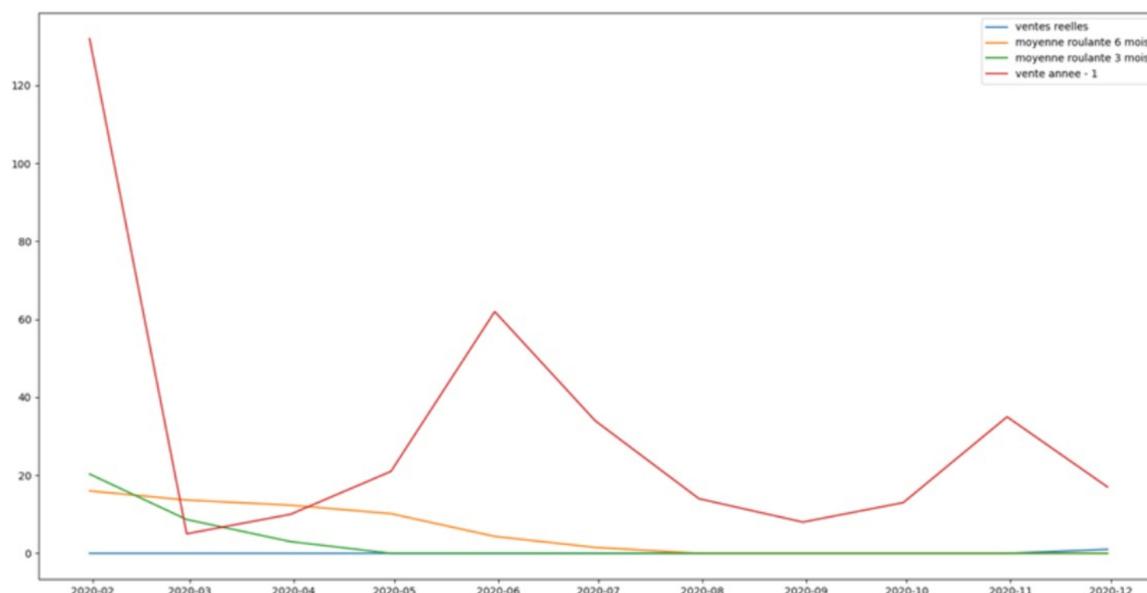


FIGURE 5.13 – Comparaison de la copie de l'année précédente, d'une moyenne roulante sur 6 mois, d'une moyenne roulante sur 3 mois et des ventes réelles pour un produit dont les ventes chutes.

5.4/ PRÉDICTION SUR LES DONNÉES DE LIVRAISONS DE LA CERP

5.4.1/ PRÉPARATION DES DONNÉES

Pour nos expériences, n'ayant pas directement accès aux données, nous avons demandé au service IT de La CERP RRM d'extraire les données de ventes, afin d'alimenter une base de données SQL que nous pouvons exploiter. La figure 5.14 montre le pipeline utilisé pour accéder aux données, montrant ainsi les différentes étapes entre les données de ventes brutes et celles disponibles en entrée des modèles de prédiction. L'ajout de ces différentes étapes a fait apparaître une complexité supplémentaire pouvant introduire de nombreux problèmes techniques, altérant la fiabilité des données (problèmes d'export, problèmes d'import ainsi qu'une gestion des doublons entre les différentes bases de données). Ces problèmes peuvent être facilement corrigés en travaillant avec un accès direct aux données de la CERP.

Une fois dans la base "IA", les données peuvent être extraites selon les agrégations voulues. Dans un premier temps, les prédictions ont été effectuées sur 4 molécules tests pour l'ensemble des ventes de la CERP RRM. Suite à la méthode statistique mise en place par la CERP, nous avons réorienté nos travaux pour pouvoir les comparer à la méthode actuelle. Des prédictions ont donc été réalisées sur la nouvelle agrégation de données : par classe ATC5, à la semaine pour un établissement CERP.

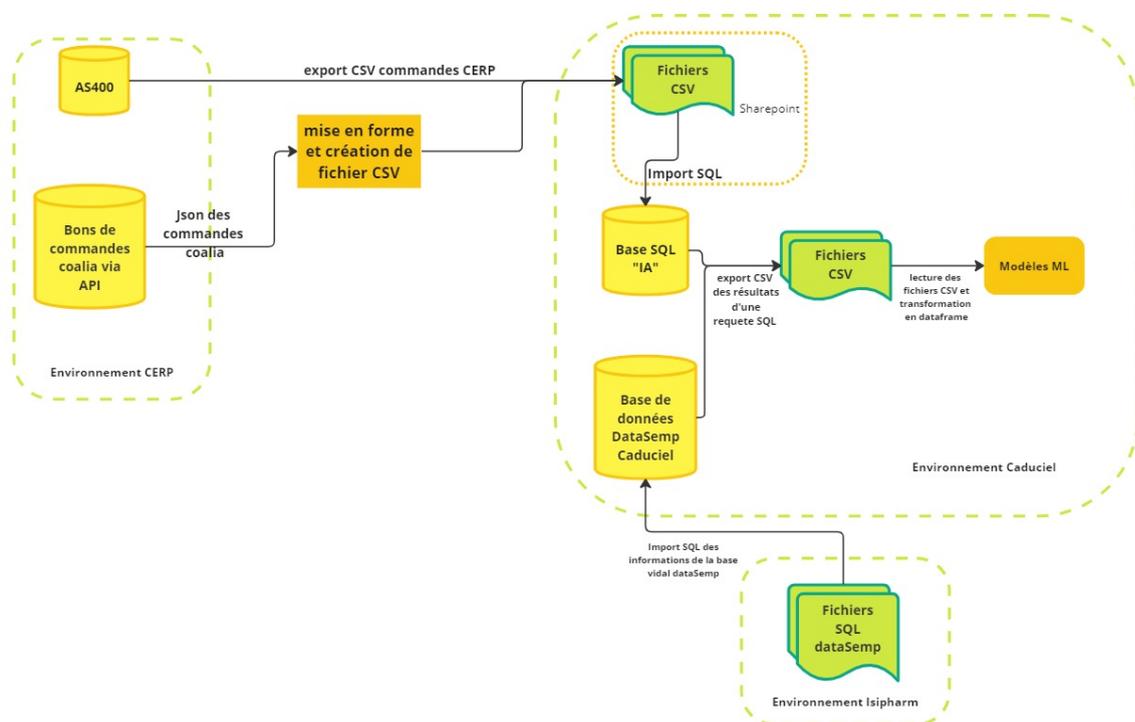


FIGURE 5.14 – Schéma simplifié du pipeline de données utilisé pour les modèles prédictifs

5.4.2/ ARCHITECTURE DE MACHINE LEARNING

Les architectures de machine learning utilisées dans ces expérimentations ont été réalisées avec le langage de programmation Python et la bibliothèque TensorFlow [80].

5.4.2.1/ LONG SHORT-TERM MEMORY (LSTM)

Ce modèle est composé d'une première couche LSTM de 300 unités avec une fonction d'activation ReLU, suivie d'une deuxième couche LSTM de 100 unités, toujours avec une activation ReLU, pour se terminer par une couche "Dense" de taille 1. Entre les deux couches LSTM, un taux de "dropout" de 20% a été ajouté pour éviter un éventuel surajustement. L'optimiseur RMSprop a été utilisé avec un taux d'apprentissage de $1e-4$ et le modèle possède les valeurs jusqu'à $N - 6$.

5.4.2.2/ LIGHTGBM

Pour la configuration du modèle LightGBM, une recherche en grille a été effectuée sur les paramètres suivants : le taux d'apprentissage avec une variation de 0,01 entre 0,01 et 0,08, le nombre d'estimateurs entre 100 et 2000 avec un pas de 100, le nombre de

feuilles entre 2 et 10 avec un pas de 1, et la profondeur maximale entre 2 et 16 avec également un pas de 1. L'arrêt anticipé a été défini à 300 époques, car le modèle avait tendance à s'arrêter au début de l'entraînement avec un arrêt anticipé plus court.

5.4.2.3/ TRANSFORMERS

L'architecture Transformers utilisée est celle implémentée dans l'API de deep learning Keras [82], pour la classification [125], mais adaptée à la régression avec une couche de sortie "dense" de taille 1 et une fonction d'activation "linéaire". L'architecture possède les paramètres suivants : une taille de tête de 256 et 4 têtes pour l'attention multi-tête, avec un taux de "dropout" de 0,25. L'optimiseur Adam a été utilisé avec un taux d'apprentissage de 1e-4. L'entraînement se déroule sur 1000 époques, avec un arrêt anticipé après 20 époques et une taille de lot de 64.

5.4.3/ PRÉDICTIONS PAR MOLÉCULES

5.4.3.1/ BENCHMARK DE L'ERREUR QUADRATIQUE MOYENNE (RMSE)

Les tableaux suivants présentent une comparaison des différentes RMSE de chaque modèle sur la prédiction des ventes des quatre molécules : Amoxicilline (tableau 5.2), Cétirizine (tableau 5.3), Colécalciférol (tableau 5.4) et Paracétamol (tableau 5.5). Comme les échelles du nombre de produits vendus ne sont pas les mêmes pour les quatre molécules, la colonne "/mean" représente la RMSE divisée par la moyenne des ventes de chaque molécule, afin de permettre une comparaison entre elles.

Amoxicilline				
Modèle	RMSE (train)	/mean	RMSE (test)	/mean
Baseline	36020	0,36	39244	0,39
Stacked-LSTM	29092	0,29	26947	0,27
LightGBM	48934	0,49	27977	0,28
Transformers	29986	0,30	23566	0,23

TABLE 5.2 – Comparaison de RMSE pour la prévision des ventes d'Amoxicilline

Cétirizine				
Modèle	RMSE (train)	/mean	RMSE (test)	/mean
Baseline	4006	0,30	3662	0,27
Stacked-LSTM	3040	0,23	6200	0,47
LightGBM	4470	0,49	3373	0,25
Transformers	3299	0,25	2103	0,16

TABLE 5.3 – Comparaison de RMSE pour la prévision des ventes de Cétirizine

Colécalciférol				
Modèle	RMSE (train)	/mean	RMSE (test)	/mean
Baseline	21397	0,22	25603	0,27
Stacked-LSTM	13308	0,14	29700	0,31
LightGBM	26358	0,28	27480	0,29
Transformers	17100	0,18	23433	0,24

TABLE 5.4 – Comparaison de RMSE pour la prévision des ventes de Colécalciférol

Paracétamol				
Modèle	RMSE (train)	/mean	RMSE (test)	/mean
Baseline	90713	0,27	106626	0,32
Stacked-LSTM	88230	0,27	63182	0,19
LightGBM	61733	0,19	102648	0,31
Transformers	106431	0,32	53491	0,16

TABLE 5.5 – Comparaison de RMSE pour la prévision des ventes de Paracétamol

Le modèle LSTM parvient à mieux approximer les courbes sur les données d'entraînement, car il essaie d'apprendre les micro-changements que l'on peut considérer comme du bruit. Cependant, le modèle Transformers est plus efficace avec les données de test : il est capable de généraliser la forme de la courbe et est donc plus précis que le modèle LSTM sur des données qu'il n'a jamais vues. Cela peut être observé sur les courbes ci-dessous. Il est également intéressant de noter que le modèle LightGBM n'apprend pas les courbes, mais obtient de meilleurs résultats sur le paracétamol en imitant le modèle de base et en prédisant des valeurs autour de la moyenne.

5.4.3.2/ COURBES DE PRÉVISION

Pour toutes les figures contenues dans cette sous-section, les courbes violettes représentent les données réelles, les courbes vertes les prédictions sur les données d'entraînement, les courbes rouges (si elles existent) représentent les prédictions sur les données de l'ensemble de validation et les courbes bleues sont les prédictions sur l'ensemble de test. L'axe des abscisses représente les semaines et l'axe des ordonnées représente le nombre de ventes.

Le modèle Transformers a tendance à suivre la tendance sans se soucier des pics de bruit (figure 5.15, figure 5.16, figure 5.17 et figure 5.18). Au contraire, le modèle Stacked-LSTM essaie d'approximer la courbe autant que possible (figure 5.19, figure 5.20, figure 5.21 et figure 5.22). Cela fait une différence dans l'ensemble de test qui n'a pas les mêmes proportions que les autres années : le Stacked-LSTM prédit une courbe similaire aux années précédentes tandis que le Transformers suit la tendance à la baisse des ventes. Le modèle LightGBM a de très mauvaises compétences prédictives sur les données d'en-

entraînement mais approxime assez bien la courbe sur les données de test (figure 5.23 et figure 5.24). Il tend également à s'approcher de la moyenne. Cela montre qu'il ne comprend pas la saisonnalité, mais permet d'obtenir de bons résultats sur les données de paracétamol qui ont une courbe plus linéaire (figure 5.25 et figure 5.26).

