



HAL
open science

Modélisation des processus de prévision de la chaîne logistique pour l'amélioration des performances industrielles

Mohamed Sameh Belaid

► **To cite this version:**

Mohamed Sameh Belaid. Modélisation des processus de prévision de la chaîne logistique pour l'amélioration des performances industrielles. Automatique. Ecole nationale supérieure Mines-Télécom Lille Douai, 2022. Français. NNT : 2022MTLD0004 . tel-03860429

HAL Id: tel-03860429

<https://theses.hal.science/tel-03860429>

Submitted on 18 Nov 2022

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.

THESE

Présentée dans le but d'obtenir le grade de

DOCTEUR

Spécialité : Automatique, Génie Informatique, Traitement du Signal et des Images

par

Mohamed Sameh Belaid

Doctorat de l'université de Lille délivré par IMT Nord Europe

Titre de la Thèse :

Modélisation des processus de prévision de la chaîne logistique pour l'amélioration des performances industrielles

Soutenue le 30 Mars 2022 devant le jury :

Président du jury & Rapporteur,	Professeur (UNISTRA)	Bertrand Rose
Rapporteur,	Directeur de Recherche (UGE)	Allou-Badara Samé
Examinatrice,	Enseignante - Chercheuse (IMT Mines-Albi)	Aurélie Montarnal
Examineur,	Professeur (EDHEC)	Michel Philippart
Directeur de Thèse,	Professeur (IMT Nord Europe)	Stéphane Lecoeuche
Co-directeur de Thèse, Enseignant - Chercheur HDR (IMT Nord Europe)		Anthony Fleury
Invité,	Chef de projets IT (elm.leblanc)	Yannick Grall

Laboratoire d'accueil :

CERI Systèmes Numériques, IMT Nord Europe

Ecole doctorale 631 - **Mathématiques, sciences du numérique et de leurs interactions (MADIS)**

Ce manuscrit est le résultat des travaux de recherche supporté par l'Association nationale de la recherche et de la technologie (ANRT) et l'entreprise elm.leblanc du groupe Bosch.



BOSCH

Remerciement

Je tiens à exprimer ma reconnaissance illimitée à tous les intervenants qui ont aidé à l'accomplissement de ces travaux par leurs conseils, leurs critiques constructives et leur soutien moral.

Cette thèse ne serait rien sans l'encadrement que j'ai reçu. C'est pourquoi je souhaite tout d'abord, remercier chaleureusement Prof. Stéphane Lecoecue et Enseignant - Chercheur HDR. Anthony Fleury pour leur implication tout au long de cette recherche, qui ont sus discuter et structurer l'ensemble de mes idées depuis les débuts et jusqu'à ma soutenance.

Je tiens à exprimer ma grande gratitude à Dr. Baptiste Hervé pour son accompagnement et ses conseils judicieux tout au long de ce projet. Je souhaite également remercier l'ensemble de mes collègues de l'entreprise elm.leblanc, et notamment le service BDO, pour la confiance qu'ils ont pu me faire pour la réalisation de ce projet. C'est grâce à leur engagement et leur disponibilité que ces travaux ont pu être réalisés.

Je souhaite remercier aussi les membres du Jury d'avoir accepté d'évaluer mes travaux de recherche et pour leurs retours qu'ils m'adresseront lors des corrections.

Merci à tous les membres du CERI pour leur accueil dans le laboratoire, pour le partage et pour la convivialité des moments passés au sein de l'équipe.

Enfin, merci à mes parents, ma femme, mes frères et mes amis pour leur soutien dans cette période particulière qu'est la thèse de doctorat.

À tous ces intervenants, je présente mes hommages, mon respect et ma gratitude.

Sommaire

Introduction Générale	14
Chapitre 1 : Positionnement et enjeux de la recherche	18
I. Introduction.....	18
II. Positionnement scientifique : Analyse prédictive et Supply Chain	19
II.1. Analyse prédictive de la Supply Chain	19
II.2. Sales and Operations Planning (S&OP)	21
II.3. L'analyse prédictive de la Supply Chain et le processus S&OP	23
III. Définition de la problématique associant Analyse prédictive et Supply Chain dans le contexte de l'entreprise elm.leblanc.....	24
III.1. Le contexte et les enjeux pour l'entreprise	24
III.2. La SC et les processus de production et de planification.....	27
III.3. Zoom sur le processus de prévisions des ventes	28
IV. Conclusion	33
Chapitre 2 : Etat de l'art.....	35
I. Introduction.....	35
II. Problématique de la prévision des ventes	35
III. Présentation de domaines d'application et de leur spécificité en termes de prévision de ventes	37
III.1. Ventes au détail	37
III.2. Secteur de la mode	39
III.3. Secteur alimentaire.....	41
III.4. Ventes en ligne.....	43
III.5. Industrie automobile.....	44
III.6. Secteur de l'énergie	46
III.7. Synthèse des besoins de prédiction associés aux différents domaines.....	49
IV. Horizon des principales techniques utilisées pour la prévision des ventes	52
IV.1. Les approches de modélisation de séries temporelles	53
IV.2. Les approches de Data Mining et de Machine Learning.....	59
IV.3. Les approches de Deep Learning	64
IV.4. Approches d'hybridation	67
V. Conclusion	72
Chapitre 3 : Méthodologie et Expérimentations	74
I. Introduction.....	74
II. Méthodologie.....	74
II.1. Formalisation du problème traité.....	74
II.2. Préparation des données	77

II.3.	Description de la méthodologie.....	79
II.4.	Modèles utilisés versus caractéristiques de la série temporelle	82
II.5.	Processus d'apprentissage et critère de compétition.....	89
III.	Expérimentations et validation de la méthodologie	90
III.1.	Expérimentations réalisées.....	90
III.2.	Analyses du choix des modèles par profil de produits	106
III.3.	Performances des modèles par horizon de prédiction.....	114
IV.	Vers une méthodologie générique	118
IV.1.	Généralisation.....	119
IV.2.	Discussion autour de la méthodologie proposée	121
V.	Exploration et ouverture	124
VI.	Conclusion.....	129
Chapitre 4 : Méthode d'hybridation : Approfondissement		130
I.	Introduction.....	130
II.	Contexte de l'approche d'hybridation	130
III.	Méthodologie.....	131
III.1.	Description de l'approche.....	131
III.2.	Description des données.....	133
III.3.	Présentation du réseau MLP utilisée dans l'approche.....	134
III.4.	Processus d'apprentissage et de test.....	135
IV.	Expérimentations	136
IV.1.	Les données utilisées pour les expérimentations.....	136
IV.2.	Choix de configuration.....	139
IV.3.	Résultats obtenus après implémentation.....	141
V.	Conclusion	144
Chapitre 5 : Industrialisation des travaux et mise en production		146
I.	Introduction.....	146
II.	Intégration de l'outil dans le processus S&OP	147
III.	Le Cloud Computing	149
III.1.	Microsoft Azure.....	151
III.2.	Présentation du Blob Storage	152
III.3.	Azure Machine Learning	152
IV.	Description du pipeline retenu pour le portage de la solution.....	154
IV.1.	Etapas du pipeline.....	154
IV.2.	Implémentation	155
V.	Mise en œuvre et Résultats	158

VI. Conclusion	161
Conclusion Générale	163
Références.....	168
Annexe A : Secteur de la vente au détail	178
Annexe B : Secteur de la mode	180
Annexe C : Secteur de l'agroalimentaire.....	183
Annexe D : Secteur de la vente en ligne.....	187
Annexe E : Secteur automobile	190
Annexe F : Secteur automobile	194
Résumé	199

Liste des figures

Figure 1 : Étapes du processus S&OP (Wagner et al., 2014).....	22
Figure 2 : Gammes de produits d'elm leblanc pour les particuliers	25
Figure 3 : Étapes et acteurs de la Supply Chain d'elm.leblanc.....	27
Figure 4 : Processus autour de la méthode de prévision actuelle	28
Figure 5 : Exemple des horizons de prédiction à partir du mois d'Août.....	30
Figure 6 : Erreur pour chaque produit pour l'horizon H1	30
Figure 7 : Exemple du comportement de ventes d'un produit.....	32
Figure 8 : Exemple de corrélation des produits pour H1	32
Figure 9 : Catégories des techniques de prévisions.....	53
Figure 10 : Approche proposée par (Thomassey & Fiordaliso, 2006).....	61
Figure 11 : Approche multi-niveau proposée par (Pavlyshenko, 2019).....	61
Figure 12 : Méthodologie proposée par (Punam et al., 2019).....	62
Figure 13 : Approche proposée par (T. Liu et al., 2020).....	63
Figure 14 : Approche proposée par (Bandara et al., 2019).....	66
Figure 15 : Structure hybride parallèle (Hajirahimi & Khashei, 2019)	68
Figure 16 : Structure hybride en série (Hajirahimi & Khashei, 2019)	68
Figure 17 : Structure hybride parallèle-série (Hajirahimi & Khashei, 2019)	69
Figure 18 : Résumé des structures hybrides (d'après (Hajirahimi & Khashei, 2019)).....	70
Figure 19 : Approche proposée par (Xu et al., 2019)	72
Figure 20 : Méthodologie de sélection des structures de modèles pour chaque série temporelle	80
Figure 21 : Comparaisons des quantités prédites automatiquement et manuellement par rapport aux ventes réelles des exemples de produits pour l'horizon H3 de 2018.....	91
Figure 22 : Comparaisons des quantités prédites automatiquement et manuellement par rapport aux ventes réelles des exemples de produits pour l'horizon H3 de 2019.....	100
Figure 23 : Exemples de comportements de ventes réelles des produits modélisés par Prophet sur leurs cycles de vie représentant des variations semblables.....	107
Figure 24 : Exemples de comportements de ventes réelles des produits modélisés par ARIMA sur leurs cycles de vie représentant des variations semblables.....	108
Figure 25 : Exemples de comportements de ventes réelles des produits modélisés par un lissage exponentiel avec une tendance additive non amortie et une saisonnalité multiplicative sur leurs cycles de vie représentant des variations semblables	110
Figure 26 : Exemples de comportements de ventes réelles des produits modélisés par Prophet pour H1 sur leurs cycles de vie représentant des variations semblables.....	111
Figure 27 : Exemples de comportements de ventes réelles des produits modélisés par SARIMA pour H1 sur leurs cycles de vie représentant des variations semblables.....	112
Figure 28 : Exemples de comportements de ventes réelles des produits modélisés par un lissage exponentiel sans tendance avec une saisonnalité multiplicative sur leurs cycles de vie représentant des variations semblables	113
Figure 29 : Répartition des modèles pour l'horizon H1 de 2019	114
Figure 30 : Répartition des modèles pour l'horizon H3 pour 2019	116
Figure 31 : Exemples de comportement de ventes des produits utilisant Prophet pour les prédictions à H1 et H3.....	117
Figure 32 : Exemples de comportement de ventes des produits utilisant le lissage exponentiel pour les prédictions à H1 et H3	118

Figure 33 : : Comparaisons des quantités prédites par rapport aux ventes réelles des produits de Drancy pour l’horizon H3 de 2020/2021.....	125
Figure 34 : Approche de combinaison des modèles	132
Figure 35 : Exemple d’un réseau de neurones de type MLP	134
Figure 36 : Exemples d’architecture de réseau de neurones (13, 6, 3,1) testée dans notre méthodologie	135
Figure 37 : Répartition des données d’entraînement et de test	135
Figure 38 : Exemple d’un modèle à un seul horizon H3	138
Figure 39 : Exemple d’un modèle multi horizons	139
Figure 40 : Architecture retenue du réseau utilisée pour la combinaison (13, 6, 1)	141
Figure 41 : Erreurs réalisées par le modèle final en modifiant à chaque une variable d’entrée	143
Figure 42 : Erreurs réalisées par le modèle final en modifiant à chaque fois une variable d’entrée	144
Figure 43 : Nouveau processus de prévisions des ventes.....	148
Figure 44 : Modèles des services Cloud	149
Figure 45 : Exemples des services Cloud d’Azure	151
Figure 46 : Etapes de création d’un projet ML sur Azure ML	153
Figure 47 : Exemple d’un pipeline sur Azure ML	154
Figure 48 : Capture du pipeline implémenté sur Azure ML pour la prédiction des ventes	156
Figure 49 : Exemples pour quelques produits des erreurs absolues enregistrées pour chaque mois	160
Figure 50 : Méthodologie proposée par (Loureiro et al., 2018)	180
Figure 51 : Approche proposée par (F. L. Chen & Ou, 2009)	184
Figure 52 : Approche proposée par(Ji et al., 2019).....	188
Figure 53 : Approche d’hybridation proposée par(M. Li et al., 2018)	189
Figure 54 : Méthodologie proposée par (Subrmanian et al., 2020)	190
Figure 55 : Approche proposée par (Sa-Ngasoongsong et al., 2012)	191
Figure 56 : Approche proposée par (Shao & Tsai, 2018)	195
Figure 57 : Approche proposée par (Zhao et al., 2015)	197
Figure 58 : Techniques par catégorie (Bourdeau et al., 2019).....	198

Liste des tables

Tableau 1 : Synthèse des domaines d'applications	51
Tableau 2 : Critère de choix des modèles possibles	83
Tableau 3 : Paramètres des modèles de lissage exponentiel	86
Tableau 4 : Erreurs de prévisions pour différents produits sur les horizons H1 et H3 de 2018 selon la méthode employée (manuelle ou automatique)	93
Tableau 5 : Mesure d'erreur (RMSE) à l'entraînement pour un horizon H3 et choix du modèle par produit pour les prévisions au mois d'octobre 2018.....	97
Tableau 6 : Mesure d'erreur (RMSE) à l'entraînement pour un horizon H1 et choix du modèle par produit pour les prévisions au mois de décembre 2018	98
Tableau 7 : Erreurs de prévisions pour différents produits sur les horizons H1 et H3 de 2019 selon la méthode employé (manuelle ou automatique)	101
Tableau 8 : Mesure d'erreur (RMSE) à l'entraînement pour un horizon H3 et choix du modèle par produit pour les prévisions au mois d'octobre 2019.....	104
Tableau 9 : Mesure d'erreur (RMSE) à l'entraînement pour un horizon H1 et choix du modèle par produit pour les prévisions au mois de décembre 2019	106
Tableau 10 : Répartition des produits par modèle pour le moyen terme (H3)	107
Tableau 11 : Répartition des produits par type de lissage exponentiel pour H3	109
Tableau 12 : Répartition des produits par modèle pour le court terme H1	110
Tableau 13 : Répartition des produits par type de lissage exponentiel pour le court terme H1	112
Tableau 14 : Erreurs de prévisions pour différents produits sur les horizons H1 et H3 de 2020/2021 selon la méthode employé (manuelle ou automatique)	126
Tableau 15 : Mesure d'erreur (RMSE) à l'entraînement pour un horizon H3 et choix du modèle par produit pour les prévisions du mois de novembre 2020.....	127
Tableau 16 : Mesure d'erreur (RMSE) à l'entraînement pour un horizon H1 et choix du modèle par produit pour les prévisions du mois de janvier 2020	128
Tableau 17 : Description des architectures testées.....	134
Tableau 18 : Exemples des données d'entrée du réseau avec 13 entrées	137
Tableau 19 : Erreurs des architectures des réseaux testées selon l'horizon de prédiction.....	140
Tableau 20 : Erreurs de prédictions des modèles par rapport à la combinaison	142
Tableau 21 : Ventes et prévisions réalisées sur le S1 2021	159
Tableau 22 : Erreurs absolues enregistrées sur le premier semestre de 2021.....	161

Introduction Générale

Dans un monde économique caractérisé par une concurrence accrue, les entreprises cherchent toujours à renforcer leur compétitivité afin de s'imposer sur le marché. Pour atteindre cet objectif, plusieurs leviers de performance sont envisagés, parmi lesquels la maîtrise des processus de planification et de production ainsi que l'amélioration de la Supply Chain. Ces thématiques sont notamment au cœur de l'émergence des nouvelles techniques de l'industrie 4.0. Cette ère industrielle est caractérisée par une tendance à la numérisation de l'usine à travers le recours aux technologies des objets connectés et de l'intelligence artificielle.

Étant donné que le gain de performance requiert des technologies de pointe croisant plusieurs domaines (mathématiques, informatique, automatique, ...), les entreprises font appel aux compétences des spécialistes et des laboratoires de recherche pour réaliser ces projets et garantir des résultats satisfaisants. C'est dans ce contexte que le projet de recherche décrit dans ce manuscrit a été mené, au travers de la collaboration entre l'entreprise elm.leblanc du groupe BOSCH et l'Institut Mines-Télécom.

L'entreprise elm.leblanc est spécialisée dans la construction de chaudières et de chauffe-eaux individuels à gaz et est l'un des leaders du marché français avec un large catalogue de produits. La thèse s'intègre dans le champ de l'amélioration des performances de cette entreprise, et vise à mettre en place une solution innovante pour améliorer la performance de sa Supply Chain. Différents apports peuvent être envisagés dans de nombreux processus de l'entreprise allant de la prévision des ventes jusqu'à l'optimisation de la livraison des clients en passant par la planification de la production. Vu la diversité des sujets qui peuvent être traités pour améliorer la Supply Chain de l'entreprise, ce projet s'est déroulé en plusieurs étapes. Après une étude de l'état existant de l'entreprise, dans le but de cerner la problématique industrielle, le projet de thèse s'est orienté vers la prévision des ventes, cela dans l'objectif d'obtenir un impact significatif sur l'amélioration de sa performance.

Ce projet de thèse concerne la mise en place d'un outil d'aide à la décision pour aider l'entreprise à mieux gérer la planification de ses processus de production, de stockage et de réapprovisionnement des matières premières, à travers une meilleure prévision des ventes

sachant que son impact peut atteindre l'ensemble de ces processus. Plus particulièrement la contribution de thèse porte sur la mise en place d'une méthodologie de prévision des ventes originale adaptée à la variété du catalogue produits d'elm.leblanc.

Les données collectées et nettoyées depuis les différents services de l'entreprise sont représentées principalement par l'ensemble des historiques de ventes des différents produits. De ce fait, notre approche proposée s'appuie sur ces historiques pour fournir un outil d'aide à la décision basé sur des techniques d'analyse de données et des séries temporelles. Ce nouvel outil de prévision vise à transformer le processus interne autour de la prévision du besoin et de la planification de la production.

Plusieurs techniques et approches ont été étudiées dans la revue de la littérature pour comprendre leurs cas d'usage et leurs spécificités d'application. Nous nous sommes particulièrement intéressés aux modèles de séries temporelles parmi lesquels plusieurs ont été étudiés et sélectionnés pour être mis en œuvre selon deux stratégies. La première portant sur une compétition entre les modèles et la deuxième sur leur coopération.

Une dimension importante du travail est sa réalisation dans un contexte de recherche industrielle. En effet, bien que ce travail soit mené dans un cadre de recherche scientifique, il a porté sur l'implémentation de la méthodologie proposée dans un environnement réaliste. Ceci se manifeste par la conception et le déploiement d'une solution industrielle qui répond aux exigences de l'usage professionnel d'un outil industriel, à savoir une facilité d'usage et de maintenance, une pérennité dans le temps ainsi que des bons résultats. L'outil proposé représente un nouveau service intégré aux processus de planification de l'entreprise.

Le manuscrit présente ainsi l'ensemble du projet mené pendant 3 années en tant que salarié ingénieur doctorant de l'entreprise elm.leblanc. Les développements scientifiques, méthodologiques et industriels y sont présentés au travers de 5 chapitres organisés comme suit.

Le premier chapitre présente une étude bibliographique des nouvelles tendances de la Supply Chain prédictive et du processus S&OP. Cette étude permet le positionnement des travaux réalisés par rapport aux avancées scientifiques et méthodologiques du domaine de l'analyse prédictive dans la Supply Chain. L'étude du processus S&OP constitue quant à elle le cadre d'intégration de l'outil développé dans une démarche d'amélioration des processus de

planification de l'entreprise. Ces deux contextes permettent donc de spécifier le positionnement des travaux dans un cadre industriel d'application et ainsi la définition de la problématique industrielle à laquelle répond ces travaux.

La réponse à la problématique identifiée passe d'abord par une recherche bibliographique, qui sera l'objet du deuxième chapitre. Une revue de la littérature présentera les principaux domaines d'application des techniques des prévisions de vente en insistant sur leurs spécificités. Ensuite plusieurs approches et méthodologies employées dans la littérature seront présentées afin de détailler cet état de l'art et d'identifier les enjeux et les contraintes liés au processus de prévision.

Le troisième chapitre présente notre première contribution autour d'une nouvelle méthodologie dédiée à la prévision des ventes portant sur la compétition entre différents modèles sélectionnés pour répondre à un large catalogue des produits de l'entreprise. D'un point de vue scientifique, nos recherches seront concentrées principalement sur les approches guidées par les données et les méthodes de modélisation des séries temporelles. Sur la base des limites actuelles des travaux de recherche et du besoin exprimé autour de l'amélioration des processus de planification de l'entreprise elm.leblanc, nous proposerons une nouvelle méthodologie de prévision des ventes, solution originale qui pourra également être reprise par des professionnels de tous domaines ayant des cas d'études similaires.

Le quatrième chapitre présente une deuxième contribution scientifique permettant de surpasser certaines limites de l'approche originale, dans le but de fournir des informations plus précises et supplémentaires pour la prédiction des ventes. Cette deuxième approche proposée porte sur une idée de coopération des modèles de prédiction. Nous détaillons dans ce chapitre cette méthodologie, et nous utilisons les données de l'entreprise elm.leblanc pour conduire plusieurs expérimentations afin de la valider. Nous présentons ainsi les résultats, les hypothèses d'implémentation et les limites de cette proposition.

Le cinquième et dernier chapitre porte sur l'industrialisation des nouvelles méthodes proposées. Étant dans un cadre industriel, ces recherches visent à mettre en place un outil d'aide à la décision pour l'entreprise afin de renforcer le processus de planification de la demande dans sa Supply Chain. Ce contexte implique l'élaboration d'un livrable qui sera exploité par l'entreprise, et qui doit donc répondre à certains critères de mise en production.

Les travaux menés durant ce projet ont permis de proposer deux méthodologies pour répondre à la problématique de prévisions des ventes dans un contexte industriel. Sur le plan scientifique, nous avons proposé une première approche permettant l'estimation des prévisions des ventes en se basant sur l'historique des données disponibles et en employant des techniques de séries temporelles. La deuxième approche représente une solution pour faire face à une limite rencontrée avec la méthodologie originale et constitue une contribution basée sur une idée de coopération des modèles de prévisions.

Sur le plan industriel, notre outil développé permet de répondre au besoin de l'entreprise elm.leblanc en proposant un outil d'aide à la décision intégré dans le processus S&OP pour améliorer le processus de planification de la demande. La solution proposée permet d'avoir un outil avec une certaine pérennité d'usage et facilité d'exploitation pour les utilisateurs finaux. Cet outil peut être aussi exploité par d'autres cas d'usage ayant les mêmes conditions d'application que celui traité dans ce manuscrit.

Chapitre 1 : Positionnement et enjeux de la recherche

I. Introduction

L'entreprise elm.leblanc représente en France la division Thermo-technologie (TTFR) du groupe Bosch. Elle développe et fabrique un large choix de solutions thermiques pour répondre aux besoins du marché à travers ses deux sites de production de Drancy et Saint-Thégonnec. Depuis sa création, elm.leblanc suit une stratégie d'amélioration continue de ses produits en intégrant de plus en plus de nouvelles technologies, mais aussi de ses processus de production, notamment en modernisant sa chaîne logistique et ses lignes de montage.

Dans le but de pousser l'efficacité de leurs sites de production, l'entreprise a mis en place une stratégie de développement de ses activités de recherche et d'innovation autant pour ses produits que pour ses processus de planification et de production, ainsi que pour les compétences des personnels. Plusieurs projets sont lancés dans l'ère de l'industrie 4.0 sur différentes thématiques allant des approches guidées par les données jusqu'aux lignes de production reconfigurables en passant par la réalité augmentée et l'opérateur du futur.