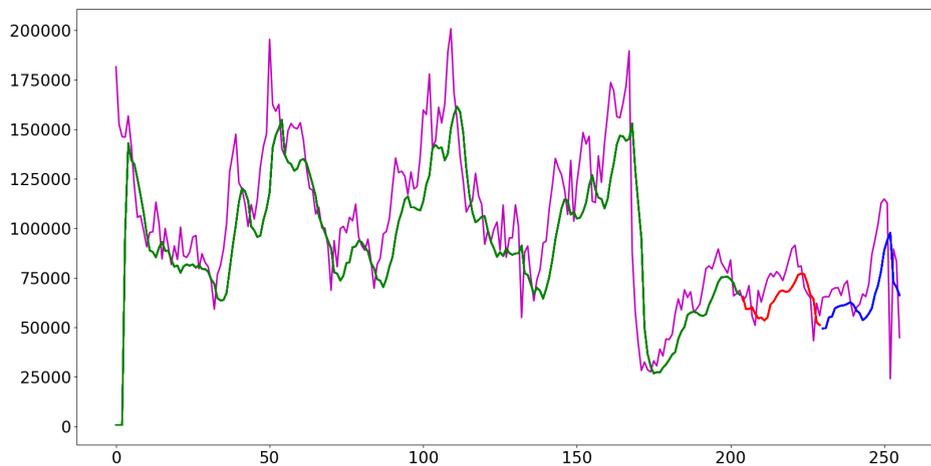


FIGURE 5.15 – Courbes de prévision du modèle Transformers pour l'Amoxicilline

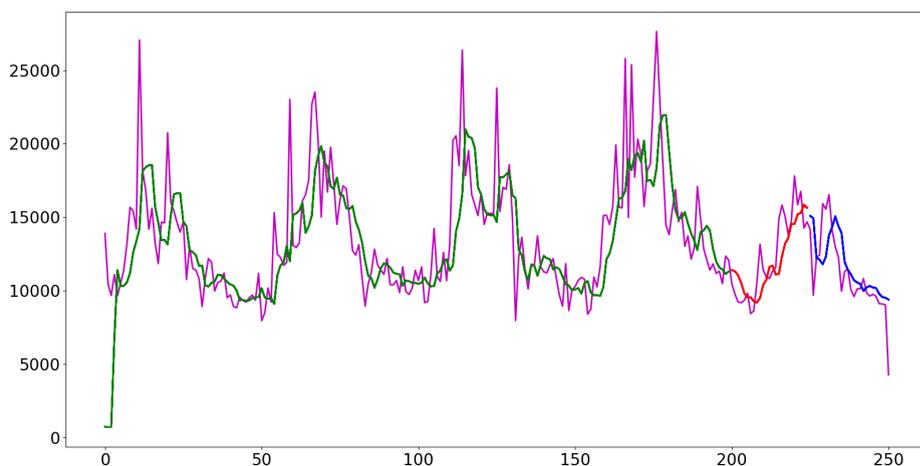


FIGURE 5.16 – Courbes de prévision du modèle Transformers pour la Cétirizine

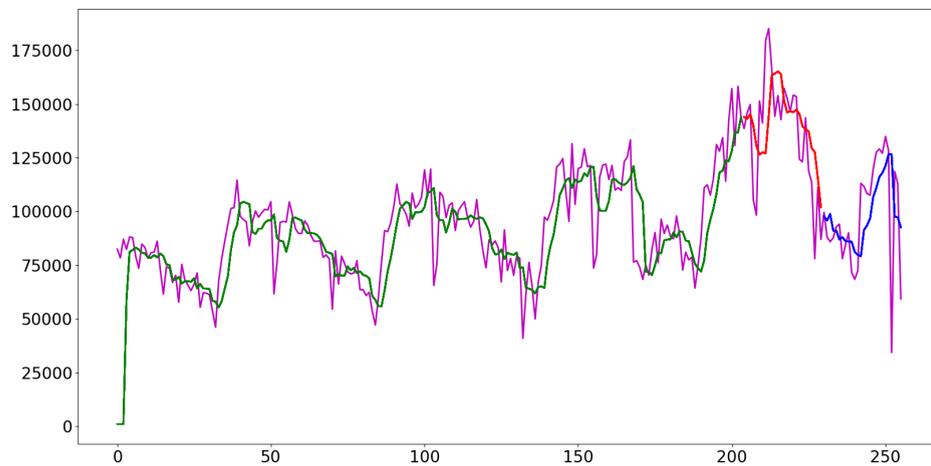


FIGURE 5.17 – Courbes de prévision du modèle Transformers pour le Colécalciférol

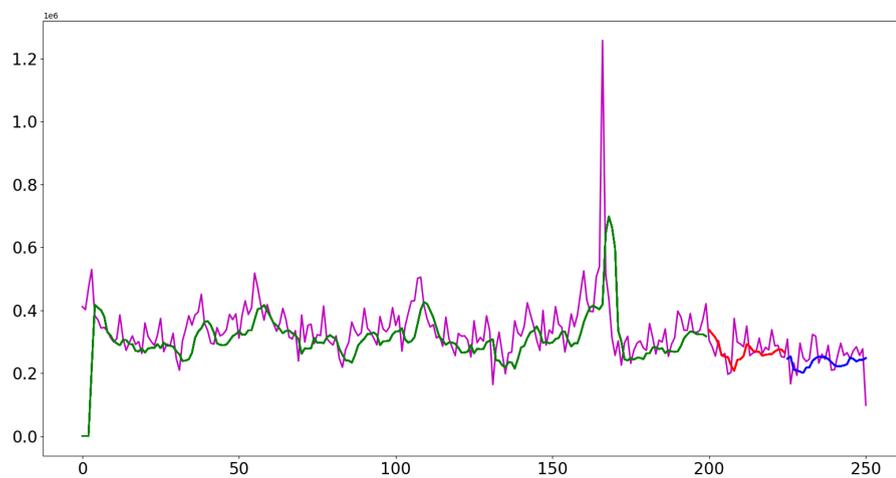


FIGURE 5.18 – Courbes de prévision du modèle Transformers pour le Paracétamol

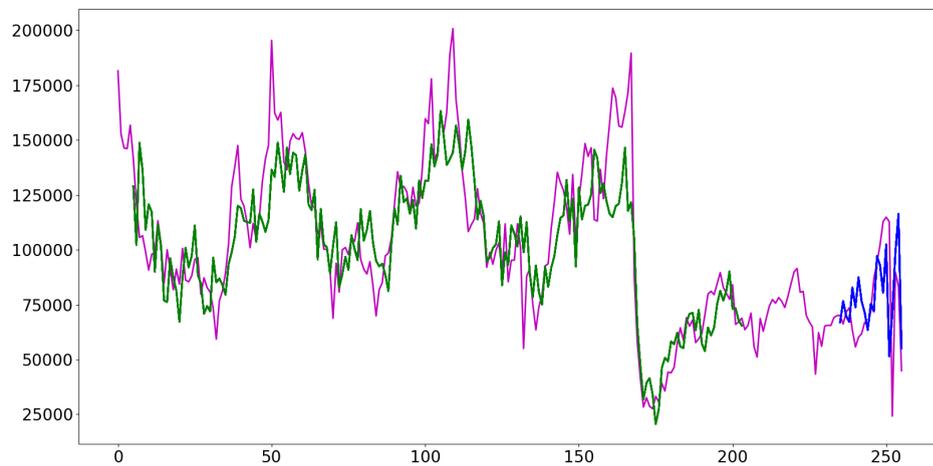


FIGURE 5.19 – Courbes de prévision du modèle LSTM pour l'Amoxicilline

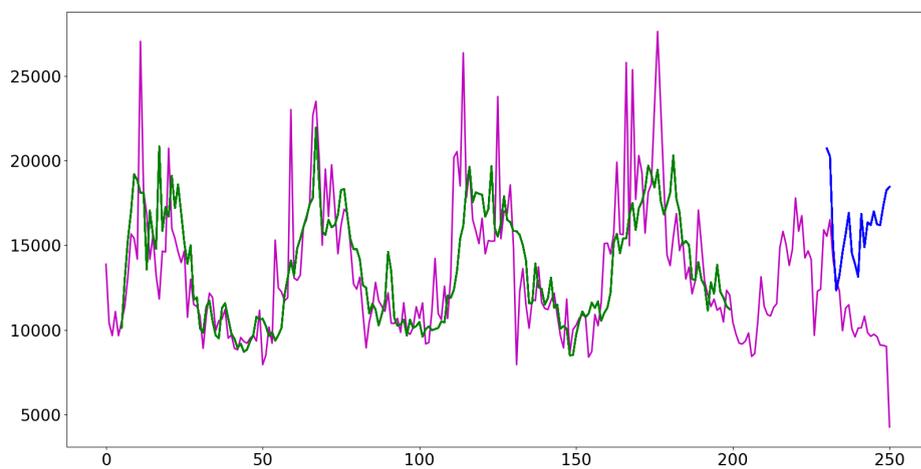


FIGURE 5.20 – Courbes de prévision du modèle LSTM pour la Cétirizine

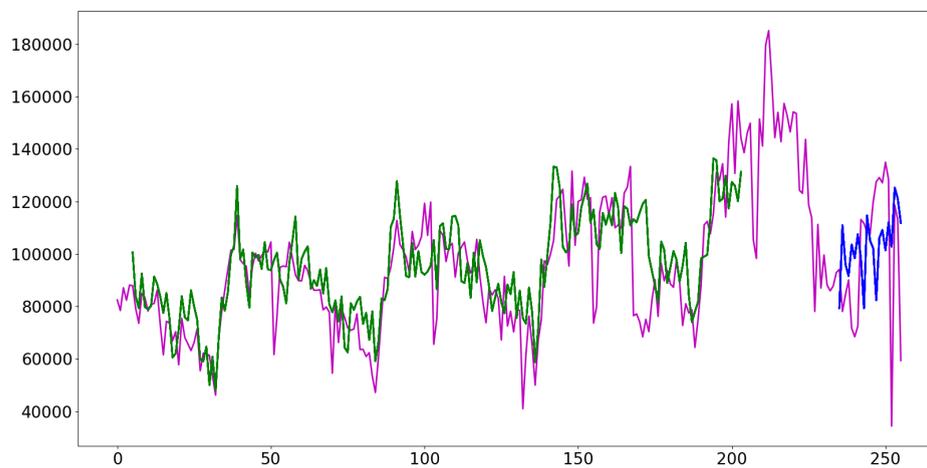


FIGURE 5.21 – Courbes de prévision du modèle LSTM pour le Colégicaliférol

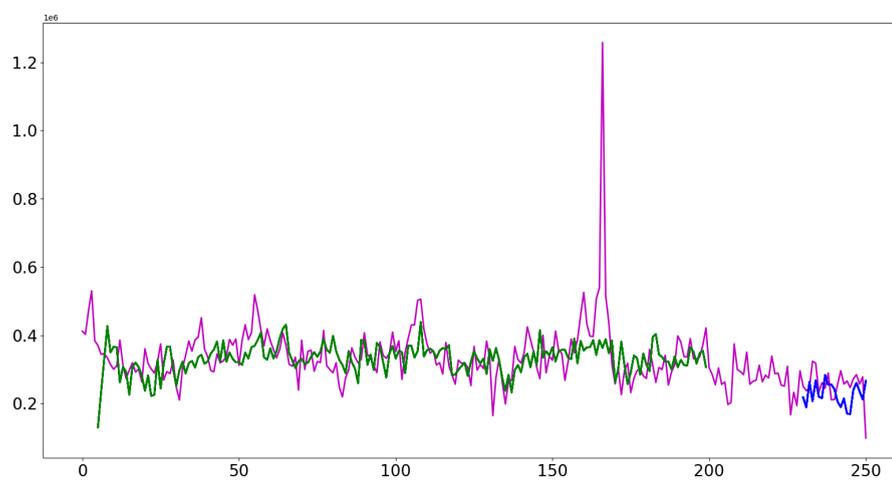


FIGURE 5.22 – Courbes de prévision du modèle LSTM pour le Paracétamol

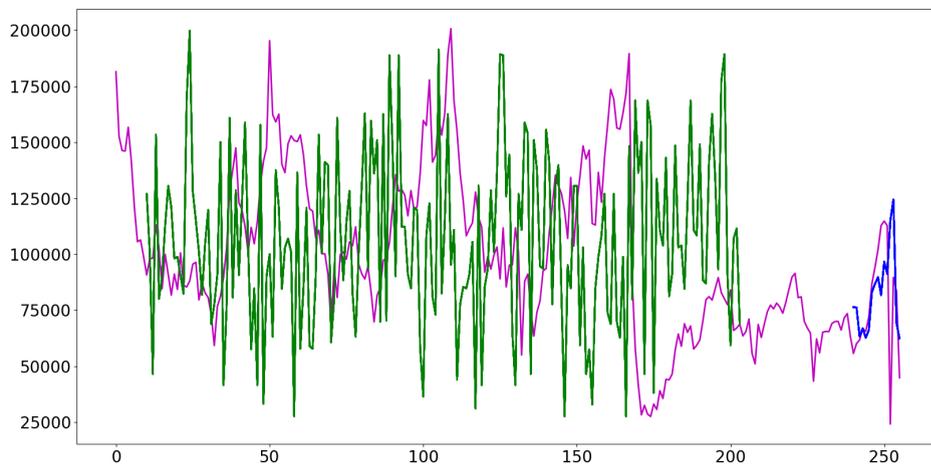


FIGURE 5.23 – Courbes de prévision du modèle LightGBM pour l'Amoxicilline

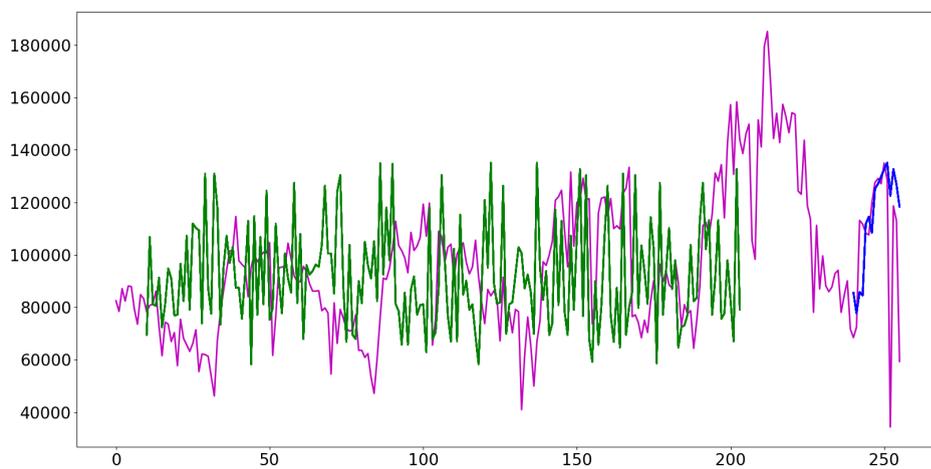


FIGURE 5.24 – Courbes de prévision du modèle LightGBM pour le Colécalciférol

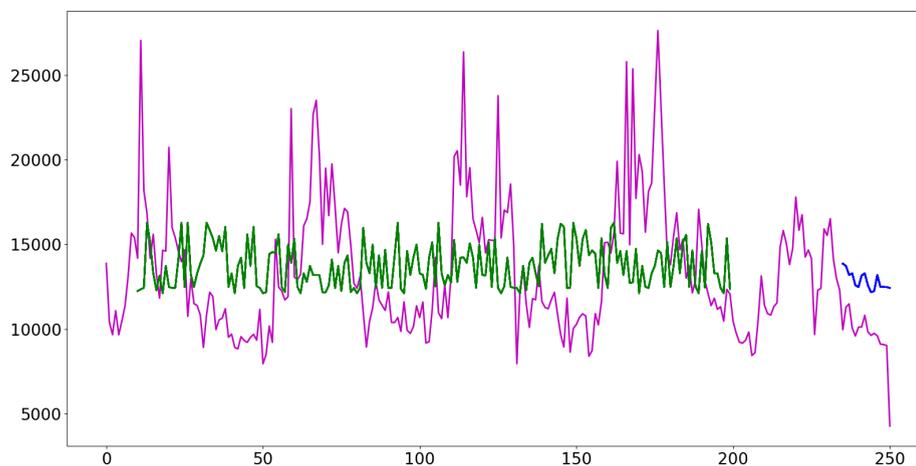


FIGURE 5.25 – Courbes de prévision du modèle LightGBM pour la Cétirizine

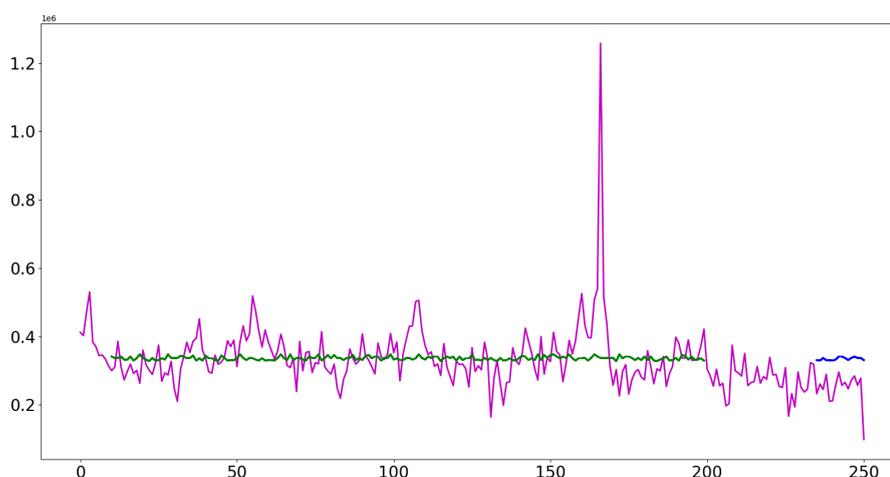


FIGURE 5.26 – Courbes de prévision du modèle LightGBM pour le Paracétamol

Ces travaux ont permis de choisir deux types de modèles (LSTM et Transformers) pour passer à l'échelle de toutes les molécules. Mais durant ces travaux, la CERP a également mis en place un outil de prédictions basé sur des méthodes statistiques.

5.4.4/ MÉTHODE STATISTIQUE MISE EN PLACE PAR LA CERP

Durant cette thèse, la CERP RRM a lancé en parallèle un projet de prévision basé sur des méthodes statistiques. Grâce à une solution logicielle connectée à leur base de données

de vente, la CERP peut utiliser différentes méthodes statistiques pour prédire les futurs besoins :

1. **Lissage exponentiel**
2. **Comportement saisonnier**
3. **Box-Jenkins**
4. **Moyenne mobile**
5. **Approximation de courbes**
6. **Copie du passé**
7. **Historique et prévision statistique**

Les prédictions sont effectuées pour l'année en cours et l'année suivante, et elles sont mises à jour mensuellement. Cependant, le calcul de l'erreur est uniquement réalisé sur la prédiction pour le mois suivant.

L'erreur est calculée sous forme de la valeur absolue moyenne (MAE).

Agrégation à l'échelle temporelle : Les prédictions sont réalisées à une échelle mensuelle, ce qui implique que les données de ventes sont agrégées au niveau mensuel.

Agrégation à l'échelle du produit : L'agrégation des données au niveau du produit présente deux cas distincts :

1. **Pour les médicaments** : Dans le cas des médicaments, le niveau 5 de la classification ATC est utilisé pour regrouper les ventes, qui sont ensuite ramenées au niveau du produit grâce à un système de ratios.
2. **Pour les produits hors médicaments** : En l'absence de regroupements pertinents pour les produits hors médicaments, l'agrégation des ventes se fait directement au niveau du produit.

Agrégation à l'échelle de la population : Les prédictions sont réalisées pour l'ensemble des ventes de la CERP. Cela implique une agrégation de l'ensemble des ventes mensuelles d'un produit pour l'ensemble des clients de la CERP. Ensuite, un système de ratios est utilisé pour rapporter les quantités estimées à l'établissement.

Avec cette méthode, la CERP RRM obtient en moyenne une précision de **85%** pour les prédictions au niveau du produit pour l'ensemble de ses clients. Cependant, cette précision diminue en moyenne à **75%** lorsque les prédictions sont rapportées à l'échelle de l'établissement.

5.4.4.1/ PRÉDICTIONS PAR CLASSIFICATION ATC5

Ce nouvel outil a permis de recentrer nos travaux sur des regroupements plus pertinents d'un point de vue des médicaments (classification ATC5) tout en offrant une base de comparaison des résultats plus précise.

Les modèles LSTM et Transformers ont été testés sur 116 catégories ATC5 disponibles dans nos données, précisément sur les ventes de l'établissement d'Aix en Provence. Pour simuler les mêmes conditions de prédiction que l'outil statistique, les modèles ont eu en entrée les ventes de la semaine N-4, ainsi que la tendance des ventes sur la période [N-5,N-8]. Les autres paramètres restent les mêmes que pour les modèles présentés pour la prédiction de molécules. Par exemple, pour la prédiction de la semaine N, le modèle LSTM aura accès aux données de ventes telles que décrites dans le schéma de la figure 5.27.

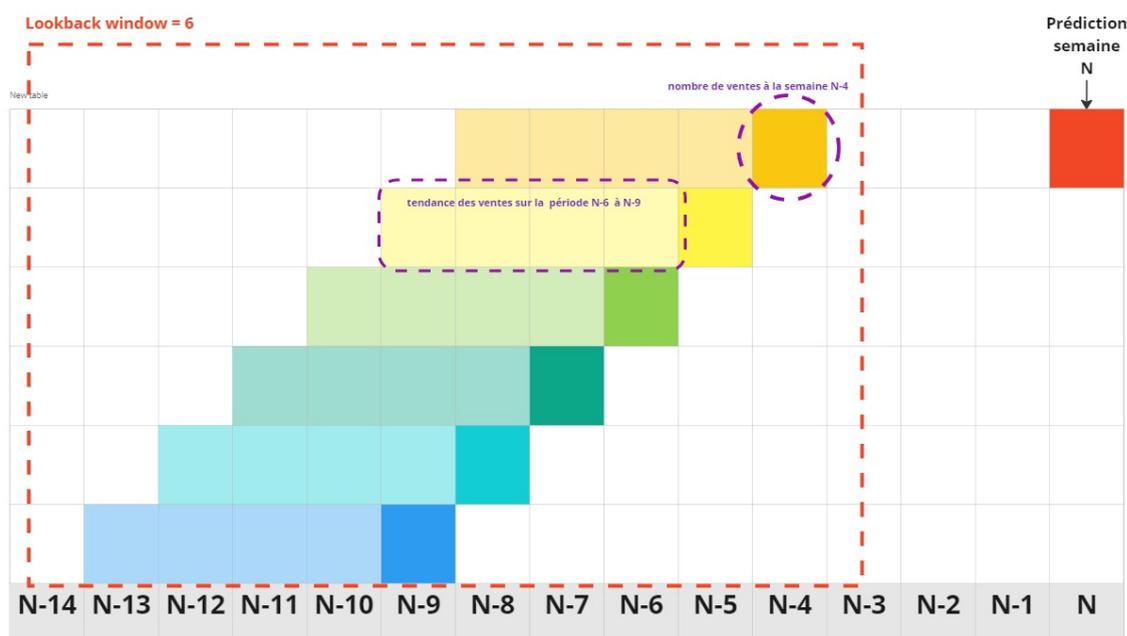


FIGURE 5.27 – Schéma d'explication des données d'entrée pour le modèle LSTM sur les prédictions à la semaine pour les classes de médicaments ATC5. Pour prédire la semaine N, on donne en entrée le nombre de vente des semaines [N-4,N-9], soit une fenêtre de lookback de 6 pas de temps. On donne également pour chacune de ces valeurs, la tendance des ventes sur les 4 pas de temps précédents : Par exemple, on accompagne le nombre de vente de la semaine N-4 de la tendance des ventes sur la période [N-5,N-8].

L'erreur a été calculée sous deux angles :

1. **Erreur absolue à la semaine** : Les modèles effectuent une prédiction à la semaine

sur 4 semaines, puis l'erreur est calculée indépendamment pour chaque semaine, pour finalement être additionnée. Cette solution regarde donc les prédictions à la semaine, ce qui est un niveau de précision plus élevé que l'outil statistique.

- 2. Erreur absolue au mois** : Les modèles effectuent toujours une prédiction à la semaine sur 4 semaines, mais cette fois-ci on additionne les 4 valeurs pour avoir une prédictions sur le mois avant de calculer l'erreur. Cela permet aux modèles de pouvoir contrebalancer l'erreur sur les 4 semaines. Ce qui correspond finalement aux prédictions faites par l'outil statistique.

Une comparaison des résultats de chaque méthodes a été effectué pour le mois de septembre 2022 :

Modèle	% MAE Mois	% MAE Semaine
LSTM	8.44	13.29
Transformers	10.42	13.50
Outil stat	19.25 *	/

TABLE 5.6 – Comparaison de la MAE pour le mois de septembre 2022, sur toutes les ventes par catégorie ATC5 pour l'établissement CERP d'Aix en Provence. (* Cette statistique prend en compte la totalité des produits et pas seulement les produits ATC (médicaments). La donnée séparée pour ce mois ci n'as pas pu être fournie. A titre de comparaison : pour le mois de février 2024 l'erreur des produits ATC est de 24,2% et de 28.01% pour les produits hors ATC)

5.4.5/ CONCLUSION DES TRAVAUX SUR LES PRÉDICTIONS

Dans cette sous-partie, nous avons pu observer que l'utilisation de méthodes basées sur des réseaux de neurones permettait d'obtenir des prédictions intéressantes avec un niveau de précision plus élevé que les méthodes actuelles. L'inconvénient de devoir entraîner de multiples modèles peut être réduit en regroupant les produits dans des groupes cohérents. Si pour les médicaments la classification ATC de niveau 5 permet de remplir ce rôle, pour les produits de parapharmacie, il n'existe pas de classification pertinente préétablie. La force des modèles basés sur des réseaux de neurones est de pouvoir ingérer un grand nombre de données et y discerner des corrélations potentielles. L'idée d'ajouter des données externes en plus des données de ventes pour améliorer les prédictions est donc prometteuse.

5.5/ AJOUT DE DONNÉES EXTERNES

Pour déterminer quelles données externes pourraient être pertinentes, des étudiants en pharmacie ont fourni leur expertise sur une étude des causes de vente dans les

catégories suivantes :

1. Antiprurigineux, incluant antihistaminiques, anesthésiques, etc.
2. Préparations pour la gorge
3. Médicaments du rhume et de la toux
4. Allergènes
5. Préparations pour le nez
6. Système cardiovasculaire
7. Vaccins
8. Toniques
9. Anthelminthiques

De cette analyse, il est ressorti des grandes catégories de causes provoquant les ventes des produits liés à ces catégories :

1. **Causes temporelles** : On y retrouve des facteurs tels que les saisons, les périodes de vacances, les périodes scolaires, etc. Ce type de données est déjà implémenté grâce aux dates présentes dans les jeux de données.
2. **Causes météorologiques** : Le principal facteur dans cette catégorie est le changement de température, qui peut influencer la saisonnalité des séries temporelles. On peut également s'intéresser à l'humidité et au vent, mais ce sont des données qui ont des impacts géographiques très locaux et qui ne sont pas pertinents à l'échelle de notre problématique.
3. **Causes épidémiologiques** : Les données d'observation du nombre de malades peuvent être un bon indicateur du nombre de ventes des traitements correspondants. Pour la France, le réseau Sentinelle [148] fournit des données et des bulletins d'observations pour la varicelle, la diarrhée aiguë et les infections respiratoires aiguës (IRA).
4. **Causes liées à la pollution** : La pollution peut être une cause d'augmentation de certaines pathologies. On peut facilement trouver des jeux de données grâce, par exemple, au Laboratoire Central de Surveillance de la Qualité de l'Air (LCSQA) [2] et sa plateforme de données Géod'air [1]. Cette plateforme possède des données historiques de mesures de la pollution, mais elle ne réalise que des prévisions jusqu'à J+2, ce qui pose problème pour nos prédictions au mois suivant. La pollution est également très influencée par les températures : l'air froid a tendance à "fixer" les particules polluantes près du sol. On peut donc observer une tendance saisonnière avec des pics de pollution en hiver.

L'ajout de données externes (températures) en entrée des modèles de prédiction de ventes a permis d'améliorer les résultats de **73** catégories ATC niveau 3 sur un total de **95**, passant d'une erreur moyenne pondérée de **16%** à **13%**. L'ajout de données externes peut donc s'avérer pertinent. La difficulté première à l'application dans un contexte industriel est d'obtenir des prédictions de ces différentes données et de pouvoir les intégrer dans un pipeline fiable et robuste. Heureusement, on remarque que la plupart de ces données externes ont une corrélation avec les températures (et donc les saisons), ce qui a motivé le choix de tester cette donnée pour améliorer la compréhension de la tendance saisonnière pour les modèles.

5.6/ CONCLUSION

Dans ce chapitre, nous avons pu observer que la corrélation entre les besoins de traitement et les ventes à la CERP RRM n'est pas nécessairement directe, étant donné la diversité des acteurs sur le marché et la répartition logique du besoin entre ces acteurs. De plus, les capacités de stockage des différents intervenants ont tendance à diluer les besoins en traitements, ce qui complexifie la prévision des ventes.

En examinant les avantages et les contraintes des différents niveaux de granularité pour l'agrégation des ventes de produits, il a été relevé que l'analyse des données de pharmacie offre des informations complémentaires précieuses par rapport aux données de ventes du grossiste-répartiteur. Toutefois, l'analyse directe des données d'une seule pharmacie pour la CERP RRM s'avère peu pertinente en raison du faible volume de données disponibles. En effet, les expériences menées sur les données de pharmacie ont démontré la possibilité d'améliorer la prédiction des besoins pour les produits vendus régulièrement, notamment par l'utilisation de techniques telles que la moyenne mobile. Cependant, il a été constaté que le volume de données d'une seule pharmacie n'est pas suffisant pour des analyses pertinentes de la part de la CERP RRM. Ne pouvant pas obtenir un jeu de données sur plus de pharmacies au moment de l'étude, cela a conduit à une réorientation des travaux sur les données de commande de cette dernière.

L'analyse des données de commande a révélé que l'utilisation de méthodes basées sur les réseaux de neurones permet d'obtenir des prédictions plus précises, notamment en regroupant les produits dans des catégories cohérentes. L'intégration de données externes, telles que les données météorologiques, a été observée comme un moyen prometteur d'améliorer les prédictions, notamment en ce qui concerne la corrélation saisonnière des ventes.

Si le regroupement des médicaments par la classification ATC a permis de faire des groupes cohérents, il reste le problème des produits de parapharmacie qui n'ont jus-

qu'alors pas de classification correcte. Le prochain chapitre de cette thèse propose une méthode pour regrouper ces produits en se basant sur l'étude de leurs données de ventes.