La digitalisation est désormais au cœur de la transformation de l'outil industriel d'elm.leblanc à travers différents paliers. On cite l'exemple de la mise en place des équipements de fabrication connectés permettant le suivi des processus industriels et leurs résultats via des écrans de suivi des indicateurs en temps réel favorisant ainsi la transparence au sein de l'usine. L'entreprise est entrée depuis plusieurs années dans une démarche de valorisation de ses données par la mise en place d'un système de gestion de bases de données rassemblant données de logistique/livraison, de production, de suivi, de SAV, etc. Ce système sera le support au développement des approches basées sur l'intelligence artificielle et l'apprentissage automatique afin d'augmenter la fiabilité et la robustesse des processus de la Supply Chain. L'entreprise utilise également des robots collaboratifs pour le déplacement des palettes des produits finis en coopérant avec les opérateurs des lignes et un robot dans une ligne de production pour réaliser des tâches difficiles et répétitives en cohabitant avec les opérateurs de la ligne. Enfin, la réalité augmentée est employée pour la modélisation des postes afin d'aider les opérateurs à se former d'une manière plus efficace et plus rapide et à assurer leurs tâches avec plus de facilité et de flexibilité.

La stratégie d'excellence industrielle mise en place par elm.leblanc et l'amélioration constante de la maîtrise des savoir-faire de l'entreprise constituent un cadre idéal pour l'application des travaux de recherche qui seront développés dans ce manuscrit.

En écho aux enjeux industriels, nous allons d'abord présenter le contexte scientifique dans lequel s'inscrit la conduite de ce projet de transformation numérique, consistant à exploiter davantage encore l'Intelligence Artificielle et l'analyse prédictive dans les processus de décision associés à la planification de la production et de la Supply Chain.

Après cet état des lieux, nous détaillerons les attentes de l'entreprise elm.leblanc en termes d'amélioration de la performance tant sur le plan de la satisfaction client et que de l'efficacité industrielle. Avec une courte présentation du catalogue produits et des différentes bases de données, une description des processus existants de prévisions dans l'entreprise sera donnée. Cette description est complétée par une étude sur l'état actuel de la performance des prévisions obtenues, l'importance de celles-ci dans le bon fonctionnement de l'entreprise et parfois leurs limites. À l'issue de la présentation des résultats de cette étude, la dernière partie de ce chapitre sera consacrée à la présentation de la problématique traitée dans ce manuscrit de thèse.

II. Positionnement scientifique : Analyse prédictive et Supply Chain

Dans la culture d'entreprise, le terme Supply Chain Management (SCM) est largement associé à la gestion de la logistique. (Christopher Martin, 2011) dans son livre, explique que ce terme est plus large et couvre plusieurs aspects de planification et de coordination entre les différentes parties prenantes, du fournisseur jusqu'au client. Il définit la Supply Chain Management comme « *La gestion des relations en amont et en aval avec les fournisseurs et les clients afin d'offrir une valeur supérieure au client à moindre coût pour l'ensemble de la Supply Chain.* ». Ce concept a été largement étudié par les spécialistes et est employé par les professionnels dans le but d'améliorer les performances des entreprises. En tenant compte des évolutions technologiques, plusieurs techniques récentes sont employées aujourd'hui pour atteindre cet objectif avec plus d'efficacité.

II.1. Analyse prédictive de la Supply Chain

Des travaux récents sur le SCM s'intéressent à l'intégration de nouvelles technologies liées à l'intelligence artificielle telles que la science des données, le Big Data, l'analyse prédictive, etc.

dans le développement et la conception de ses processus. Ces tendances ouvrent des nouvelles opportunités aux professionnels. (Waller & Fawcett, 2013) suggèrent différentes définitions et concepts à cet effet, à savoir l'analyse prédictive de la SCM et la science des données pour la SCM. Ce dernier est défini comme « ... *l'application des méthodes quantitatives et qualitatives provenant de diverses disciplines en combinaison avec la théorie de la SCM pour résoudre ses problèmes et fournir des résultats pertinents, en tenant compte de la qualité et la disponibilité des données.* ». Ainsi, ils ont appelé au développement des activités de recherche sur ces sujets, vue la rareté des travaux.

L'analyse des données massives (Big Data Analytics) est définie comme l'application des techniques analytiques avancées, notamment la fouille des données, les analyses statistiques et prédictives sur une grande quantité de données (Tiwari et al., 2018). D'après l'enquête de (Schoenherr & Speier-pero, 2015), plusieurs avantages sont générés par l'application de l'analyse prédictive à la SCM. Ces techniques offrent des opportunités aux entreprises qui visent à accroître l'efficacité de leur Supply Chain, à améliorer leurs stratégies de gestion des risques et à améliorer les processus de prise de décision.

Pendant longtemps, la SCM a utilisé les statistiques et la recherche opérationnelle pour optimiser les processus visant à équilibrer l'offre et la demande (Tiwari et al., 2018). Actuellement, la science des données, l'analyse prédictive et le Big data, désignés par l'acronyme DPB (Waller & Fawcett, 2013), offrent plus de possibilités pour créer des outils d'aide à la décision basés sur les données afin d'améliorer la profitabilité de la Supply Chain. C'est pourquoi il existe une demande croissante pour les professionnels ayant ces compétences, étant donné que de tels outils permettent d'améliorer la visibilité et les bénéfices sur toute la Supply Chain, de diriger la stratégie de l'entreprise, et de renforcer l'efficacité des opérations et la performance financière. Les DPB peuvent être un élément stratégique complémentaire qui permet la mise en place d'aperçus prédictifs avancés dans les processus d'exécution de la stratégie de l'entreprise (G. Wang et al., 2016). C'est un domaine qui évolue rapidement pour répondre à différentes problématiques industrielles.

Cependant, (Schoenherr & Speier-pero, 2015) soulignent les barrières qui peuvent survenir et limitent l'utilisation et l'exploitation de ces techniques. Selon leur enquête, les difficultés et problèmes les plus importants concernent principalement le manque des données, et l'incapacité à identifier les données les plus appropriées à utiliser pour effectuer les analyses.

D'autres obstacles sont également identifiés, tels que les problèmes de sécurité et le manque de soutien de la direction d'une entreprise, voir les problèmes d'interopérabilité entre les différents systèmes d'information.

Un autre concept très important et qui peut également être considéré comme un obstacle, est la qualité des données. Évidemment, les informations et les décisions dérivées de l'analyse des données sont aussi bonnes que les données sur lesquelles elles reposent. (Hazen et al., 2014) étudient le rôle de la qualité des données dans l'exploitation des techniques de DPB dans le contexte de SCM. Ils insistent sur le fait que la surveillance et le contrôle de la qualité des données sont une priorité face à la complexité croissante de ces dernières (Kache et al., 2017). Pour faire face à ce problème, (Hazen et al., 2014) proposent des méthodes de suivi et de contrôle de la qualité des données.

Au-delà de disposer de données suffisantes et de qualité, l'amélioration de l'efficacité du SCM passe aussi par des processus normalisés et permettant de coordonner l'ensemble des activités des différents services concernés, par la collecte de ces données, leur analyse et la mise en place des actions qui découlent des processus de décision guidés par les données.

II.2. Sales and Operations Planning (S&OP)

Depuis son apparition, le S&OP a toujours été un processus pilote pour la Supply Chain. De nos jours, il a été largement étudié dans la littérature et mis en œuvre par plusieurs entreprises. Il est considéré comme le processus de planification clé dans la Supply Chain (L. Lapede, 2011). Il s'agit d'un processus intégral qui vise à équilibrer l'offre et la demande et à faire le lien entre les plans stratégique et opérationnel de l'entreprise (Thome et al., 2012a). Ces objectifs sont réalisables par le déploiement d'un alignement vertical entre les niveaux de planification (stratégique, tactique, opérationnel) (Thome et al., 2012b) ainsi que par un alignement horizontal ou inter fonctionnel entre les différentes fonctions de l'entreprise (Kathuria et al., 2007; Thome et al., 2012a). La plupart des experts dans ce domaine d'étude placent le rôle du processus S&OP à un niveau de planification tactique sur une période de moins de 3 mois à 18 mois (Grimson & Pyke, 2007; Thome et al., 2012b).

Ce processus comprend généralement cinq étapes (Grimson & Pyke, 2007; Thome et al., 2012a; Wagner et al., 2014; Wallace & Stahl, 2008) comme le montre la figure 1.

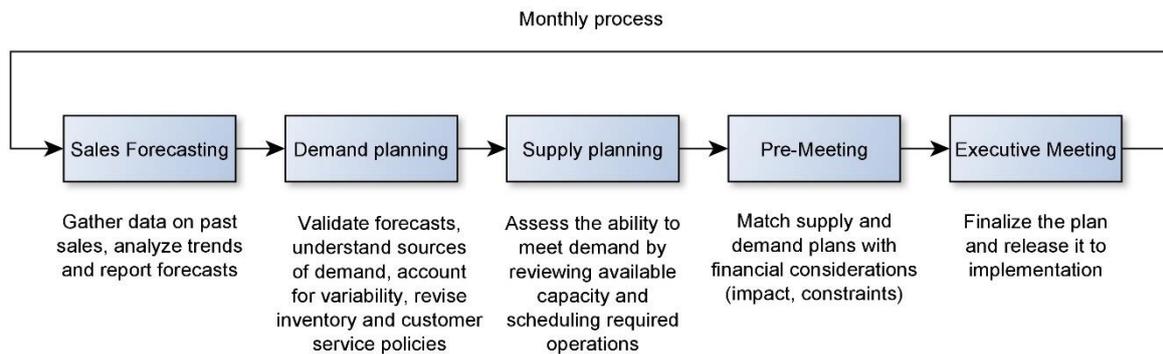


Figure 1 : Étapes du processus S&OP (Wagner et al., 2014)

La première étape consiste à rassembler les données historiques et à les utiliser pour préparer un premier plan prévisionnel de ventes basé sur les techniques statistiques, sans la contrainte de la capacité de production de l'usine (Grimson & Pyke, 2007). (Wagner et al., 2014) ont ajouté à cela le fait de générer des rapports de performances pour évaluer les événements passés. La planification de la demande est la deuxième étape et permet de préparer le plan de la demande sur la base des prévisions prévues à l'étape précédente. La troisième étape consiste à définir le plan d'approvisionnement et de production, en connaissant les données caractérisant les capacités de l'usine, les stocks de composants et le plan de demande préalablement défini. La « pré-réunion » correspond à la quatrième étape au cours de laquelle les différents membres représentant leurs services dans l'entreprise se réunissent pour ajuster et valider les plans qui ont été définis. Le recours à l'IA comme agent virtuel participant à cette délibération est évidemment une perspective à moyen terme, dès lors que l'automatisation des tâches de collecte, de traitement, de prédiction et de décision aura été garantie. La dernière étape consiste à approuver ce qui a été décidé lors de la pré-réunion, à gérer les exceptions qui peuvent survenir et à exécuter les plans après un consensus entre les parties prenantes.

Lors de la mise en place d'un processus S&OP, plusieurs éléments doivent être pris en considération. Dans son article, (B. L. Lapide, 2004a) présente les facteurs de succès d'un processus S&OP idéal comme la routine des réunions structurées, la collaboration et la coopération entre les différentes fonctions de l'entreprise ainsi que l'utilisation d'outils technologiques de planification appropriés. Un autre point à prendre en considération, qui a été mentionné dans les travaux récents de (Kristensen & Jonsson, 2018), est que les variables contextuelles d'application du processus S&OP (comme le type d'industrie, la taille de

l'entreprise, la stratégie de production, etc...) influencent sa conception et ses performances. (B. L. Lapide, 2004b) met l'accent sur l'utilisation des outils technologiques appropriés pour chaque étape du processus et propose une architecture technologique de planification intégrée entre l'offre et la demande pour bien implémenter le processus.

Le processus S&OP a un impact positif sur la performance de l'entreprise, permettant d'atteindre l'excellence opérationnelle (B. L. Lapide, 2004a) et une Supply Chain intégrée (Thome et al., 2012b). Plusieurs bénéfices ont été identifiés par (Wagner et al., 2014) à travers leur enquête menée auprès des professionnels et experts en gestion des opérations : amélioration de la précision des prédictions, amélioration de la visibilité de la Supply Chain et réduction des risques liés à la perturbation de la Supply Chain, réduction des stocks, etc. Ce processus permet de renforcer la coopération entre les collaborateurs de l'entreprise et d'établir un alignement horizontal entre les services (Grimson & Pyke, 2007; Wagner et al., 2014). Dans ce contexte, (Tuomikangas & Kaipia, 2014) ont proposé huit mécanismes qui constituent un cadre conceptuel permettant de pousser le processus S&OP en termes de coordination comme étant un facteur de réussite de ce processus.

Afin de faire un diagnostic de l'état actuel d'une entreprise et de définir l'état souhaité à terme, plusieurs modèles de maturité ont été proposés dans la littérature (Cecere et al., 2009; Grimson & Pyke, 2007; B. L. Lapide, 2005; Wagner et al., 2014). Chaque modèle offre plusieurs niveaux de maturité du processus S&OP dans les entreprises selon différents critères. Cette variété ne permet pas d'avoir une base commune pratique pour mettre en place un tel processus (Thome et al., 2012a), alors les travaux de (Wagner et al., 2014) tentent de synthétiser les modèles précédents et de développer un cadre plus détaillé enrichi de leurs investigations réalisées par des professionnels.

II.3. L'analyse prédictive de la Supply Chain et le processus S&OP

Dans ce paragraphe, nous présentons l'intérêt d'associer le processus S&OP avec une analyse prédictive, notamment autour de la prévision des ventes.

Un aspect clé du processus S&OP est l'équilibre entre l'offre et la demande. Ce rend nécessaire de faire face à la variabilité de la demande. Selon (Wallace & Stahl, 2008), l'étape de planification de la demande peut être la plus difficile des cinq étapes du S&OP pour de nombreuses entreprises. (G. Wang et al., 2016) soulignent plusieurs indicateurs pour l'analyse

de la Supply Chain, parmi lesquelles l'analyse et la mesure de la performance de la planification de la demande. Comme nous l'avons mentionné précédemment, cette étape est précédée de la prévision des ventes dans laquelle des techniques statistiques peuvent être utilisées. (L. Lapide, 2011) insiste sur l'importance des prévisions de la demande pour le bon déroulement du processus S&OP puisque les plans générés ensuite seront pilotés par cette étape. Par conséquent, une planification efficace des capacités nécessite une planification précise de la demande (G. Wang et al., 2016).

(G. Wang et al., 2016) affirment qu'il existe deux techniques quantitatives largement utilisées dans l'analyse de la Supply Chain qui sont « l'analyse et la prévision de séries temporelles » et « l'analyse de régression ». Le premier vise à déterminer les modèles existants dans les données historiques et à prédire les événements futurs sur cette base, et le second vise à comprendre les relations et la causalité entre les variables influant la série. Comme le souligne (G. Wang et al., 2016), l'utilisation des techniques de prévision et d'optimisation pour soutenir le processus S&OP fournit une capacité de gestion inter-fonctionnelle entre les différentes parties de la Supply Chain.

III. Définition de la problématique associant Analyse prédictive et Supply Chain dans le contexte de l'entreprise elm.leblanc

III.1. Le contexte et les enjeux pour l'entreprise

L'entreprise elm.leblanc est spécialisée dans la fabrication des systèmes de chauffage résidentiels et industriels. Le métier de l'entreprise repose sur une expertise dans la conception de ses produits à travers ses bureaux d'études, leur production grâce à ses lignes de montages modernisées, et leur maintenance par l'intermédiaire des équipes de techniciens compétents, pour fournir des produits plus efficaces répondant aux besoins du marché. Le site de Drancy dispose de plusieurs lignes d'assemblage spécifiques à chaque famille de produits permettant de produire près de 100 000 chaudières chaque année. Il dispose également d'une zone de stockage optimale accessible par un robot permettant une gestion efficace d'un grand nombre des composants approvisionnés. Le déplacement des produits finis à la fin des lignes de montage se fait à travers des véhicules autonomes sans interventions humaines.

Pour l'usine de Drancy, on trouve principalement trois catégories de produits : les produits finis, les pièces détachées et les accessoires vendus généralement en kits avec les produits finis. Ces derniers correspondent surtout à des chaudières et des chauffe-bains. La figure 2 présente les gammes des produits proposées par l'entreprise elm.leblanc.

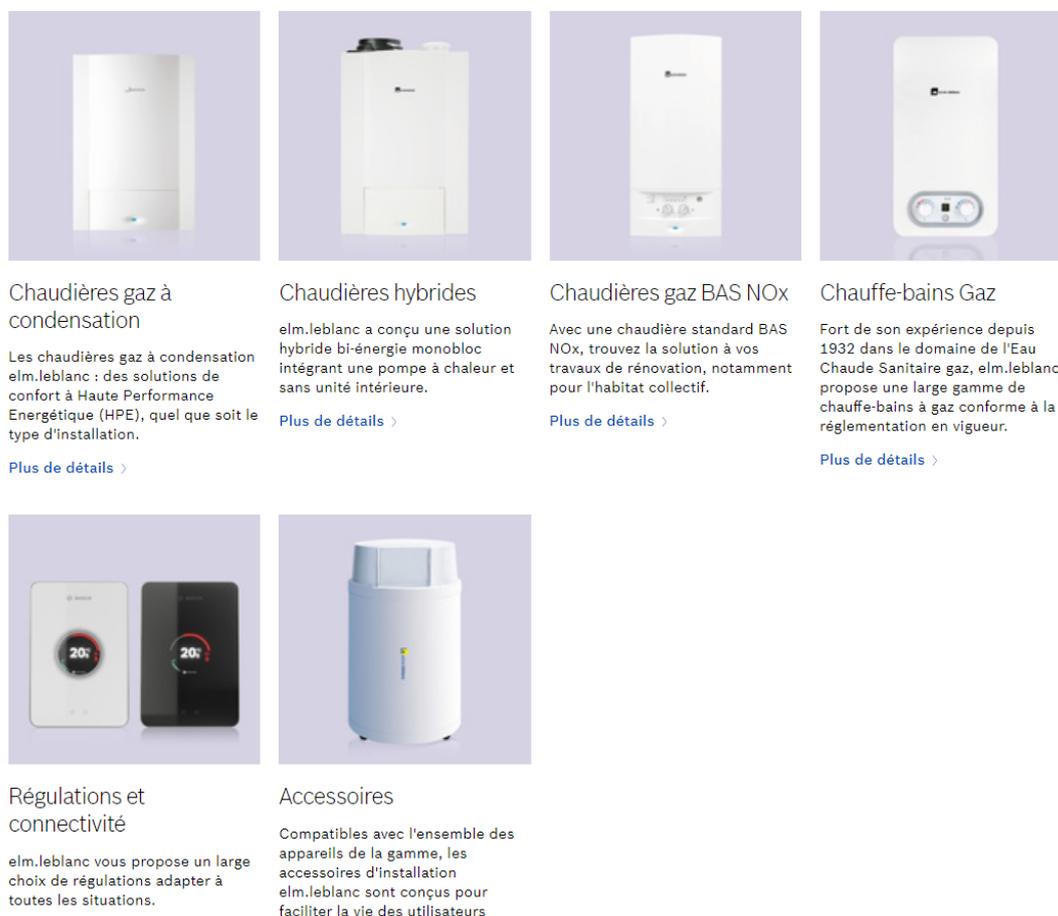


Figure 2 : Gammes de produits d'elm leblanc pour les particuliers

Dans chacune des gammes de produits se trouvent plusieurs références qui correspondent à des produits avec différentes caractéristiques comme la technologie employée pour le chauffage ou encore la variation des puissances. La plupart des produits qui sont fabriqués sur le site de Drancy sont destinés au marché français. Pour faciliter la livraison sur tout le territoire, elm.leblanc dispose d'un entrepôt de stockage des produits finis à Montbeugny (département de l'Allier). Le reste des produits destinés à l'export est fabriqué à la commande et expédié directement depuis l'usine. Pour le marché local, l'entreprise a une stratégie de make-to-stock visant à garantir un délai de livraison d'une semaine pour ses clients. La production assure également une couverture des stocks de 3 semaines pour un nombre important de produits en cas de rupture de matières premières ou des défauts de qualité des

fournisseurs. Toutefois, on se trouve parfois avec des ruptures de stocks ou un surstockage important conduisant à la non satisfaction des commandes clients ou à un coût supplémentaire de stockage.

Le marché d'elm.leblanc se décompose principalement en deux types : les chantiers qui sont caractérisés par des commandes généralement importantes et qui ont une nature plutôt aléatoire puisqu'ils se sont établis suite à un appel d'offre, et les ventes chez les grossistes et les installateurs qui sont plus stables et régulières. Pour répondre aux besoins de ces deux types, l'entreprise doit avoir les quantités nécessaires des bonnes références et la capacité de pouvoir faire des livraisons dans les délais attendus. Bien que le marché des grossistes et installateurs soit plus stable, la satisfaction de leurs demandes représente un enjeu important pour l'entreprise si l'on considère la variété des produits demandés ainsi que les délais nécessaires pour la fabrication et la livraison. Les chantiers représentent un challenge encore plus considérable étant donné le manque de visibilité pour ce type de demande et les délais parfois assez courts de livraison.

Comme les deux types de marchés diffèrent de par leurs caractéristiques, les références de produits diffèrent également dans leur comportement de vente en termes de quantité et stabilité des ventes. On se trouve avec des produits ayant une forte saisonnalité d'une année sur l'autre, tandis que d'autres ont des ventes moins stables avec plus de variabilité. Certaines références sont caractérisées par une augmentation importante des quantités d'une année sur l'autre, alors que d'autres restent dans le même intervalle depuis leur lancement.

Plusieurs contraintes représentent des défis à relever par l'entreprise pour répondre à la demande des clients. Ces contraintes sont liées aux différentes étapes de la Supply Chain comme les délais d'approvisionnement des matières premières, les coûts de stockage et la qualité de ces derniers, la disponibilité des ressources et des équipements de production, les coûts de livraison et de stockage des produits finis, etc. Afin d'encore mieux satisfaire ces contraintes et réduire davantage ses coûts, elm.leblanc dispose d'une stratégie globale autour de la digitalisation de ses processus et du suivi des indicateurs clés. Cette stratégie vise à optimiser les processus de production ainsi que le respect des délais de livraison des commandes avec la bonne quantité qui est l'un des indicateurs clés les plus importants pour refléter les performances globales de la Supply Chain.

III.2. La SC et les processus de production et de planification

Dans l'objectif de produire les quantités nécessaires et respecter les délais de livraison, plusieurs processus sont organisés dans la Supply Chain grâce à différents acteurs. Le schéma de la figure 3 représente les principales composantes de la SC d'elm.leblanc.