TSCAPE : UNE NOUVELLE MÉTHODE DE CLUSTERING DE DONNÉES TEMPORELLES

6.1/ INTRODUCTION

Les recherches axées sur l'analyse des cycles de vie des produits à partir des données provenant d'une pharmacie ont suscité un intérêt considérable au sein de la CERP. En effet, cette entreprise est confrontée à des défis de classification de certains de ses produits, en particulier ceux de la parapharmacie qui ne bénéficient pas de classifications normalisées similaires à celles des médicaments. Cette situation soulève des problématiques complexes en ce qui concerne la prédiction et le regroupement de ces produits.

Prenons l'exemple du regroupement des "huiles essentielles". Certaines de ces huiles se vendent davantage pendant les mois d'été, tandis que d'autres connaissent une demande accrue en hiver, présentant ainsi des saisonnalités inverses. Cette variation saisonnière complexe ajoute une dimension supplémentaire à la tâche de regroupement. Actuellement, la CERP effectue des prédictions de regroupement en fonction des données historiques, puis répartit les quantités en fonction de la part de marché de chaque produit. Cependant, ce problème de saisonnalité contraire compromet considérablement l'efficacité de ce système de prédiction.

En prenant en compte ces observations, il est devenu impératif d'adopter une nouvelle approche pour le regroupement des produits. Cela a conduit à la reprise de nos travaux sur l'étude des formes de courbes de vente des produits. Cette démarche vise à mieux comprendre les motifs saisonniers, les tendances et les variations dans les schémas de vente. Une analyse approfondie de ces caractéristiques servira de base solide pour le développement d'une nouvelle technique de regroupement plus efficace et adaptée à la

diversité des produits de la parapharmacie.

Dans cette nouvelle approche, nous prévoyons d'utiliser des méthodes d'apprentissage automatique et de fouille de données pour extraire des motifs saisonniers et des tendances significatifs à partir des données de vente. En combinant ces informations avec des techniques avancées de regroupement, nous espérons obtenir des clusters plus cohérents et représentatifs des similarités et des différences entre les produits.

6.2/ ÉTAT DE L'ART

L'objectif principal de cette contribution consiste à examiner la structure des courbes des séries temporelles en vue de les regrouper en fonction de leurs similarités. La littérature propose différentes techniques pour comparer ces séries temporelles, comme illustré dans le travail de [51]. Les auteurs catégorisent trois principaux types de regroupement pour les séries temporelles : basé sur les données brutes dans [28], utilisant le regroupement par logique floue sur des données d'imagerie par résonance magnétique fonctionnelle ; basé sur les caractéristiques dans [16], employant l'analyse de cluster pour identifier les oscillateurs locaux ; et basé sur le modèle dans [38], comparant le cepstrum de prédiction linéaire des séries temporelles ARIMA. Les techniques exploitant les distances euclidiennes (*par exemple*, [36]) consistent à créer une matrice de similarité à travers la distance euclidienne pour l'analyse en composantes principales (ACP). Cependant, l'exploration de techniques plus efficaces, telles que la Dynamic Time Warping (DTW) proposée par [12], reste moins explorée.

6.2.1/ TECHNIQUES DE COMPARAISON DE DISTANCES DANS L'ANALYSE DE SÉRIES TEMPORELLES

La Dynamic Time Warping (DTW) est une méthode bien connue pour calculer la plus petite distance entre deux séries temporelles, permettant une flexibilité dans l'alignement des points pour prendre en compte les variations de vitesse et de temps. La DTW a trouvé des applications majeures dans les tâches de regroupement de séries temporelles et au-delà.

L'efficacité de la DTW dans le regroupement de séries temporelles est évidente dans diverses études, illustrant son adaptabilité et son utilité en reconnaissance de formes, comme illustré dans [66]. L'application de la DTW s'est étendue aux études aviaires où elle a identifié des motifs dans les chants d'oiseaux, offrant des insights sur la communication et le comportement des oiseaux dans le travail de [71].

De plus, la Longest Common SubSequence (LCSS), telle qu'utilisée dans [43], fonctionne

en identifiant la plus longue sous-séquence commune à deux séries temporelles, fournissant une mesure robuste de similarité. La distance d'édition avec pénalité réelle (ERP), voir [46], offre une technique de calcul de la distance entre les séries temporelles tout en tenant compte de leur alignement et de leurs différences d'amplitude. La Distance Invariante à la Complexité (CID), proposée par [73], se concentre sur la capture du motif de complexité des séries temporelles, la rendant robuste contre les variations de longueur et d'échelle.

De plus, les mesures basées sur la forme comme la Slope Distance (voir [110]) et l'Analyse de Forme Procrustes, comme dans le travail de [141], ont été utilisées pour aligner des matrices. Ces techniques fournissent différentes méthodologies pour comparer des séries temporelles et sont utilisées dans divers domaines, contribuant significativement à la reconnaissance de motifs dans [140], le traitement du signal dans [145], l'analyse de données ([143]), et l'identification ([131]).

6.2.2/ TECHNIQUES D'INCORPORATION ET DE RÉDUCTION DE LA DIMENSIONNALITÉ

Les techniques d'incorporation et de réduction de la dimensionnalité jouent un rôle crucial dans la visualisation de données haute dimension en dimensions inférieures tout en préservant des informations cruciales.

Le MultiDimensional Scaling (MDS) est une technique fondamentale utilisée pour transformer les matrices de distances en nuages de points sur des plans euclidiens prédéfinis, comme proposé dans [60]. Ses applications couvrent divers domaines, démontrant sa polyvalence et son efficacité. Par exemple, le MDS a été appliqué dans l'analyse de profil, révélant des motifs distincts dans les profils individuels, comme dans [32]. De plus, le MDS a été utilisé pour prédire la faillite, fournissant une visualisation succincte de la situation financière des entreprises, comme dans le travail de [47]. Le MDS a également été utilisé pour analyser la santé financière des banques, offrant des aperçus de la dynamique du secteur bancaire, voir, par exemple, [39]. De plus, le MDS a trouvé application dans l'analyse du comportement des virus en fournissant une compréhension structurée des motifs de comportement des virus, voir [86]. La technique collabore souvent avec des algorithmes de regroupement, comme le montre l'étude analysant les chants des épaulards dans [55], en utilisant le MDS dans le cadre d'une méthodologie de regroupement non supervisée s'appuyant sur une séquence d'algorithmes.

En plus du MDS, plusieurs autres algorithmes de réduction de la dimensionnalité existent avec des caractéristiques et des applications distinctes. L'ACP (Analyse en Composantes Principales, utilisée par exemple dans [85]) et le KernelPCA sont des techniques linéaires largement utilisées qui projettent les données sur un nouveau sous-espace pour maxi-

miser la variance, offrant une représentation simplifiée mais informative, et peuvent être utilisées avant un algorithme de regroupement comme dans [135]. L'algorithme t-SNE (t-Distributed Stochastic Neighbor Embedding) se spécialise dans la visualisation de données haute dimension en minimisant la divergence entre les points de données dans différentes dimensions, voir [62]. Isomap (Isometric Mapping, [34]) maintient les relations géométriques dans l'espace haute dimension, préservant les propriétés locales dans la représentation en dimensions inférieures. Locally Linear Embedding (LLE), comme dans [33], se concentre sur la linéarité locale, projetant les points de données de manière à maintenir les relations entre les points voisins. UMAP (Uniform Manifold Approximation and Projection), comme dans le travail de [99], est doué pour préserver à la fois les structures locales et globales dans l'espace des données.

Ces techniques offrent des stratégies diverses pour la réduction de la dimensionnalité et sont applicables dans une myriade de domaines, contribuant de manière significative à la visualisation et à l'analyse des données.

6.2.3/ ALGORITHMES DE REGROUPEMENT

Le regroupement englobe divers algorithmes, chacun avec des méthodologies distinctes et des applications dans l'analyse de données. K-means, comme dans [45], est une technique omniprésente reposant sur des centroïdes, impliquant l'assignation itérative et le remplacement des centroïdes de regroupement. En revanche, DBSCAN (Density-Based Spatial Clustering of Applications with Noise, [23]) identifie des clusters basés sur des régions denses séparées par des zones de densité plus faible. Son évolution : HDBSCAN ([111]) se concentre sur la similarité entre les membres des clusters, en donnant la priorité à l'association de voisinage plutôt qu'à la proximité du centre du cluster. Cela conduit à la création de clusters plus précis mais de tailles inégales, une caractéristique particulièrement avantageuse pour les données distribuées de manière non uniforme. Bien que HDBSCAN excelle dans le traitement des clusters basés sur la similarité, il peut rencontrer des défis avec des nuages de points non compacts et des points de données qui n'adhèrent pas facilement aux formations de cluster traditionnelles.

Plusieurs autres algorithmes de regroupement remarquables méritent attention. OPTICS (Ordering Points To Identify the Clustering Structure), proposé par [29], est une extension de DBSCAN, fournissant une représentation de regroupement hiérarchique. Mean Shift ([50]) utilise une technique non paramétrique pour identifier des clusters en localisant les maxima de densité dans l'espace des données. Cet algorithme a récemment été amélioré avec l'utilisation de GPUs comme dans [137]. Le regroupement hiérarchique agglomératif ([13]) fonctionne en fusionnant successivement des points de données individuels ou des clusters en fonction de leur similarité, créant un dendrogramme pour illustrer le processus

de fusion.

Enfin, BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies, [26]) utilise une approche de regroupement hiérarchique pour gérer de grands ensembles de données, produisant un résumé compact pour l'analyse ultérieure. Le regroupement spectral, comme dans [40, 144, 130], fonctionne sur des graphes de similarité et la théorie spectrale pour partitionner les données en clusters. Enfin, l'Affinity Propagation ([58]) utilise le passage de messages pour trouver les exemples les plus appropriés de points de données pour le regroupement, équilibrant la qualité de l'exemple et le nombre de clusters.

Ces algorithmes démontrent des approches diverses et ont des applications dans divers domaines, adaptés à différents types d'ensembles de données et de structures.

6.3/ DESCRIPTION DE LA MÉTHODE TSCAPE

La solution TSCAPE qui a été développée au cours de cette thèse, est composée de trois étapes principales (figure 6.1) :

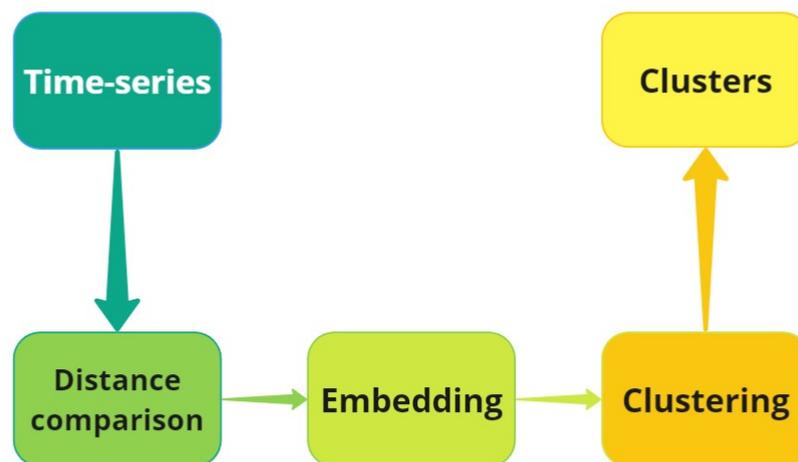


FIGURE 6.1 – Les différentes étapes de la solution TSCAPE

6.3.1/ DÉFORMATION TEMPORELLE DYNAMIQUE (DTW)

Tout d'abord, la Déformation temporelle dynamique ou Dynamic Time Warping (DTW) en anglais, calcule la plus petite distance entre deux séries temporelles. Contrairement à un simple calcul de distance euclidienne, il ne compare pas point par point les deux séries temporelles et offre un certain degré de flexibilité pour aligner les points. Cet algorithme ($DTW(x, y)$) est appliqué sur les données (P) comparées deux par deux pour créer une matrice M de taille $N \times N$, où N correspond au nombre de séries temporelles à regrouper. La matrice M est définie de la manière suivante :

$$M(i, j) = DTW(i, j) \quad (6.1)$$

avec i, j appartenant à P .

L'algorithme $DTW(X, Y)$ est défini comme suit. Pour deux séries temporelles X et Y de tailles respectives $N1$ et $N2$, la première étape consiste à créer une matrice de distance D de taille $N1 \times N2$ avec l'équation suivante :

$$D(k, p) = |X_k - Y_p| + \min\{D(k-1, p), D(k, p-1), D(k-1, p-1)\} \quad (6.2)$$

La prochaine étape consiste à reconstituer le chemin entre $D(N1-1, N2-1)$ et $D(0, 0)$ en utilisant la plus petite valeur de distance. Dans ce but, on ajoute les valeurs des cellules traversées en utilisant la règle de déplacement suivante :

$$PourLaCelluleActuelle = D(k, p), \quad (6.3)$$

$$CelluleSuivante = \min\{D(k-1, p), D(k, p-1), D(k-1, p-1)\}. \quad (6.4)$$

Une fois que la matrice M a été créée et remplie, elle est utilisée comme entrée pour l'algorithme de Positionnement multidimensionnel.

Des tests ont également été réalisés en remplaçant l'algorithme classique de Dynamic Time Warping par le soft-DTW [88], qui est une variante de l'algorithme DTW avec un paramètre additionnel "gamma" (tableau 6.1).

6.3.2/ PROJECTION EUCLIDIENNE : POSITIONNEMENT MULTIDIMENSIONNEL (MDS)

MDS est un algorithme utilisé en science des données pour visualiser les dissimilarités entre les objets d'un ensemble de données en les cartographiant dans un espace de dimensions inférieures tout en préservant autant que possible leurs distances mutuelles

d'origine. Le principal objectif de MDS est de représenter des données haute dimension dans un nombre réduit de dimensions, généralement deux ou trois, pour faciliter la visualisation et l'interprétation.

6.3.2.1/ FORMULATION MATHÉMATIQUE

Considérons un ensemble de données de n objets, où les dissimilarités entre chaque paire d'objets sont représentées par une matrice de dissimilarités $n \times n$ symétrique D . MDS vise à incorporer ces objets dans un espace de dimensions inférieures, généralement p dimensions.

L'objectif de MDS est de trouver une configuration de points $X = \{x_1, x_2, \dots, x_n\}$ dans un espace p -dimensionnel tel que les distances euclidiennes entre ces points approximent au mieux les dissimilarités dans la matrice originale D .

La fonction de contrainte dans MDS classique, également connue sous le nom de contrainte de Kruskal, est conçue pour minimiser la différence entre les dissimilarités d'origine D_{ij} et les dissimilarités reconstruites en utilisant les points incorporés x_i et x_j dans l'espace réduit.

La fonction de contrainte est donnée par :

$$E(X) = \sum_{i=1}^n \sum_{j=i+1}^n w_{ij} (d_{ij} - \|x_i - x_j\|)^2 \quad (6.5)$$

où :

- $E(X)$ est la fonction de contrainte à minimiser,
- w_{ij} est un poids associé à la dissimilarité entre les objets i et j ,
- d_{ij} est la dissimilarité d'origine entre les objets i et j ,
- $\|x_i - x_j\|$ représente la distance euclidienne entre les points x_i et x_j dans l'espace de dimensions inférieures.

Le problème de minimisation est résolu à l'aide de techniques d'optimisation telles que la descente de gradient, la décomposition en valeurs propres ou des méthodes d'optimisation non linéaires.

6.3.2.2/ DÉTAILS OPÉRATIONNELS

1. **Entrée** : L'entrée de MDS est une matrice de dissimilarités représentant les dissimilarités entre objets deux à deux.

2. **Calcul des distances** : Calculez les distances euclidiennes deux à deux à partir de la matrice de dissimilarités.

3. **Optimisation** : Minimisez la fonction de contrainte en trouvant la configuration optimale de points dans l'espace de dimensions inférieures.

4. **Visualisation** : Visualisez les points incorporés dans l'espace 2D ou 3D pour explorer les relations entre les objets.

6.3.3/ LA RÉDUCTION MULTIDIMENSIONNELLE DANS TSCAPE

La réduction multidimensionnelle sert plusieurs objectifs essentiels dans le prétraitement des données de séries temporelles pour l'analyse de regroupement. Principalement, MDS s'efforce de maintenir les relations originales entre les séries temporelles en conservant autant que possible les distances ou dissimilarités mutuelles lors de la réduction de dimensionnalité. Cette préservation de la proximité garantit la conservation d'informations cruciales indispensables pour le regroupement ultérieur. De plus, MDS aide à réduire le bruit en filtrant les détails superflus dans les données de séries temporelles, mettant en évidence les motifs et structures plus importants, améliorant ainsi l'efficacité des algorithmes de regroupement ultérieurs. Par ailleurs, en représentant les données de séries temporelles dans un espace de dimensions réduit, MDS permet une application plus efficace des algorithmes de regroupement, générant des regroupements plus instructifs et précis en reflétant les relations originales entre les points de données. Enfin, MDS contribue à l'efficacité computationnelle en réduisant la complexité des algorithmes de regroupement ultérieurs grâce à la réduction de dimensionnalité, rendant le processus de regroupement plus gérable sur le plan computationnel et facilitant l'application des algorithmes ayant des difficultés avec des dimensions plus élevées.

6.3.3.1/ UTILISATION DANS TSCAPE

La deuxième étape de TSCAPE consiste à appliquer un algorithme de réduction multidimensionnelle sur la matrice de distance générée par le DTW. Cet algorithme permet, en utilisant une matrice de distance, de créer des points dans un espace euclidien de dimension d . Chaque point représente une série temporelle contenue dans la matrice M . Si $d \leq 3$, la visualisation des positions des points est possible graphiquement. Mais cela ne constitue pas une limite restrictive. La dernière étape consiste à effectuer un regroupement sur ces points.

6.3.4/ REGROUPEMENT (CLUSTERING)

Pour l'étape de regroupement, quelques algorithmes ont été appliqués sur les points représentant les séries temporelles.

L'algorithme K-Means est une méthode de regroupement largement utilisée en analyse de données pour partitionner un ensemble de données en K clusters distincts. Il fonctionne en minimisant la somme des distances au carré entre les points de données et leurs centres de cluster attribués, formant ainsi K clusters autour de centres appelés centroïdes.

6.3.4.1/ FORMULES MATHÉMATIQUES

Soit $X = \{x_1, x_2, \dots, x_n\}$ un ensemble de n points de données et K le nombre de clusters à former.

1. **Initialisation des Centroïdes** : Sélectionnez de manière aléatoire K points comme centroïdes initiaux $C = \{c_1, c_2, \dots, c_K\}$.

2. **Assignment des Points aux Clusters** : Pour chaque point x_i , calculez la distance à chaque centroïde c_j et assignez x_i au cluster le plus proche du centroïde. Ceci peut être formulé comme :

$$\text{Cluster}(x_i) = \arg \min_j \|x_i - c_j\|^2 \quad (6.6)$$

3. **Mise à Jour des Centroïdes** : Recalculer les centroïdes de chaque cluster comme la moyenne de tous les points assignés à ce cluster :

$$c_j = \frac{1}{|S_j|} \sum_{x_i \in S_j} x_i \quad (6.7)$$

où S_j est l'ensemble des points assignés au cluster j .

4. Répétez les étapes 2 et 3 jusqu'à convergence, c'est-à-dire jusqu'à ce qu'il n'y ait aucun changement dans l'allocation des points aux clusters ou jusqu'à ce qu'une tolérance prédéfinie soit atteinte.

6.3.4.2/ UTILISATION DANS TSCAPE

Le regroupement n'est donc pas appliqué sur les formes de courbe mais sur une représentation sur un plan euclidien de leurs dissimilarités. Appliquer K-means sur les données transformées par MDS aide à former des clusters distincts en exploitant les représentations réduites et structurées fournies par MDS. La combinaison permet à K-means de fonctionner sur les dimensions réduites tout en cherchant à former des clusters en fonction des relations spatiales dérivées parmi les points de données de séries temporelles. Cette intégration permet à K-means de capturer des motifs et des similarités si-

gnificatifs dans l'espace réduit créé par MDS, conduisant potentiellement à des résultats de regroupement plus instructifs et précis pour les données de séries temporelles.

Le principal avantage de l'algorithme K-means par rapport à d'autres méthodes de regroupement, en particulier lorsqu'il s'agit de former des clusters très concentrés, réside dans sa capacité à minimiser la variance intra-cluster. K-means cherche à minimiser la somme des distances au carré des points à leurs centroides respectifs, regroupant efficacement les points autour de centres (centroides) pour minimiser la dispersion dans chaque cluster. En comparaison avec d'autres algorithmes de regroupement, K-means est efficace pour former des clusters compacts car il vise explicitement à réduire la variance intra-cluster en utilisant une approche itérative de réaffectation des points aux clusters en fonction de leur proximité avec les centroides, favorisant ainsi la concentration des données au sein de chaque cluster. Ceci est particulièrement pertinent pour notre travail car nous visons à créer des clusters avec les distances intra-cluster les plus petites, correspondant à des groupes de produits avec les courbes les plus similaires.

6.4/ EXPÉRIMENTATIONS

6.4.1/ DONNÉES UTILISÉES

Dans cette étude, la méthode sera appliquée à deux ensembles de données de la CERP :

- Le premier ensemble de données correspond aux ventes des 800 produits de parapharmacie les plus vendus dans le catalogue CERP entre 2017 et fin 2022.
- Le deuxième ensemble de données correspond aux ventes de tous les produits référencés dans le catalogue COALIA (catalogue des produits vendus par la centrale d'achat COALIA) entre 2017 et fin 2022.

Cette méthode à également fait l'objet d'une publication scientifique. Pour cette publication, nous avons utilisé les deux jeux de données "open sources" suivant pour simuler des séries temporelles de ventes :

- Le premier est un ensemble de données ouvertes [97] correspondant aux données des échanges d'entreprises, entre 2006 et 2018. Pour cet article, la colonne "Volume" (correspondant au nombre d'actions échangées) a été utilisée.
- Le deuxième ensemble de données correspond aux données sur les prix de vente de maisons selon leur quartier et leur nombre de chambres [103]. Les données ont été regroupées par numéro de quartier, type de propriété et nombre de chambres : par exemple, "2603_5.house" contiendra toutes les maisons de 5 chambres dans le quartier 2603.

Les résultats pour ces deux jeux de données sont disponibles sur le papier présentant la

méthode TSCAPE [146].

Chaque ensemble de données est transformé pour obtenir trois colonnes : la date, le nom du produit et un nombre de vente correspondant au volume livré par la CERP pour ce produit. Les données sont agrégées par mois et par produit. Cela nous donne, pour chaque produit, une série temporelle de ventes mensuelles. Les valeurs manquantes sur la période sont remplacées par zéro.

Avant d'être utilisées dans les algorithmes de DTW, les données sont également normalisées entre 0 et 1 afin d'éviter que le volume des ventes n'influence l'échelle de distance de la matrice de dissimilarité. Cette normalisation permet d'établir une comparaison équitable entre les séries temporelles, indépendamment de leurs volumes de ventes respectifs..

6.4.2/ MÉTRIQUES

Pour comparer les différentes solutions, trois métriques ont été utilisées :

- Le score de silhouette (SilScore) vérifie la densité des clusters et leur séparation. L'intervalle de solution pour le score de silhouette se situe entre 1 (meilleure solution) et -1 (pire solution).
- Le score de Calinski-Harabasz (CHScore) est quelque peu similaire : c'est un rapport entre la somme des variances inter-clusters et la somme des variances intra-clusters. Il varie de 0 (pire) à l'infini (meilleur), et sa plage est liée à l'ensemble de données.
- L'indice de Davies-Bouldin (DbScore) mesure la similarité avec le cluster le plus proche. Il s'agit de la moyenne du rapport maximal entre la distance d'un point au centre de son cluster et la distance entre deux centres de clusters. Il varie entre 0 (meilleur) et l'infini (pire).

6.4.3/ REGROUPEMENT PAR K-MEANS AVEC UNE MÉTRIQUE DTW (TSLEARN)

Dans une première étape, les données ont été regroupées à l'aide de la fonction `TimeSeriesKmeans` de la bibliothèque Python `TSlearn` [114]. Cette bibliothèque est spécialisée dans le traitement des séries temporelles. La fonction `TimeSeriesKmeans` dispose d'une option dans son paramètre `metric` qui utilise directement le calcul de DTW. Le nombre de clusters a été choisi par une technique du coude (figures 6.2 et 6.3). Le nombre de clusters a été fixé à 10 pour les deux jeux de données (Parapharmacie et COALIA). Les paramètres de K-means sont les mêmes que ceux décrits dans la section "4.4.3 Configuration de K-means" de cet article, à l'exception du paramètre spécifique à cette solution : `metric`, qui a été initialisé à `dtw`.

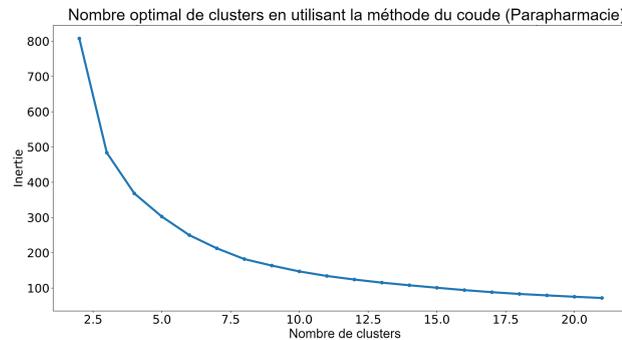


FIGURE 6.2 – Inertie des clusters par rapport à leur nombre pour le jeu de données "Parapharmacie"

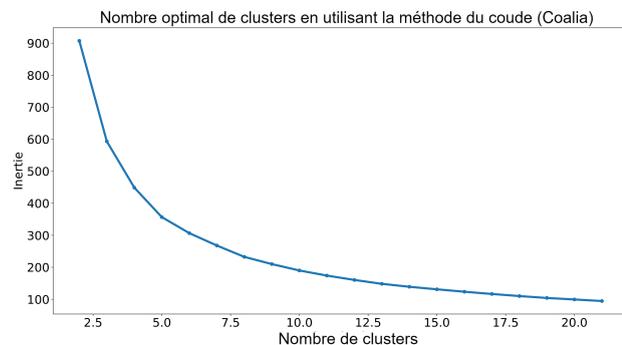


FIGURE 6.3 – Inertie des clusters par rapport à leur nombre pour le jeu de données "COALIA"

Les résultats montrent que cette première solution n'est pas très efficace pour nos données. Le score de silhouette est faible, ce qui indique une faible cohérence des clusters (faible séparation des clusters ou distances trop grandes entre les individus d'un cluster). En effet, les courbes de ventes sont assez différentes les unes des autres, et donc la détection des motifs ne semble pas fonctionner très bien (tableaux 6.4 et 6.5).

6.4.4/ ANALYSE EN COMPOSANTES PRINCIPALES (PCA) POUR LE REGROUPEMENT DE SÉRIES TEMPORELLES

L'Analyse en Composantes Principales (PCA) est une technique de réduction de dimension largement utilisée dans le regroupement de séries temporelles. Dans ce contexte, la PCA améliore l'efficacité des algorithmes de regroupement ultérieurs, tels que K-means. Le processus implique de centrer les données de séries temporelles et d'extraire les composantes principales qui capturent la variance maximale.

Après le centrage des moyennes, la matrice de covariance est calculée et la PCA est appliquée pour obtenir les composantes principales ordonnées par variance. La décision

sur le nombre de composantes à conserver est cruciale, souvent basée sur un seuil de variance prédéfini. Ensuite, les séries temporelles sont reconstruites en utilisant les composantes principales sélectionnées, ce qui donne une représentation de dimension réduite.

La représentation réduite sert d'entrée pour le regroupement K-means, où les séries temporelles sont regroupées en clusters en fonction de leur similarité dans l'espace de caractéristiques réduit. La PCA offre des avantages tels que la réduction du bruit, l'efficacité computationnelle et l'amélioration des performances de regroupement.

L'utilisation de la PCA en conjonction avec un algorithme de regroupement tel que K-means semble être l'une des méthodes les plus efficaces pour le regroupement de séries temporelles, comme le montrent les travaux de [49, 105, 53]. Ces approches diffèrent de notre travail car elles ne reposent pas sur la comparaison des distances entre les séries temporelles. Malgré l'obtention de résultats supérieurs (tableau 6.4 et tableau 6.5), cela se fait au détriment de l'interprétabilité de la réduction de dimensionnalité : dans notre méthode, deux points sont nécessairement proches car ils ont une courte distance calculée par le DTW, c'est-à-dire que leurs courbes sont similaires, contrairement à la méthode PCA, qui réduit les dimensions en créant ses propres axes, rendant l'analyse de projection plus complexe.

Nous avons également réalisé des tests en intégrant la PCA dans l'étape d'incorporation de notre méthodologie, afin d'évaluer si elle améliore les résultats (tableau 6.4 et tableau 6.5).

6.4.5/ EXPÉRIMENTATION DE TSCAPE

Pour comparer les deux solutions, les paramètres utilisés sont conservés, ainsi que le nombre de clusters et les paramètres de K-means (à l'exception de "*metric = dtw*").

1. Configuration de la Déformation temporelle dynamique

Le paramètre "gamma" du soft-DTW a été testé avec différentes valeurs. Les scores des différentes valeurs ont été comparés les uns aux autres, ainsi qu'avec les résultats obtenus avec une Déformation temporelle dynamique classique (tableau 6.1).

2. Configuration du Positionnement multidimensionnel

Pour l'étape de Positionnement multidimensionnel, la variation du paramètre de dimension (*n_components*) a été testée. Un exemple pour les deux jeux de données, dans un espace de dimension $d = 2$, est présenté dans les figures 6.4 et 6.5.

Augmenter le nombre de dimensions permet d'offrir davantage de degrés de liberté à l'algorithme de Positionnement multidimensionnel. Cependant, en examinant le

	Moyenne de 10 exécutions avec les données "Para"			Moyenne de 10 exécutions avec les données "COALIA"		
SoftDTW gamma	SilScore Moyenne	CHScore Moyenne	DbScore Moyenne	SilScore Moyenne	CHScore Moyenne	DbScore Moyenne
DTW	0.37	880.83	0.83	0.34	1076.94	0.86
10	0.36	933.8	0.76	0.39	900.34	0.78
0.1	0.35	624.77	0.82	0.34	842.21	0.88
0.01	0.37	1017.08	0.8	0.35	1056.08	0.81

TABLE 6.1 – Variation des métriques en fonction des paramètres gamma du SoftDTW.

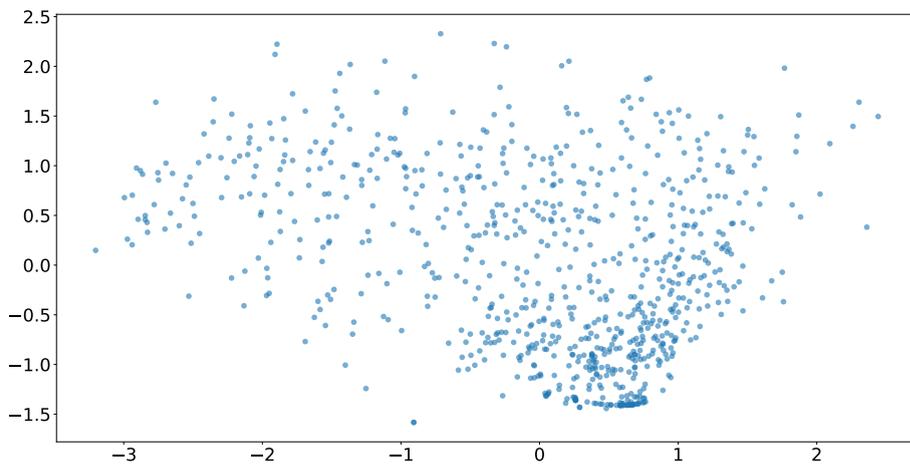


FIGURE 6.4 – Projection 2D grâce au MDS pour le jeu de données "Parapharmacie". Chaque point correspond à la représentation d'une série temporelle. Deux points proches l'un de l'autre signifient que les deux séries temporelles ont une courbe similaire.

tableau 6.2, on peut constater que l'augmentation de la dimension $d \leq 4$ n'améliore pas les différents scores.

L'algorithme MDS (Multidimensional Scaling) dispose d'un paramètre appelé *random_state* qui permet de l'initialiser avec différentes valeurs de graine. La variation de ce paramètre ne semble pas avoir un impact significatif sur le résultat (tableau 6.3). L'observation de ces résultats démontre que le changement de la graine d'initialisation du MDS engendre seulement une légère variation du résultat de regroupement pour les ensembles de données comportant de nombreux échantillons. Toutefois, dans le cas d'ensembles de données comportant seulement quelques échantillons, ce paramètre peut s'avérer utile pour améliorer le score de silhouette.

Des tests ont été effectués avec d'autres algorithmes de réduction de dimensions.

L'algorithme t-SNE [62] est l'une des meilleures méthodes de cette catégorie. Il a

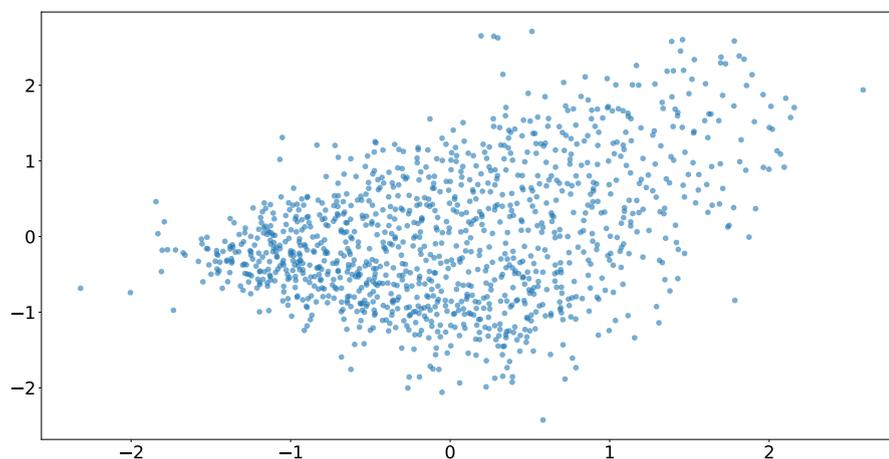


FIGURE 6.5 – Projection 2D grâce au MDS pour le jeu de données "COALIA". Chaque point correspond à la représentation d'une série temporelle. Deux points proches l'un de l'autre signifient que les deux séries temporelles ont une courbe similaire.

	Moyenne de 10 exécutions avec les données "Parapharmacie"			Moyenne de 10 exécutions avec les données "COALIA"		
NDim	SilScore Moyenne	CHScore Moyenne	DbScore Moyenne	SilScore Moyenne	CHScore Moyenne	DbScore Moyenne
2	0.37	1017.08	0.8	0.34	1021.72	0.82
3	0.36	971.49	0.85	0.33	939.88	0.85
4	0.35	965.94	0.85	0.32	923.35	0.86
5	0.35	965.84	0.85	0.32	923.24	0.87
6	0.35	966.27	0.85	0.32	920.89	0.87
7	0.35	966.13	0.85	0.32	922.77	0.86
8	0.35	966.43	0.85	0.32	922.46	0.86
9	0.35	966.44	0.85	0.32	922.78	0.86
10	0.35	966.24	0.85	0.32	922.89	0.86

TABLE 6.2 – Variation des métriques par rapport au nombre de dimensions du MDS.

été utilisé juste après le MDS pour vérifier s'il était capable de placer les points correspondant aux produits de manière plus efficace que le MDS. Les résultats (tableau 6.4 et tableau 6.5) montrent une légère amélioration du score de silhouette pour les ensembles de données comportant de nombreux échantillons. Il semble que lorsque le nombre d'échantillons et la taille de la matrice M sont plus importants, le MDS a moins de degrés de liberté pour placer les points entre eux, ce qui conduit à des résultats plus convergents quelle que soit la graine utilisée. Il peut alors être assisté par le t-SNE pour améliorer ces résultats (figures 6.6 et 6.7).

	Moyenne de 10 exécutions avec les données "Parapharmacie"			Moyenne de 10 exécutions avec les données "COALIA"		
Numéro de graine	SilScore Moyenne	CHScore Moyenne	DbScore Moyenne	SilScore Moyenne	CHScore Moyenne	DbScore Moyenne
1	0.37	1019.73	0.66	0.34	1018.83	0.82
2	0.38	1019.03	0.8	0.34	1023.42	0.81
3	0.37	1017.46	0.81	0.34	1018.7	0.82
4	0.37	1020.88	0.82	0.34	1018.31	0.82
5	0.37	1024.15	0.82	0.34	1015.03	0.81
6	0.37	1013.24	0.82	0.34	1029.9	0.81
7	0.37	1020.42	0.83	0.34	1021.5	0.81
8	0.37	1012.41	0.8	0.34	1018.71	0.83
9	0.37	1017.1	0.8	0.34	1021.72	0.82

TABLE 6.3 – Variation des métriques par rapport au numéro de graine du MDS.

Des tests ont également été réalisés en remplaçant l'algorithme MDS par un algorithme de spectral embedding (SE) [44]. On obtiens des résultat assez similaires notamment lorsque qu'on ajoute l'algorithme t-SNE à la suite (tableau 6.4 et tableau 6.5).

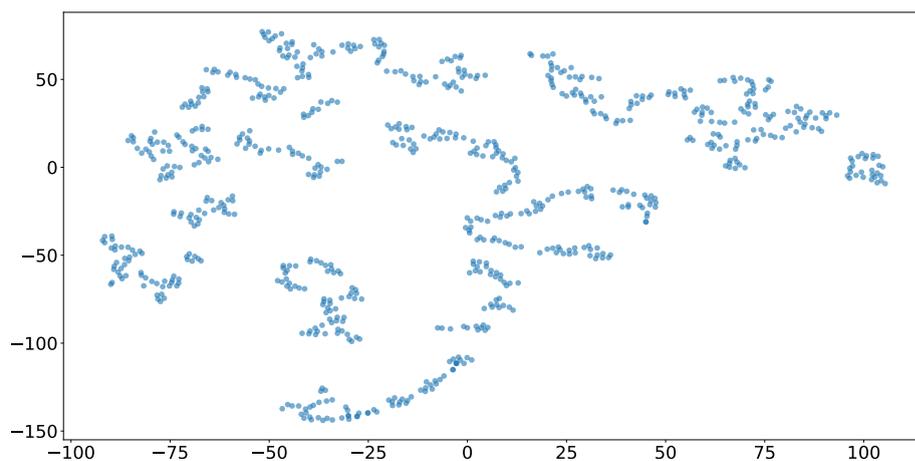


FIGURE 6.6 – Projection 2D grâce au MDS suivie d'un remplacement de points à l'aide de l'algorithme t-SNE pour l'ensemble de données "Parapharmacie". Chaque point correspond à la représentation d'une série temporelle. Deux points proches l'un de l'autre signifient que les deux séries temporelles ont une courbe similaire.

3. Configuration du K-means

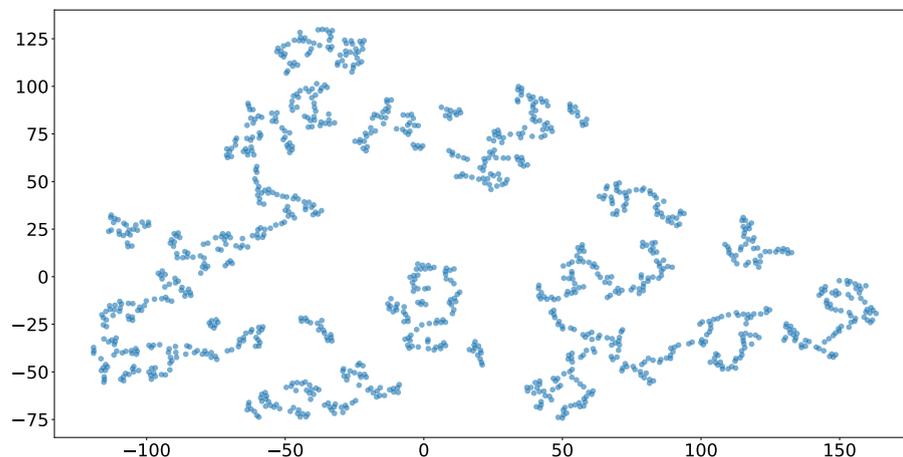


FIGURE 6.7 – Projection 2D grâce au MDS suivie d'un remplacement de points à l'aide de l'algorithme t-SNE pour l'ensemble de données "COALIA". Chaque point correspond à la représentation d'une série temporelle. Deux points proches l'un de l'autre signifient que les deux séries temporelles ont une courbe similaire

Les paramètres du K-means sont les mêmes que pour la partie tslearn : le paramètre *max_iter* a été fixé à 400, *n_init* à 100, *init* à *K-means++*, et enfin *random_state* a été initialisé et varié de 0 à 20. Le nombre de clusters est également le même que celui défini dans la partie tslearn de cet article : 10 pour les ensembles de données de Parapharmacie et 10 pour l'ensemble de données COALIA. Tous les points sont regroupés dans un cluster, mais parfois des points proches les uns des autres se retrouvent dans 2 clusters différents. Cela est dû à la méthode de regroupement K-means, c'est la seule conséquence problématique de ce choix. Ce problème peut être corrigé en augmentant suffisamment le nombre de clusters afin de mieux répartir les points dans chaque groupe, au détriment de la taille du nombre d'individus par cluster et de la facilité d'analyse (figures 6.8 et 6.9).

6.4.6/ RÉSULTATS

Dans notre méthodologie, les deux premières étapes impliquant le dynamic time warping et le multidimensional scaling représentent la similarité entre deux séries temporelles en établissant une distance entre deux points. Notre objectif est de regrouper des séries temporelles ayant des courbes les plus similaires, ce qui nécessite le regroupement des points les plus proches. Chaque individu au sein d'un cluster devrait être aussi proche que possible des autres membres du même cluster et idéalement loin des membres des autres clusters. Cela implique la création de clusters autour de leurs centroïdes, une tâche très bien réalisée par l'algorithme K-means. Des métriques

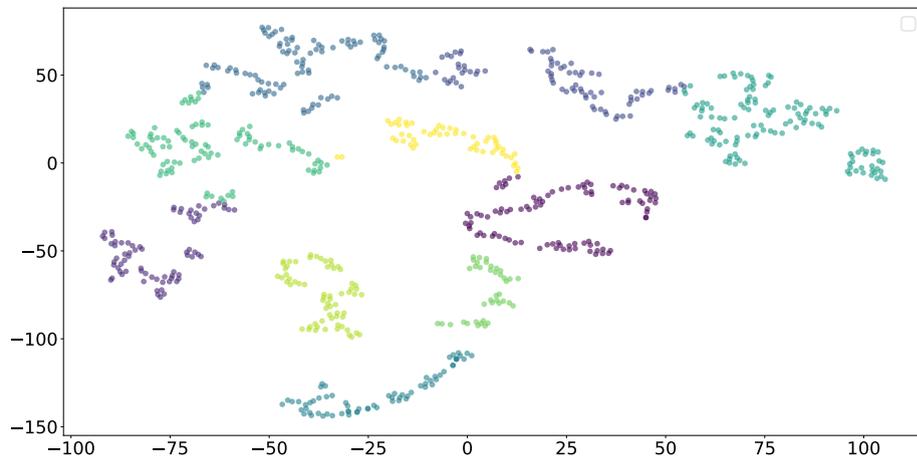


FIGURE 6.8 – Clusters pour le jeu de données "Parapharmacie". Les clusters sont représentés par des couleurs différentes : deux points de même couleur sont dans le même cluster.