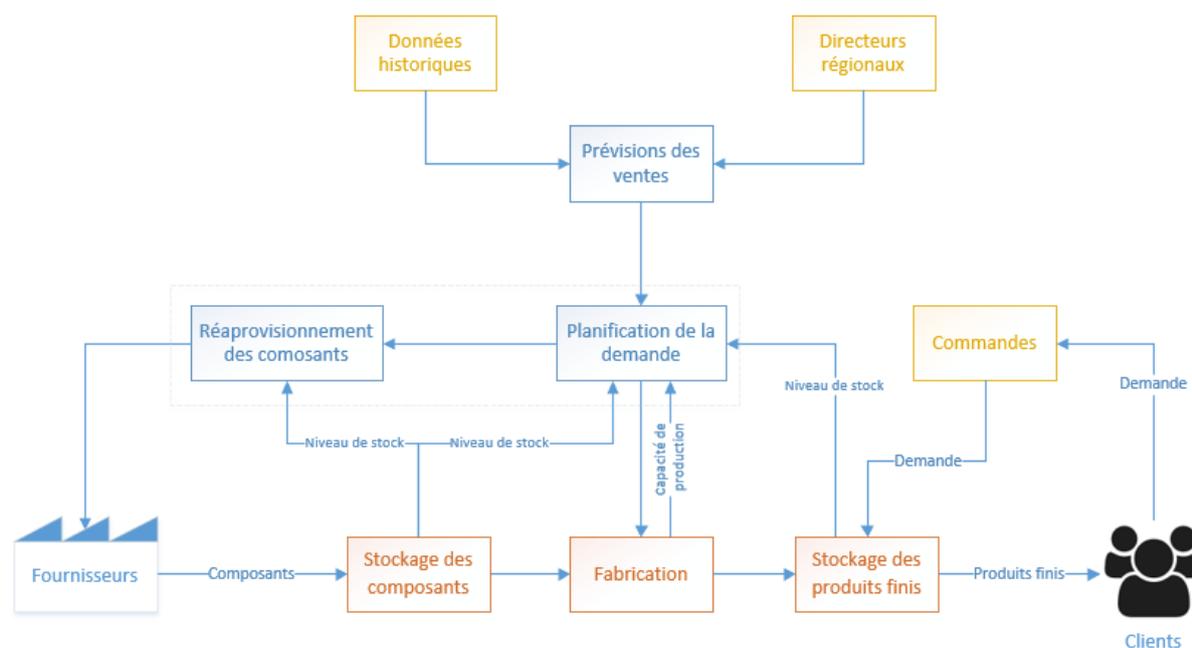


Figure 3 : Étapes et acteurs de la Supply Chain d'elm.leblanc

La première étape consiste à préparer un plan de prévisions des ventes en se basant sur les remontées des directeurs commerciaux régionaux et les données historiques des ventes. Ce processus représente l'étape la plus importante à partir de laquelle découle le reste des plans. La deuxième étape est la planification de la production en prenant en considération les ressources et les équipements disponibles. Ce plan permet le réapprovisionnement des matières premières auprès des fournisseurs en prenant en considération les niveaux des stocks disponibles dans le magasin. L'étape de fabrication permet l'exécution du plan de production défini pour ensuite envoyer les produits finis au dépôt de stockage. Les commandes clients (France) sont préparées et envoyées directement à partir de cet entrepôt. Comme mentionné plus haut, l'étape des prévisions des ventes est déterminante pour la bonne préparation des étapes de planification. Des prévisions non précises impliquent un plan de production non adéquat avec les attentes du marché entraînant soit un surstockage ou au contraire une rupture des produits finis et des composants, donc des pertes financières et des coûts supplémentaires pour l'entreprise. Par conséquent, l'anticipation des ordres de

fabrication avec un prévisionnel des bons de commande constitue un facteur d'amélioration considérable dans l'objectif d'optimiser les performances globales de la Supply Chain.

III.3. Zoom sur le processus de prévisions des ventes

Ce processus est décrit tel qu'il était au démarrage du projet. Le processus de prévision des ventes est piloté par un prévisionniste qui entre en interaction principalement avec les directeurs régionaux sur le terrain pour récupérer les données liées aux demandes de leurs clients. La personne en charge de ces prévisions interagit également avec le service marketing pour obtenir les informations liées aux nouveaux produits lancés ou encore les produits remplaçant ceux qui ont été arrêtés et les promotions qui sont mises en place. La présentation de l'organisation actuelle ci-après est complétée par une analyse de sa performance.

III.3.1. Le processus de prévision de ventes

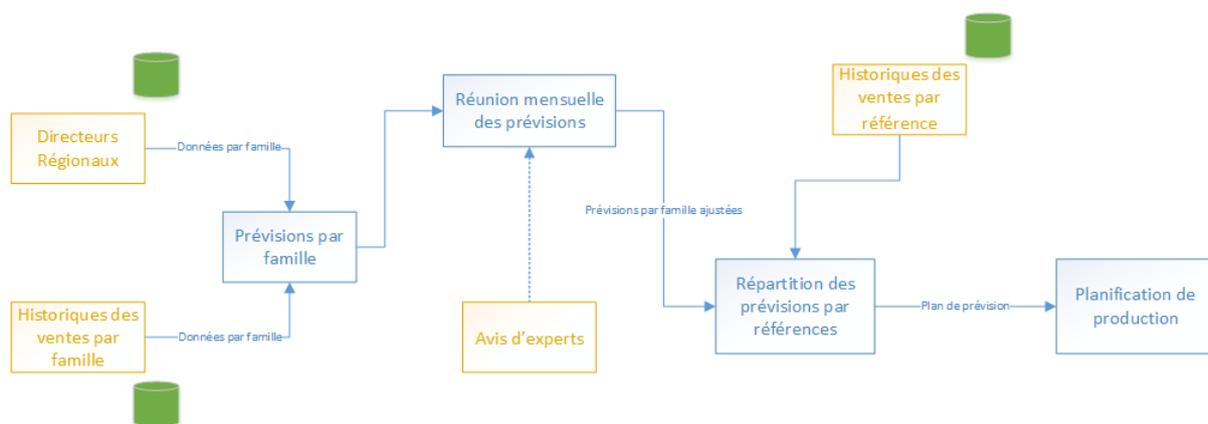


Figure 4 : Processus autour de la méthode de prévision actuelle

Le processus de prévision de ventes se fait mensuellement sur deux étapes. D'abord, les directeurs régionaux renvoient leurs prévisions de 6 mois par famille de produits au responsable des prévisions, qui s'occupe de les rassembler et de les affiner, selon son expertise et en regardant les ventes de l'année N-1. Ces prévisions par famille sont discutées ensuite par les responsables des autres services (marketing, logistique et production) pour les ajuster durant la réunion mensuelle en prenant en considération les informations comme les promotions qui sont en cours ou les campagnes publicitaires mises en place. L'analyste des prévisions prépare par la suite les prévisions par référence en partageant les quantités par famille et en regardant les ventes de l'année N-1 pour chaque référence. Finalement, ces

prévisions seront envoyées aux responsables sur le site de production pour la planification des ordres de fabrication et l'approvisionnement en matières premières.

III.3.2. Les données utilisées pour la prévision et les résultats actuels

Comme mentionné ci-dessus, les données employées pour préparer les prévisions sont issues de plusieurs sources. Les informations des directeurs régionaux sont principalement des bons de commandes des clients, l'historique des ventes mensuelles est issu du système de gestion des ventes de l'entreprise et les informations du service marketing correspondent aux promotions effectuées sur certains produits, des périodes de campagnes publicitaires ou encore sur l'évolution du catalogue produits (nouveautés lancées, suppression de références).

Ces données doivent être utilisées et traitées selon l'ensemble des références de produits du catalogue afin de préparer les prévisions par catégorie et par produit pour différents horizons de prédiction. Ce choix de préparation de deux prévisions est justifié d'abord par le besoin du service de production à connaître les quantités par référence pour pouvoir planifier les lignes d'assemblage pour chaque produit, et ensuite par l'appartenance des produits à des familles différentes du service des ventes à celles du service de production. Les prévisions par catégorie de produit peuvent être utilisées également pour définir les objectifs commerciaux pour chaque région par la direction commerciale.

Au démarrage du projet en 2019, une étude a été menée en interne afin d'analyser la performance du processus de prévision. Cette analyse a consisté à comparer les ventes réalisées avec les données de prévisions des produits représentant le plus grand volume des ventes qui correspondent aux chaudières murales assemblées à l'usine de Drancy et vendues sur le marché français durant les années 2017 et 2018. L'analyse des performances des prévisions actuelles est réalisée selon quatre horizons de temps (Hi). La notion d'horizon de prédiction est définie par le nombre de mois qui suit le dernier mois M ayant des ventes connues : l'horizon 1 (H1) correspond à une prédiction faite pour le mois M+1 lors du mois M, l'horizon 2 (H2) correspond à deux prédictions faites pour le mois M+1 et M+2 lors du mois M-2, etc.



Figure 5 : Exemple des horizons de prédiction à partir du mois d’Août

L’erreur réalisée pour chaque produit, c’est-à-dire l’écart, en valeur absolue, entre la vente réelle et la valeur prédite a été calculée pour différents horizons de prédiction. Nous avons normalisé les erreurs pour pouvoir comparer les produits avec des quantités différentes. Ensuite, nous avons analysé les 20 % des produits les mieux prédits et les 20 % des produits les moins bien prédits comme le montre l’histogramme ci-dessous. La figure 6 présente les erreurs constatées pour chaque produit pour l’horizon H1, des produits anonymisés de l’usine de Drancy pour garder la confidentialité des données.

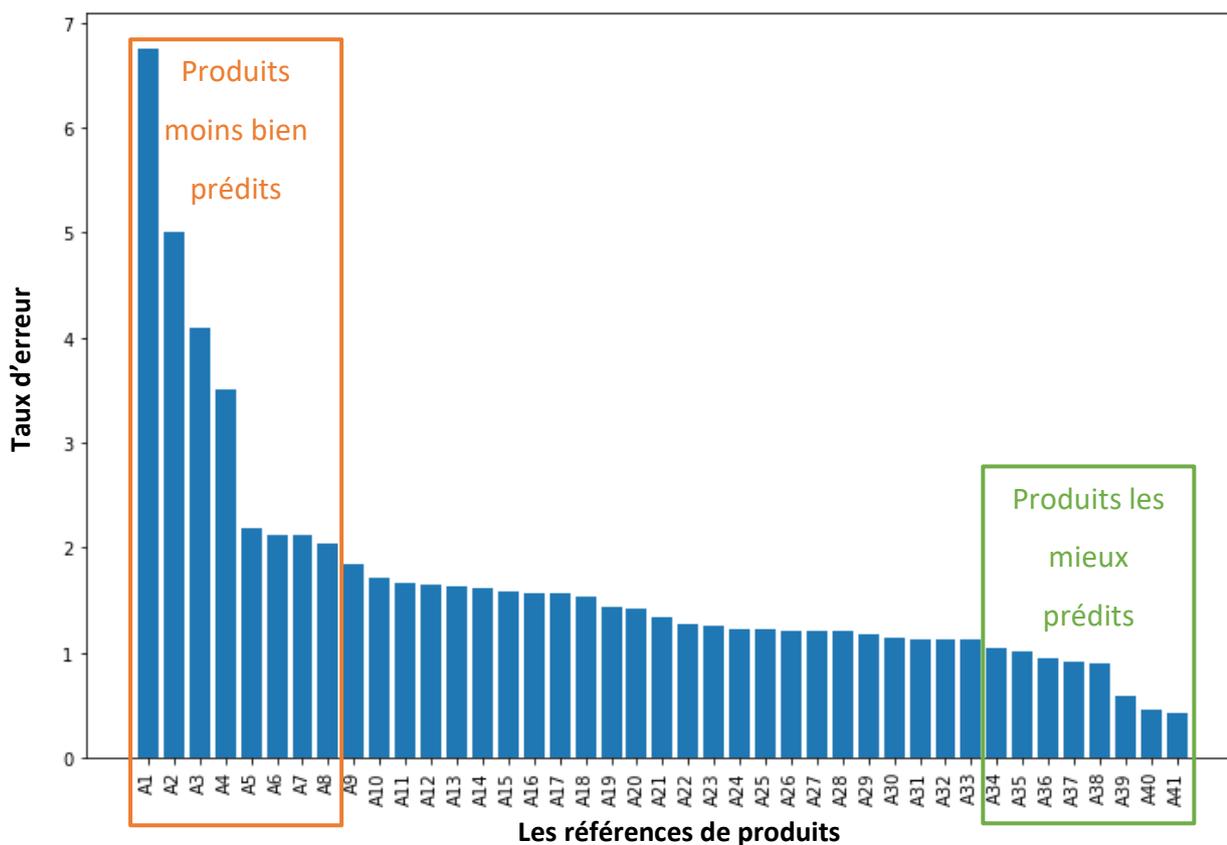


Figure 6 : Erreur pour chaque produit pour l’horizon H1

Nous avons également étudié le nombre d'occurrences des produits bien et mal prédits pour les différents horizons :

- Dans les 20 % des produits bien prédits, se trouvent quelques produits qui se répètent dans tous les horizons et donc sont généralement bien prédits, mais on trouve également certains produits qui apparaissent seulement à l'horizon H1 alors leurs prévisions ont été ajustées tardivement sans avoir des informations tracées.
- Dans les 20 % des produits moins bien prédits, il y a une difficulté constante à prédire ces produits puisque la plupart de ces références apparaissent dans la liste des moins bien prédits pour tous les horizons. De même que le premier cas, on trouve certains produits qui apparaissent seulement à l'horizon H1 donc leurs prévisions ont été ajustées tardivement sans avoir des informations tracées.

→ On remarque, dans cette étude d'occurrence sur les horizons, qu'il y a un manque de données explicatives qui justifient les modifications réalisées sur les prévisions notamment d'une façon tardive. Le système actuel est manuel et est basé principalement sur l'intuition et la connaissance des directeurs régionaux sur le marché, ainsi que sur l'expertise de l'analyste.

→ On remarque aussi que généralement, on trouve des produits correctement prédits à tous les horizons ce qui montre qu'il y a certaines références qui sont bien connues par l'analyste de l'entreprise. A contrario, il y a d'autres produits qui figurent toujours dans les moins bien prédits et donc l'analyste a toujours du mal à les prédire.

Pour poursuivre l'analyse des résultats, nous avons cherché à caractériser la ressemblance entre produits en considérant la forme de leur signal de vente ou encore leur comportement de ventes. Dans la figure 7, on présente un exemple de signal de ventes d'un produit caractérisé surtout par une saisonnalité hivernale durant la période de chauffe.

Cette étude des corrélations a été menée dans le but de savoir si les produits bien prédits actuellement ont le même comportement ou pas que certains produits mal prédits.

Dans la matrice ci-dessus, l'axe des abscisses représente les produits bien prédits et l'axe des ordonnées correspond aux produits mal prédits. Les valeurs de corrélation proche de 1 (bleu foncé) représentent les produits ayant un signal de ventes similaires (ou encore corrélation positive), les valeurs proches de -1 (rouge foncé) correspondent aux produits qui ont un comportement inverse (lorsqu'un produit est vendu l'autre ne se vend pas) (corrélation négative), et les valeurs proches de 0 pour les produits dont les ventes ne sont pas corrélées.

→ On remarque que certains produits (comme Y1 et Y4), bien que mal prédits, ont une corrélation positive significative avec d'autres produits (comme X1) qui sont bien prédits. Ce constat pose une question importante : pourquoi réussit-on à prédire quelques-uns, mais pas les autres sachant qu'ils ont un comportement similaire ? Cette question peut également faire écho au fait que, vu le processus actuel mis en place, ces prévisions semblent plus basées sur l'intuition que sur des méthodes quantitatives exactes.

→ La même remarque peut être constatée pour les produits ayant une forte corrélation négative pour lesquels nous savons en prédire uniquement certains alors que d'autres comportements peuvent être inférés par des comportements bien prédits.

III.3.3. Potentiel d'amélioration

Suite à l'étude de l'état existant, nous avons constaté que les prévisions des ventes sont faites selon un processus ayant recours à beaucoup d'expertise humaine et que les résultats ne sont parfois pas assez satisfaisants. Concernant les processus de planification et de prévision, l'étude a également montré quelques écueils en termes de communication et de coordination entre l'entité industrielle et l'entité des ventes. En effet, la structure de l'entreprise au sein du groupe est organisée en deux entités sous différentes directions. Cette structuration rend la collaboration moins agile, notamment pour les sujets transversaux à ces deux entités. Cela nous mène aux axes de recherches proposés dans la partie suivante.

IV. Conclusion

A partir de l'étude de la littérature (dans la partie II) et de l'analyse des processus actuels de l'entreprise elm (partie III), différentes pistes d'amélioration de la Supply Chain peuvent être proposées, notamment en s'appuyant sur des nouvelles techniques guidées par les données.

Plus précisément, il a été souligné l'importance des prévisions des ventes dans la planification de la Supply Chain et dans le processus S&OP comme étape primordiale pour la planification des activités de production au sein de l'entreprise. Il est aussi démontré que cette étape est complexe et peut être améliorée par l'emploi de nouvelles méthodes d'analyse prédictive.

Ce constat permet de construire la problématique suivante qui guidera la suite de ce manuscrit :

Comment améliorer les performances de la Supply Chain de l'entreprise à travers l'automatisation et la fiabilisation des prévisions des ventes par une approche guidée par les données ?

Dans ce cadre, nous présentons dans le deuxième chapitre une étude bibliographique sur les principaux domaines d'application des techniques de prévisions des ventes, ainsi que les approches et les méthodologies proposées dans la littérature pour répondre à cette problématique.

Chapitre 2 : Etat de l'art

I. Introduction

Nous avons évoqué dans le premier chapitre la relation entre la prévision des ventes, comme étant un processus de planification de la Supply Chain, et le maintien de l'équilibre entre l'offre et la demande. La prévision des ventes tient son importance du fait qu'elle constitue la première brique dans les étapes d'élaboration des plans de la Supply Chain. Une bonne qualité de prévision des ventes permet l'établissement d'un plan de demande précis permettant ainsi, la maîtrise de la planification de la production et l'anticipation des quantités nécessaires en stock. D'après (Huang et al., 2014), des prévisions précises atténuent l'effet du coup de fouet et sont d'une importance cruciale pour un processus juste à temps. De ce fait, les industriels dans plusieurs domaines d'activités cherchent toujours à réduire les erreurs de leurs prévisions.

La première partie de ce chapitre sera consacrée à la présentation de quelques domaines d'applications des prévisions des ventes : leurs enjeux, spécificités et les techniques employées. La deuxième partie présentera les principales méthodes de la littérature afin de comprendre leurs principes mathématiques, les hypothèses d'utilisation ainsi que leurs avantages.

II. Problématique de la prévision des ventes

La préparation d'un plan de prévisions des ventes est une tâche difficile pour les entreprises dans de nombreux secteurs car les ventes sont très souvent influencées par plusieurs facteurs. Le besoin des consommateurs, les effets météorologiques, les périodes de l'année, la localisation des achats, les tendances du marché, l'évolution de la concurrence, les délais d'approvisionnement et de livraison et les changements de réglementation sont quelques exemples de ces variables qui peuvent impacter les achats des consommateurs. Du fait de leur variété et parfois de leur dépendance, l'analyste des prévisions doit chercher à sélectionner parmi ces différentes informations les plus pertinentes permettant d'obtenir des prévisions fiables et précises tout en évitant d'une complexité trop importante de l'outil de prévision.

Face à ce problème, de nombreuses solutions existent dans la littérature pour différents domaines d'application. Dans les travaux de (Mccarthy et al., 2006), une enquête a été réalisée auprès des professionnels sur les techniques et les pratiques employées pour les prévisions des ventes. Ils ont regroupé les techniques en deux niveaux : des techniques qualitatives, basées sur l'analyse des avis d'experts ou de clients : comité de direction, forces des ventes, vendeurs, consommateurs, usages et des techniques quantitatives, basées sur l'utilisation de données numériques et représentées par les méthodes analytiques telles que la régression et l'analyse du cycle de vente des produits. Ces travaux montrent que, jusqu'à une période récente (début des années 2000), les méthodes qualitatives, telles que les remontées du terrain auprès des clients, sont largement utilisées et ont un impact important sur la préparation des prévisions des ventes. Progressivement, les techniques statistiques comme la régression et le lissage exponentiel sont de plus en plus employées surtout en travaillant à une échelle plus détaillée (des produits et des composants).

Plusieurs travaux se sont focalisés sur la combinaison des techniques qualitatives et quantitatives. (Franses & Legerstee, 2013) montrent que, lorsque le modèle quantitatif donne de bons résultats, l'usage d'avis d'expert comme une variable supplémentaire pour modifier les prévisions entraîne une dégradation de la qualité des prévisions en comparaison avec celles du modèle uniquement. Cependant, dans le cas où les prévisions du modèle et celles de l'expert ne performant pas bien, l'usage de l'avis d'expert comme entrée supplémentaire pour le modèle peut améliorer les résultats finaux des prévisions. (Baecke et al., 2017) étudient l'impact de l'intégration du jugement humain dans un système statistique de prévisions et montrent qu'il existe un effet positif sur l'amélioration de la qualité des prévisions comparant avec les prédictions basées uniquement sur un modèle ou sur le jugement humain.

Dans la littérature, plusieurs techniques quantitatives traitent la problématique de prévision des ventes comme un sujet de prédiction des séries temporelles. Les travaux de (De Gooijer & Hyndman, 2006) proposent un état de l'art intéressant sur des techniques de modélisation des séries chronologiques. Ils présentent les modèles les plus importants jusqu'à cette date comme la famille des lissages exponentiels et le modèle ARIMA et ses dérivés, ainsi que des modèles non linéaires comme les réseaux de neurones et les méthodes de combinaison. Avec les avancements technologiques et l'apparition du Big Data, des nouvelles techniques sont

apparues pour répondre à la problématique des prévisions des ventes notamment l'apprentissage automatique (ou Machine Learning en anglais).

Au-delà des évolutions des techniques de prédiction, avec des modèles et des techniques mathématiques de plus en plus performants, les outils de prévision se sont également adaptés et spécialisés selon le domaine d'application et la nature des ventes à prédire. La suite de cette partie présente quelques domaines d'application et les spécificités des outils de prévision qui y sont employés.

III. Présentation de domaines d'application et de leur spécificité en termes de prévision de ventes

Nous verrons, dans ce qui suit, l'emploi des techniques de prévisions selon la nature des données et des besoins de prévision distincts pour quelques grands domaines économiques.

III.1. Ventes au détail

D'après (Ramos et al., 2015), la prévision des ventes est l'élément le plus important parmi tous les plans de décisions stratégiques dans le domaine de la vente au détail. Ce domaine d'application a attiré l'attention de nombreuses équipes, vu la grande compétitivité de ce domaine d'activités, pour satisfaire les préférences et les attentes des clients et pour gagner en performance économique. Les entreprises actives dans ce domaine cherchent à prédire plusieurs variables comme la demande saisonnière, les taux d'évolution du marché, la demande des nouveaux produits, etc. permettant ainsi la mise en place d'une stratégie appropriée (Krishna et al., 2018) pour faire face aux changements du marché et aux concurrents.

Des prévisions précises permettent aux entreprises de ventes au détail de gagner des avantages à plusieurs niveaux de l'organisation : les financiers seront capables d'estimer les coûts des projets pour le lancement de nouveaux produits et les budgets nécessaires, les approvisionneurs pourront faire des plans court et long terme des achats de matières premières, le service marketing sera capable d'estimer l'impact de leurs actions auprès des clients et le service logistique sera capable de bien planifier leurs besoins pour le transport des produits (Ramos et al., 2015).

Le secteur de ventes au détail concerne plusieurs types d'industrie et donc différents types de produits. Dans les grandes surfaces, notamment, se trouve une diversité de produits

alimentaires, textiles, d'équipements de la maison de loisirs ou de high-tech. Chaque type de produit est caractérisé par un cycle de vie, un marché auquel il est adressé, un comportement de ventes spécifique, des événements influenceurs, etc. Les événements spéciaux comme les jours fériés, les fêtes nationales et religieuses ainsi que les promotions jouent un rôle important dans la détermination des profils de ventes de certains produits, notamment ceux qui sont liés à ces événements. Ce domaine est ainsi caractérisé par une grande variété de produits et donc de profils de ventes très différents.

Vu la multitude des produits concernés par ce secteur de ventes au détail, les données exploitées pour calculer les prévisions des ventes diffèrent d'une manière importante d'une entreprise à l'autre. Des travaux comme (CATAL et al., 2019) reposent sur différentes données représentées principalement par les données de ventes du magasin Walmart et des variables économiques tels que le taux de chômage, le prix du pétrole, la température, etc.... Les historiques des ventes concernent différentes catégories de produits entre 2012 et 2018. Les travaux de (Behera & Nain, 2019) utilisent des données de vente, avec un historique qui varie entre 1985 et 2009, de produits de différentes natures (alimentaire, médicale, etc...), et organisés selon 16 catégories pour un total de 1559 produits répartis sur 10 magasins de l'enseigne Big Mart. Les travaux de (Jeong, 2017) emploient des données mensuelles de vente au détail de deux types de magasins au Corée de Sud entre janvier 2010 et février 2017 pour l'entraînement et l'évaluation de leurs techniques de prévisions. Ils reposent alors sur les quantités historiques vendues sur cette période pour réaliser des prédictions.