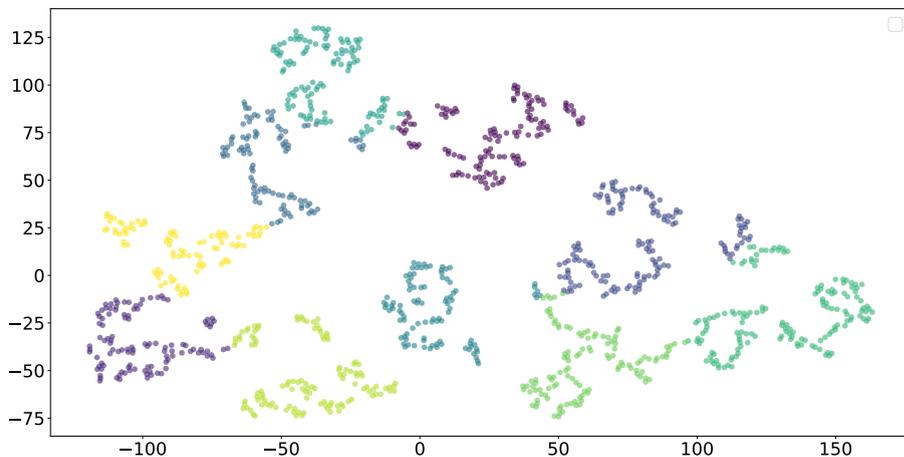


FIGURE 6.9 – Clusters pour le jeu de données "COALIA". Les clusters sont représentés par des couleurs différentes : deux points de même couleur sont dans le même cluster.

telles que le score silhouette, le score Calinski-Harabasz et l'indice Davies-Bouldin sont pertinents pour évaluer le regroupement de données lorsqu'on cherche à regrouper les points les plus proches. On rappelle que le score silhouette mesure la similarité des points au sein des clusters et leur dissimilarité par rapport aux autres clusters. Le score Calinski-Harabasz évalue la compacité et la séparation des clusters, tandis que l'indice Davies-Bouldin quantifie à la fois la dispersion intra-cluster et la séparation inter-cluster. Ces métriques offrent une évaluation complète des clusters en termes de cohésion intra-cluster, de séparation inter-cluster et de différenciation entre les clusters.

Le score silhouette des clusters prouve l'amélioration que notre solution TSCAPE a apportée (tableau 6.4 & tableau 6.5). L'ajout du multidimensional scaling améliore le score silhouette par rapport à une solution composée uniquement du dynamic time warping et de K-means (avec des paramètres égaux). Ajouter un algorithme t-SNE entre MDS et le regroupement améliore le score silhouette dans certains cas. Remplacer l'algorithme DTW par son Soft-DTW permet d'améliorer les résultats avec certaines valeurs de paramètres "gamma".

6.5/ CONCLUSION

À travers ce chapitre, nous avons présenté une technique pour regrouper des séries temporelles. Ce regroupement est basé sur l'analyse et la comparaison des formes de courbe. Les différentes techniques pourraient être testées en comparant les scores silhouette, le score Calinski-Harabasz et l'indice Davies-Bouldin de tous les clusters. Les résultats ont montré que la solution proposée dans cet article est plus efficace pour créer des clusters que les méthodes existantes basées sur le dynamic time warping et K-Means. Bien que notre méthode soit moins efficace que les méthodes basées sur PCA au niveau des métriques, elle offre une meilleure interprétation en termes de réduction de dimension. La prochaine étape de ce travail consiste à utiliser ces clusters pour faire des prédictions en regroupant des produits dans le contexte de données d'un grossiste-détaillant pour les pharmacies françaises. Des travaux d'amélioration du regroupement peuvent également être réalisés, par exemple en testant d'autres algorithmes de regroupement plus puissants que K-means, ou en couplant cette méthode avec des approches de regroupements basées sur des connaissances métiers.

Dataset	Sil Score mean	CH Score mean	Db Score mean	Sil Score STD	CH Score STD	Db Score STD
Parapharmacie						
DTW + MDS + K-means	0.37	880.83	0.83	0	0	0
DTW + MDS + t-SNE + K-means	0.49	1332.75	0.66	0	0	0
DTW + K-means	0.29	573.2	1.04	0	0	0
DTW + t-SNE + K-means	0.49	1688.34	0.66	0	0	0
DTW + Spectral Embedding + K-means	0.36	651.24	0.84	0	0	0
DTW + SpectralEmbedding + t-SNE + K-means	0.51	1589.85	0.66	0	39.11	0.1
SoftDTW + Spectral Embedding + K-means	0.36	724.49	0.85	0	0	0
SoftDTW + SpectralEmbedding + t-SNE + K-means	0.49	1287.42	0.75	0	20.38	0
SoftDTW + K-means	0.31	609.29	1	0	0	0
SoftDTW + t-SNE + K-means	0.48	1537.75	0.66	0	0	0
SoftDTW + MDS + K-means	0.37	1017.08	0.8	0	0	0
SoftDTW + MDS + t-SNE + K-means	0.51	1943.72	0.64	0	0	0
SoftDTW + PCA + K-means	0.44	1784.61	0.71	0	0	0
SoftDTW + PCA + t-SNE + K-means	0.54	1891.24	0.59	0	0	0
SoftDTW + MDS + K-Medoids	0.37	985.43	0.84	0	0	0
SoftDTW + MDS + t-SNE + K-Medoids	0.46	1624.37	0.71	0	0	0
SoftDTW + MDS + SpectralClustering	0.28	642.80	0.80	0	0	0
SoftDTW + MDS + t-SNE + SpectralClustering	-0.15	72.60	2.91	0	0	0
SoftDTW + SpectralClustering	0.30	692.32	0.81	0	0	0
SoftDTW + K-Medoids	0.30	757.06	0.90	0	0	0
tslearn	0.053	13.65	5.49	0	0.34	0.08

TABLE 6.4 – Analyse comparative (moyenne de 10 exécutions) de toutes les solutions et options du présent document pour l'ensemble de données Parapharmacie (la meilleure solution est en gras).

Dataset COALIA	Sil Score mean	CH Score mean	Db Score mean	Sil Score STD	CH Score STD	Db Score STD
DTW + MDS + K-means	0.34	1076.94	0.86	0	0	0
DTW + MDS + t-SNE + K-means	0.49	2090.51	0.72	0	0	0
DTW + K-means	0.23	709.5	0.1.16	0	0	0
DTW + t-SNE + K-means	0.43	2075.94	0.77	0	0	0
DTW + Spectral Embedding + K-means	0.35	815.51	0.83	0	0	0
DTW + SpectralEmbedding + t-SNE + K-means	0.48	1852.56	0.68	0	14.11	0
SoftDTW + Spectral Embedding + K-means	0.44	1857.87	0.69	0	0	0
SoftDTW + Spectral Embedding + t-SNE + K-means	0.51	2359.93	0.61	0	21.07	0
SoftDTW + K-means	0.27	628.72	1.06	0	0	0
SoftDTW + t-SNE + K-means	0.46	1519.71	0.72	0	0	0
SoftDTW + MDS + K-means	0.35	1056.08	0.81	0	0	0
SoftDTW + MDS + t-SNE + K-means	0.5	1696.74	0.67	0	0	0
SoftDTW + PCA + K-means	0.40	1980.78	0.70	0	0	0
SoftDTW + PCA + t-SNE + K-means	0.52	1994.34	0.63	0	0	0
SoftDTW + MDS + K-Medoids	0.33	941.54	0.86	0	0	0
SoftDTW + MDS + t-SNE + K-Medoids	0.45	1581.05	0.71	0	0	0
SoftDTW + MDS + SpectralClustering	0.31	818.65	0.84	0	0	0
SoftDTW + MDS + t-SNE + SpectralClustering	0.47	1676.94	0.68	0	0	0
SoftDTW + SpectralClustering	0.27	789.48	0.92	0	0	0
SoftDTW + K-Medoids	0.30	816.31	0.94	0	0	0
tslearn	0.27	21.95	1.07	0.01	0.64	0.06

TABLE 6.5 – Analyse comparative (moyenne de 10 exécutions) de toutes les solutions et options du présent document pour l'ensemble de données COALIA (la meilleure solution est en gras).

COMPARAISON DE DIFFÉRENTS LLM'S POUR L'ASSISTANCE AUX PHARMACIENS

7.1/ INTRODUCTION

Les récentes avancées dans le domaine des grands modèles de langage (LLM) ont permis de faire émerger des solutions d'agents conversationnels et autres outils utilisant le langage naturel, jusqu'alors impossible. Dans la vie de tous les jours, les pharmaciens sont souvent obligés d'aller rechercher de nombreuses informations sur des sujets assez variés (réglementaires, informations sur les produits, procédures sur le fonctionnement de leur LGO, etc.). Cette activité peut s'avérer chronophage et empêcher le pharmacien d'être disponible pour son cœur de métier. Est donc apparue l'idée d'utiliser ces nouvelles technologies qu'offrent les LLM pour créer un assistant capable de comprendre une question, rechercher des informations et synthétiser une réponse dans un langage agréable pour l'utilisateur. La pharmacie étant un secteur sensible, les technologies utilisées pour ce genre d'outils doivent respecter certains critères selon l'utilisation. Dans ce chapitre, nous allons donc nous intéresser à l'étude réalisée pour comparer les différents modèles et techniques possibles pour créer ce type de solution, et voir si cela est applicable dans le cadre d'un agent conversationnel pouvant répondre aux questions couramment posées au service client du LGO LEO.

7.2/ APPORT DE CONNAISSANCES EXPERTES AU LLM

Pour apporter de la connaissance à un modèle de langage, on peut créer un modèle expert en réentraînant un modèle généraliste sur des données d'un domaine spécifique. Cette technique nécessite d'importantes ressources de calcul ainsi qu'une quantité im-

portante de données. Une autre solution est d'utiliser un modèle généraliste et de lui fournir en entrée des informations pertinentes, puis de le laisser uniquement synthétiser les informations dans une réponse agréable pour l'utilisateur. C'est sur cette deuxième technique que nos travaux se sont portés.

7.2.1/ CRÉATION DE LA BASE DE CONNAISSANCES

Dans un premier temps, il faut créer la base de connaissances qui sera utilisée par le modèle de langage. Pour ce faire, la documentation PDF du logiciel LEO a été récupérée. Pour chaque document, le processus suivant est appliqué (cf. figure 7.1) :

1. Comme le nombre de tokens en entrée des modèles est limité par un maximum, et que tous les documents ont une taille variable, la première étape est de découper les documents en plusieurs parties (chunks).
2. Pour chaque partie du document, la date du document, la source et le texte de la partie sont stockés en base de données.
3. Aux trois entités précédemment nommées sont ajoutés un vecteur obtenu grâce au passage du texte dans un algorithme qui va créer un vecteur représentant l'idée de cette partie du document.

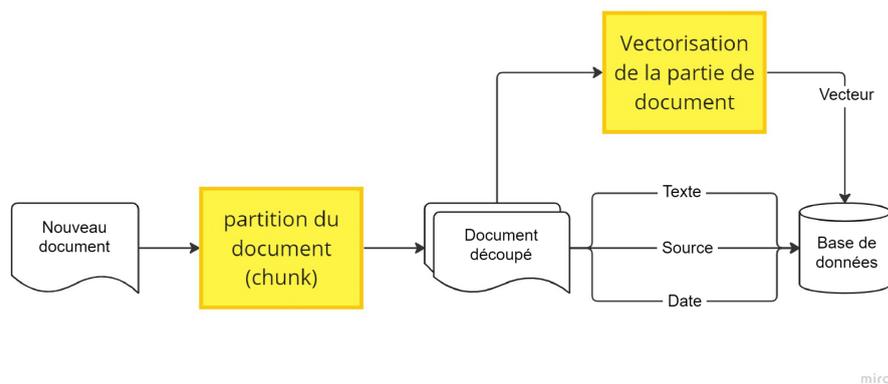


FIGURE 7.1 – Schéma expliquant le processus d'ajout d'un document dans la base vectorielle de connaissances

7.2.2/ RECHERCHE DE CONNAISSANCES ET RÉPONSE AUX QUESTIONS

Lorsqu'un utilisateur pose une question à l'agent conversationnel, le texte de la question est passé dans l'algorithme de vectorisation, puis le vecteur de la question est comparé à tous les autres vecteurs de la base. Les vecteurs les plus similaires sont ensuite

sélectionnés. Si un vecteur de la base est proche de celui de la question, cela signifie que le texte qu'il représente a des points sémantiques similaires à la question qui a été posée. Ce qui induit que le texte doit apporter des éléments de réponse à la question, ou du moins parler du sujet. L'étape suivante sera de pouvoir sélectionner parmi ces vecteurs proches ceux qui viennent d'une source fiable et de préférence les plus récents. Ce filtrage pourra se faire grâce à un système de coefficients (les données provenant d'une source fiable et les plus récentes seront priorisées). Une fois les parties de textes restantes filtrées, elles sont envoyées en tant que contexte, en entrée du modèle de langage en tant que base de connaissances. Un prompt système demande de répondre à la question à partir des informations qui lui sont fournies (figure 7.2).

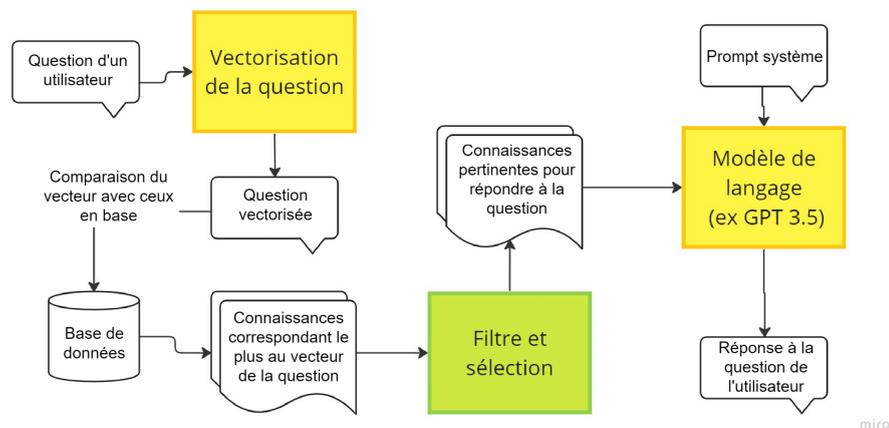


FIGURE 7.2 – Schéma expliquant toutes les étapes qui permettent à l'agent conversationnel de répondre à la question de l'utilisateur. On peut voir en vert la brique de filtre et sélection qui n'est pas encore implémentée.

7.2.2.1/ COMPARAISON SÉMANTIQUE DES VECTEURS

Pour comparer la question et les parties de textes vectorisées en base, on utilise la technique de similarité des cosinus. Cette technique mesure l'angle entre deux vecteurs dans un espace vectoriel.

La similarité des cosinus entre deux vecteurs de documents, souvent désignés comme **A** et **B**, est calculée en mesurant le cosinus de l'angle formé entre ces vecteurs dans l'espace vectoriel. Cela peut être exprimé par la formule suivante :

$$\text{Similarité}(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} \quad (7.1)$$

où $\mathbf{A} \cdot \mathbf{B}$ représente le produit scalaire entre les vecteurs **A** et **B**, et $\|\mathbf{A}\|$ et $\|\mathbf{B}\|$ représentent les normes euclidiennes des vecteurs respectifs.

La similarité des cosinus produit un score compris entre -1 et 1, où 1 indique une similarité parfaite entre les documents, 0 indique une absence de similarité et -1 indique une dissimilarité parfaite. Plus le score de similarité des cosinus est proche de 1, plus les documents sont considérés comme similaires sur le plan sémantique.

7.3/ TYPE DE MODÈLES LLM

Il existe différents modèles de LLM, que nous allons classer dans 2 grandes catégories.

7.3.1/ MODÈLES PROPRIÉTAIRES

Ce sont les modèles qui appartiennent à une entreprise et qui sont accessibles via une interface (exemple de ChatGPT pour les gammes de modèles GPT 3.5 et GPT4) ou via un appel à une API (exemple de l'API OpenAI pour ces mêmes modèles). Ils sont pour la plupart très performants et ont les meilleurs résultats. Le problème principal de ces modèles, c'est que pour les utiliser, il faut envoyer les informations au serveur de l'entreprise qui les possède. Lorsque l'on manipule des données sensibles, cela peut être problématique car il y a un risque d'interceptions ou d'utilisation de ces données. Une autre problématique est que notre service devient dépendant de l'entreprise qui possède le modèle utilisé : si l'entreprise ferme sa connexion API ou dégrade les résultats du modèle, les utilisateurs de ce service n'ont pas de moyens de contrôle sur ces points. Un dernier point, l'utilisation de ces modèles est souvent payante.

7.3.2/ MODÈLES OPEN SOURCE

Ces modèles ont leurs architectures et leurs poids d'entraînement en libre accès. Ce qui permet de les récupérer et de les faire tourner sur nos propres ressources locales, pour les utiliser directement ou les ré-entraîner sur nos propres données. Cela nécessite cependant des ressources plus ou moins conséquentes en fonction du nombre de paramètres du modèle. La plupart des modèles ne peuvent pas tourner sur un ordinateur personnel classique et demandent des serveurs de calcul dédiés. Les performances de ces modèles sont pour le moment moins efficaces que celles des modèles propriétaires, mais certains d'entre eux ont des résultats assez satisfaisants pour être une alternative viable aux modèles propriétaires.

Le choix du type de modèle va donc être un arbitrage entre le coût (abonnement/investissement en serveur), la confidentialité des données manipulées et les performances du modèle.

7.4/ AGENT CONVERSATIONNEL POUR LE SERVICE CLIENT D'UN LGO

Le service client du logiciel de gestion d'officine LEO a deux missions principales :

1. Résoudre des incidents techniques rencontrés par les clients pharmaciens.
2. Répondre à des demandes d'informations sur le fonctionnement du logiciel.

La valeur ajoutée du service client se trouve principalement sur la première de ces missions car les informations sur le logiciel sont disponibles dans une base documentaire. Pour qu'il puisse se concentrer sur sa mission avec le plus de valeur ajoutée, un projet d'agent conversationnel pour assister le service clients dans l'apport d'informations aux pharmaciens a été mis en place. Cet agent aurait accès à la documentation du LGO LEO, qui aura été préalablement transformée en une base de connaissances vectorielles, avec la méthode décrite précédemment dans ce chapitre. Dans un premier temps, cette solution permettra de remonter des informations, voire des réponses pour faire gagner du temps aux techniciens du service client, tout en gardant un contrôle sur la réponse fournie au pharmacien. Quand l'outil aura été assez éprouvé et que la confiance dans ces réponses sera assurée, on peut imaginer une évolution vers son utilisation directe par les pharmaciens.

7.4.1/ DONNÉES DE TEST

Pour tester l'efficacité de cette méthode, le service client a fourni les 10 questions les plus posées par les pharmaciens.

1. Comment facturer les actes de vaccination ?
2. Comment facturer une contraception d'urgence et des contraceptifs gratuits ?
3. Comment créer un nouveau produit avec LPP¹ ?
4. Comment recycler des factures ?
5. Comment retransmettre lot FS² ou PJ³ en anomalie (démarquer une facture, reconstituer un lot) ?
6. Comment retrouver des ventes abandonnées ou supprimées ?
7. Comment retrouver un patient par numéro de lot médicament (en cas de rappel fabricant) ?

1. Liste des Produits et Prestations remboursables. Cela correspond à tout les remboursables hors médicaments.

2. La mise en lot est une procédure pour envoyer les facturations à l'assurance maladie. Le FS (pour feuille de soins numérique) correspond à un lot avec une ordonnance numérique.

3. Le lot PJ (pour pièce jointe) correspond à un lot avec une ordonnance scannée

8. Comment faire une procédure de fin d'exercice ?
9. Comment éditer un listing de ventes ?
10. Comment créer, facturer, clôturer et recycler des contrats de location ?

Ces 10 questions vont être soumises aux différents modèles. Ces modèles auront accès à une base de connaissances du fonctionnement du logiciel LEO. Cette base est composée des informations contenues dans **46 fichiers PDF différents**. Ces PDF sont composés de texte mais également d'images, ces dernières ne sont donc pas considérées par notre solution.

7.4.2/ MODÈLES UTILISÉS

Le modèle pour vectoriser la base de connaissances ainsi que les questions est un modèle d'OpenAI : **Ada 2**. Comme les données utilisées pour cette expérimentation ne sont pas sensibles, ce n'est pas problématique. Dans le cadre de la solution pour des données sensibles, il faudrait prévoir un modèle de vectorisation open source en local.

Au niveau des LLM pour l'agent conversationnel, deux modèles propriétaires de chez OpenAI ont été testés : **GPT 3.5 turbo 16k** et **GPT4**. Ces deux modèles, comme pour le modèle de vectorisation, sont utilisés via l'API d'OpenAI.

Un modèle open source a également été testé, pour voir si ces petits modèles qui peuvent être lancés avec des ressources assez limitées peuvent être suffisamment efficaces pour être viables pour notre solution. Ce modèle est la version 7 milliards de paramètres de Mistral : **Mistral 7b**, par la société Mistral AI.

Pour comparer ces modèles, nous avons également un "**modèle de référence**" qui correspond aux réponses faites par un technicien expert sur le logiciel LEO.