Selon les données à disposition et leurs spécificités, les techniques utilisées pour réaliser la prévision des ventes peuvent varier d'un secteur/d'un catalogue application à l'autre. Dans les travaux de (Behera & Nain, 2019), les auteurs emploient un modèle de XGBoost pour calculer les prévisions de ventes et prouvent ses performances meilleures par rapport aux modèles de Régression Linéaire, Arbres de Décision et Ridge Regression. (Jeong, 2017) comparent dans leurs travaux deux techniques de prédiction des séries temporelles. Ils emploient des modèles SARIMA avec différents paramètres, et plusieurs méthodes de Lissage Exponentiel avec une nature différente de leurs composantes (tendance et saisonnalité). Les travaux de (CATAL et al., 2019) comparent plusieurs modèles de régression comme Bayesian Regression, Decision Forest Regression, Neural Network Regression et Boosted Decision Tree

Regression, et des techniques de séries temporelles comme ARIMA et des techniques de Lissage Exponentiel.

Une description plus détaillée de quelques travaux de la littérature pour le secteur de la vente au détail est donnée en annexe A.

D'après les exemples étudiés dans la littérature, nous remarquons que les modèles ARIMA sont souvent présents dans les travaux de prévision des ventes au détail. Cependant, d'autres techniques comme le Lissage Exponentiel et les techniques de combinaisons et de « Boosting » sont présentes et donnent parfois de meilleurs résultats. Ces approches sont utilisées à la fois pour la prédiction à court et long terme.

III.2. Secteur de la mode

La prévision des ventes dans le domaine de la mode constitue un rôle clé pour assurer la rentabilité des marques étant donné qu'elle représente un axe fondamental pour satisfaire une demande très changeante et très sectorisée, éviter le surstock ou la rupture du stock des produits (Loureiro et al., 2018). Ce domaine est caractérisé par une grande complexité car il est fortement impacté par les attentes et les préférences des clients qui varient énormément, et donc une entreprise doit adopter une stratégie de flux tiré, axée sur le marketing et la demande des consommateurs, d'où la criticité de la tâche de prédiction des ventes (N. Liu et al., 2013). Ce secteur est largement affecté par plusieurs variables externes (tendances globales, climat politique, météo, etc.) ainsi qu'une longue Supply Chain commençant par l'agriculture du coton en passant par les usines de fabrication souvent très éloignées des consommateurs ciblés, jusqu'aux détaillants ou aux plateformes de vente en ligne (N. Liu et al., 2013).

Pour l'industrie de la mode, les produits sont caractérisés généralement par un cycle de vie court car ils seront remplacés rapidement par les produits de la nouvelle saison, entraînant ainsi peu d'historique pour une référence précise (Choi et al., 2012), ainsi qu'une grande variété de produits pour chaque entreprise (Loureiro et al., 2018). Les produits de ce secteur sont généralement de différents types et dépendent des goûts des consommateurs, influencés par les tendances et les événements qui affectent le marché. Les conditions météorologiques et les saisons impactent également les ventes de ce secteur et le type des produits proposés sur le marché, affectant ainsi les besoins et les goûts des consommateurs.

Les compagnies publicitaires et de marketing impactent également beaucoup les choix des clients et donc influencent par la suite les ventes des articles de ce secteur.

Les techniques de prévision dans le secteur de la mode sont caractérisées par leur capacité à traiter une grande variété de produits ainsi que par leur court horizon de prévision. (Craparotta et al., 2019) utilisent dans leurs travaux les données de ventes d'un détaillant européen des produits de la mode. Pour l'entraînement de leur approche, ils emploient les données des collections pour les femmes des saisons estivales 2015 et 2016 ayant 380 produits, et pour le test de leur méthodologie, les données de 80 produits de la collection de 2017 ont été utilisées. Les données correspondent aux historiques des ventes, les photos des produits ainsi que leurs caractéristiques telles que le prix initial, la date de sortie, le cycle de vie, etc... Dans les travaux de (Armando & Craparotta, 2017), utilisent des modèles de séries temporelles et donc ils emploient des données d'historiques de ventes hebdomadaires de plusieurs produits avec des transformations supplémentaires telles que la création d'une base de données avec des moyennes mobiles.

Pour ce domaine, les techniques des prévisions classiques (ARIMA, SARIMA, Lissage Exponentiel, ...) sont les plus utilisées (Loureiro et al., 2018), mais plusieurs autres approches ont été étudiées et utilisées également. Les travaux de (Craparotta et al., 2019) emploient un modèle d'apprentissage profond appelé « Siamese Neural Network », construit à travers deux réseaux de neurones convolutifs (CNN) qui permettent l'exploitation des photos et des caractéristiques des produits, ainsi que les historiques des ventes pour calculer les prévisions pour les nouveaux produits lancés. Dans les travaux de (Armando & Craparotta, 2017), les auteurs proposent une méthodologie de combinaison des modèles de séries temporelles tels que Lissage Exponentiel, Prophet ou encore un réseau de neurones autorégressif pour la prédiction des ventes. Leur approche consiste à choisir les deux meilleurs modèles qui permettent de donner les meilleures prédictions des produits, pour ensuite combiner leurs résultats à travers une somme pondérée. D'autres travaux s'intéressent aux techniques de prédiction pour les produits récemment lancés ou les nouveaux produits à travers d'autres approches proposées dans la littérature comme le « New Product Forecast System (NPFS) » (Ching-Chin et al., 2010) et la méthode NM (Tanaka, 2010).

Une description plus détaillée des travaux évoqués pour caractériser les données et les techniques utilisées pour le secteur de la mode est donnée en annexe B.

D'après l'étude de la littérature, il s'est avéré que les approches récentes basées sur des modèles non-linéaires de réseaux de neurones et des arbres de décision sont plus adaptées à ce domaine d'application, bien que dans certains travaux, des techniques plus classiques aient produit des résultats satisfaisants.

III.3. Secteur alimentaire

Les entreprises du domaine de l'industrie alimentaire emploient aussi les techniques de prévisions des ventes pour assurer l'approvisionnement de leurs clients au bon moment et avec les bonnes quantités. Ces entreprises doivent planifier leurs Supply Chain au mieux pour ne pas avoir des produits de validité expirée en stock et donc des pertes financières et alimentaires (Da Veiga et al., 2016). La présence de plusieurs intermédiaires augmente la longueur de la Supply Chain et la planification des approvisionnements doit être de plus en plus réalisée sur des horizons longs. Enfin, dans ce secteur surtout, les conditions de stockage obligent les industriels à employer un matériel spécifique et coûteux, ce qui augmente encore les coûts engendrés et donc la bonne gestion de volumes stockés aura un grand impact sur la rentabilité de l'entreprise, d'où la nécessité d'avoir des prévisions de vente précises à partir desquelles découlent les plans d'approvisionnement et de gestion des stocks.

(F. L. Chen & Ou, 2009) expliquent que, pour les détaillants de la vente des produits alimentaires, des prévisions imprécises impliquent une augmentation des déchets alimentaires ou des ruptures des stocks de certains produits étant donné leur courte durée de vie et leurs différentes conditions de stockage. Plusieurs autres facteurs sont à prendre en considération pour ce type de prévisions : les fluctuations à court terme comme les promotions et la météorologie, les fluctuations saisonnières à moyen terme comme les fruits et légumes de saison, les périodes de fêtes ainsi que les fluctuations long terme comme la situation économique ou les recommandations de santé (Arunraj & Ahrens, 2015).

Les conditions de conservation d'autres produits notamment laitiers ou congelés sont également assez coûteuses pour l'entreprise. Néanmoins, il existe également des produits tels que la conserverie qui ont un cycle de vie important et leur stockage pour des mois voire des années ne causera pas de problème de conservation, mais plutôt de zone de stockage.

Dans les travaux de (Elmasdotter & Nyströmer, 2018), les données de ventes journalières de produits alimentaires d'une chaîne d'épicerie Equatorienne sont utilisées. Les données d'entraînement et d'évaluation des modèles correspondent aux ventes de plusieurs magasins

entre 2013 et 2017, pour plus de 4000 produits. (X. Liu & Ichise, 2017) étudient dans leurs travaux l'effet des facteurs météorologiques sur la vente des produits alimentaires. Ils étudient la prédiction des ventes de 5 types de boissons pour une enseigne japonaise. Pour ce fait, ils emploient les données de ventes historiques de ces produits ainsi que les données météorologiques pour la période entre janvier 2012 et décembre 2013. (Adithya Ganesan et al., 2019) traitent le problème de prévision des ventes de 7 types de produits alimentaires vendus dans les cinémas indiens. Pour l'entraînement et l'évaluation de leurs approches, ils emploient dans leurs travaux des données de ventes journalières de trois ans entre 2015 et 2017 avec 13 millions transactions enregistrées sur cette période, ainsi que d'autres variables liées aux caractéristiques des films comme le temps d'occupation dans les cinémas ou encore la langue utilisée.

Comme pour les autres secteurs, la prévision des ventes pour le domaine de l'agroalimentaire est aussi caractérisée par une variété de techniques rencontrées. (Elmasdotter & Nyströmer, 2018) comparent des approches de séries temporelles, en utilisant la méthode non linéaire LSTM et le modèle linéaire ARIMA pour la prédiction des ventes journalières. Ils concluent que pour la prédiction d'une seule journée, les deux méthodes présentent des résultats similaires, tandis que pour la prédiction de 7 jours, le modèle LSTM a fourni des meilleures performances. Dans les travaux de (X. Liu & Ichise, 2017), les auteurs emploient le modèle LSTM basé sur les réseaux de neurones couplés avec un auto encodeur qui permet la réduction du nombre des variables météorologiques utilisées. Cette approche utilisée donne des meilleurs résultats comparant avec d'autres techniques d'apprentissage automatique comme SVM, forêts aléatoires, AdaBoost ou encore la Régression Logistique. (Adithya Ganesan et al., 2019) proposent dans leurs travaux l'usage d'un modèle de réseau de neurones dynamiques pour améliorer les prévisions et comparent leur approche avec d'autres techniques comme ARIMA, LSTM, XGBoost ou encore Extra Trees.

Une description plus détaillée des travaux évoqués pour caractériser les données et les techniques utilisées pour le secteur de l'agroalimentaire est donnée en annexe C.

Comme les secteurs étudiés précédemment, les approches de prédiction pour ce domaine d'activités reposent sur les modèles classiques tels que SARIMA et Lissage Exponentiel mais aussi sur des techniques basées sur les réseaux de neurones lorsqu'on dispose de suffisamment de données.

III.4. Ventes en ligne

Grâce à l'énorme évolution numérique et logistique, la vente en ligne s'est largement développée et nous trouvons aujourd'hui des géants d'internet à l'échelle internationale spécialisés dans ce domaine comme Amazon, eBay ou encore Ali Express. Les entreprises actives dans le e-commerce, comme le commerce traditionnel, cherchent toujours à anticiper les demandes clients en se reposant sur des prévisions des ventes plus solides pour éviter les ruptures de stock, et pour optimiser les services logistiques à partir des entrepôts surtout pour les entreprises transfrontalières (Ji et al., 2019). Le e-commerce s'est développé rapidement ces dernières années et couvre aujourd'hui la plupart des domaines commerciaux surtout avec la période de crise sanitaire du COVID-19.

Ce secteur se caractérise par la grande quantité d'informations qui concernent les catalogues produits, les différents commerçants, les stocks, les choix des utilisateurs, leurs données de navigation et les historiques d'achat qui peuvent être récupérés par ces enseignes. Ce domaine est influencé par plusieurs facteurs externes liés aux tendances générales et des actualités ainsi que des périodes spécifiques dans l'année comme les soldes ou encore la période de Noël, mais aussi par une concurrence très rude en termes d'offres et de délais. Les produits vendus sur ces plateformes sont généralement caractérisés par une très grande variété, de l'alimentaire au bricolage et de l'électronique au multimédia rendant la prédiction des ventes une tâche ambitieuse. Ces produits sont généralement stockés dans de multiples entrepôts de distribution et transportés par différents moyens, pour garantir et faciliter leur acheminement vers les consommateurs le plus rapidement possible, ce qui engendre pour chaque produit un coût important, mais complexe à estimer pour les entreprises.

La richesse des données collectées par les entreprises comme le géant du e-commerce américain Amazon, permettent l'exploitation des nouvelles technologies d'intelligence artificielle pour optimiser le management de leurs Supply Chains. Ce secteur de ventes est aussi largement affecté par les avis des clients postés sur les sites de ventes en ligne permettant de décrire les expériences personnelles avec un certain produit, et donc influencer des potentiels acheteurs à faire une transaction (Z. P. Fan et al., 2017; Sharma et al., 2019). Les entreprises exploitent ainsi ces informations pour améliorer leurs prévisions. L'historique de navigation des utilisateurs sur Internet peut être exploité également pour identifier les champs d'intérêts des consommateurs et donc leur proposer des produits qui

sont susceptibles d'être achetés par eux. L'ensemble de ces données est ainsi exploité pour réaliser les prévisions de ventes. Étant données les grandes quantités de données récoltées, des techniques avancées d'analyse de données sont employées pour extraire des informations utiles (Ji et al., 2019).

Prédire les ventes dans le e-commerce représente un véritable challenge vu la demande stochastique (C. Zhang et al., 2020) et les multiples paramètres qui entrent en jeu comme le comportement des acheteurs ou encore les campagnes publicitaires. (C. Zhang et al., 2020) emploient un modèle Autorégressif logarithmique pour leur cas d'étude. Pour la détection des augmentations brusques dans la quantité des ventes, (R. Chen & Liu, 2017) combinent plusieurs méthodes de séries temporelles et de classification tel que SVM. Leur approche permet de déterminer la date de début, du maximum atteint et la fin de ces augmentations, et fournit des meilleurs résultats que d'autres méthodes comme LSTM et ARIMA. (Sharma et al., 2019) comparent dans leurs travaux trois types d'approches à savoir des techniques de régression, des techniques d'arbres de décision et des réseaux de neurones, pour la prédiction des ventes des livres sur la plateforme Amazon, en prenant en considération plusieurs variables influentes. Il est à noter que les géants d'internet ont su se structurer à la mise en place d'équipes de R&D dédiées à la mise au point de nouveaux services basés sur la donnée et sur l'IA et que nombreuses innovations sont issus de ses équipes.

Une description plus détaillée des travaux évoqués pour caractériser les données et les techniques utilisées pour le secteur de la vente en ligne est donnée dans l'annexe D.

Dans ce secteur d'activités, on remarque généralement l'emploi des techniques combinant plusieurs modèles tels que les approches d'hybridation et les approches à plusieurs niveaux. Ceci peut être expliqué par la grande disponibilité à la fois des volumes et des variétés des données sur les plateformes de ventes en ligne.

III.5. Industrie automobile

Un autre domaine d'application aussi concerné par les techniques de prévision des ventes est l'industrie automobile. L'industrie automobile est l'une des plus larges industries au monde avec plus de 90 millions de véhicules produits chaque année (depuis 2015) selon l'Organisation Internationale des Constructeurs Automobiles (OICA). Ce secteur regroupe plusieurs entreprises et organisations impliquées dans la construction des véhicules cherchant ainsi à optimiser leurs Supply Chains et processus de fabrication pour répondre aux

besoins du marché. Par conséquent, il est primordial que les entreprises puissent bien planifier leurs activités en prenant en considération les attentes des clients, l'accessibilité et la disponibilité des ressources de production qui caractérisent ce secteur et son économie globale (Shahabuddin, 2009).

Les prédictions des ventes, notamment au long terme, permettront à une entreprise de prendre des décisions cruciales comme l'élargissement ou la réduction des équipes de travail, ou l'implémentation des nouvelles lignes de production. L'industrie automobile est en effet caractérisée par une chaîne logistique très étendue vu que les composants sont fabriqués dans plusieurs pays, par de nombreux équipementiers, ensuite transportés et assemblés dans les usines d'assemblage et donc une durée de conception et de mise en production assez longue pouvant aller de 12 à 60 mois (Sa-Ngasoongsong et al., 2012). Cette contrainte nécessite la préparation d'un plan de demande et donc des prévisions des ventes à un horizon de 6 à 24 mois, ce qu'implique l'usage des techniques permettant de fournir des prévisions précises sur un long terme.

Ce secteur est également caractérisé par une forte personnalisation des véhicules pour satisfaire les attentes des différents clients. Cet aspect rend la prédiction de la demande plus difficile et favorise plutôt la fabrication après réception de la commande. Les produits sont également influencés par d'autres facteurs externes comme les délais de livraison des composants, les défauts de qualité ou encore l'apparition des nouvelles technologies comme les véhicules électriques.

Le calcul des prévisions pour le secteur de l'automobile repose sur l'historique des ventes de l'entreprise, mais aussi sur différentes autres données comme les taux des parts de marché et les facteurs externes influant les ventes telles que les aides de l'état ou encore les réglementations des carburants. Les travaux de (Abdellatif et al., 2019) utilisent un historique des ventes mensuelles de février 2015 à mars 2018 pour la prédiction des ventes d'un modèle de voiture dans le marché égyptien. Ils reposent également sur des données décrivant la situation économique du pays comme le prix du pétrole, le taux de change avec le dollar américain ou encore le taux d'inflation. Dans la même idée, (Vahabi et al., 2016) emploient des données historiques annuelles dans leur application de prédiction des ventes d'une marque du marché iranien. Les données correspondent aux chiffres de ventes annuelles historiques entre 1990 et 2016 ainsi que des facteurs externes tels que le taux

d'intérêt de la location des véhicules, les tarifs d'importation, les revenus moyens par personne, etc...

Différentes techniques sont employées dans la littérature pour la prévision des ventes dans ce domaine qui se basent sur les avis et les expériences des experts, et aussi sur des méthodes quantitatives. (Abdellatif et al., 2019) comparent l'usage de différentes techniques à savoir un réseau de neurones, régression logistique multiple et le modèle ANFIS pour la prédiction des ventes d'un modèle de véhicule. Leurs résultats montrent des meilleures performances du réseau de neurones par rapport aux autres approches. (Vahabi et al., 2016) combinent le modèle ANFIS avec l'algorithme génétique (GA) dans leur cas d'étude. Le modèle ANFIS est utilisé pour l'entraînement et la prédiction, tandis que l'algorithme génétique est utilisé pour optimiser les résultats de prédiction. Cette approche montre des meilleurs résultats par rapport au modèle ANFIS uniquement et à un réseau de neurones.

Une description plus détaillée des travaux évoqués pour caractériser les données et les techniques utilisées pour le secteur automobile est donnée dans l'annexe E.

D'après les travaux étudiés, on remarque l'usage des techniques basées sur des réseaux de neurones ainsi que le modèle ANFIS pour surpasser les limites qui peuvent être rencontrées avec les méthodes classiques (Lissage Exponentiel, Régression Linéaire, ARIMA), ainsi que l'exploitation des données liées aux variables externes et pas seulement les historiques de ventes.

III.6. Secteur de l'énergie

Le domaine de l'énergie constitue un secteur important pour l'étude des prévisions des ventes. Plusieurs travaux ont traité la problématique de prédiction de l'énergie et plus précisément dans la prédiction de la consommation énergétique des bâtiments (Ahmad et al., 2018; Bourdeau et al., 2019; Deb et al., 2017) dans le but de réduire les coûts engendrés et améliorer l'efficacité énergétique de ces derniers (Somu et al., 2020). Le terme énergie employé dans ce domaine désigne l'ensemble des énergies électrique ou combustible pour le chauffage, la climatisation, l'éclairage, les sanitaires et différents équipements qui peuvent contribuer aux usages d'un bâtiment (Ahmad et al., 2018).

La production et la vente de l'électricité représentent également un pilier important de ce domaine d'étude où des méthodes de prévision y sont appliquées pour anticiper les besoins

de consommation des clients. Ce secteur a attiré l'attention des chercheurs depuis longtemps vu son importance en tant que source d'énergie vitale, mais aussi pour le vaste marché en profonde mutation auquel s'adressent les compagnies actives dans ce secteur. Ce secteur présente plusieurs enjeux pour les industriels vu les changements importants qui marquent le marché comme l'augmentation de l'usage du numérique, d'une mobilité électrique, de la décarbonation de l'industrie et de l'augmentation de la part des énergies renouvelables (Nagbe et al., 2018).

Les prévisions de la consommation d'électricité sont essentielles pour la gestion des sources de production et des réseaux de transport, la gestion des risques et la mise en place des politiques par les fournisseurs de ce domaine (Elamin & Fukushige, 2018). Vu le volume considéré, les erreurs de prévision peuvent engendrer des coûts très élevés. Ainsi de nombreux travaux de recherche sont proposés dans la littérature pour mieux prédire la demande. Certains travaux comme (Shao & Tsai, 2018) et (Zhao et al., 2015) se sont intéressés à la problématique de prévisions des ventes dans le secteur de l'électricité où il est nécessaire d'estimer en avance les quantités d'énergie à produire sur chaque période de l'année pour satisfaire la demande..

La vente de l'électricité est caractérisée par une forte saisonnalité, avec en hiver l'usage intensif des chauffages et en été celui de la climatisation. La demande en électricité est toujours plus forte surtout avec la croissance démographique et le développement économique caractérisés par l'augmentation constante des zones urbaines, mais aussi d'autres facteurs très significatifs, comme le développement du numérique et des datacenters, mais aussi la transformation de l'industrie et de la mobilité avec l'arrivée des véhicules électriques. Ce produit (l'électricité) subit également les fluctuations d'autres indicateurs économiques comme le prix du pétrole et des matières premières nécessaires pour sa génération.

Avec l'augmentation du nombre des compteurs électriques connectés qui renvoient régulièrement les données, il est apparu différents techniques et modèles permettant d'exploiter ces données pour la prédiction. Les séries temporelles étudiées dans ce domaine ont différentes granularités : on trouve des données où les mesures de la consommation sont prises chaque minute alors que pour d'autres les relevés se font chaque année. Dans les travaux de (K. Li et al., 2020), les données utilisées pour l'entraînement et l'évaluation des

modèles correspondent à l'historique des ventes de l'électricité journalières entre janvier 2018 et juin 2020, et mensuelles de janvier 2011 jusqu'au juin 2020 dans une région du marché chinois. Ils emploient également d'autres variables qui caractérisent la consommation électrique comme les dates, les semaines et les mois de ventes, ainsi que des données climatiques de la région. (Cao & Wu, 2016) testent leur approche en employant des données de consommation mensuelle historique de l'électricité en Chine entre janvier 2010 et décembre 2015, et des données historiques de ventes mensuelles d'un producteur de l'électricité dans le marché des Etats Unis entre janvier 1994 et décembre 2014. Pour les deux cas d'étude, les auteurs procèdent à la normalisation des données brutes avant les utiliser pour l'entraînement et la validation des modèles.

Le secteur de l'énergie est caractérisé par des techniques de prédiction adaptés à différentes problématiques, pour la planification, la gestion voire l'optimisation. Des techniques de régression et de prédiction des séries temporelles sont utilisées dans les travaux de (K. Li et al., 2020). Les données d'entrée sont décomposées à travers la méthode X11, et ensuite, les techniques de XGBoost, Light GBM et Gradient Boosting Decision Tree (GBDT) sont utilisées pour la prédiction du court et moyen terme. Pour le long terme, les auteurs proposent d'utiliser le modèle Prophet. Dans les travaux de (Cao & Wu, 2016), les auteurs emploient un modèle de Support Vector Regression (SVR) couplé avec l'algorithme d'optimisation « Fruit Fly » (FOA) et un ajustement de la saisonnalité présentée dans les données, pour calculer des prévisions mensuelles de la consommation et de la vente de l'électricité.

Une description plus détaillée des travaux évoqués pour caractériser les données et les techniques utilisées pour le secteur de l'énergie est donnée dans l'annexe F.

D'après les travaux présentés, les modèles MARS, SVR et les modèles basés sur les réseaux de neurones sont les plus employés dans ce champ d'étude. Ceci peut être justifié par la disponibilité d'une grande quantité de données à exploiter et la capacité de ces modèles à identifier les aspects non linéaires dans les séries temporelles.

III.7. Synthèse des besoins de prédiction associés aux différents domaines

Nous avons étudié dans les paragraphes précédents, quelques principaux domaines d'applications des techniques de prévisions des ventes. Nous avons présenté les caractéristiques de chaque secteur, les données généralement employées ainsi que les techniques employées dans chaque cas d'étude. Comme nous pouvons le remarquer, certaines techniques sont plus employées dans un domaine que d'autres, mais plusieurs autres modèles, notamment les méthodes classiques comme le modèle ARIMA et le lissage exponentiel sont communes entre les différents secteurs. Nous présentons dans le tableau 1 un résumé des spécificités de chaque domaine étudié.

SECTEUR	PRODUITS	TYPE DE CLIENT/ACHAT	BESOINS DE PREVISION/ENJEUX	HORIZON DE PREDICTON	ALGORITHMES
VENTES AU DETAIL	<ul style="list-style-type: none"> - Produits diversifiés - Différents commerces concernés 	<ul style="list-style-type: none"> - Grand public - Dédié par spécialité - Achats à fréquence variable 	<ul style="list-style-type: none"> - Satisfaire les clients - Garantir la disponibilité des produits 	1 jour - 6 mois	ARIMA et ses variétés, Techniques de Boosting, Méthodes de combinaison
SECTEUR DE LA MODE	<ul style="list-style-type: none"> - Produits saisonniers - Sectorisés par clients - Influencés par les tendances 	<ul style="list-style-type: none"> - Grand public - Achats mensuels / trimestriels 	<ul style="list-style-type: none"> - Répondre aux attentes du marché - Satisfaire les goûts très différents des consommateurs 	1 semaine - 4 mois	ARIMA et ses variétés, Lissage Exponentiel, Techniques de Boosting, LSTM et MLP
ALIMENTAIRE	<ul style="list-style-type: none"> - Produits périssables avec court cycle de vie - Affectés par les saisons 	<ul style="list-style-type: none"> - Grand public - Achats quotidiens 	<ul style="list-style-type: none"> - Limiter les déchets et le gaspillage - Gérer des stocks périssables 	1 jour - 14 jours	ARIMA et ses variétés, Lissage Exponentiel, Multivariate BPN, MFLN (Multilayer Function Link Network)
VENTE EN LIGNE	<ul style="list-style-type: none"> - Produits très diversifiés 	<ul style="list-style-type: none"> - Utilisateurs des plateformes des ventes 	<ul style="list-style-type: none"> - Réduire les stocks 	1 jour - 3 mois	ARIMA et ses variétés, Techniques de Boosting,

			- Assurer une livraison rapide		Réseaux de neurones, Techniques d'hybridation
SECTEUR AUTOMOBILE	- Produits très personnalisables - Délais de fabrication importants	- Consommateurs avec un certain niveau de vie	- Réduire les délais d'attente - Réduire les stocks - Satisfaire les attentes des consommateurs	2 ans	Régression, VAR, ANFIS, Lissage Exponentiel, Réseaux de neurones
ENERGIE	- Consommation continue sans interruption	- Particuliers et industriels - Consommation quotidienne	- Répondre aux besoins de consommation - Satisfaire une demande croissante - Piloter des sources multiples	0.5 heure - 1 mois	MARS, SVM, SVR, Techniques de Régression, Réseaux de neurones

Tableau 1 : Synthèse des domaines d'applications

Dans la deuxième partie de ce chapitre, nous allons nous intéresser plus spécifiquement à la description mathématique des modèles employés dans la littérature pour répondre à la problématique des prévisions des ventes. Il s'agira de présenter les spécificités des modèles, de leurs méthodes d'estimation et de leurs usages.

IV. Horizon des principales techniques utilisées pour la prévision des ventes

Comme présenté dans le paragraphe précédent, il existe plusieurs techniques employées dans la littérature pour la prévision des ventes. Ces approches diffèrent dans leur façon de calculer les prédictions et les algorithmes utilisés pour réaliser cette tâche. Le choix de la technique et du modèle à utiliser dépend de plusieurs critères et propriétés du cas d'application.

Bien que les entreprises cherchent aujourd'hui à automatiser cette tâche, l'intervention humaine reste souvent indispensable dans certains cas d'application où de très nombreux facteurs, dont les facteurs humains peuvent influencer les ventes. Néanmoins, plusieurs techniques sont développées dans le but de réduire cette intervention en utilisant des approches de plus en plus sophistiquées. Avec les différents facteurs influant les ventes, les techniques doivent être capables de répondre précisément aux besoins de prédiction malgré des comportements de vente très variables. Cet aspect a marqué la nécessité de développer des modèles capables de détecter la nature linéaire mais aussi la nature non linéaire et les fluctuations dans les ventes.

L'évolution récente des techniques d'apprentissage automatique et d'apprentissage profond a permis de développer de nouvelles approches pour les prévisions des ventes. Nous pouvons regrouper ces approches sous quatre catégories principales. Des techniques considérant le problème de prévision des ventes comme une problématique de séries temporelles et emploient ainsi des modèles dédiés à la modélisation des séries chronologiques. Des méthodologies employant des approches de Data Mining et d'apprentissage automatique permettent d'exploiter davantage les dépendances statistiques des variables et facteurs externes afin d'améliorer la qualité des prévisions. La troisième catégorie est basée sur les approches d'apprentissage profond. En effet, ces techniques ont vu le jour avec le développement des modèles sophistiqués basés surtout sur des réseaux de neurones qui peuvent être employés tant qu'une grande quantité de données est disponible. Une dernière catégorie constitue des méthodes et approches d'hybridation des modèles, généralement

linéaires et non linéaires, proposés dans les catégories précédentes. La figure 9 présente un schéma global des différentes catégories.

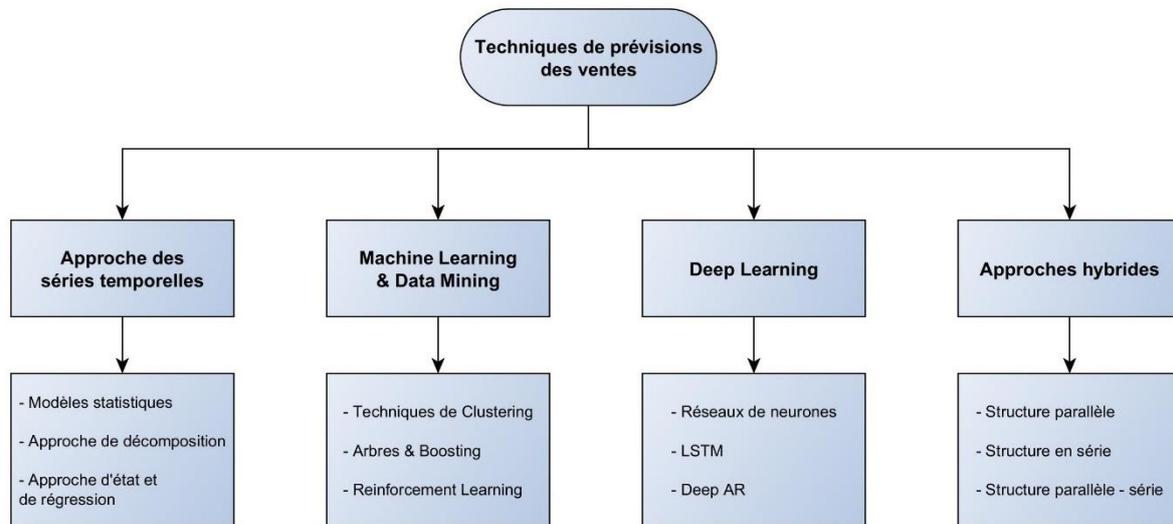


Figure 9 : Catégories des techniques de prévisions

Dans les paragraphes ci-dessous, nous présentons les différentes catégories de techniques de prévision ainsi que les modèles appartenant à ces familles, en suivant la catégorisation précédente.

IV.1. Les approches de modélisation de séries temporelles

IV.1.1. Modèles statistiques de prévision

La problématique de prévisions des ventes peut être traitée par une approche de modélisation des séries temporelles. Ces techniques sont plus classiques par comparaison avec les méthodes d'apprentissage automatique et de Deep Learning, mais elles sont toujours très utilisées dans plusieurs domaines. Le modèle ARIMA « Autoregressive Integrated Moving Average » est l'un des modèles les plus rencontrés dans la littérature. Ce modèle est composé d'un terme autorégressif (AR) décrivant la relation entre la variable à modéliser et ses anciennes valeurs, un terme moyenne mobile (MA) décrivant la relation linéaire entre les résidus de la série, et un terme de différenciation (I) pour rendre la série stationnaire. On parle aussi du modèle SARIMA qui est une extension du modèle ARIMA en y intégrant des termes saisonniers permettant de modéliser la saisonnalité dans les séries temporelles. Ces modèles sont principalement utilisés pour les séries temporelles univariées et stationnaires (Panigrahi & Behera, 2017).

D'après (Hyndman, R.J. Athanasopoulos, 2019), une série chronologique stationnaire est une série dont les propriétés statistiques telles que la moyenne sont indépendantes du temps. Ainsi, les séries chronologiques avec une tendance ou une saisonnalité, ne sont pas stationnaires puisque leurs propriétés statistiques varient au fil du temps. Pour rendre une série temporelle stationnaire, une technique de différenciation est employée généralement. Cette technique est utilisée pour stabiliser la série. Elle consiste à remplacer chaque valeur de la série par la différence avec la valeur précédente afin de réduire les effets de la tendance et de la saisonnalité. Une série temporelle peut être différenciée plusieurs fois jusqu'à sa stabilisation. D'autres transformations, telles qu'une transformation logarithmique, peuvent être employées pour stabiliser la variance de la série.

Comme présenté dans le paragraphe précédent, les modèles ARIMA et SARIMA sont employés dans plusieurs domaines d'application pour la prévision des ventes. Dans les travaux récents, on trouve rarement l'usage unique de ces modèles, par contre on les trouve souvent dans des approches de combinaison de plusieurs techniques. Les travaux de (Pongdatu & Putra, 2018) représentent une étude de comparaison des modèles SARIMA et Holt-Winters pour la prédiction des transactions des clients dans un magasin. L'estimation des paramètres du modèle SARIMA est effectuée à travers les graphiques d'autocorrélation partielle (ACF) et le choix entre les modèles candidats est fait en choisissant le candidat ayant la mesure d'erreur la plus faible. Les résultats montrent que SARIMA donne des meilleurs résultats pour les prévisions à court terme alors que le lissage exponentiel est plus performant sur le long terme. Le modèle de Holt-Winters permet aussi de mieux modéliser les séries présentant des formes de saisonnalité que le modèle SARIMA.

Bien que les modèles ARIMA et SARIMA soient largement utilisés pour la prédiction des séries temporelles, ils ne permettent pas la modélisation de la variance des séries au fil du temps. Cette limite peut être contournée en utilisant une transformation logarithmique ou une transformation de Box-Cox lorsque la variance subit des faibles changements. Néanmoins, dans les applications avec une grande volatilité où la variance évolue constamment comme la prédiction des cours des actions en bourse, ces transformations sont moins performantes. Dans ces cas, les modèles « Autoregressive Conditional Heteroskedasticity (ARCH) », et « Generalized Autoregressive Conditional Heteroskedasticity (GARCH) » sont employés. Dans

notre recherche bibliographique, nous avons constaté que l'usage de ces modèles pour la prédiction des ventes reste très limité.

IV.1.2. Approche de décomposition des séries temporelles

La décomposition des séries temporelles consiste à considérer la série comme une combinaison (additive ou multiplicative) de différents éléments comme une tendance, une saisonnalité et un bruit. La technique de décomposition est employée pour mieux comprendre, analyser et identifier les motifs cachés dans une série chronologique. Ces techniques peuvent être combinées aussi avec d'autres modèles de prédiction pour améliorer les résultats comme le cas des travaux de (Gurnani et al., 2017).

Il existe dans la littérature plusieurs techniques pour effectuer la décomposition d'une série temporelle. (Hyndman, R.J. Athanasopoulos, 2019) présentent dans leur livre les méthodes souvent utilisées pour cet objectif. La technique classique peut être additive ou multiplicative et permet de décomposer la série en tendance, saisonnalité et bruit. Cette technique estime que la saisonnalité est constante d'un an à l'autre et la détermination du cycle de la tendance (« trend-cycle ») est basée sur l'usage des moyennes mobiles dans les données saisonnières.

Une méthode populaire pour la décomposition est X11, développée par « US Census Bureau and Statistics Canada » (Bee Dagum & Bianconcini, 2016). Cette méthode est basée sur la méthode classique avec certaines améliorations comme la possibilité de prendre en considération des variations faibles de la saisonnalité. Cette méthode est employée sur des données mensuelles ou trimestrielles et permet de prendre en considération les variations journalières et les effets des facteurs externes connus (Hyndman, R.J. Athanasopoulos, 2019).

La méthode « Seasonal Extraction in ARIMA Time Series » connue sous le terme SEATS est une procédure de décomposition développée par la banque d'Espagne et est largement utilisée. Comme la méthode X11, elle est applicable sur des données mensuelles ou trimestrielles aussi (Bee Dagum & Bianconcini, 2016).

Une autre méthode largement utilisée dans la littérature est « Seasonal and Trend decomposition using Loess » connue par la décomposition STL. Cette méthode est plus robuste et a plusieurs avantages par rapport aux autres méthodes. Elle permet de travailler avec différentes granularités de données (non seulement mensuelles ou trimestrielles), elle est capable de modéliser une saisonnalité variable et elle est robuste aux valeurs aberrantes

de la série (Hyndman, R.J. Athanasopoulos, 2019). Cette technique est généralement employée pour une décomposition additive.

Le lissage exponentiel est une famille de modèles classiques de prévision de séries temporelles largement utilisée dans les entreprises et l'industrie (De Gooijer & Hyndman, 2006). Ce type de modèle permet de modéliser les séries en les décomposant en niveau (ou moyenne), tendance et saisonnalité (Vandepuut, 2018). La nature additive ou multiplicative de ses composantes permet de créer plusieurs variétés de modèles de cette famille. Les travaux de (Robin Hyndman et al., 2008) représentent un état de l'art pertinent de cette famille de produits tout en introduisant une composante d'erreur supplémentaire et un Framework de prédiction automatique sous le langage R.

Cette famille de modèles est largement utilisée pour la prédiction des ventes, mais généralement en combinaison avec d'autres techniques notamment dans les travaux récents. (Panigrahi & Behera, 2017) emploient cette technique avec un réseau de neurones en proposant une approche pour la prédiction des séries temporelles. Le modèle de lissage exponentiel est utilisé pour modéliser l'aspect linéaire ou non linéaire de la série au départ, ensuite, les résidus sont prédits à travers le réseau de neurones. Les données brutes des séries temporelles sont introduites dans le modèle de lissage exponentiel après une étape de normalisation avec la méthode min-max.

Une autre famille de modèles basée sur la décomposition des séries temporelles est la famille des modèles additifs généralisés. Ce type de modèle a été d'abord proposé par (Hastie & Tibshirani, 1987) présentant la possibilité de modéliser une variable sous la forme d'une sommation de plusieurs prédicteurs linéaires ou non linéaires comme suit :

$$g(y) = s_1(x_1) + \dots + s_p(x_p) + \varepsilon$$

Avec y est la variable à prédire, g est la fonction liant y avec les prédicteurs x_1, \dots, x_p et s_1, \dots, s_p les fonctions de chaque prédicteur.

Selon (LARSEN, 2015), ces modèles sont utilisés grâce à leurs interprétabilité, flexibilité et automatisation. Les travaux de (Thouvenot et al., 2016) et (S. Fan & Hyndman, 2012) représentent des exemples d'étude de la problématique de prédiction de la consommation électrique à travers des modèles additifs. (Thouvenot et al., 2016) adoptent une approche basée sur les modèles additifs semi-paramétriques. D'abord, un algorithme LASSO est

employé pour la sélection des variables à utiliser pour la prédiction telle que les prédicteurs météorologiques et les effets du calendrier. Ensuite, les paramètres des estimateurs du modèle additif (des Splines) sont calculés afin de minimiser la mesure d'erreur de prédiction.

Une des plus récentes méthodes de décomposition des séries temporelles est le modèle « Prophet ». En 2018, l'équipe de recherche de Facebook a publié (Taylor & Letham, 2018) un nouveau modèle de prévision des séries temporelles basé sur les modèles additifs généralisés (GAM). Ce modèle est constitué de trois composantes : tendance, saisonnalité et évènements. Il permet de prendre en considération plusieurs saisonnalités dans la série à prédire ou encore d'autres variables externes influençant les ventes et il est robuste envers les données manquantes. Les travaux de (Zunic et al., 2020) constituent un cadre d'application de ce modèle pour la prédiction des ventes au détail employant Prophet. Le but était de calculer les prévisions mensuelles et trimestrielles de plusieurs produits. Les données brutes sont introduites dans le modèle sans modification, néanmoins, dans ces travaux, on trouve une intervention humaine pour paramétrage du modèle, notamment pour la composante événementielle. Les résultats de prédiction sont exploités ensuite pour classer les produits du portfolio de l'entreprise.

Bien que les techniques classiques de séries temporelles présentées dans ce paragraphe soient largement utilisées, certaines présentent quelques limites. Pour les modèles saisonniers, nous avons besoin des données couvrant au moins deux saisons pour pouvoir capturer l'aspect saisonnier de la série. Cependant, ce n'est pas toujours le cas surtout lors du lancement des nouveaux produits. En plus, ces techniques ne sont généralement pas robustes si certaines données sont manquantes dans la série ou si des valeurs aberrantes sont présentes et donc une étape de nettoyage et de préparation des données est nécessaire avant de procéder au calcul des prévisions. Enfin, l'usage des variables exogènes représentant des facteurs externes qui peuvent influencer les ventes n'est pas pris en compte dans les techniques classiques. Ce dernier point peut être résolu en s'appuyant sur des approches de régression qui seront le sujet du paragraphe suivant.

IV.1.3. Approche dynamique de régression

Une autre approche proposée dans la littérature pour la prédiction des séries temporelles correspond à des modèles de régression. L'idée est de prédire les valeurs futures de la série temporelles en se basant sur d'autres informations externes ou encore d'autres séries

temporelles qui influencent la série à prédire. Ceci passe par l'exploitation des variables externes dites exogènes ayant une relation avec la série originale.

Le modèle le plus basique est le modèle de régression linéaire simple qui permet de décrire une relation linéaire entre une variable à prédire et une variable explicative. Lorsqu'on dispose de deux ou plusieurs variables exogènes, on peut employer le modèle de régression linéaire multivarié reliant la variable à prédire avec ces prédicteurs externes. Bien qu'on suppose généralement une relation linéaire entre les variables à prédire et les variables externes, il existe de nombreux cas dans lesquels une relation non linéaire est plus appropriée. Pour pouvoir traiter cet aspect, on peut procéder à une transformation de la variable à prédire et/ou les prédicteurs avant de construire le modèle de régression. La transformation logarithmique est la plus couramment employée.

La plupart des modèles de séries temporelles présentés dans les paragraphes précédents 3.1.a et 3.1.b permettent de modéliser une série temporelle en fonction de ses valeurs précédentes sans avoir recours à d'autres facteurs externes qui peuvent avoir un impact sur l'évolution de la série. Par opposition, les modèles de régression permettent d'exploiter ces facteurs externes sous forme de variables exogènes, mais avec une capacité plus limitée modéliser les variabilités dans les séries temporelles comme les modèles classiques. La combinaison de ces deux propriétés des modèles de régression et de séries temporelles représente l'introduction des modèles dynamiques de régression (Pankratz, 1991).

Les travaux de (Pinho et al., 2016) représentent une application réussie de cette technique pour la prévision des ventes au détail de différents produits. Les auteurs emploient plusieurs modèles dynamiques avec différentes configurations à chaque fois. Le premier modèle développé représente une structure du modèle ARIMA incorporant des séries de Fourier pour identifier les formes (ou patterns) de saisonnalité dans la série, reposant uniquement sur l'historique des ventes du produit. Un autre modèle utilisé emploie seulement des variables externes comme le prix, les périodes de promotions et les événements du calendrier. Un troisième modèle emploie ces mêmes facteurs externes mais d'un autre produit de la même catégorie. Une version logarithmique de ces mêmes modèles a été développée également pour comparer leurs résultats également. Leurs résultats montrent que les modèles dynamiques testés donnent des meilleurs résultats par rapport au modèle ARIMA n'exploitant pas ces facteurs externes.

Les techniques de régression présentées dans ce paragraphe représentent des approches intéressantes pour contourner les limites constatées chez les modèles classiques. Néanmoins, pour ces techniques, la préparation des données des variables externes à utiliser comme prédicteurs peut être une tâche parfois difficile. En effet, pour prédire des valeurs futures d'une série, il faut avoir celles des prédicteurs à utiliser ce qui est parfois plus difficile à prédire que la série originale en soi.

Dans le paragraphe suivant, nous présentons d'autres approches et techniques utilisées pour la prédiction des ventes.

IV.2. Les approches de Data Mining et de Machine Learning

Comme présentés dans la première partie de ce chapitre, plusieurs travaux sont présentés dans la littérature pour la prédiction des ventes avec des techniques d'apprentissage automatique (ou Machine Learning (ML)). Certains travaux utilisent des techniques de clustering pour un prétraitement des données avant d'employer des modèles de prédiction. On trouve, généralement, dans l'état de l'art deux façons d'emploi des modèles ML pour la prédiction : l'usage d'un seul modèle de régression pour faire les prévisions, et la combinaison de plusieurs techniques pour construire un modèle à plusieurs niveaux plus puissant et robuste. Bien que la première façon soit plus directe et plus facile à implémenter et à exploiter, son usage reste assez limité surtout dans les travaux récents. La deuxième approche est plus utilisée dans la littérature sous le nom de « Ensemble Learning » et différentes techniques sont employées pour assurer la mise en place de cette idée à savoir le « Bagging », « Boosting » et « Stacking » (Polikar, 2012).

IV.2.1. Clustering

Le regroupement des données ou « clustering » est l'une des techniques de description des données les plus fondamentales. Le but du clustering est d'identifier une structure dans un ensemble de données non caractérisées au préalable en organisant les données en groupes homogènes (clusters) où la similitude est maximisée entre les éléments du même groupe, et est minimisée entre les éléments appartenant à des groupes différents.

Ces techniques sont employées également pour étudier les séries temporelles dans le but d'identifier et d'extraire des informations utiles à partir des bases de données complexes et massives. D'après (Aghabozorgi et al., 2015), les techniques de clustering sont utilisées avec les séries temporelles pour quatre applications principales : détection des anomalies et des

motifs inattendus dans une série, détection des changements dynamiques dans la série, prédiction et recommandation à travers des combinaisons avec des techniques de prédiction, et découverte de nouveaux motifs dans des nouvelles bases de données.

Les travaux de (Dai et al., 2015) constituent un cadre d'application du clustering avec le modèle « Support Vector Regression (SVR) » pour la prédiction des ventes des serveurs informatiques. Pour la phase d'entraînement, les auteurs proposent de normaliser les données à travers la méthode min-max et puis d'employer le modèle de k-means pour regrouper les produits similaires en 3 groupes en utilisant la distance euclidienne comme mesure de similarité entre les données. Un modèle SVR est entraîné ensuite pour chaque cluster pour avoir les paramètres les plus adéquats à chacun. Pour la phase de prédiction, ils utilisent 3 mesures de similarité pour associer chaque nouvelle donnée avec le cluster le plus adéquat. Les mesures employées sont la moyenne, la médiane et la méthode MinMax. Ensuite, ils emploient le modèle SVR associé au cluster pour générer des prédictions. Les résultats montrent que cette approche est plus performante par rapport à l'usage d'un modèle SVR directement sur les données sans clustering.