7.4.3/ MÉTHODE DE COMPARAISONS DES PERFORMANCES

Pour comparer les performances des modèles, les réponses ont été anonymisées puis mélangées avant d'être soumises à des techniciens du service client. Pour chaque question, le technicien doit classer les réponses par ordre de préférence (tableau 7.1).

Ce classement est ensuite utilisé pour calculer plusieurs métriques (voir tableau 7.2) :

1. Le nombre de fois où la réponse **X** est arrivée en choix numéro 1. Ce score est appelé **top 1** et est exprimé en pourcentage par rapport au total des questions.
2. Le score de performance de la réponse **X**, calculé comme suit :

$$\text{Score}(x) = N - n(x) + 1 \quad (7.2)$$

où N représente le nombre de réponses à la question et $n(x)$ représente le classement pour la réponse X . Ce score est appelé **score moyen**.

7.4.4/ RÉSULTATS

Question	tech 1	tech 2	tech 3
1	2-3-1-4	2-3-1-4	3-2-1-4
2	2-1-3-4	3-2-1-4	3-2-1-4
3	2-3-1-4	2-3-1-4	3-2-1-4
4	1-3-4-2	1-2-3-4	3-1-4-2
5	1-3-4-2	1-3-2-4	3-1-4-2
6	1-4-3-2	1-2-3-4	1-3-4-2
7	1-3-(2-4)	1-2-4-3	3-1-2-4
8	3-1-4-2	1-2-4-3	3-4-1-2
9	1-(2-3-4)	1-2-3-4	1-4-3-2
10	3-4-1-2	1-2-3-4	4-3-1-2

TABLE 7.1 – Classements des réponses par préférences pour chaque technicien interrogé. Quand plusieurs réponses sont classées ex aequo, elles sont indiquées entre parenthèses.

Modèle	top 1 (%)	score moyen
Expert humain	47	31
GPT 3.5 turbo 16k	17	23
GPT 4	33	30
Mistral 7b	3	17.33

TABLE 7.2 – Score moyen de chaque modèle en fonction du classement fait par les techniciens.

On peut observer dans le tableau 7.2 que la réponse de l'expert humain reste préférée par les techniciens dans 47% des cas. Les modèles propriétaires se partagent 50% des réponses arrivées en premier choix. Le modèle open source n'arrive pas à rivaliser avec les autres modèles et a été sélectionné comme meilleure réponse que 3% du temps, ce qui est assez compréhensible au vu de sa taille.

De manière générale, les réponses obtenues par les différents modèles propriétaires sont satisfaisantes. Les erreurs observées sont dues à une mauvaise remontée des données techniques par le système en amont du LLM. Le modèle open source, quant à lui, a des problèmes à générer une réponse même avec les bonnes connaissances techniques passées en entrée.

7.4.5/ PERSPECTIVES D'AMÉLIORATIONS

Ces premières expérimentations peuvent être améliorées de différentes façons.

7.4.5.1/ UTILISATION DE MODÈLES PLUS PERFORMANTS

L'utilisation de meilleurs modèles peut améliorer la solution, notamment au niveau des modèles open sources. Un candidat potentiel serait le modèle **Mixtral 8x7b**, qui est l'un des meilleurs modèles open source selon le leaderboard du site ChatBot Arena [6].

Une autre façon d'améliorer les résultats est sur l'apport des connaissances.

7.4.5.2/ LLM : AJOUT D'UN SYSTÈME D'ÉVALUATION DES SOURCES DE CONNAISSANCES

L'évaluation de la fiabilité des sources documentaires s'effectue dans un mode spécifique de l'application. Ce mode, dit d'entraînement, va permettre à un utilisateur de juger les réponses de l'agent conversationnel. Si une réponse ne semble pas satisfaisante car l'information apportée n'est pas correcte, l'utilisateur va pouvoir le signaler et la note de fiabilité de la source va être légèrement abaissée. Dans le cas contraire, la note de la source va être augmentée en cas de réponse satisfaisante. On peut imaginer un mécanisme similaire pour la date de l'information apportée (cf. figure 7.3).

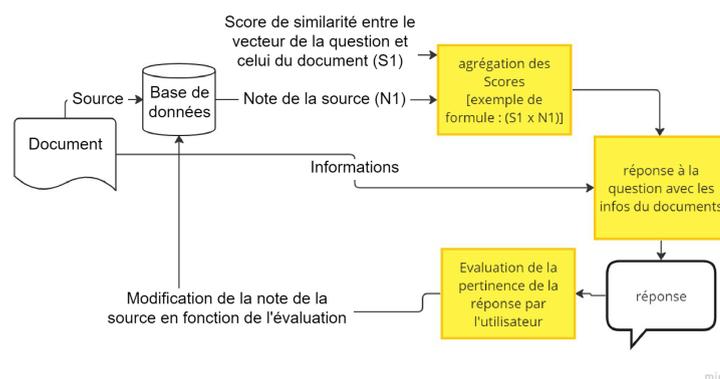
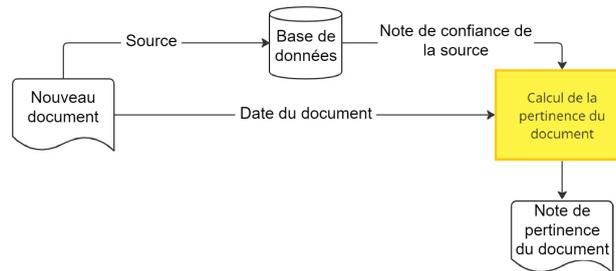


FIGURE 7.3 – Schéma expliquant le processus d'évaluation des sources de connaissances

7.4.5.3/ SÉLECTION DES CONNAISSANCES PERTINENTES PAR RAPPORT À LA QUESTION

Comme expliqué précédemment, chaque source de données a une note de fiabilité qui est stockée en base. Lorsqu'une partie de document est remontée par la proximité de son

vecteur vis-à-vis de celui de la question utilisateur, la note de fiabilité de sa source est récupérée de la base. À cette note, il est également ajouté la date du document. Cette agrégation va nous donner un poids qui va nous permettre d'affiner la pertinence des morceaux de texte remontés (cf. fig. 7.4).



miro

FIGURE 7.4 – Schéma expliquant la méthode de sélection des connaissances les plus pertinentes à donner à l'agent conversationnel

7.5/ CONCLUSION

Ce chapitre a présenté l'étude qui a été réalisée pour impliquer des modèles de langage dans un outil d'aide pour le service client du logiciel de gestion d'officine LEO. L'idée est d'apporter des connaissances techniques extraites de la documentation du logiciel à un modèle de langage capable de synthétiser une réponse. Les expérimentations réalisées ont montré que malgré un choix en première position à 47% du temps pour la réponse donnée par l'expert humain, les modèles d'OpenAI arrivent à obtenir un score moyen assez proche de celui-ci. Ces résultats montrent que l'utilisation de cette technique peut créer un outil efficace pour aider le service client dans son travail de tous les jours. Des idées d'améliorations sont à creuser pour augmenter la performance des agents conversationnels notamment au niveau de la récupération de connaissances liées à la question posée. Une fois cette technique améliorée, elle pourra être utilisée pour d'autres services au sein de la CERP, par exemple pour synthétiser des informations réglementaires, ou de pharmacovigilance, voire même créer des indicateurs à intégrer en entrée des systèmes prédictifs. À noter que selon la sensibilité des données traitées, il faudra développer un outil totalement open source à exécuter en local ou utiliser des API propriétaires certifiant la sécurité et la confidentialité de ces dernières.



CONCLUSION & PERSPECTIVES

CONCLUSION & PERSPECTIVES

8.1/ CONCLUSION

Cette thèse a présenté les travaux de recherche réalisés au sein de la société CERP Rhin Rhone Méditerranée. Ces travaux ont eu pour objectifs d'améliorer les prédictions du besoin en produits pharmaceutiques par les clients pharmaciens de la CERP RRM, et plus largement de développer les techniques de science des données dans le secteur des officines de pharmacie et de la répartition pharmaceutique.

Dans un premier temps, pour bien comprendre les travaux présentés, ce mémoire a fait une présentation du monde de la répartition pharmaceutique, en passant par l'historique et une présentation du secteur de nos jours. Il a ensuite présenté, sur l'état de l'art, des techniques pour la prédiction des séries temporelles, et a fini par un focus sur les travaux déjà réalisés dans le cadre d'un déploiement de techniques d'intelligence artificielle dans le cadre de problématiques industrielles en lien avec le métier de grossiste-répartiteur.

Dans une seconde partie, cette thèse a présenté en détails les contributions réalisées. Une première a porté sur l'amélioration des prédictions de commandes. Cette contribution a commencé par étudier les relations entre un besoin en produit pharmaceutique et une commande effectuée à la CERP RRM, puis elle a étudié les différentes possibilités d'agrégation de données. L'étape suivante de cette contribution a été d'apporter des méthodes d'analyse de données sur les données de ventes d'une officine de pharmacie. Cette étude a permis de montrer la valeur que peut apporter l'analyse des données de son officine au pharmacien. Elle a également pu améliorer les commandes automatiques en proposant une méthode simple à base de moyenne roulante pour réduire l'erreur de la méthode actuellement utilisée. Le dernier apport de cette analyse a été de dévoiler une limite : l'analyse d'un client isolé n'est pas pertinente, travailler avec des données d'officine est intéressant car elles apportent plus de détails que les données de ventes de la CERP, mais il faudrait agréger plusieurs bases de données d'officine pour avoir des résultats pertinents. Ce constat a permis de recentrer les travaux sur les données de

vente de la CERP. Les travaux ont été orientés sur l'application de techniques de machine learning pour prédire les besoins de 4 molécules, avant de s'aligner sur le nouvel outil statistique mis en place par la CERP durant la thèse. Les travaux se sont alors redirigés sur la prédiction à la classe ATC 5 (médicaments), par établissement, sur deux échelles de temps : le mois et la semaine. Les techniques de machine learning utilisées ont obtenu de meilleurs résultats que les méthodes statistiques actuelles. Enfin, la première contribution a étudié l'ajout de données externes pour améliorer les prédictions.

Les travaux de la première contribution ont pu avoir des résultats sur l'agrégation de médicaments grâce à une classification pertinente préexistante : la classification ATC. Mais il n'existe aucune classification de ce genre pour les produits de parapharmacie. La CERP RRM a émis la problématique d'avoir des difficultés à regrouper efficacement ces derniers dans son outil prédictif. La deuxième contribution de cette thèse a donc été de proposer une technique de classification basée non pas sur des données de connaissance métier mais sur l'analyse des courbes de ventes. Cette contribution a permis de comparer différentes techniques existantes, que ce soit dans les algorithmes de comparaison de distances, de réduction de dimensions et de clustering. La méthode développée a obtenu des résultats intéressants tout en gardant une facilité d'interprétation de ces derniers, ce que ne font pas les techniques ayant de meilleurs résultats dans l'état de l'art.

La dernière contribution de cette thèse est née des avancées dans le domaine des modèles de traitement du langage naturel. Ces avancées ont permis d'imaginer de nouveaux outils qui permettraient de résoudre plusieurs problématiques liées à l'activité de la CERP RRM. L'objectif de cette contribution a été de tester la faisabilité d'un outil de synthèse de connaissances grâce à des techniques de traitement du langage naturel. Cet outil doit être capable de répondre à une question en allant interroger une base de connaissances créée à partir de différentes sources non structurées. Les travaux ont donc consisté à réaliser une étude de faisabilité en mettant en place un outil qui, dans un premier temps, permet d'extraire des informations de divers documents PDF et de les organiser dans une base de connaissances. Dans un deuxième temps, l'outil est capable de comprendre une question posée par un utilisateur, d'aller chercher les informations liées à cette question dans la base de connaissances et de synthétiser une réponse à l'utilisateur. Le cas d'usage pour cette expérimentation a été la création d'un agent conversationnel pouvant répondre à des questions sur le fonctionnement d'un logiciel de gestion d'officine. Cet agent utilise une base de connaissances issue de la documentation de ce logiciel. Il a été testé sur les 10 questions les plus posées au service client du logiciel, et ces réponses ont été évaluées par des techniciens de ce même service client. Dans 47% des cas, la réponse humaine a été préférée, ce qui signifie que les techniciens interrogés ont préféré la réponse d'un des différents modèles de langage sur 53% des réponses. Ce résultat montre que cette méthode est prometteuse et qu'un travail d'amélioration des performances peut amener à un outil efficace pouvant répondre aux

divers besoins de l'entreprise pour traiter des informations non structurées.

Les différentes contributions de cette thèse ont montré que l'utilisation de techniques d'intelligence artificielle et de science des données permettent non seulement d'améliorer les prédictions de commandes, mais également de proposer des solutions innovantes à des problématiques plus larges qu'uniquement les prédictions de commandes. Tous ces travaux ouvrent également la porte à de nombreuses réflexions et perspectives de recherche et développement que nous allons aborder dans la prochaine partie de ce mémoire.

8.2/ PERSPECTIVES

Comme évoqué précédemment, les contributions de cette thèse donnent lieu à des pistes de réflexion et de nouveaux sujets de travaux.

8.2.1/ LA DONNÉE DE PHARMACIE : UN ENJEU POUR L'AVENIR DU SECTEUR

Dans le monde de l'intelligence artificielle, et plus largement dans le périmètre d'action d'un data scientist, la donnée est une ressource primordiale. Les travaux présentés dans ce mémoire ont montré que les secteurs de la répartition pharmaceutique et des officines de pharmacies possèdent beaucoup de données exploitables, et de l'autre côté un intérêt pour les informations que ces dernières peuvent apporter au côté métier. On va pouvoir distinguer deux type de données : celles internes à la CERP RRM et les données dites externes. Dans le cas des données internes, la CERP a mis en place la chaîne de vie de la donnée dans ses grandes lignes, notamment grâce à leur entrepôt de données. L'étape à parfaire est peut être de fournir des outils de visualisation pour les métiers. Dans un deuxième temps, un intérêt scientifique serait d'exploiter les données externes pour améliorer les différents outils prédictifs. Dans ces données externes on peut déjà imaginer les données agrégées de plusieurs officines. En effet, chaque pharmacien est propriétaire des données de son officine, mais nombres d'entre eux les revendent à des entreprises spécialisées en analyse de données. Un rapprochement de la CERP vers ces entreprises, ou plus directement un partenariat avec les pharmaciens pour avoir accès à ces données, permettrait d'enrichir les entrées des modèles prédictifs. Une autre voie pour l'amélioration des prédictions grâce à des données externes est de poursuivre les travaux présentés à la fin du chapitre 5 de cette thèse, avec l'ajout de données externes non générées par l'activité des officines ou du secteur de la répartition. Dans tous les cas, l'ajout de ces données externes nécessite de suivre certaines étapes :

Collecte : Le travail de la collecte de données, doit avant tout passer par plusieurs réflexions : de quelles données ai-je besoin ? Où est-ce que je peux trouver ces données ? Et enfin, comment les collecter ?

Stockage : Une fois les données récoltées, selon les données et leurs sensibilités, il faut également avoir une réflexion sur les moyens mis en place pour les stocker : durée de conservation, anonymisation des données sensibles, type de stockage, etc.

Traitement et analyse : Une fois notre base de données prête, il est temps d'exploiter cette ressource pour répondre aux problématiques. L'utilisation de ces données peut servir à alimenter des techniques d'intelligence artificielle ou interagir avec différentes applications.

Visualisation : Il est également important de développer des outils de visualisation pour permettre aux utilisateurs non techniques d'accéder aux données. Cette étape peut se présenter par exemple sous la forme de tableaux de rapports. L'aide d'UX/UI designers peut être un plus pour faciliter la compréhension de ces outils de visualisation par les utilisateurs.

8.2.2/ UN SECTEUR CONFRONTÉ À DES PROBLÉMATIQUES DE RUPTURES

Selon une commission d'enquête du Sénat français [4], on constate une hausse des pénuries de médicaments depuis 2018. Si les causes peuvent être multiples, il est intéressant d'étudier le sujet. Si tous les acteurs de la chaîne doivent être mis à contribution, les grossistes-répartiteurs, par leur rôle et leur place entre les laboratoires, ont une position de choix pour être une des pierres angulaires de l'analyse et de la résolution de ces problèmes. Une initiative a été mise en place : le portail DP rupture créé par l'Ordre national des pharmaciens (CNOP). Cette plate-forme permet aux pharmaciens de déclarer automatiquement un produit en rupture et au laboratoire concerné de répondre à cette déclaration en apportant des informations.

L'analyse des données générées par le portail DP-rupture, additionnée à d'autres données externes comme par exemple des informations sur les compositions des produits, peut permettre de comprendre les causes de ruptures et d'anticiper de futures ruptures. L'utilisation de techniques d'intelligence artificielle sur ces données est une perspective de recherche et développement intéressante. Par exemple, il semble pertinent d'utiliser des techniques de machine learning pour prédire les futures ruptures grâce aux données historiques. Analyser les causes de ruptures peut également être intéressant, qu'elle soit liée aux matières premières, à la production, au laboratoire, etc. L'analyse de

ces causes peut également être couplée à une analyse des interactions entre les produits ; par exemple, est-ce que la rupture d'un produit peut entraîner une pression sur un autre (ce qui causerait une rupture en chaîne) ?

8.2.3/ LES MODÈLES DE LANGAGES EN ASSISTANCE

L'amélioration de la technique présentée dans la troisième contribution peut permettre de nombreuses applications...

8.2.3.1/ LES LLM'S EN ASSISTANCE AUX PHARMACIENS

Appliquer un système de synthèse de base de connaissances basé sur une agrégation de données pharmaceutiques (réglementaires, informations produits, pharmacovigilance, etc.) permettrait de créer un agent conversationnel d'aide aux pharmaciens. En effet, les pharmaciens passent beaucoup de temps à prendre connaissance de ce genre de documents, mais ensuite doivent également prendre du temps pour former leurs équipes à ces connaissances. L'équipe de la pharmacie pourrait s'assister de cet agent au lieu de devoir toujours être dépendante du pharmacien titulaire. Il y a par contre un grand point sur la fiabilité des données à mettre en place et ce genre d'outil doit dans tous les cas être fiabilisé par un humain sur tout ce qui touche à la santé. L'utilisation de modèles dit "Open sources", exécuté sur une machine locale pour permettre une sécurité des données, peut être une piste de travail intéressante. Une autre idée pourrait être de créer un jeu de données pour entraîner un modèle de langage spécialiste de la pharmacie française. Ce modèle pourrait ensuite être relié à une base de connaissances qui lui servirait de mise à jour. L'intérêt de ce type de modèle spécialisé serait de pouvoir apporter des réponses plus axées "pharmacien".

8.2.3.2/ UTILISATIONS DES LLM'S POUR AUGMENTER LES DONNÉES DE PRÉDICTIONS

Une autre partie intéressante de recherche est d'utiliser des modèles de langage pour transformer une information textuelle en indicateurs à ajouter aux données d'entrée des modèles prédictifs. On peut imaginer un LLM analyser des informations textuelles par exemple un article indiquant le retrait du marché d'un médicament, pour transformer ça en un indicateur en entrée d'un modèle prédisant les commandes de ce médicament, et qui lui indiquerait un arrêt des commandes de ce médicament dans le futur. Un exemple concret serait d'utiliser un modèle de langage pour analyser les articles de la revue Prescrire [8], puis décerner aux produits cités dans les articles un coefficient qui pourrait influencer sur les prédictions. Par exemple, si un article annonce un retrait de la vente d'un

produit à une certaine date, on pourrait avoir ce coefficient qui indiquerait aux modèles prédictifs que les ventes vont s'arrêter à cette date.

8.2.4/ CLASSIFICATION ET DÉPENDANCES INTER-PRODUITS

Un autre type de travaux à développer est de continuer la classification de produits, ainsi que de travailler sur les interactions inter-produits : créer une base permettant de mettre en lumière quels produits se concurrencent ou se vendent de pair. Cela permettrait de coupler des prédictions ou potentiellement d'améliorer ces dernières (par exemple voir comment l'arrêt de commercialisation d'un produit va affecter les produits liés, de même pour l'arrivée d'un nouveau traitement, ou du passage en générique d'un produit). De plus, le regroupement entre la CERP RRM et la CERP ROUEN va sûrement bousculer les classifications produits, car les deux entités n'utilisent pas les mêmes bases de références. De gros travaux pour construire une base pertinente sont donc à prévoir dans le futur. Une fois cette base et ces dépendances inter-produits créées, il pourrait être intéressant de regarder du côté des modèles prédictifs à base de graphes. En effet si l'on considère un graphe modélisant tout ou une partie du marché pharmaceutique, avec chaque sommet représentant un produit et les arêtes correspondant aux relations entre les produits, il est donc possible d'étudier les techniques de prédiction à base de réseaux de neurones en graphe, comme déjà utilisé pour prédire le trafic routier [132] ou les pandémies [126].

PUBLICATIONS

JOURNAL PAPERS

- Jeremy Renaud, Raphaël Couturier, Christophe Guyeux & Benoit Courjal (2024) *TSCAPE : time series clustering with curve analysis and projection on an Euclidean space*, **Connection Science**, 36 :1, DOI : 10.1080/09540091.2024.2312119
- Jeremy Renaud, Ralph Karam, Michel Salomon, Raphaël Couturier, *Deep learning and gradient boosting for urban environmental noise monitoring in smart cities*, **Expert Systems with Applications**, Volume 218, 2023, 119568, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2023.119568>.

CONFERENCE PAPERS

- Jeremy Renaud, Raphaël Couturier, Christophe Guyeux, Benoit Courjal and Christine Giot, "A Comparative Study of Predictive Models for Pharmaceutical Sales Data", **2022 2nd International Conference on Computer, Control and Robotics (ICCCR)**, Shanghai, China, 2022, pp. 191-195, doi : 10.1109/ICCCR54399.2022.9790232.