IV.2.2. Arbres de décision et Ensemble Learning

Les arbres de décision sont l'une des approches de machine learning utilisées à la fois pour des problèmes de classification et de régression. Plusieurs travaux dans la littérature ont employé ces techniques comme modèle de prédiction des ventes ou comme une étape de prétraitement des données avant l'usage d'un autre modèle. Les arbres de décision peuvent être utilisés comme des modèles séparés ou en les combinant ensemble ou avec d'autres modèles pour construire des approches d'apprentissage plus robuste, d'où l'introduction du paradigme « ensemble learning ».

(Cadavid et al., 2018) indiquent dans leurs travaux que les méthodes des forêts aléatoires et des arbres de décision présentent des bons résultats dans plusieurs cas d'étude avec une importante interprétabilité, et les approches floues « fuzzy approaches » conviennent mieux aux données présentant une incertitude et des informations imprécises comme la météo.

Les travaux de (Thomassey & Fiordaliso, 2006) proposent une approche combinant une technique de clustering avec les arbres de décision pour la prédiction des profils de ventes des produits. Les données utilisées dans cette étude correspondent à des historiques de ventes ainsi que des caractéristiques des produits. D'abord, un modèle de K-means est utilisé

pour regrouper les produits selon des profils de ventes spécifiques. Ensuite, un modèle d'arbres de décision est entraîné afin d'apprendre à classer les produits selon les profils identifiés. Pour chaque nouveau produit à prédire, le modèle d'arbre de décision assignera ce dernier au profil de ventes le plus adéquat avec ses caractéristiques. Le diagramme de la figure 10 présente la méthodologie présentée dans ces travaux.

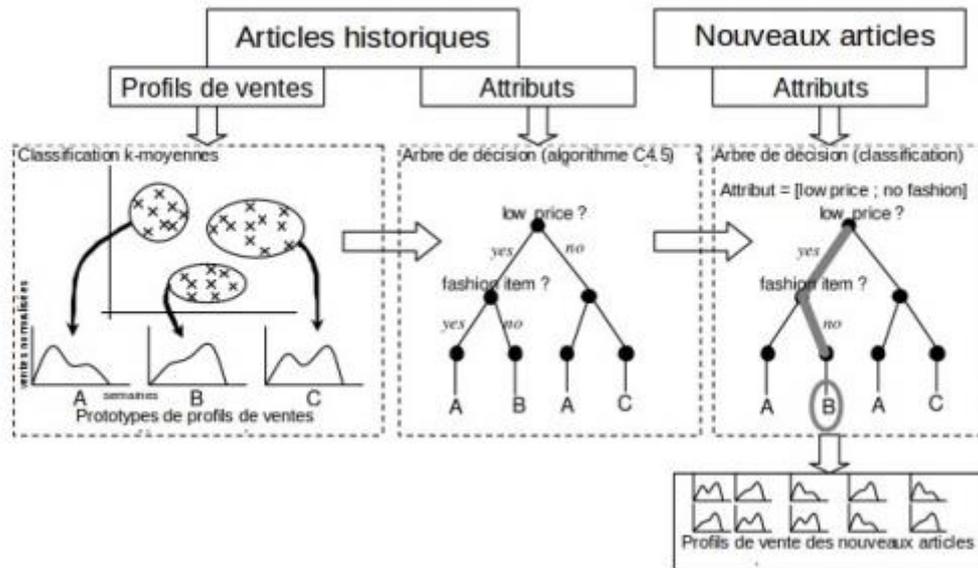


Figure 10 : Approche proposée par (Thomassey & Fiordaliso, 2006)

La technique de combinaison de plusieurs modèles ou encore le « Stacking » a été utilisée dans plusieurs problèmes de prévision. Les travaux de (Pavlyshenko, 2019) proposent une approche à plusieurs niveaux composés de différents modèles d'apprentissage automatique comme le montre la figure 11 suivante.

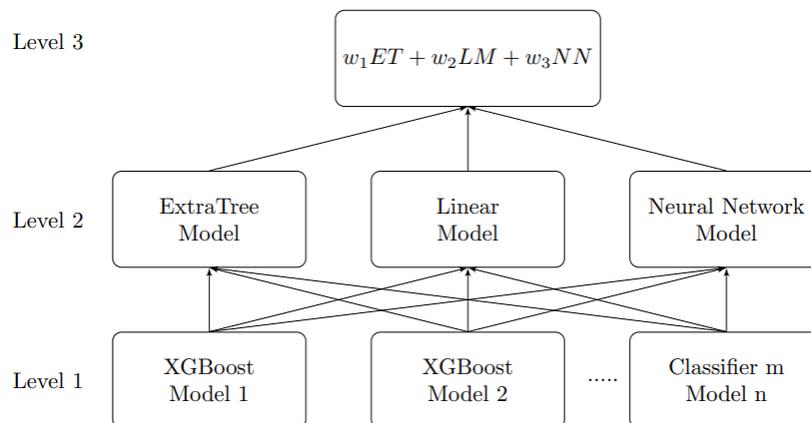


Figure 11 : Approche multi-niveau proposée par (Pavlyshenko, 2019)

Le premier niveau correspond à plusieurs modèles d'arbres entraînés par un « boosting de gradient XGBoost » en parallèle avec différentes séries temporelles en entrée. Ensuite, les résultats de chaque modèle du premier niveau seront des entrées des modèles du deuxième niveau. Dans cette approche, les auteurs utilisent les modèles « ExtraTree », un modèle linéaire et un réseau de neurones à ce stade. Les prédictions finales du troisième niveau correspondent à la somme pondérée des résultats des modèles du deuxième niveau. Notant que cette méthodologie a été adaptée par les auteurs pour gagner la première place de la compétition « Grupo Bimbo Inventory Demand » sur « Kaggle » ayant comme objectif de prédire la demande de la plus grande entreprise mexicaine de pain et de distribution.

Dans la même idée, (Punam et al., 2019) présentent dans leurs travaux un modèle à deux niveaux pour la prévision des ventes.

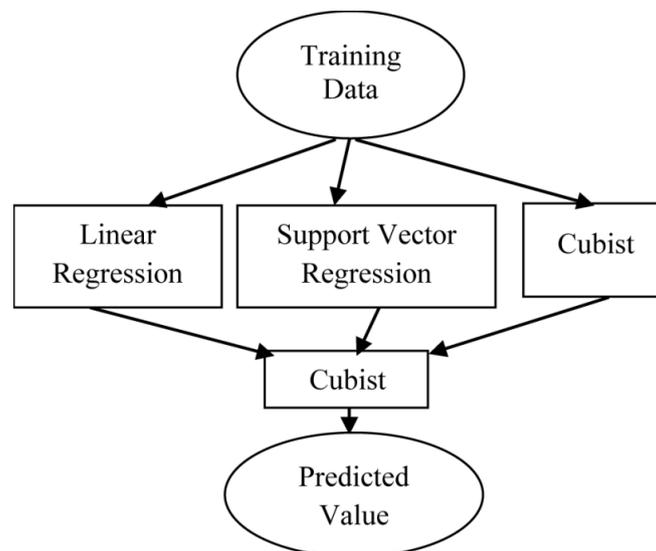


Figure 12 : Méthodologie proposée par (Punam et al., 2019)

Premièrement, les prédictions sont faites individuellement par trois différents modèles à savoir la régression linéaire, le « Support Vector Regression (SVR) » et le modèle « Cubist Regression ». Ce dernier est basé sur une extension de l’algorithme M5 de Quinlan utilisé pour la construction d’arbres de régression. Ensuite, ces prédictions seront des entrées pour un autre modèle « Cubist Regression » pour réaliser des prédictions finales plus précises que les résultats des modèles utilisés séparément.

Dans d’autres travaux comme (Cheriyen et al., 2019), les auteurs évaluent les performances de trois modèles : modèle linéaire généralisé, arbre de décision et « Gradient Boosted Tree », pour la prédiction de la tendance des demandes clients. Les résultats montrent que le modèle

« Gradient Boosted Tree » basé sur un ensemble des arbres de décision performe mieux que les autres modèles de comparaison.

IV.2.3. L'apprentissage par renforcement (Deep Reinforcement Learning)

C'est un champ d'études assez récent et peu d'applications employant ces techniques sont développées dans la littérature, mais (Liu et al., 2020) proposent une approche appartenant à ces techniques d'apprentissage pour la prévision de la consommation énergétique immobilière. Dans cette étude, les auteurs développent 3 techniques de DRL à savoir « Asynchronous Advantage Actor-Critic (A3C) », « Deep Deterministic Policy Gradient (DDPG) » et « Recurrent Deterministic Policy Gradient (RDPG) », et les comparent avec 3 autres techniques d'apprentissage supervisé à savoir Régression linéaire multiple (MLR), Réseau de neurones à rétropropagation (BPNN) et forêts aléatoires (RF) comme le montre le de la figure 13.

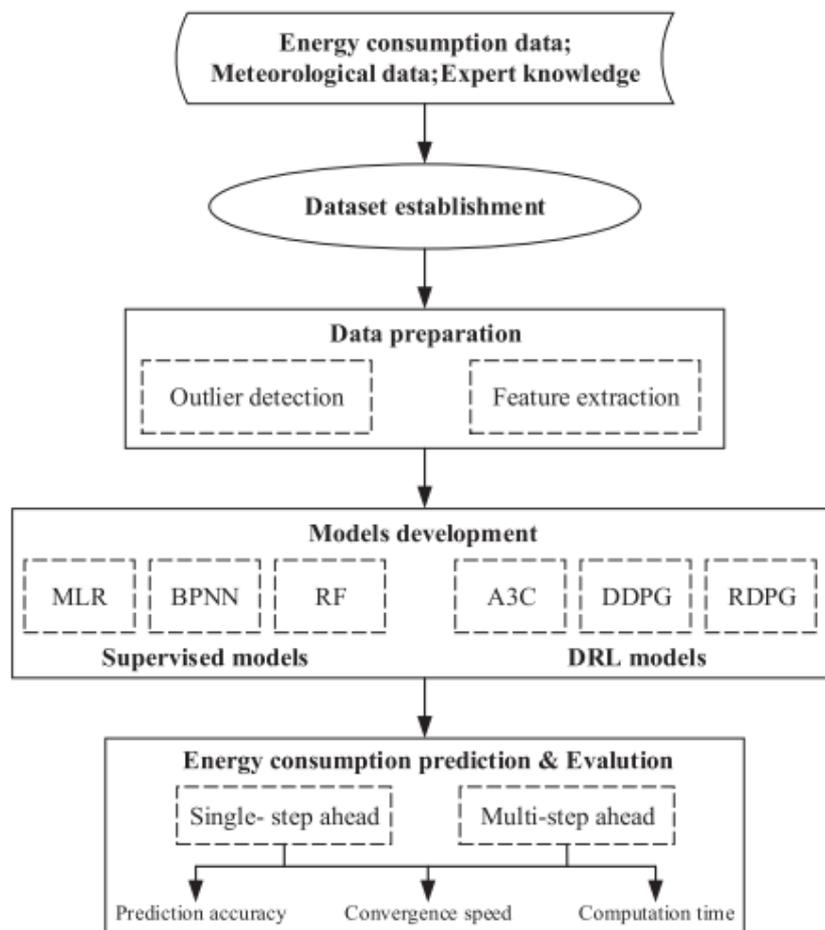


Figure 13 : Approche proposée par (T. Liu et al., 2020)

La mise en application de cette méthodologie montre que les modèles DDPG et RDPG ont des bonnes performances et permettent ainsi d'améliorer les prévisions mais ils ont un temps de calcul assez important par comparaison avec les méthodes d'apprentissage supervisé. La technique A3C ne présente pas d'avantages pour cette application.

IV.3. Les approches de Deep Learning

Dans les travaux de (Cadavid et al., 2018), une enquête est réalisée sur les techniques d'apprentissage automatique employées récemment (entre 2009 et 2017) dans des travaux de prédiction des ventes et de la demande client. D'après leurs résultats, le réseau de neurones artificiel (ANN) avec rétropropagation du gradient représente un outil attractif vu sa capacité de généralisation et de détection des patterns non-linéaires dans la série temporelle.

Le « Multi Layer Perceptron » est un réseau de neurones à propagation avant (« feedforward ») qui constitue un des premiers essais d'application de ce type de modèle pour la prédiction des séries temporelles. Il peut être construit à partir d'une ou plusieurs couches de neurones cachées imbriquant chacune une fonction d'activation pour transformer les données d'entrées en sorties. Le réseau est utilisé comme une fonction d'approximation entre la valeur à prédire et les valeurs précédentes de la série. Pour entraîner ce type de modèle, l'algorithme de rétropropagation est généralement utilisé dans la littérature.

Les travaux de (Kong & Martin, 1995) représentent une des premières applications de ce type de réseau de neurones pour la prévision des ventes. Les auteurs utilisent les historiques de vente comme données d'apprentissage pour le réseau, et testent différentes configurations afin d'identifier la structure meilleure à travers des implémentations empiriques de chacune. La configuration donnant les meilleurs résultats est constituée de 12 neurones en entrée, 6 neurones dans la couche cachée et un neurone pour la couche de sortie avec une fonction d'activation Sigmoïde. Les résultats montrent de meilleures performances par rapport à la méthode conventionnelle utilisée dans l'entreprise.

Les applications de ce type de réseau de neurones restent toutefois limitées pour la prédiction des séries temporelles et les approches statistiques sont généralement plus performantes. De plus, ce type de réseau nécessite des larges bases de données et parfois un temps de calcul important. Néanmoins, il peut être utilisé comme un modèle de régression lorsqu'on dispose d'autres variables exogènes.

D'autres types de réseaux de neurones constitueront des évolutions intéressantes dans les techniques de prévisions. En effet, avec le développement des réseaux de neurones profonds et les résultats intéressants obtenus dans diverses applications, plusieurs techniques basées sur ces modèles ont été proposées pour répondre à la problématique de prévisions des ventes. Les réseaux de neurones sont employés pour effectuer des prévisions comme une technique de régression en se basant sur l'historique des données et/ou d'autres variables explicatives des ventes (Loureiro et al., 2018).

Les travaux de (Loureiro et al., 2018) représentent un cadre d'application de plusieurs techniques de prévision des ventes, notamment le réseau de neurones profond. Les auteurs utilisent les données de ventes historiques avec les caractéristiques des produits et les avis d'experts comme variables exogènes, comme entrées. Les modèles employés sont la régression linéaire, les arbres de décision, les forêts aléatoires, la technique de « Support Vector Machines », le réseau de neurones artificiels (ANN) et le réseau de neurones profond complètement connecté et composé de plusieurs couches de neurones (DNN). Leur approche consiste à mettre en compétition ces différentes techniques lors de la prédiction de chaque produit et utiliser le modèle le plus performant entre eux. Les résultats montrent que le DNN performe le mieux entre ces différentes techniques, mais le processus d'entraînement de ce modèle est assez complexe. Il nécessite une expertise dans le domaine contrairement à d'autres techniques plus simples comme les forêts aléatoires qui donnent des résultats satisfaisants.

Une autre approche intéressante basée sur les réseaux de neurones récurrents a été largement utilisée dans l'état de l'art qui est le « Long Short Term Memory (LSTM) » représentant ainsi une autre application réussie des réseaux de neurones artificiels (ANN) pour la prévision de séries temporelles (Hsu, 2017; Pankaj Malhotra et al., 2015).

Comme exemple des travaux employant cette technique, (Marino et al., 2016) comparent deux structures du modèle « Long Short Term Memory (LSTM) » à savoir une architecture standard et une architecture basée sur la structure « Sequence to Sequence (S2S) » sur des données de consommation horaire. Les résultats montrent que le LSTM standard donne des bons résultats pour prédire la prochaine heure mais ne réussit pas à prédire la minute suivante, tandis que la deuxième architecture LSTM-S2S performe dans les deux cas d'étude. La même technique de LSTM a été employée également dans les travaux de (Somu et al.,

2020) pour prédire la consommation électrique sur le court, moyen et long terme. Cette technique a donné de bons résultats dans ce cas d'étude. De plus, les auteurs ont étudié l'impact de changement des paramètres du réseau comme le nombre des couches sur la précision du LSTM.

Les travaux de (Bandara et al., 2019) représentent également une application de LSTM pour la prédiction des ventes. La figure 14 présente leur méthodologie proposée.

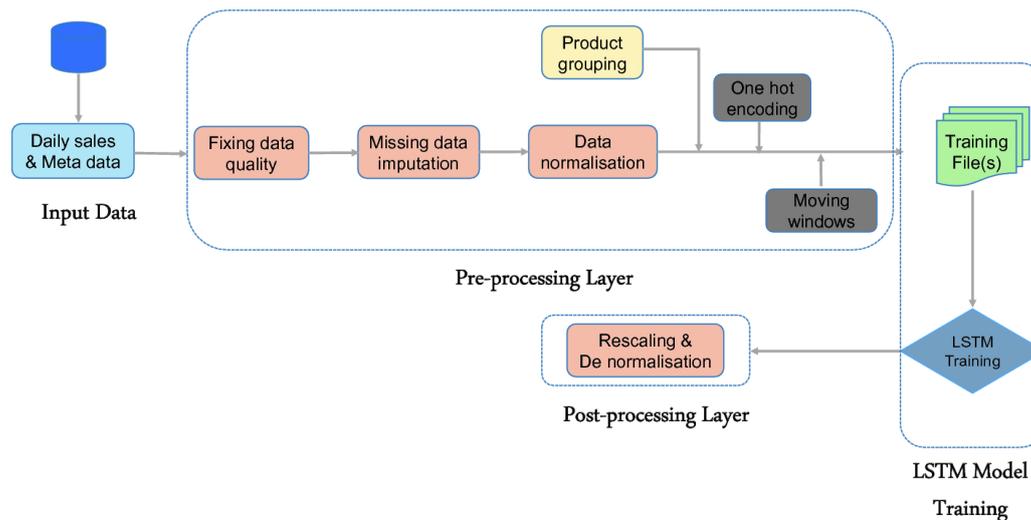


Figure 14 : Approche proposée par (Bandara et al., 2019)

Leur approche est constituée de 3 couches principales : la première consiste en la préparation et le prétraitement des données, la deuxième est l'usage du modèle LSTM pour la prédiction, et la troisième couche concerne le post-traitement des données à travers une dénormalisation et une remise à l'échelle. Les résultats de l'application de cette approche montrent que le modèle LSTM performe mieux par rapport au Lissage Exponentiel et au modèle Naïf Saisonnier.

Cependant, le LSTM présente certains inconvénients, en particulier leur complexité et leur sur-paramétrage (De Gooijer & Hyndman, 2006). De plus, en général, ils ont besoin de très longues séries temporelles pour produire de bonnes prévisions, de sorte que pour les courtes séries, les méthodes classiques fonctionnent mieux. En développant ce modèle, (Salinas et al., 2020) proposent un nouveau modèle appelé « DeepAR » basée sur un réseau de neurones récurrent autorégressif pour la prédiction des séries temporelles, permettant de contourner quelques inconvénients de LSTM.

Pour faire face à certaines limites que présentent les approches et les techniques présentées jusqu'à présent, les travaux récents traitant la problématique de prévision des ventes et de séries temporelles proposent d'employer des approches hybrides.

IV.4.Approches d'hybridation

Pour la prédiction des séries temporelles, des approches de plus en plus complexes sont apparues dans la littérature en visant à combiner la puissance des modèles de différentes natures linéaire et non-linéaire, tout en palliant certaines de leurs faiblesses. Selon (G. P. Zhang, 2003), ces méthodes appelées « hybrides » sont développées pour pallier plusieurs difficultés constatées avec les méthodes classiques : il est difficile de déterminer la nature du processus étudié, il est souvent compliqué de trouver des séries temporelles purement linéaires ou non-linéaires dans des données réelles et le fait qu'il n'y a pas une méthode qui peut performer dans toutes les situations. Les réseaux de neurones sont généralement employés pour construire des modèles hybrides en les combinant avec des modèles des séries temporelles (Ratnayaka et al., 2015) ou d'autres modèles d'apprentissage automatique (Weng et al., 2019), vu leur capacité à identifier les formes non linéaires dans les données.

Selon (Hajirahimi & Khashei, 2019), on peut résumer les avantages des approches hybrides en trois points :

- Amélioration des prévisions grâce à une détection et modélisation complète des motifs,
- Réduction du risque d'un mauvais choix de modèle grâce à la combinaison des résultats,
- Procédure de sélection de modèle plus simple grâce à l'usage de différentes composantes.

D'après l'enquête réalisée par (Hajirahimi & Khashei, 2019), il existe dans la littérature trois principales structures d'approches hybrides : structure parallèle, structure en série, structure parallèle-série.

La structure parallèle consiste à entraîner les modèles simultanément avec les données brutes et calculer ensuite les prédictions en combinant les résultats de chaque modèle à travers une fonction de pondération.

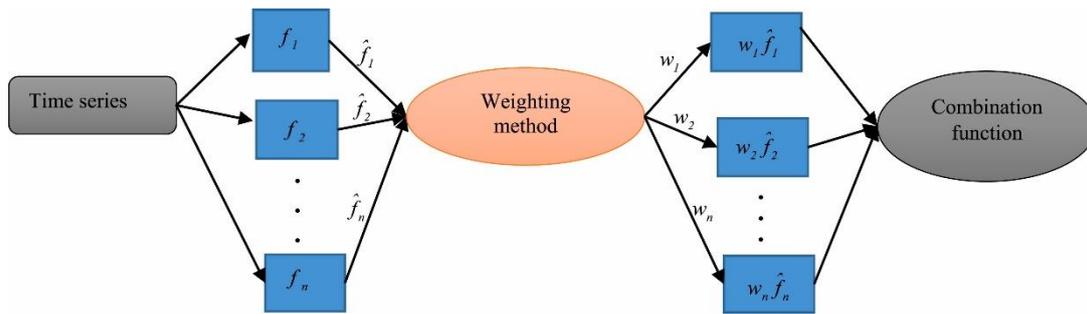


Figure 15 : Structure hybride parallèle (Hajirahimi & Khashei, 2019)

La structure en série consiste à entraîner le premier modèle avec les données brutes, ensuite son résultat sera l'entrée du deuxième modèle ainsi de suite. À la fin, les prédictions finales consistent à sommer les résultats des différents modèles.

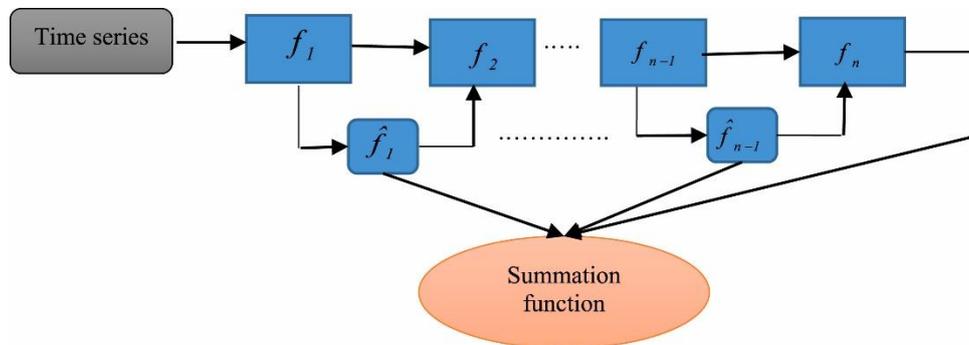


Figure 16 : Structure hybride en série (Hajirahimi & Khashei, 2019)

La dernière structure combine les deux approches en parallèle-série. Cette structure consiste à fournir au premier modèle les données brutes et récupérer ses résultats qui seront des prévisions (de la partie linéaire dans les données) et des résidus. Ensuite, le deuxième modèle sera entraîné en prenant comme entrée les données brutes et les résultats du premier modèle. Dans ce cas, il y aura trois possibilités pour les entrées du deuxième modèle : données brutes avec les prévisions, données brutes avec les résidus, données brutes avec les prévisions et avec les résidus.

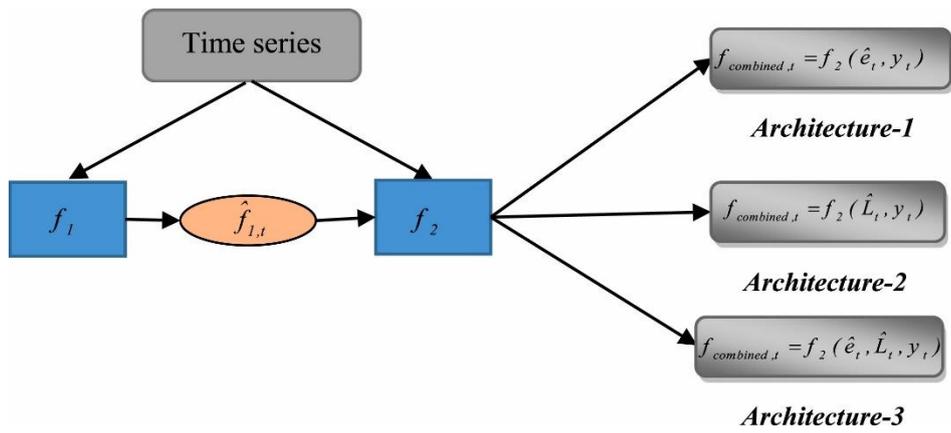


Figure 17 : Structure hybride parallèle-série (Hajirahimi & Khashei, 2019)

Le schéma de la figure 18 résume les différents résultats identifiés dans cette revue de l'état de l'art.

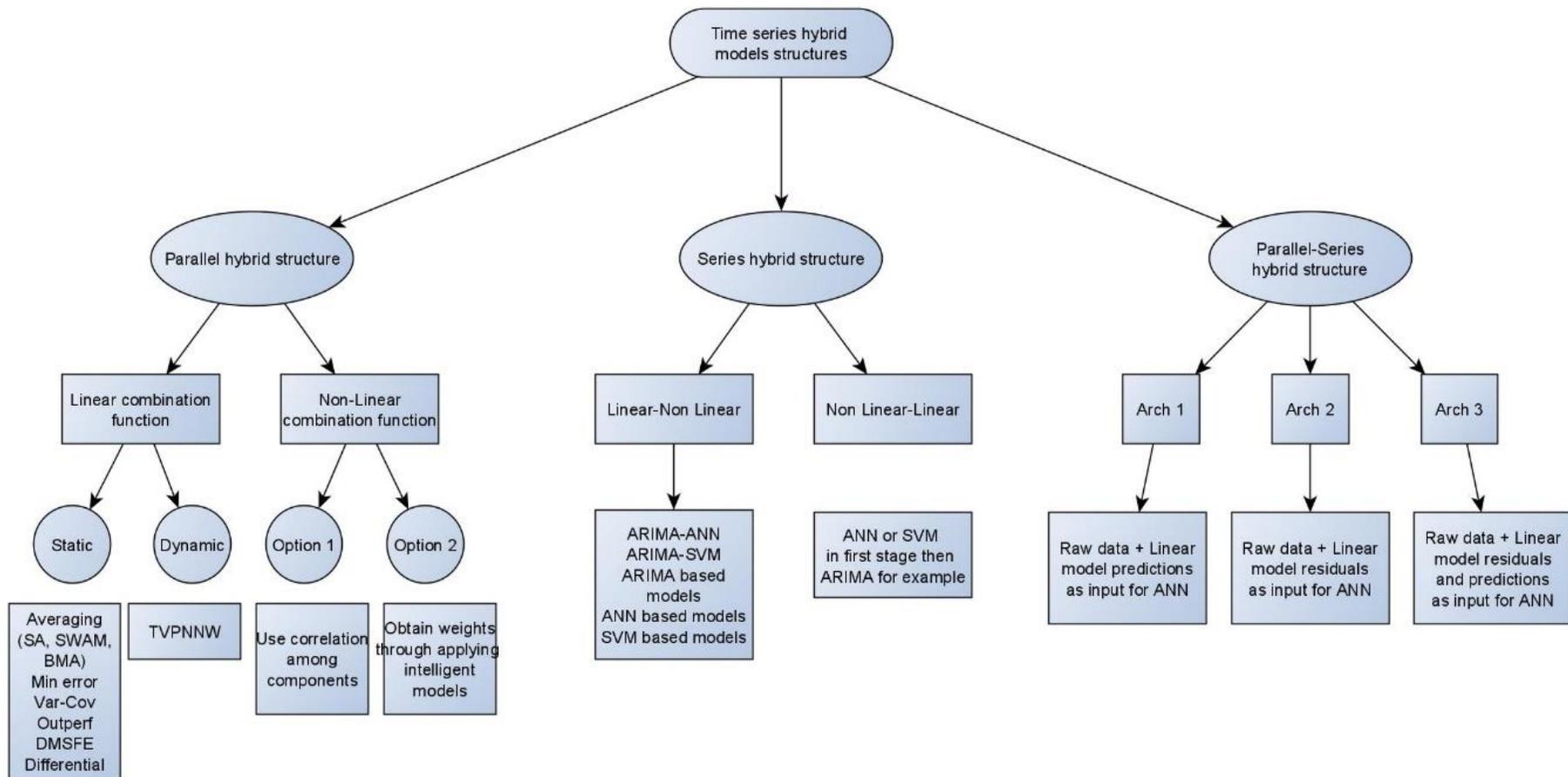


Figure 18 : Résumé des structures hybrides (d'après (Hajirahimi & Khashei, 2019))

Dans la structure hybride parallèle, la fonction de combinaison peut être linéaire ou non-linéaire. Les fonctions linéaires statiques sont des techniques de sommation pondérée qui ne varient pas au fil du temps, et les fonctions linéaires dynamiques, notamment la méthode TVPNNW (L. Wang et al., 2018), permettent un changement des poids accordés aux modèles combinés lorsque des nouvelles observations sont introduites dans la série. Les fonctions non-linéaires distribuent les poids sur les modèles combinés en utilisant une fonction d'optimisation qui permet de réduire l'erreur en choisissant les poids adéquats, ou à travers un modèle intelligent tel qu'un réseau de neurone.

Pour la structure hybride en série, les deux principales possibilités consistent à utiliser un modèle linéaire suivi d'un modèle non-linéaire ou l'inverse. Pour les modèles linéaires, ARIMA ou le Lissage Exponentiel sont généralement utilisés, et pour les modèles non-linéaires, on emploie généralement un réseau de neurones (ANN) ou un modèle SVM.

Dans la littérature, plusieurs travaux ont adopté ces structures pour répondre à la problématique de prévision des ventes. (G. P. Zhang, 2003) emploient le modèle ARIMA pour capturer la partie linéaire des données et un réseau de neurones (ANN) pour la partie non-linéaire. Leur application sur des données réelles montre que cette approche hybride performe mieux que les modèles utilisés individuellement. Dans la même idée, (Panigrahi & Behera, 2017) proposent une hybridation du modèle de lissage exponentiel avec un ANN. L'application de leur approche sur plusieurs séries temporelles montre de meilleures performances comparant avec les modèles ARIMA, Lissage Exponentiel, MLP et d'autres méthodes hybrides existantes.

Les travaux de (Xu et al., 2019) proposent une hybridation du modèle SARIMA avec le modèle Support Vector Regression (SVR) pour la prédiction de la demande dans l'industrie aéronautique. La figure 19 présente l'approche proposée dans ces travaux.

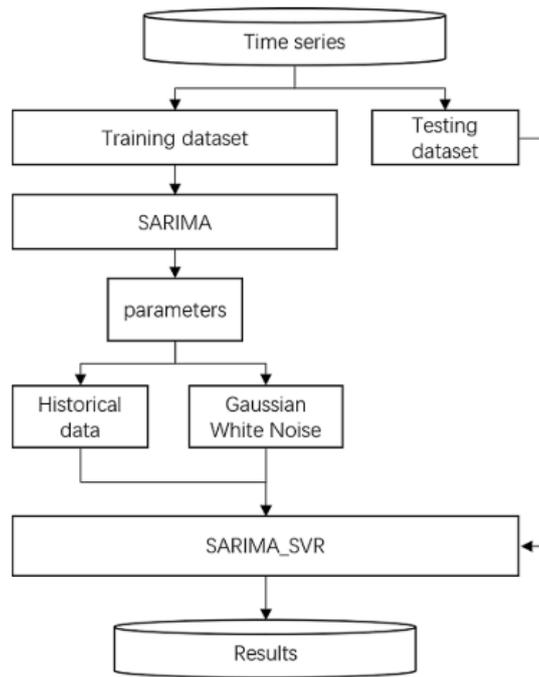


Figure 19 : Approche proposée par (Xu et al., 2019)

Dans cette étude, le modèle SARIMA est entraîné pour avoir les paramètres optimaux, ensuite, ces paramètres sont rectifiés par un modèle de bruit blanc gaussien avant de les envoyer au modèle SVR. Ce dernier prendra en entrée les paramètres rectifiés du modèle SARIMA ainsi que les données historiques de ventes. Les auteurs comparent différentes configurations du modèle SVR pour identifier celle la plus adéquate à leur cas d'étude.

Dans la même démarche, (Gurnani et al., 2017) comparent divers modèles d'apprentissage automatique à savoir « XGBoost », « Support Vector Machines (SVM) » et « AutoRegressive Neural Network (ARNN) » avec ARIMA, des méthodes hybrides et des méthodes basées sur la décomposition STL afin d'identifier le modèle le plus précis pour prévoir les ventes dans l'industrie pharmaceutique. Ils ont conclu que les modèles non linéaires tels que SVM et XGBoost fonctionnaient mieux que ARIMA, cependant, les modèles hybrides donnaient de meilleures performances que les modèles utilisés individuellement.

V. Conclusion

Ce chapitre a été consacré à l'étude de l'état de l'art des techniques de prévisions des ventes qui est une tâche complexe. Dans la première partie, nous avons présenté les principaux domaines d'application de la prévision des ventes avec les spécificités de chaque domaine en termes de besoins et de produits. Nous avons présenté également quelques travaux de la

littérature comme des cas d'études de cette problématique dans différents secteurs. Nous avons essayé d'associer les techniques les plus appropriées à chaque domaine d'application afin de construire une synthèse claire sur les usages des principales techniques de prévision des ventes.

La deuxième partie de ce chapitre a été consacrée à l'investigation et la compréhension des méthodes de prévisions des ventes dans la littérature en présentant les principaux types de modèles rencontrés. Plusieurs travaux ont été développés pour répondre à cette problématique selon plusieurs approches. Ces techniques diffèrent dans leur traitement des données et leur modélisation des variables d'entrée. Certaines approches se sont focalisées sur des techniques de séries temporelles tandis que d'autres reposent sur méthodes d'apprentissage automatique de régression. Les techniques plus sophistiquées des réseaux de neurones profonds, des modèles ensemblistes et des approches hybrides ont été étudiées également.

Nous avons remarqué que dans un grand nombre de cas d'études la même démarche est employée. Les chercheurs testent des modèles séparément et évaluent leurs performances au départ, ensuite on trouve souvent une combinaison des résultats de plusieurs modèles, ou encore des approches d'hybridation à travers la modélisation de la partie linéaire des données par un modèle linéaire et la partie non linéaire par un autre modèle adéquat.

Nous pouvons conclure que, dans la plupart des travaux, les auteurs cherchent souvent à identifier le meilleur modèle ou approche de prédiction qui correspond à leur cas d'étude et donc aux données disponibles. Ceci nous ramène à répondre à la problématique suivante :

Comment faire pour obtenir un modèle de prédiction fiable pour un grand nombre de séries temporelles de différentes natures et en se basant seulement sur les historiques de vente, parfois avec peu de données ?

Dans ce cadre, nous allons développer dans le chapitre suivant une méthodologie basée sur des modèles de séries temporelles permettant de répondre à la problématique de prédiction des ventes de plusieurs produits ayant différents comportements de vente, mais aussi avec peu d'historiques de vente disponibles. Nous verrons que cette approche ne sera pas conçue pour être spécifique au cas d'étude considéré, mais pourra être appliquée à différents champs d'études.

Chapitre 3 : Méthodologie et Expérimentations

I. Introduction

Nous avons évoqué, dans le premier chapitre, le rôle des techniques prédictives et des approches guidées par les données dans l'amélioration des performances industrielles des entreprises. Nous avons vu également l'importance de la prévision des ventes qui est une étape clé dans la planification des processus de production, d'approvisionnement et de stockage en lien avec le management de la Supply Chain. Cette importance est justifiée par la nécessité des entreprises commerciales, quel que soit leur domaine, de savoir à l'avance le besoin du marché pour pouvoir le satisfaire au bon moment et avec les bonnes quantités.

Dans le second chapitre, un état de l'art est réalisé sur les techniques employées dans plusieurs domaines d'application pour traiter la problématique de prévision des ventes posée par l'entreprise elm.leblanc. Une étude a été menée sur les caractéristiques de chaque approche ainsi que les spécificités des différents cas d'applications.

Dans ce chapitre nous allons présenter une première contribution autour d'une méthodologie originale pour répondre à la problématique d'une meilleure prévision des ventes dans le contexte du catalogue produits d'elm.leblanc. Cette méthodologie consistant à mettre en place une représentation multi-modèles pour la prévision des ventes, nous détaillerons ainsi chaque structure de modèle retenu. Ce chapitre se terminera par une présentation des résultats des expérimentations réalisées et par une discussion sur les avantages et les limites de cette approche.

II. Méthodologie

II.1. Formalisation du problème traité

Dans l'optique d'améliorer les performances de sa Supply Chain, l'entreprise elm.leblanc, spécialisée dans le domaine des systèmes de chauffage résidentiels et industriels, vise à automatiser le processus de prévision et améliorer leur qualité par une approche guidée par les données. Cet objectif peut être atteint à travers la mise en place d'une méthodologie qui répond à la spécificité de ses données, à la problématique actuelle et aux objectifs de gain de performance.

Comme montré dans l'analyse de l'état de l'art de l'existant présenté dans le précédent chapitre, le processus de prévision actuel repose sur l'intuition et la connaissance des experts du terrain. Nous avons remarqué qu'il y a un manque de données explicatives permettant de justifier les modifications des prévisions à différents horizons. L'étude de corrélation préliminaire nous a montré que, malgré le comportement similaire de certains produits, la méthode actuelle ne permet pas de prédire les ventes de ces produits avec la même exactitude. Cette méthode ne présente pas des résultats satisfaisants et répétables, ce qui explique la motivation pour la mise en place d'une nouvelle méthodologie.

Dans ce travail, nous utiliserons comme support les données de ventes mensuelles de plusieurs produits avec différents horizons de prédiction. Ces horizons assurent la capacité à fournir des prévisions pour le court et le moyen terme. Les données à disposition correspondent à des séries temporelles représentant chacune une référence de produit : le produit le plus ancien présent parmi les données utilisées a été lancé en 2015 et le plus récent début de l'année 2021.

Les données couvrent un catalogue de produits de l'entreprise appartenant à des familles distinctes. Certaines ventes de produits présentent des comportements saisonniers tandis que d'autres répondent à des événements plus aléatoires, ce qui conduit à des séries chronologiques de nature différente. De plus, le nombre d'observations présent dans les séries temporelles considérées varie d'un produit à l'autre. La création de nouvelles références peut résulter du lancement d'un tout nouveau produit sur le marché et donc sans données d'historique, ou bien correspondre à un produit qui serait le successeur d'une ou plusieurs références et dont il va occuper l'ancienne part de marché. Cette variété des données implique de trouver un modèle adéquat de séries temporelles pour chaque type de vente suivant le catalogue produit existant.

Pour bien formuler le problème, soit R l'ensemble des références produits du catalogue. Chaque référence de produit i peut être représentée par une série temporelle de N_i observations passées représentant l'historique des ventes, sachant que N_i varie d'un produit à l'autre. Nous cherchons à calculer les prévisions de ventes de chaque référence pour plusieurs horizons H de prédictions. Le schéma ci-dessous montre une représentation graphique de la problématique pour une référence de produit avec t la dernière observation enregistrée :



La prédiction de plusieurs horizons peut se faire selon différentes stratégies. D'après (Ben Taieb et al., 2012), il existe dans la littérature principalement 3 stratégies pour atteindre cet objectif :

- **Une stratégie directe** : l'idée est de développer un modèle spécifique pour chaque étape de prédiction, donc un modèle différent pour $t+1$, $t+2$, etc. La création d'un modèle pour chaque horizon à prédire constitue un temps de calcul supplémentaire, notamment en travaillant sur plusieurs références, et d'autant plus lorsque le nombre d'horizons à prédire augmente. En plus, l'usage des modèles séparés ne permet pas de modéliser la dépendance entre deux étapes de prédictions consécutives comme c'est souvent le cas dans les séries chronologiques.

$$\text{prédiction}(t+1) = \text{modèle1} (\text{obs}(t), \text{obs}(t-1), \text{obs}(t-2), \dots, \text{obs}(t-N_i))$$

$$\text{prédiction}(t+2) = \text{modèle2} (\text{obs}(t), \text{obs}(t-1), \text{obs}(t-2), \dots, \text{obs}(t-N_i))$$

- **Une stratégie récursive** : l'idée est d'utiliser le même modèle pour prédire les différentes étapes de prédiction en utilisant à chaque fois la valeur prédite à l'étape précédente comme entrée pour faire la prédiction de l'étape suivante. En employant les valeurs prédites comme entrées, cette stratégie peut permettre aux erreurs de prédiction de s'accumuler et donc les performances peuvent se dégrader surtout pour les longs horizons.

$$\text{prédiction}(t+1) = \text{modèle} (\text{obs}(t), \text{obs}(t-1), \text{obs}(t-2), \dots, \text{obs}(t-N_i))$$

$$\text{prédiction}(t+2) = \text{modèle} (\text{prédiction}(t+1), \text{obs}(t), \text{obs}(t-1), \dots, \text{obs}(t-N_i))$$

- **Une stratégie RECTIFY** : c'est une stratégie hybride combinant les deux approches précédentes en utilisant un modèle spécifique pour chaque étape de prédiction et en utilisant les prédictions du modèle précédent comme entrée au modèle suivant. Cette

stratégie peut surmonter le problème de dégradation des performances sur les longs horizons de l'approche récursive mais le temps de calcul restera toujours important du fait de la création d'un nouveau modèle pour chaque étape de prédiction.

$$\text{prédiction}(t+1) = \text{modèle1} (\text{obs}(t), \text{obs}(t-1), \text{obs}(t-2), \dots, \text{obs}(t-N_i))$$

$$\text{prédiction}(t+2) = \text{modèle2} (\text{prédiction}(t+1), \text{obs}(t), \text{obs}(t-1), \dots, \text{obs}(t-N_i))$$

Dans ces travaux, nous avons adopté une stratégie récursive pour la prédiction de plusieurs horizons étant donné le temps d'exécution trop important des autres stratégies suite à la création d'un nouveau modèle à chaque itération. La problématique de prévisions des ventes peut être formulée comme suit :

$$\hat{v}_i(t+H) = f(\hat{v}_i(t+H-1), \dots, \hat{v}_i(t+1), v_i(t), v_i(t-1), \dots, v_i(t-N_i)) ; \forall i \in R$$

Avec $\hat{v}_i(t)$ la prédiction du produit i à l'instant t , H est l'horizon de prédiction, $v_i(t)$ les observations des ventes du produit i , f le modèle expliquant les prédictions futures en fonction des observations historiques des ventes et N_i le nombre des observations passées du produit i .

Comment réaliser la prédiction pour l'ensemble des produits commercialisés dans un tel catalogue correspondant à des séries temporelles avec peu de données et d'une façon automatique ?

Nous visons dans ces travaux à proposer une méthodologie qui permet d'automatiser et d'améliorer la qualité des prévisions des ventes d'une entreprise. Ce travail est effectué dans l'objectif de gagner en performance et en efficacité autour de la Supply Chain, tout en prenant en considération les caractéristiques des produits et de leurs données disponibles. Dans le paragraphe suivant, nous détaillons notre approche permettant de répondre à la problématique énoncée.

II.2. Préparation des données

Une série chronologique, comme son nom l'indique, est une séquence de points de données représentant une évolution d'une quantité dans le temps. Ce domaine a attiré l'attention des chercheurs au cours des dernières décennies en raison de ses différentes applications, notamment dans les domaines économique, financier et industriel. L'importance accordée à

ce domaine repose principalement sur l'objectif de prédire le comportement futur d'une série chronologique.

Dans cette étude de cas, nous avons principalement des séries chronologiques mensuelles de nature non stationnaire, et de longueur différente (de 72 mois pour les produits les plus anciens à 1 mois pour les produits les plus récents).

Nous avons mis en place quelques critères de choix à respecter pour identifier les produits à prédire :

- Le produit doit avoir au moins 8 mois d'ancienneté : cette contrainte est nécessaire pour que les modèles aient un minimum d'historique pour être entraînés.
- Le produit doit être présent sur le marché pendant les 6 derniers mois : cette contrainte évite les produits en fin de vie et les produits arrêtés et ne considère que les produits qui sont en vente actuellement.
- Le produit a été vendu au moins 10 fois depuis sa création et une fois par mois en moyenne : cette contrainte assure que la prédiction ne se fera pas sur des produits qui sont vendus rarement et pour lesquelles la mesure des ventes peut être perçue comme « fortement bruitée ».

Ces critères représentent des contraintes logiques à respecter et assurent le bon fonctionnement de la méthodologie et le bon usage des algorithmes mis en place.

Pour les produits récemment lancés, dans le premier cas, pour lequel nous ne disposons pas de données historiques, il est nécessaire de se reposer sur les avis des experts du terrain (les directeurs régionaux et les grossistes) étant donné que nous ne connaissons pas encore la réaction du marché envers ce nouveau produit. Dans le deuxième cas où le produit en remplace un ou des plus anciens, le nouveau peut se reposer sur l'historique des références remplacées. Ce remplacement des produits est proposé par le service marketing ou le bureau d'étude de l'entreprise.

Les références de produits de l'entreprise elm.leblanc peuvent être réparties principalement en 3 catégories. Les produits finis que sont les chaudières et les chauffe-eaux / chauffe-bains, les pièces détachées et enfin les accessoires. En respectant les critères cités, le jeu de données

correspond à environ 1500 références de produits (entre 2015 et 2021). Les données de chaque produit sont représentées par une série temporelle.

II.3. Description de la méthodologie

D'après l'étude bibliographique du chapitre précédent, les modèles employés par les auteurs des travaux dans leurs études de cas ont été choisis pour répondre à une problématique particulière et modéliser des données de ventes ayant certaines spécificités. L'objectif de ce travail est de proposer une méthodologie de prédiction des ventes en se basant sur les techniques des séries temporelles permettant de prendre en considération différentes natures et particularités des séries.

D'après le jeu de données à disposition, nous avons remarqué que les produits et leur façon de se vendre peuvent être très divers. Ainsi, l'utilisation d'une structure de modèle de prédiction unique risque d'introduire beaucoup de complexité dans celle-ci ou la rendre difficilement estimable, voire non estimable. Le choix d'un seul type de modèle peut être réussi pour quelques produits mais nous ne pouvons pas garantir des performances de qualité sur l'ensemble du catalogue. Notre réflexion nous amène alors à l'usage de plusieurs types de modèles afin de pouvoir correctement prédire les ventes pour l'ensemble du catalogue d'elm.leblanc. Plutôt que d'utiliser une seule structure de modèle additif avec de nombreuses composantes, autant utiliser plusieurs modèles avec moins de composantes.

Nous allons utiliser, dans notre méthodologie, quatre structures de modèles différentes. Le modèle ARIMA ainsi que sa dérivée SARIMA sont des techniques qui sont évaluées dans plusieurs applications et ont prouvé leurs performances. La famille des modèles de Lissage Exponentiel (LE) ayant différentes variétés qui permet de modéliser plusieurs formes de séries temporelles à travers ses composantes de tendance et de saisonnalité. Prophet est un modèle récent basé sur un modèle additif permettant la modélisation des tendances et des saisonnalités multiples dans une série chronologique.