BIBLIOGRAPHIE

- [1] **Accueil — Geod’air : données et statistiques sur la qualité ; de l’air en France — geodair.fr.** <https://www.geodair.fr/>. [Accessed 21-02-2024].
- [2] **Accueil — LCSQA — lcsqa.org.** <https://www.lcsqa.org/fr/>. [Accessed 21-02-2024].
- [3] **Brochure CSRP 2023 — CSRP — csrp.fr.** <https://www.csrp.fr/documentation/brochure-csrp-2023>. [Accessed 03-10-2023].
- [4] **Commission d’enquête sur la pénurie de médicaments et les choix de l’industrie pharmaceutique française — senat.fr.** <https://www.senat.fr/travaux-parlementaires/structures-temporaires/commissions-denquete/2022-2023-ce-penurie-de-medicaments.html>. [Accessed 06-03-2024].
- [5] **Le marché — CSRP — csrp.fr.** <https://www.csrp.fr/le-marche>. [Accessed 03-10-2023].
- [6] **LMSys Chatbot Arena Leaderboard - a Hugging Face Space by lmsys — huggingface.co.** <https://huggingface.co/spaces/lmsys/chatbot-arena-leaderboard>. [Accessed 04-03-2024].
- [7] **Marché français — Leem — leem.org.** <https://www.leem.org/marche-francais>. [Accessed 03-10-2023].
- [8] **prescrire.org — prescrire.org.** <https://www.prescrire.org/fr/>. [Accessed 02-04-2024].
- [9] **Un secteur en crise — CSRP — csrp.fr.** <https://www.csrp.fr/un-secteur-en-crise>. [Accessed 03-10-2023].
- [10] ROSENBLATT, F. **The perceptron : a probabilistic model for information storage and organization in the brain.** *Psychological review* 65, 6 (1958), 386.
- [11] DURBIN, J. **The fitting of time-series models.** *Revue de l’Institut International de Statistique* (1960), 233–244.
- [12] SAKOE, H., AND CHIBA, S. **Dynamic programming algorithm optimization for spoken word recognition.** *IEEE transactions on acoustics, speech, and signal processing* 26, 1 (1978), 43–49.
- [13] DAY, W. H., AND EDELSBRUNNER, H. **Efficient algorithms for agglomerative hierarchical clustering methods.** *Journal of classification* 1 (1984), 7–24.
- [14] HANSEN, L. K., AND SALAMON, P. **Neural network ensembles.** *IEEE transactions on pattern analysis and machine intelligence* 12, 10 (1990), 993–1001.

- [15] SPECHT, D. F., AND OTHERS. **A general regression neural network.** *IEEE transactions on neural networks* 2, 6 (1991), 568–576.
- [16] SHAW, C., AND KING, G. **Using cluster analysis to classify time series.** *Physica D : Nonlinear Phenomena* 58, 1-4 (1992), 288–298.
- [17] WOLPERT, D. H. **Stacked generalization.** *Neural networks* 5, 2 (1992), 241–259.
- [18] AMARI, S.-I. **Backpropagation and stochastic gradient descent method.** *Neurocomputing* 5, 4-5 (1993), 185–196.
- [19] KROGH, A., AND VEDELSBY, J. **Neural network ensembles, cross validation, and active learning.** *Advances in neural information processing systems* 7 (1994).
- [20] CORTES, C., AND VAPNIK, V. **Support-vector networks.** *Machine Learning* 20, 3 (Sept. 1995), 273–297.
- [21] HAN, J., AND MORAGA, C. **The influence of the sigmoid function parameters on the speed of backpropagation learning.** In *International workshop on artificial neural networks* (1995), Springer, pp. 195–201.
- [22] BREIMAN, L. **Bagging predictors.** *Machine learning* 24 (1996), 123–140.
- [23] ESTER, M., KRIEGEL, H.-P., SANDER, J., XU, X., AND OTHERS. **A density-based algorithm for discovering clusters in large spatial databases with noise.** In *kdd* (1996), vol. 96, pp. 226–231.
- [24] ORGANIZATION, W. H., AND OTHERS. **Guidelines for atc classification and ddd assignment.**
- [25] ROJAS, R., AND ROJAS, R. **The backpropagation algorithm.** *Neural networks : a systematic introduction* (1996), 149–182.
- [26] ZHANG, T., RAMAKRISHNAN, R., AND LIVNY, M. **Birch : an efficient data clustering method for very large databases.** *ACM sigmod record* 25, 2 (1996), 103–114.
- [27] KROGH, A., AND SOLLICH, P. **Statistical mechanics of ensemble learning.** *Physical Review E* 55, 1 (1997), 811.
- [28] GOLAY, X., KOLLIAS, S., STOLL, G., MEIER, D., VALAVANIS, A., AND BOESIGER, P. **A new correlation-based fuzzy logic clustering algorithm for fmri.** *Magnetic resonance in medicine* 40, 2 (1998), 249–260.
- [29] ANKERST, M., BREUNIG, M. M., KRIEGEL, H.-P., AND SANDER, J. **Optics : Ordering points to identify the clustering structure.** *ACM Sigmod record* 28, 2 (1999), 49–60.
- [30] MASON, L., BAXTER, J., BARTLETT, P., AND FREAN, M. **Boosting algorithms as gradient descent.** *Advances in neural information processing systems* 12 (1999).

- [31] BROWNE, M. W. **Cross-validation methods**. *Journal of mathematical psychology* 44, 1 (2000), 108–132.
- [32] DING, C. S. **Profile analysis : Multidimensional scaling approach**. *Practical Assessment, Research, and Evaluation* 7, 1 (2000), 16.
- [33] ROWEIS, S. T., AND SAUL, L. K. **Nonlinear dimensionality reduction by locally linear embedding**. *science* 290, 5500 (2000), 2323–2326.
- [34] TENENBAUM, J. B., SILVA, V. D., AND LANGFORD, J. C. **A global geometric framework for nonlinear dimensionality reduction**. *science* 290, 5500 (2000), 2319–2323.
- [35] BREIMAN, L. **Random forests**. *Machine learning* 45 (2001), 5–32.
- [36] ELMORE, K. L., AND RICHMAN, M. B. **Euclidean distance as a similarity metric for principal component analysis**. *Monthly Weather Review* 129, 3 (2001), 540 – 549.
- [37] FRIEDMAN, J. H. **Greedy function approximation : a gradient boosting machine**. *Annals of statistics* (2001), 1189–1232.
- [38] KALPAKIS, K., GADA, D., AND PUTTAGUNTA, V. **Distance measures for effective clustering of arima time-series**. In *Proceedings 2001 IEEE international conference on data mining* (2001), IEEE, pp. 273–280.
- [39] MAR-MOLINERO, C., AND SERRANO-CINCA, C. **Bank failure : a multidimensional scaling approach**. *The European Journal of Finance* 7, 2 (2001), 165–183.
- [40] NG, A., JORDAN, M., AND WEISS, Y. **On spectral clustering : Analysis and an algorithm**. *Advances in neural information processing systems* 14 (2001).
- [41] DREISEITL, S., AND OHNO-MACHADO, L. **Logistic regression and artificial neural network classification models : a methodology review**. *Journal of biomedical informatics* 35, 5-6 (2002), 352–359.
- [42] FRIEDMAN, J. H. **Stochastic gradient boosting**. *Computational statistics & data analysis* 38, 4 (2002), 367–378.
- [43] VLACHOS, M., KOLLIOS, G., AND GUNOPULOS, D. **Discovering similar multi-dimensional trajectories**. In *Proceedings 18th international conference on data engineering* (2002), IEEE, pp. 673–684.
- [44] BELKIN, M., AND NIYOGI, P. **Laplacian eigenmaps for dimensionality reduction and data representation**. *Neural Computation* 15, 6 (2003), 1373–1396.
- [45] AGARWAL, P. K., AND MUSTAFA, N. H. **K-means projective clustering**. In *Proceedings of the Twenty-Third ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* (New York, NY, USA, 2004), PODS '04, Association for Computing Machinery, p. 155–165.

- [46] CHEN, L., AND NG, R. **On the marriage of lp-norms and edit distance**. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30* (2004), pp. 792–803.
- [47] NEOPHYTOU, E., AND MOLINERO, C. M. **Predicting corporate failure in the uk : A multidimensional scaling approach**. *Journal of Business Finance & Accounting* 31, 5-6 (2004), 677–710.
- [48] WICHARD, J. D., AND OGORZALEK, M. **Time series prediction with ensemble models**. In *2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541)* (2004), vol. 2, IEEE, pp. 1625–1630.
- [49] YANG, K., AND SHAHABI, C. **A pca-based similarity measure for multivariate time series**. In *Proceedings of the 2nd ACM international workshop on Multimedia databases* (2004), pp. 65–74.
- [50] DERPANIS, K. G. **Mean shift clustering**. *Lecture Notes* 32 (2005), 1–4.
- [51] LIAO, T. W. **Clustering of time series data—a survey**. *Pattern recognition* 38, 11 (2005), 1857–1874.
- [52] ROKACH, L., AND MAIMON, O. **Decision trees**. *Data mining and knowledge discovery handbook* (2005), 165–192.
- [53] SINGHAL, A., AND SEBORG, D. E. **Clustering multivariate time-series data**. *Journal of Chemometrics : A Journal of the Chemometrics Society* 19, 8 (2005), 427–438.
- [54] ANTOINE, A. **Système coopératif et répartition pharmaceutique**. Master's thesis, UHP - Université Henri Poincaré, June 2006.
- [55] BROWN, J. C., HODGINS-DAVIS, A., AND MILLER, P. J. O. **Classification of vocalizations of killer whales using dynamic time warping**. *The Journal of the Acoustical Society of America* 119, 3 (2006), EL34–EL40.
- [56] DISMUKE, C., AND LINDROOTH, R. **Ordinary least squares**. *Methods and designs for outcomes research* 93, 1 (2006), 93–104.
- [57] CURCIO, D., LONGO, F., MIRABELLI, G., AND PAPOFF, E. **Pharmaceutical routes optimization using artificial intelligence techniques**. In *2007 4th IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems : Technology and Applications* (2007), IEEE, pp. 238–242.
- [58] FREY, B. J., AND DUECK, D. **Clustering by passing messages between data points**. *science* 315, 5814 (2007), 972–976.
- [59] CHAHARSOOGHI, S. K., HEYDARI, J., AND ZEGORDI, S. H. **A reinforcement learning model for supply chain ordering management : An application to the beer game**. *Decision Support Systems* 45, 4 (2008), 949–959.

- [60] COX, M. A. A., AND COX, T. F. **Multidimensional Scaling**. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 315–347.
- [61] HYNDMAN, R. J., AND KHANDAKAR, Y. **Automatic time series forecasting : the forecast package for r**. *Journal of statistical software* 27 (2008), 1–22.
- [62] VAN DER MAATEN, L., AND HINTON, G. **Visualizing data using t-sne**. *Journal of machine learning research* 9, 11 (2008).
- [63] KANNAN, G., NOORUL HAQ, A., AND DEVIKA, M. **Analysis of closed loop supply chain using genetic algorithm and particle swarm optimisation**. *International journal of production research* 47, 5 (2009), 1175–1200.
- [64] JARA, A. J., BELCHI, F. J., ALCOLEA, A. F., SANTA, J., ZAMORA-IZQUIERDO, M. A., AND GÓMEZ-SKARMETA, A. F. **A pharmaceutical intelligent information system to detect allergies and adverse drugs reactions based on internet of things**. In *2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)* (2010), IEEE, pp. 809–812.
- [65] NAIR, V., AND HINTON, G. E. **Rectified linear units improve restricted boltzmann machines**. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (2010), pp. 807–814.
- [66] PETITJEAN, F., KETTERLIN, A., AND GANÇARSKI, P. **A global averaging method for dynamic time warping, with applications to clustering**. *Pattern recognition* 44, 3 (2011), 678–693.
- [67] WALLACH, D. **Numerus clausus : Pourquoi la France va manquer de médecins**. Springer, 2011.
- [68] BERGSTRA, J., AND BENGIO, Y. **Random search for hyper-parameter optimization**. *J. Mach. Learn. Res.* 13, null (feb 2012), 281–305.
- [69] DURAND, M. **Le marché de la transaction d’officines : Une opportunité de croissance pour le répartiteur pharmaceutique**. PhD thesis, Université de Lorraine, 2013.
- [70] IANNONE, R., INGENITO, A., MARTINO, G., MIRANDA, S., PEPE, C., AND RIEMMA, S. **Merchandise and replenishment planning optimisation for fashion retail**. *International Journal of Engineering Business Management* 5, Godište 2013 (2013), 5–26.
- [71] JANČOVIČ, P., KÖKÜER, M., ZAKERI, M., AND RUSSELL, M. **Unsupervised discovery of acoustic patterns in bird vocalisations employing dtw and clustering**. In *21st European Signal Processing Conference (EUSIPCO 2013)* (2013), pp. 1–5.
- [72] ANUSHA, S. L., ALOK, S., AND SHAIK, A. **Demand forecasting for the indian pharmaceutical retail : A case study**. *Journal of Supply Chain Management Systems* 3, 2 (2014), 1–8.

- [73] BATISTA, G. E., KEOGH, E. J., TATAW, O. M., AND DE SOUZA, V. M. **Cid : an efficient complexity-invariant distance for time series**. *Data Mining and Knowledge Discovery* 28 (2014), 634–669.
- [74] DA VEIGA, C. P., DA VEIGA, C. R. P., CATAPAN, A., TORTATO, U., AND DA SILVA, W. V. **Demand forecasting in food retail : A comparison between the holt-winters and arima models**. *WSEAS transactions on business and economics* 11, 1 (2014), 608–614.
- [75] DAHL, G. E., JAITLY, N., AND SALAKHUTDINOV, R. **Multi-task neural networks for qsar predictions**. *arXiv preprint arXiv :1406.1231* (2014).
- [76] KINGMA, D. P., AND BA, J. **Adam : A method for stochastic optimization**. *arXiv preprint arXiv :1412.6980* (2014).
- [77] LI, M., ZHANG, T., CHEN, Y., AND SMOLA, A. J. **Efficient mini-batch training for stochastic optimization**. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (2014), pp. 661–670.
- [78] LOUPPE, G. **Understanding random forests : From theory to practice**. *arXiv preprint arXiv :1407.7502* (2014).
- [79] ZHANG, X., TRMAL, J., POVEY, D., AND KHUDANPUR, S. **Improving deep neural network acoustic models using generalized maxout networks**. In *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (2014), IEEE, pp. 215–219.
- [80] ABADI, M., AGARWAL, A., BARHAM, P., BREVDO, E., CHEN, Z., CITRO, C., CORRADO, G. S., DAVIS, A., DEAN, J., DEVIN, M., GHEMAWAT, S., GOODFELLOW, I., HARP, A., IRVING, G., ISARD, M., JIA, Y., JOZEFOWICZ, R., KAISER, L., KUDLUR, M., LEVENBERG, J., MANÉ, D., MONGA, R., MOORE, S., MURRAY, D., OLAH, C., SCHUSTER, M., SHLENS, J., STEINER, B., SUTSKEVER, I., TALWAR, K., TUCKER, P., VANHOUCHE, V., VASUDEVAN, V., VIÉGAS, F., VINYALS, O., WARDEN, P., WATTENBERG, M., WICKE, M., YU, Y., AND ZHENG, X. **TensorFlow : Large-scale machine learning on heterogeneous systems**, 2015. Software available from tensorflow.org.
- [81] BOX, G. E., JENKINS, G. M., REINSEL, G. C., AND LJUNG, G. M. **Time series analysis : forecasting and control**. John Wiley & Sons, 2015.
- [82] CHOLLET, F., AND OTHERS. **Keras**. <https://keras.io>, 2015.
- [83] KARIMI, S., WANG, C., METKE-JIMENEZ, A., GAIRE, R., AND PARIS, C. **Text and data mining techniques in adverse drug reaction detection**. *ACM Computing Surveys (CSUR)* 47, 4 (2015), 1–39.
- [84] CHEN, T., AND GUESTRIN, C. **Xgboost : A scalable tree boosting system**. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (2016), pp. 785–794.

- [85] JOLLIFFE, I. T., AND CADIMA, J. **Principal component analysis : a review and recent developments**. *Philosophical transactions of the royal society A : Mathematical, Physical and Engineering Sciences* 374, 2065 (2016), 20150202.
- [86] LOPES, A. M., ANDRADE, J. P., AND TENREIRO MACHADO, J. **Multidimensional scaling analysis of virus diseases**. *Computer Methods and Programs in Biomedicine* 131 (2016), 97–110.
- [87] YUEHONG, Y., ZENG, Y., CHEN, X., AND FAN, Y. **The internet of things in healthcare : An overview**. *Journal of Industrial Information Integration* 1 (2016), 3–13.
- [88] CUTURI, M., AND BLONDEL, M. **Soft-DTW : a differentiable loss function for time-series**. In *Proceedings of the 34th International Conference on Machine Learning (06–11 Aug 2017)*, D. Precup and Y. W. Teh, Eds., vol. 70 of *Proceedings of Machine Learning Research*, PMLR, pp. 894–903.
- [89] KE, G., MENG, Q., FINLEY, T., WANG, T., CHEN, W., MA, W., YE, Q., AND LIU, T.-Y. **Lightgbm : A highly efficient gradient boosting decision tree**. *Advances in neural information processing systems* 30 (2017).
- [90] RUDER, S. **An overview of gradient descent optimization algorithms**, 2017.
- [91] AGARAP, A. F. **Deep learning using rectified linear units (relu)**. *arXiv preprint arXiv :1803.08375* (2018).
- [92] CARPENTER, K. A., COHEN, D. S., JARRELL, J. T., AND HUANG, X. **Deep learning and virtual drug screening**. *Future Medicinal Chemistry* 10, 21 (2018), 2557–2567. PMID : 30288997.
- [93] CHEVILLARD, G., LUCAS-GABRIELLI, V., AND MOUSQUES, J. **«déserts médicaux» en france : état des lieux et perspectives de recherches**. *L'Espace géographique* 47, 4 (2018), 362–380.
- [94] MENGER, V., SCHEEPERS, F., AND SPRUIT, M. **Comparing deep learning and classical machine learning approaches for predicting inpatient violence incidents from clinical text**. *Applied Sciences* 8, 6 (2018), 981.
- [95] PETROPOULOS, F., HYNDMAN, R. J., AND BERGMEIR, C. **Exploring the sources of uncertainty : Why does bagging for time series forecasting work ?** *European Journal of Operational Research* 268, 2 (2018), 545–554.
- [96] SELLWOOD, M. A., AHMED, M., SEGLER, M. H., AND BROWN, N. **Artificial intelligence in drug discovery**, 2018.
- [97] SZRLEE. **Djia 30 stock time series**, 2018. data retrieved from Kaggle, <https://www.kaggle.com/datasets/szrlee/stock-time-series-20050101-to-20171231?resource=download>.
- [98] BASILE, A. O., YAHY, A., AND TATONETTI, N. P. **Artificial intelligence for drug toxicity and safety**. *Trends in pharmacological sciences* 40, 9 (2019), 624–635.

- [99] BECHT, E., MCINNES, L., HEALY, J., DUTERTRE, C.-A., KWOK, I. W., NG, L. G., GINHOUX, F., AND NEWELL, E. W. **Dimensionality reduction for visualizing single-cell data using umap.** *Nature biotechnology* 37, 1 (2019), 38–44.
- [100] BOTZUNG, A. **La dispensation de médicaments aux mineurs à l’officine : étude de droit pharmaceutique.** Master’s thesis, Université de Lorraine, 2019.
- [101] DA SILVA, D. A., TEN CATEN, C. S., DOS SANTOS, R. P., FOGLIATTO, F. S., AND HSUAN, J. **Predicting the occurrence of surgical site infections using text mining and machine learning.** *PloS one* 14, 12 (2019), e0226272.
- [102] FREEDMAN, D. H., AND OTHERS. **Hunting for new drugs with ai.** *Nature* 576, 7787 (2019), S49–S53.
- [103] HOLDINGS, H. **House property sales time series**, 2019. data retrieved from Kaggle, <https://www.kaggle.com/datasets/htagholdings/property-sales?resource=download>.
- [104] LAFORGE, P. **Le dossier médical partagé.** *Actualités Pharmaceutiques* 58, 584 (2019), 29–30.
- [105] LI, H. **Multivariate time series clustering based on common principal component analysis.** *Neurocomputing* 349 (2019), 239–247.
- [106] WU, J., CHEN, X.-Y., ZHANG, H., XIONG, L.-D., LEI, H., AND DENG, S.-H. **Hyperparameter optimization for machine learning models based on bayesian optimization.** *Journal of Electronic Science and Technology* 17, 1 (2019), 26–40.
- [107] BAGOT, M. **Plateforme adaptative pour le suivi de l’état de santé de patients mobiles.** PhD thesis, Lorient, 2020.
- [108] BERNERT, R. A., HILBERG, A. M., MELIA, R., KIM, J. P., SHAH, N. H., AND ABNOUSI, F. **Artificial intelligence and suicide prevention : a systematic review of machine learning investigations.** *International journal of environmental research and public health* 17, 16 (2020), 5929.
- [109] ESKILDSEN, N. K., ERIKSSON, R., CHRISTENSEN, S. B., AGHASSIPOUR, T. S., BYGSØ, M. J., BRUNAK, S., AND HANSEN, S. L. **Implementation and comparison of two text mining methods with a standard pharmacovigilance method for signal detection of medication errors.** *BMC Medical Informatics and Decision Making* 20 (2020), 1–11.
- [110] KAMALZADEH, H., AHMADI, A., AND MANSOUR, S. **Clustering time-series by a novel slope-based similarity measure considering particle swarm optimization.** *Applied Soft Computing* 96 (2020), 106701.
- [111] MALZER, C., AND BAUM, M. **A hybrid approach to hierarchical density-based cluster selection.** In *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)* (2020), pp. 223–228.

- [112] MÉNDEZ-LUCIO, O., BAILLIF, B., CLEVERT, D.-A., ROUQUIÉ, D., AND WICHARD, J. **De novo generation of hit-like molecules from gene expression signatures using artificial intelligence.** *Nature communications* 11, 1 (2020), 10.
- [113] PUNIA, S., NIKOLOPOULOS, K., SINGH, S. P., MADAN, J. K., AND LITSIU, K. **Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail.** *International journal of production research* 58, 16 (2020), 4964–4979.
- [114] TAVENARD, R., FAOUZI, J., VANDEWIELE, G., DIVO, F., ANDROZ, G., HOLTZ, C., PAYNE, M., YURCHAK, R., RUSSWURM, M., KOLAR, K., AND WOODS, E. **Tslearn, a machine learning toolkit for time series data.** *Journal of Machine Learning Research* 21, 118 (2020), 1–6.
- [115] ZEMIHI, M. **L'utilisation de l'intelligence de artificielle dans la découverte de nouveaux candidats-médicaments et dans l'imagerie médicale.** Master's thesis, Université de Lorraine, 2020.
- [116] CHEN, P.-F., WANG, S.-M., LIAO, W.-C., KUO, L.-C., CHEN, K.-C., LIN, Y.-C., YANG, C.-Y., CHIU, C.-H., CHANG, S.-C., LAI, F., AND OTHERS. **Automatic icd-10 coding and training system : deep neural network based on supervised learning.** *JMIR Medical Informatics* 9, 8 (2021), e23230.
- [117] COUSIN, T. **Déploiement de la télémédecine en france, des opportunités pour l'officine.** *Actualités Pharmaceutiques* 60, 607 (2021), 20–25.
- [118] DE RYCK, T., LANTHALER, S., AND MISHRA, S. **On the approximation of functions by tanh neural networks.** *Neural Networks* 143 (2021), 732–750.
- [119] JOHNSON, K. B., WEI, W.-Q., WEERARATNE, D., FRISSE, M. E., MISULIS, K., RHEE, K., ZHAO, J., AND SNOWDON, J. L. **Precision medicine, ai, and the future of personalized health care.** *Clinical and translational science* 14, 1 (2021), 86–93.
- [120] JUMPER, J., EVANS, R., PRITZEL, A., GREEN, T., FIGURNOV, M., RONNEBERGER, O., TUNYASUVUNAKOOL, K., BATES, R., ŽÍDEK, A., POTAPENKO, A., AND OTHERS. **Highly accurate protein structure prediction with alphafold.** *Nature* 596, 7873 (2021), 583–589.
- [121] LEE, C. Y., AND CHEN, Y.-P. P. **Prediction of drug adverse events using deep learning in pharmaceutical discovery.** *Briefings in Bioinformatics* 22, 2 (2021), 1884–1901.
- [122] MROZOVSKI, J.-M., AND DE CHALAIN, A. **L'intelligence artificielle au service du patient de l'officine.** *Actualités Pharmaceutiques* 60, 611 (2021), 28–29.
- [123] NAHR, J. G., NOZARI, H., AND SADEGHI, M. E. **Green supply chain based on artificial intelligence of things (aiot).** *International Journal of Innovation in Management, Economics and Social Sciences* 1, 2 (2021), 56–63.

- [124] NGOC, T. T., LE VAN DAI, C. M. T., AND THUYEN, C. M. **Support vector regression based on grid search method of hyperparameters for load forecasting.** *Acta Polytechnica Hungarica* 18, 2 (2021), 143–158.
- [125] NTAKOURIS, T. **Timeseries classification with a transformer model.** https://keras.io/examples/timeseries/timeseries_classification_transformer/, 2021. [Accessed 13-01-2022].
- [126] PANAGOPOULOS, G., NIKOLENTZOS, G., AND VAZIRGIANNIS, M. **Transfer graph neural networks for pandemic forecasting.** In *Proceedings of the AAAI Conference on Artificial Intelligence* (2021), vol. 35, pp. 4838–4845.
- [127] PAUL, D., SANAP, G., SHENOY, S., KALYANE, D., KALIA, K., AND TEKADE, R. K. **Artificial intelligence in drug discovery and development.** *Drug discovery today* 26, 1 (2021), 80.
- [128] WALTERS, M., AND KALINOVA, E. **Virtualized care systems, medical artificial intelligence, and real-time clinical monitoring in covid-19 diagnosis, screening, surveillance, and prevention.** *American Journal of Medical Research* 8, 2 (2021), 37–50.
- [129] YAO, X., RUSHLOW, D. R., INSELMAN, J. W., MCCOY, R. G., THACHER, T. D., BEHNKEN, E. M., BERNARD, M. E., ROSAS, S. L., AKFALY, A., MISRA, A., AND OTHERS. **Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction : a pragmatic, randomized clinical trial.** *Nature Medicine* 27, 5 (2021), 815–819.
- [130] ABBASIMEHR, H., AND BAHRINI, A. **An analytical framework based on the recency, frequency, and monetary model and time series clustering techniques for dynamic segmentation.** *Expert Systems with Applications* 192 (2022), 116373.
- [131] HITTMEIR, M., MAYER, R., AND EKELHART, A. **Distance-based techniques for personal microbiome identification.** In *Proceedings of the 17th International Conference on Availability, Reliability and Security* (2022), pp. 1–13.
- [132] JIANG, W., AND LUO, J. **Graph neural network for traffic forecasting : A survey.** *Expert Systems with Applications* 207 (2022), 117921.
- [133] JOSEPH, R. V., MOHANTY, A., TYAGI, S., MISHRA, S., SATAPATHY, S. K., AND MOHANTY, S. N. **A hybrid deep learning framework with cnn and bi-directional lstm for store item demand forecasting.** *Computers and Electrical Engineering* 103 (2022), 108358.
- [134] MODGIL, S., SINGH, R. K., AND HANNIBAL, C. **Artificial intelligence for supply chain resilience : learning from covid-19.** *The International Journal of Logistics Management* 33, 4 (2022), 1246–1268.

- [135] MOHAMMADI, Y., MIRAFTABZADEH, S. M., BOLLEN, M. H., AND LONGO, M. **An unsupervised learning schema for seeking patterns in rms voltage variations at the sub-10-minute time scale.** *Sustainable Energy, Grids and Networks* 31 (2022), 100773.
- [136] VALLÉS-PÉREZ, I., SORIA-OLIVAS, E., MARTÍNEZ-SOBER, M., SERRANO-LÓPEZ, A. J., GÓMEZ-SANCHÍS, J., AND MATEO, F. **Approaching sales forecasting using recurrent neural networks and transformers.** *Expert Systems with Applications* 201 (2022), 116993.
- [137] YOU, L., JIANG, H., HU, J., CHANG, C. H., CHEN, L., CUI, X., AND ZHAO, M. **Gpu-accelerated faster mean shift with euclidean distance metrics.** In *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)* (2022), IEEE, pp. 211–216.
- [138] ADULYASAK, Y., BENOMAR, O., CHAOUACHI, A., COHEN, M. C., AND KHERN-AMNUAI, W. **Using ai to detect panic buying and improve products distribution amid pandemic.** *AI & society* (2023), 1–30.
- [139] ALIPER, A., KUDRIN, R., POLYKOVSKIY, D., KAMYA, P., TUTUBALINA, E., CHEN, S., REN, F., AND ZHAVORONKOV, A. **Prediction of clinical trials outcomes based on target choice and clinical trial design with multi-modal artificial intelligence.** *Clinical Pharmacology & Therapeutics* (2023).
- [140] AN, S., BHAT, G., GUMUSSOY, S., AND OGRAS, U. **Transfer learning for human activity recognition using representational analysis of neural networks.** *ACM Transactions on Computing for Healthcare* 4, 1 (2023), 1–21.
- [141] ANDREELLA, A., DE SANTIS, R., VESELY, A., AND FINOS, L. **Procrustes-based distances for exploring between-matrices similarity.** *Statistical Methods & Applications* (2023), 1–16.
- [142] BAS, T. G., ASTUDILLO, P., ROJO, D., AND TRIGO, A. **Opinions related to the potential application of artificial intelligence (ai) by the responsible in charge of the administrative management related to the logistics and supply chain of medical stock in health centers in north of chile.** *International Journal of Environmental Research and Public Health* 20, 6 (2023), 4839.
- [143] EJEGWA, P. A., AND AGBETAYO, J. M. **Similarity-distance decision-making technique and its applications via intuitionistic fuzzy pairs.** *Journal of Computational and Cognitive Engineering* 2, 1 (2023), 68–74.
- [144] LEE, Y., AND PARK, S. **Spectral clustering of weighted variables on multi-omics data.** *The Korean Journal of Applied Statistics* 36, 3 (2023), 175–196.
- [145] LIU, B., LIU, C., ZHOU, Y., WANG, D., AND DUN, Y. **An unsupervised chatter detection method based on ae and merging gmm and k-means.** *Mechanical Systems and Signal Processing* 186 (2023), 109861.

- [146] JEREMY RENAUD, RAPHAËL COUTURIER, C. G., AND COURJAL, B. **Tscape : time series clustering with curve analysis and projection on an euclidean space.** *Connection Science* 36, 1 (2024), 2312119.
- [147] VOLANT, S. **Un premier enfant 28,5 ans en 2015 : 4,5 ans plus tard qu'en 1974 - Insee Première - 1642 — insee.fr.** <https://www.insee.fr/fr/statistiques/2668280>, 27/03/2017. [Accessed 05-12-2023].
- [148] IPLESP, INSERM - SORBONNE UNIVERSITÉ. **Réseau sentinelles.** <http://www.sentiweb.fr>.

TABLE DES FIGURES

2.1	Exemple d'une classification ATC complète	10
2.2	Grossistes-répartiteurs adhérents à la CSRP et leurs parts sur le marché français. Source : "https ://www.csrp.fr/le-marche"	14
2.3	Répartitions des agences de la CERP RRM sur son territoire	19
2.4	Malgré une rehausse sur 2020-2021, on observe une baisse de la marge du secteur de la répartition depuis 2007, alors que l'objectif national des dépenses de l'assurance maladie est en augmentation constante	21
2.5	Parts de marché de la répartition pharmaceutique dans la distribution des médicaments en 2022. On peut voir que les grossistes-répartiteurs, en plus d'une concurrence entre eux, sont aussi concurrencés par la vente directe des laboratoires. Source : Leem.org [7]	22
3.1	Architecture d'un Neurone Artificiel	43
3.2	Architecture d'une cellule RNN, chaque cellule reçoit en entrée l'état caché du pas de temps précédent en plus de la variable du pas de temps actuel.	48
3.3	Architecture d'une cellule LSTM, chaque cellule reçoit en entrée l'état caché et l'état cellulaire du pas de temps précédent en plus de la variable du pas de temps actuel.	49
3.4	Architecture d'une cellule GRU, chaque cellule reçoit en entrée l'état caché du pas de temps précédent en plus de la variable du pas de temps actuel. Sur cette représentation \tilde{h}_t représente le candidat pour l'état caché : \tilde{h}_t	51
3.5	Architecture des modèles de transformers comme décrite dans les travaux de Vaswani et al.	64
5.1	Schéma décisionnel au niveau du prescripteur	74
5.2	Schéma décisionnel au niveau du patient	74
5.3	Schéma décisionnel au niveau du pharmacien	76
5.4	MindMap des différentes échelles d'agregations possibles pour les données de ventes	79

5.5	Population de la ville de Gap comparée à la patientèle d'une pharmacie de Gap, regroupées par tranche d'âge et sexe, en nombre de personnes, pour l'année 2020	84
5.6	Pourcentage de la population de la ville de Gap comparé au pourcentage de la patientèle d'une pharmacie de Gap, regroupées par tranche d'âge et sexe, pour l'année 2020	85
5.7	Pourcentages cumulés de la population de la ville de Gap comparés au pourcentage de la patientèle d'une pharmacie de Gap, regroupés par tranche d'âge et sexe, pour l'année 2020. On peut observer que la répartition de la patientèle est plus faible que celle de la population sur les tranches d'âge. Cet écart se comble plus rapidement chez les femmes que chez les hommes	86
5.8	Comparaison entre les pourcentages de la patientèle, du chiffre d'affaires (CA), de la marge et des ventes d'une pharmacie de Gap regroupés par âge du patient sur la période 2008-2021. On peut observer que les moins de 40 ans ont une part moins importante dans les ventes/CA/Marge que leur représentativité dans le pourcentage de la patientèle.	87
5.9	Comparaison des pourcentages cumulés de la patientèle et des ventes de la pharmacie, regroupés par âge.	87
5.10	Pourcentages des ventes d'une pharmacie de Gap, regroupés par type de ventes (remboursées, non-remboursées, mixtes) et par âge du patient.	88
5.11	Comparaison de la copie de l'année précédente, d'une moyenne roulante sur 6 mois, d'une moyenne roulante sur 3 mois et des ventes réelles pour un produit stable.	91
5.12	Comparaison de la copie de l'année précédente, d'une moyenne roulante sur 6 mois, d'une moyenne roulante sur 3 mois et des ventes réelles pour un produit en croissance.	91
5.13	Comparaison de la copie de l'année précédente, d'une moyenne roulante sur 6 mois, d'une moyenne roulante sur 3 mois et des ventes réelles pour un produit dont les ventes chutes.	92
5.14	Schéma simplifié du pipeline de données utilisé pour les modèles prédictifs	93
5.15	Courbes de prévision du modèle Transformers pour l'Amoxicilline	96
5.16	Courbes de prévision du modèle Transformers pour la Cétirizine	96
5.17	Courbes de prévision du modèle Transformers pour le Colécalciférol	97
5.18	Courbes de prévision du modèle Transformers pour le Paracétamol	97

5.19	Courbes de prévision du modèle LSTM pour l'Amoxicilline	98
5.20	Courbes de prévision du modèle LSTM pour la Cétirizine	98
5.21	Courbes de prévision du modèle LSTM pour le Colécalciférol	99
5.22	Courbes de prévision du modèle LSTM pour le Paracétamol	99
5.23	Courbes de prévision du modèle LightGBM pour l'Amoxicilline	100
5.24	Courbes de prévision du modèle LightGBM pour le Colécalciférol	100
5.25	Courbes de prévision du modèle LightGBM pour la Cétirizine	101
5.26	Courbes de prévision du modèle LightGBM pour le Paracétamol	101
5.27	Schéma d'explication des données d'entrée pour le modèle LSTM sur les prédictions à la semaine pour les classes de médicaments ATC5. Pour prédire la semaine N, on donne en entrée le nombre de vente des semaines [N-4,N-9], soit une fenêtre de lookback de 6 pas de temps. On donne également pour chacune de ces valeurs, la tendance des ventes sur les 4 pas de temps précédents : Par exemple, on accompagne le nombre de vente de la semaine N-4 de la tendance des ventes sur la période [N-5,N-8].	103
6.1	Les différentes étapes de la solution TSCAPE	113
6.2	Inertie des clusters par rapport à leur nombre pour le jeu de données "Parapharmacie"	120
6.3	Inertie des clusters par rapport à leur nombre pour le jeu de données "COALIA"	120
6.4	Projection 2D grâce au MDS pour le jeu de données "Parapharmacie". Chaque point correspond à la représentation d'une série temporelle. Deux points proches l'un de l'autre signifient que les deux séries temporelles ont une courbe similaire.	122
6.5	Projection 2D grâce au MDS pour le jeu de données "COALIA". Chaque point correspond à la représentation d'une série temporelle. Deux points proches l'un de l'autre signifient que les deux séries temporelles ont une courbe similaire.	123
6.6	Projection 2D grâce au MDS suivie d'un remplacement de points à l'aide de l'algorithme t-SNE pour l'ensemble de données "Parapharmacie". Chaque point correspond à la représentation d'une série temporelle. Deux points proches l'un de l'autre signifient que les deux séries temporelles ont une courbe similaire.	124

6.7	Projection 2D grâce au MDS suivie d'un remplacement de points à l'aide de l'algorithme t-SNE pour l'ensemble de données "COALIA". Chaque point correspond à la représentation d'une série temporelle. Deux points proches l'un de l'autre signifient que les deux séries temporelles ont une courbe similaire	125
6.8	Clusters pour le jeu de données "Parapharmacie". Les clusters sont représentés par des couleurs différentes : deux points de même couleur sont dans le même cluster.	126
6.9	Clusters pour le jeu de données "COALIA". Les clusters sont représentés par des couleurs différentes : deux points de même couleur sont dans le même cluster.	126
7.1	Schéma expliquant le processus d'ajout d'un document dans la base vectorielle de connaissances	132
7.2	Schéma expliquant toutes les étapes qui permettent à l'agent conversationnel de répondre à la question de l'utilisateur. On peut voir en vert la brique de filtre et sélection qui n'est pas encore implémentée.	133
7.3	Schéma expliquant le processus d'évaluation des sources de connaissances	138
7.4	Schéma expliquant la méthode de sélection des connaissances les plus pertinentes à donner à l'agent conversationnel	139

LISTE DES TABLES

5.1	Comparatif de l'erreur absolue moyenne (MAE) des prédictions mensuelles pour les techniques de moyenne roulante sur 3 mois et 6 mois, par rapport aux ventes réelles pour l'année 2020. (meilleurs résultats en gras)	90
5.2	Comparaison de RMSE pour la prévision des ventes d'Amoxicilline	94
5.3	Comparaison de RMSE pour la prévision des ventes de Cétirizine	94
5.4	Comparaison de RMSE pour la prévision des ventes de Colécalciférol	95
5.5	Comparaison de RMSE pour la prévision des ventes de Paracétamol	95
5.6	Comparaison de la MAE pour le mois de septembre 2022, sur toutes les ventes par catégorie ATC5 pour l'établissement CERP d'Aix en Provence. (* Cette statistique prend en compte la totalité des produits et pas seulement les produits ATC (médicaments). La donnée séparée pour ce mois ci n'as pas pu être fournie. A titre de comparaison : pour le mois de février 2024 l'erreur des produits ATC est de 24,2% et de 28.01% pour les produits hors ATC)	104
6.1	Variation des métriques en fonction des paramètres gamma du SoftDTW.	122
6.2	Variation des métriques par rapport au nombre de dimensions du MDS.	123
6.3	Variation des métriques par rapport au numéro de graine du MDS.	124
6.4	Analyse comparative (moyenne de 10 exécutions) de toutes les solutions et options du présent document pour l'ensemble de données Parapharmacie (la meilleure solution est en gras).	128
6.5	Analyse comparative (moyenne de 10 exécutions) de toutes les solutions et options du présent document pour l'ensemble de données COALIA (la meilleure solution est en gras).	129
7.1	Classements des réponses par préférences pour chaque technicien interrogé. Quand plusieurs réponses sont classées ex aequo, elles sont indiquées entre parenthèses.	137

7.2	Score moyen de chaque modèle en fonction du classement fait par les techniciens.	137
-----	--	-----

Titre : Amélioration de la prédiction des commandes des pharmacies auprès de la CERP RRM

Mots-clés : Réseaux de neurones, Prédiction, Clustering, Répartition pharmaceutique, Traitement du langage naturel

Résumé :

La CERP Rhin Rhone Méditerranée (CERP RRM) est un grossiste répartiteur qui doit assurer l'approvisionnement de pharmacies. Le secteur du médicament, malgré l'évolution récente de la logistique hospitalière, manque considérablement d'outils d'aide à la décision. Le but de la thèse est de mettre en place un système prédictif pour l'ensemble des clients CERP afin de prévoir les commandes avec la meilleure précision possible. Les données sont essentiellement sous forme de séries temporelles.

Dans un premier temps, les travaux de cette thèse ont été de faire un état de l'art des technologies de prédictions de séries temporelles, ainsi que de l'implémentation de systèmes d'IA dans des secteurs industriels liés aux métiers de grossiste répartiteur. La principale contribution de cette thèse

a été d'améliorer les prédictions de la CERP RRM à plusieurs niveaux grâce à des techniques de machine learning. Nos résultats montrent une amélioration des prédictions par rapport à la méthode actuelle. La deuxième contribution a été de proposer une nouvelle méthode basée sur l'étude des courbes de ventes pour regrouper les produits entre eux. Cette méthode a été développée pour répondre à une problématique de regroupement des produits de parapharmacie au sein de la CERP RRM. La dernière contribution de cette thèse est une étude comparative de différents modèles de traitement du langage naturel implémentés dans un assistant conversationnel pour le service technique d'un logiciel de gestion d'officine. Cette solution a montré des résultats intéressants en se rapprochant des résultats d'un humain expert.

Title: Improvement of prediction of pharmacy orders with CERP RRM

Keywords: Deep Learning, Forecasting, Clustering, Pharmaceutical distribution, Natural language processing

Abstract:

The CERP Rhin Rhone Méditerranée (CERP RRM) is a wholesale distributor responsible for ensuring pharmacies' supply. Despite recent advancements in hospital logistics, the pharmaceutical sector notably lacks decision support tools. The thesis aims to establish a predictive system for all CERP clients to forecast orders with the highest possible accuracy. The data primarily consists of time series.

Initially, the thesis focused on conducting a state-of-the-art review of time series prediction technologies, as well as implementing AI systems in industrial sectors related to wholesale distribution professions. The main contribution of this thesis was to enhance CERP RRM predictions at multiple

levels using machine learning techniques. Our results demonstrate an improvement in predictions compared to the current method. The second contribution was to propose a new method based on sales curve analysis to group products together. This method was developed to address the issue of grouping parapharmacy products within CERP RRM. The final contribution of this thesis is a comparative study of different natural language processing models implemented in a conversational assistant for the technical service of a pharmacy management software. This solution has shown promising results, approaching those of an expert human.