La première étape consiste en la préparation et le nettoyage des données brutes qui sont l'ensemble des données de ventes, pour en extraire une base de données propre contenant l'ensemble des séries temporelles représentant chacune les ventes mensuelles pour une référence de produit. La deuxième étape est de sélectionner le type de modèle le plus adéquat

pour chaque série selon sa nature et selon le besoin de l'utilisateur, à savoir prédire le long ou le court terme.

Le diagramme de la figure 20 présente notre approche de choix de modèle de prédiction pour chaque produit du catalogue de l'entreprise elm.leblanc. Cette étape de sélection du type du modèle est la plus importante dans ce processus de prédiction.

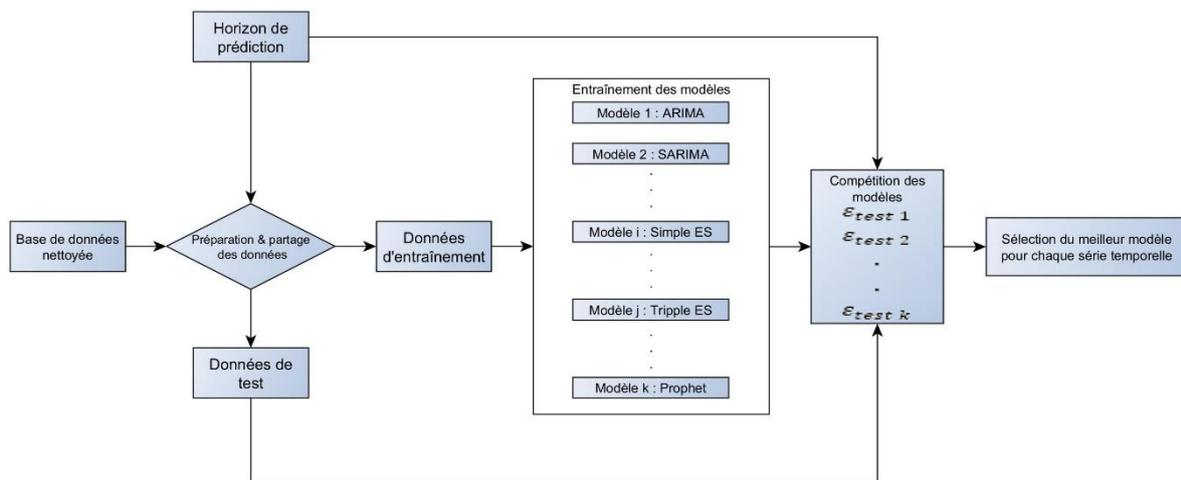


Figure 20 : Méthodologie de sélection des structures de modèles pour chaque série temporelle

Pour chaque série temporelle, la première phase consiste à séparer les données disponibles en base d'entraînement et base de test. Dans le catalogue de produits, nous disposons des produits récents ayant un nombre limité d'observations (quelques mois) ainsi que d'autres produits ayant plus de deux années d'historique. Pour les produits récents, nous allons entraîner uniquement des modèles non saisonniers. En effet, la saisonnalité telle que nous pourrions la définir pour les ventes de ce type de produits serait annuelle, et pour pouvoir capter la saisonnalité dans la série, il nous faudrait au minimum deux réalisations de la période complète (deux ans). Pour les produits les plus anciens, nous allons exploiter des modèles saisonniers et non saisonniers pour modéliser le comportement de vente des produits. Il est nécessaire de vérifier que, pour les modèles saisonniers, nous avons au moins un nombre d'observations strictement supérieur au nombre de leurs hyper-paramètres (Hyndman & Kostenko, 2007).

Pour chaque produit, nous allons entraîner les modèles sur la base de données d'entraînement et évaluer leurs performances sur la base de test dans le but de choisir le meilleur modèle qui correspond à chaque produit. D'abord, nous allons comparer des

modèles ayant la même structure pour identifier la meilleure en termes de performance et ensuite, comparer la meilleure structure de chaque type de modèle afin de sélectionner le plus performant pour chaque produit.

Pour le modèle ARIMA (paragraphe I.4.1), les paramètres « p » et « q » peuvent varier de 0 à 5 et le terme de différenciation « d » peut prendre les valeurs 0, 1 ou 2. Nous effectuons d'abord un entraînement de différentes structures en faisant varier les paramètres dans leurs intervalles selon l'algorithme Step-Wise présenté dans (Hyndman & Khandakar, 2008) et ensuite, identifier le modèle ayant l'AIC (« Akaike Information Criterion ») le plus faible. Pour le modèle SARIMA, les paramètres « p », « P », « q » et « Q » peuvent varier entre 0 et 5, les termes de différenciation « d » et « D » peuvent prendre les valeurs 0, 1 ou 2, et le paramètre de saisonnalité aura pour valeur $m = 12$. Le choix du meilleur modèle est effectué selon la même procédure que pour le modèle ARIMA. Pour les modèles de lissage exponentiel (LE), nous comparons plusieurs structures également en faisant varier la nature des composantes. Un modèle LE peut avoir une tendance additive qui peut être amortie ou non amortie, ou sans composante de tendance, et une saisonnalité qui peut être additive, multiplicative ou sans composante saisonnière. L'estimation des hyperparamètres de chaque structure est réalisée en maximisant la vraisemblance du modèle comme mentionné dans l'état de l'art. Nous choisissons ensuite la structure ayant les meilleures performances lors de l'évaluation. Pour le modèle Prophet, les hyperparamètres sont estimés par la méthode d'optimisation L-BFGS (« Limited-memory Broyden-Fletcher-Goldfarb-Shanno »).

Ensuite, pour la seconde étape, les meilleures structures de chaque type de modèle sont évaluées entre elles pour choisir celle qui performe le mieux. L'étape de test consiste à mettre en compétition ces modèles entraînés sur la base des données de test pour choisir la meilleure structure (Win_{model}) en minimisant l'erreur de test calculée pour chaque série (ε_{test}) selon l'équation suivante :

$$Win_{model} = argmin_k(\varepsilon_{test k})$$

Où k est le nombre de tests réalisés et ε_{test} la mesure d'erreur utilisée pour chaque modèle.

Le critère de compétition (i.e. la mesure d'erreur) et la technique utilisée pour l'apprentissage et le test des modèles seront expliqués au paragraphe II.5.

Ensuite, une fois que le meilleur modèle a été sélectionné pour chaque produit, nous allons réentraîner ce modèle sur l'ensemble des données (y compris les données utilisées en base de test lors de la compétition des modèles) dans le but d'affiner ses hyper paramètres afin d'avoir une meilleure modélisation du produit. Finalement, les prévisions sont calculées et renvoyées à l'utilisateur selon l'horizon demandé. Pour faire des prédictions à différents horizons, nous avons employé une stratégie récursive de prédiction.

Dans le paragraphe suivant, nous détaillons les structures de modèles employés dans notre approche ainsi que les critères de choix de ces techniques.

II.4. Modèles utilisés versus caractéristiques de la série temporelle

Nous allons présenter dans ce paragraphe les modèles retenus pour notre méthodologie ainsi que les hypothèses considérées permettant de modéliser les séries chronologiques correspondant aux signaux de vente.

Comme nous l'avons évoqué dans le chapitre précédent, la problématique de prévision des ventes peut se rapporter à des travaux sur les séries temporelles. Dans ce cadre, plusieurs modèles ont été développés dans la littérature afin de modéliser et prédire ces séries. Nous présentons ci-après les modèles qui ont été étudiés et utilisés dans nos travaux.

Le choix des modèles à utiliser dans notre méthodologie repose sur plusieurs critères liés à la nature des produits du catalogue d'elm.leblanc, à leurs ventes et aux caractéristiques des modèles. Pour pouvoir couvrir les comportements de vente de toutes les références du catalogue, nous avons considéré les spécificités de ces derniers. Nous disposons ainsi de produits vendus régulièrement sur toute l'année sans motif répétitif précis et qui ont un historique suffisant (supérieur à deux ans), des produits avec une forte saisonnalité annuelle surtout durant la saison hivernale, des produits dont les ventes sont impactées par des événements occasionnels comme les jours de fermeture de l'usine, ou encore des produits récemment lancés avec moins de 2 ans d'historique de vente. Dans le tableau 2, nous présentons les modèles choisis pour effectuer la prédiction des ventes permettant de prendre en compte les différents comportements.

Besoins Modèles	Ventes avec forte saisonnalité	Ventes des produits récents	Ventes évènementielles	Ventes sans motif répétitif
ARIMA		X		X
Prophet	X		X	
SARIMA	X			
LE Simple		X		X
LE Double		X		X
LE Triple	X			

Tableau 2 : Critère de choix des modèles possibles

Pour les produits avec un historique supérieur à deux ans n'ayant pas de motif répétitif clair et qui sont vendus régulièrement sur toute l'année, les modèles ARIMA et Lissage Exponentiel (LE) Simple et Double peuvent être utilisés. Le choix dépendra davantage du comportement de chaque produit, s'il représente une tendance croissante ou décroissante, ou encore une particularité spécifique durant une période de temps. Pour les produits ayant une forte saisonnalité, les modèles Prophet, SARIMA et le Lissage Exponentiel Triple sont les plus appropriés. Pour les produits récents (ayant moins de deux ans d'historique), les modèles ARIMA et Lissage Exponentiel Simple et Double permettent de mieux représenter ce type de série temporelle. Pour les ventes influencées par des événements spécifiques, Prophet est plus adéquat grâce à sa composante permettant de capturer cet aspect.

Comme (Hyndman & Kostenko, 2007) l'expliquent dans leur article, l'utilisation de modèles saisonniers nécessite un nombre minimum d'observations sur les données puisqu'elles doivent être au moins supérieures au nombre de paramètres du modèle à estimer. Ils soulignent que l'utilisation de la méthode Holt-Winters (LE Triple), par exemple, nécessite au moins 17 observations pour les données mensuelles, et pour un modèle SARIMA (0,1,1)(0,1,1)₁₂ au moins 16 observations. De plus, comme pour les séries temporelles présentant plusieurs variations, les modèles saisonniers nécessitent également davantage de données pour bien capturer leurs comportements. En se basant sur cet aspect, nos données de vente peuvent être séparées en deux groupes : séries chronologiques courtes avec moins de deux ans d'observations à prévoir avec des modèles non saisonniers et séries chronologiques longues avec plus de deux ans d'observations pouvant être représentées par

les modèles saisonniers. Comme mentionné dans (Hyndman, R.J. Athanasopoulos, 2019), les méthodes de lissage exponentiel avec tendances multiplicatives produisent plus souvent des prévisions de mauvaise qualité, nous avons ainsi décidé de les écarter.

Nous n'avons également pas considéré les modèles ARIMA(0,0,0) et ARIMA(0,1,0) puisque le premier correspond à une valeur au hasard et le deuxième correspond à un processus de marche aléatoire, donc ces deux modèles ne seront pas intéressants. De même pour les modèles SARIMA(0,0,0)(0,0,0), SARIMA(0,1,0)(0,0,0) et SARIMA(0,0,0)(0,1,0) qui ne seront pas intéressants pour la modélisation des ventes des produits.

II.4.1. Le modèle ARIMA

ARIMA (AutoRegressive Integrated Moving Average) est un modèle de prévision linéaire largement utilisé dans la littérature pour la prédiction des séries temporelles univariées. Ce modèle est principalement utilisé pour les séries temporelles généralement stationnaires c. à. d. séries chronologiques avec des propriétés statistiques indépendantes du temps, telles que la moyenne et la variance. Il s'agit d'une combinaison des modèles autorégressifs (AutoRegressive ou AR) de paramètre p , et de moyenne mobile (Moving Average ou MA) de paramètre q , avec un terme de différenciation (I) de paramètre d utilisé pour surmonter la non-stationnarité de certaines séries temporelles.

D'une façon générale, il est possible de modéliser les ventes $v_{t,i}$ d'un produit i au mois t à travers ARIMA(p, d, q) pour différentes valeurs de différenciation d en intégrant un opérateur de retard B défini comme $Bv_{t,i} = v_{t-1,i}$ selon l'équation suivante :

$$(1 - \varphi_1 B - \dots - \varphi_p B^p)(1 - B)^d v_{t,i} = c + (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_{t,i}$$

Où $\varepsilon_{t,i}$ est une variable d'erreur au mois t du produit i indépendante et identiquement distribuée avec une moyenne nulle et une variance constante, et φ_p et θ_q sont respectivement les coefficients d'auto-régression et de moyenne mobile du modèle.

La phase la plus importante du processus de prévision à l'aide du modèle ARIMA consiste à identifier les meilleurs paramètres permettant de bien modéliser les données et de fournir des prévisions précises. Dans la littérature, différentes méthodes sont proposées pour trouver ces paramètres. La technique la plus courante est de suivre la méthodologie Box-Jenkins. Elle

repose sur une approche itérative d'identification d'un modèle, d'estimation de ses paramètres et de vérification diagnostique pour déterminer le meilleur modèle.

Le choix des paramètres des modèles ARIMA est généralement basé sur les graphes de l'ACF (« AutoCorrelation Function » ou fonction d'autocorrélation) et la PACF (« Partial AutoCorrelation Function » ou fonction d'autocorrélation partielle) ou le calcul de l'AIC (« Akaike Information Criterion ») / BIC (« Bayesian Information Criterion »). Une autre méthode consiste simplement à essayer toutes les valeurs possibles des paramètres via « Grid Search », puis à les comparer par une métrique comme l'AIC. Une méthode plus intéressante et automatisée est l'algorithme Step-Wise de Hyndman-Khandakar décrit dans (Robin Hyndman et al., 2008). Il s'agit d'un processus automatisé d'identification du meilleur ordre en combinant des tests de racine unitaire, la minimisation de l'AICc (AIC corrigée) et de la MLE (Maximum Likelihood Estimation).

II.4.2. ARIMA Saisonnier : SARIMA

Il s'agit d'une extension du modèle ARIMA qui permet la modélisation et la prévision des aspects saisonniers. Nous employons ce modèle dans notre méthodologie pour prédire les ventes des produits représentant des motifs saisonniers. En plus des ordres ARIMA(p, d, q), ce modèle est caractérisé par des ordres supplémentaires (P, D, Q) m représentant la partie saisonnière du modèle.

Les ventes $v_{t,i}$ d'un produit i sur le mois t sont modélisées à travers SARIMA selon l'équation suivante :

$$\Phi_P(B^m) \varphi_p(B)(1 - B)^d (1 - B^m)^D v_{t,i} = \Theta_Q(B^m) \theta_q(L)\varepsilon_{t,i}$$

Les paramètres supplémentaires sont Φ d'ordre P et Θ d'ordre Q représentant respectivement les coefficients d'auto-régression saisonnière et de moyenne mobile saisonnière du modèle. B étant l'opérateur de rétrogradation défini auparavant, $\varepsilon_{t,i}$ est l'erreur au mois t du produit i , et m le nombre de périodes dans une saison.

Malgré le fait qu'il y ait plus de combinaisons possibles et de paramètres à rechercher les meilleurs paramètres pour SARIMA peuvent être choisis par les mêmes méthodes que celles présentées précédemment pour ARIMA.

11.4.3. Le lissage exponentiel

Le lissage exponentiel est une famille de modèles classiques de prévision de séries temporelles largement utilisée dans l'industrie. Ces modèles comprennent différents variants qui se distinguent par la nature de leurs composantes. Ils se caractérisent par plusieurs paramètres comme présentés dans le tableau 3.

		COMPOSANTE SAISONNIERE		
COMPOSANTE DE TENDANCE		Sans	Additive	Multiplicative
	Sans	α	α, γ	α, γ
	Additive	α, β	α, β, γ	α, β, γ
	Additive Amortie	α, β, ϕ	$\alpha, \beta, \phi, \gamma$	$\alpha, \beta, \phi, \gamma$

Tableau 3 : Paramètres des modèles de lissage exponentiel

Le lissage exponentiel simple, c'est-à-dire sans tendance ni saisonnalité, consiste à calculer des valeurs futures basées sur des moyennes pondérées des valeurs passées avec une décroissance exponentielle des poids. Il est caractérisé par le paramètre α représentant un taux d'apprentissage c'est-à-dire l'importance qui sera donnée à la dernière observation. Il est utilisé pour prédire le niveau (« level »), c'est-à-dire la valeur moyenne de la série.

Le lissage exponentiel double permet la modélisation de la tendance avec un deuxième paramètre β et de donner plus ou moins d'importance aux dernières observations comme le paramètre α pour le niveau. La tendance estimée peut-être constante ou avec un changement de monotonie en introduisant un paramètre d'amortissement ϕ . En effet, ce paramètre force la tendance à converger vers une valeur constante dans le futur à long terme.

Le triple lissage exponentiel ajoute une troisième composante permettant de capturer les effets saisonniers. Comme pour les autres composantes, l'effet saisonnier est déterminé par une méthode de pondération exponentielle avec un taux d'apprentissage γ . Les composantes de la tendance et de la saisonnalité peuvent être de nature additive ou multiplicative selon le comportement présent dans la série à modéliser. Dans le tableau 3 ci-dessus, nous avons présenté les paramètres associés aux modèles utilisés dans notre méthodologie.

Nous employons, dans notre méthodologie, plusieurs variétés de cette famille de modèles pour représenter les comportements de ventes des produits ayant différentes natures. Nous

présentons ci-dessous un exemple de modèle additif avec une tendance et une saisonnalité pour illustrer les composantes d'un tel modèle et les paramètres à estimer lors de la modélisation. Ce modèle est connu sous le nom : la méthode additive de Holt-Winters.

Nous désignons par $v_{t,i}$ les ventes d'un produit i au mois t , et $\hat{v}_{t+h,i}$ les prédictions recherchées pour le produit i pour un horizon $(t + h)$ avec h le nombre de mois à prédire.

$$\hat{v}_{t+h,i} = l_{t,i} + hb_{t,i} + s_{t+h-m(k+1),i}$$

Avec $l_{t,i}$: la composante du niveau (« level ») de la série i au mois t

$b_{t,i}$: la composante de la tendance de la série i au mois t

$s_{t+h-m(k+1),i}$: la composante de la saisonnalité de la série i

h : le nombre des mois prédire

m : le nombre des mois dans la saison ($m=12$ pour notre cas)

k : la partie entière du quotient $(h - 1)/m$

Les équations pour calculer les différentes composantes sont explicitées ci-dessous :

$$l_{t,i} = \alpha(v_{t,i} - s_{t-m,i}) + (1 - \alpha)(l_{t-1,i} + b_{t-1,i})$$

$$b_{t,i} = \beta(l_{t,i} - l_{t-1,i}) + (1 - \beta)b_{t-1,i}$$

$$s_{t,i} = \gamma(v_{t,i} - l_{t-1,i} - b_{t-1,i}) + (1 - \gamma)s_{t-m,i}$$

L'estimation des meilleurs paramètres $(\alpha, \beta, \phi, \gamma)$ peut se faire manuellement en essayant différentes valeurs possibles et en choisissant les meilleures, ou automatiquement via un outil d'optimisation (Hyndman, R.J. Athanasopoulos, 2019). La sélection du meilleur modèle peut se faire en prenant le modèle ayant le minimum d'erreur ou à travers la maximisation de la vraisemblance (le meilleur modèle aura la plus large vraisemblance).

II.4.4. Le modèle Prophet

Dans notre approche, nous avons également employé un modèle de prévision des séries temporelles basé sur les modèles additifs généralisés appelé « Prophet » qui est développé par l'équipe de recherche de Facebook. Il propose une structure additive intéressante du fait de son interprétabilité, sa flexibilité et son automatisation. Les développeurs de ce modèle

adoptent une approche « analyst-in-the-loop » permettant à l'analyste d'intervenir et de modifier les paramètres du modèle en cas de besoin.

Prophet possède trois composantes : tendance, saisonnalité et événements. Pour un signal de ventes $v_{t,i}$ du produit i au mois t , on peut le modéliser comme suit :

$$v_{t,i} = g_{t,i} + s_{t,i} + h_{t,i} + \varepsilon_{t,i}$$

Où $g_{t,i}$ la composante de la tendance, $s_{t,i}$ la composante de saisonnalité, $h_{t,i}$ la composante des événements liés au produit i , et $\varepsilon_{t,i}$ l'erreur de calcul.

L'aspect linéaire ou non-linéaire de la tendance $g_{t,i}$ peut-être représenté par deux types de modèles. Un modèle non-linéaire de croissance logistique par morceaux (« Piecewise logistic growth model ») utilisé lorsque l'analyste sait que la tendance atteindra un maximum à un moment donné. Le deuxième type représente une tendance linéaire par le biais d'un modèle linéaire par morceaux « Piecewise linear model » qui permet de prendre en considération une tendance avec plusieurs taux de croissance constants.

Une caractéristique intéressante de la modélisation de la tendance est l'aspect « points de changement » où l'utilisateur peut spécifier des dates spécifiques où la tendance peut changer en fonction de ses connaissances sur la série étudiée. Cela inclut notamment les promotions lors des ventes de certains produits spécifiques ou lorsqu'il y a une campagne publicitaire pour augmenter les ventes.

La composante de saisonnalité $s_{t,i}$ est modélisée par une série de Fourier standard qui peut gérer la saisonnalité annuelle, mensuelle et hebdomadaire. Le modèle Prophet peut gérer l'aspect additif ou multiplicatif de la saisonnalité de la série temporelle.

La troisième composante $h_{t,i}$ de Prophet représente la modélisation des vacances et des événements spéciaux. L'utilisateur insère les dates passées et futures directement et pour chaque date, associe un paramètre afin de le prendre en compte avec plus ou moins d'importance lors des prévisions. Cette composante est intéressante lors de la prévision des ventes permettant d'insérer les jours fériés et les jours de fermeture annuelle de l'usine et des services commerciaux.

Dans notre cas d'étude, nous exploitant le modèle Prophet avec une tendance linéaire pour les différents produits. La composante saisonnière est estimée pour chaque produit en

prenant en considération sa saisonnalité annuelle. Nous n'exploitons pas la composante des vacances et événements spéciaux vu que les dates de la fermeture de l'usine sont reflétées dans la baisse saisonnière des ventes durant cette période de chaque année, et les dates des promotions sont plutôt spécifiques à chaque produit, et par la suite doivent être introduites séparément pour chacun.

II.5. Processus d'apprentissage et critère de compétition

Comme présenté ci-dessus, le choix de modèle adéquat pour chaque série temporelle passe par une phase d'entraînement et de mise en compétition des modèles. Dans la littérature, on trouve plusieurs possibilités (dépendant fortement de la quantité de données disponibles) mais généralement, on a un partage de 75% des données pour l'entraînement et 25% pour le test. Dans notre cas d'étude, vu la quantité faible de données à disposition, les données sont divisées en ensembles d'entraînement de 90% et de test de 10%. Ceci est justifié également par les expérimentations qui ont été conduites.

Pour le catalogue de produits de l'entreprise elm.leblanc, nous avons effectué l'étape d'apprentissage de chaque produit sur 90 % des données et la phase de test et de comparaison des modèles sur les 10 % restant. Le choix du modèle gagnant est effectué dans la phase de mise en compétition des modèles sur les données de test. La mesure d'erreur employée est la racine de l'erreur quadratique moyenne (RMSE) définie comme :

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

Avec v_i les valeurs observées, \hat{v}_i les valeurs prédites et n le nombre de prédictions.

Le choix du meilleur type de modèle pour chaque série se fait en deux phases. La première consiste à choisir le meilleur modèle correspondant à chaque structure utilisée, et la deuxième consiste à mettre en compétition ces modèles choisis entre eux en calculant l'erreur réalisée sur les données de test, dans le but de déterminer le modèle le plus adéquat à chaque produit. L'entraînement des modèles se fait selon la stratégie récursive présentée dans le paragraphe II.1.

Une fois le modèle pour chaque série choisie, nous le réentraînons sur l'ensemble de ses données historiques afin de mettre à jour ses paramètres en ajoutant les données utilisées

durant la phase de test dans la base d'entraînement, et nous calculons ensuite les prédictions selon l'horizon demandé. Pour calculer les prédictions pour plusieurs horizons, la valeur prédite par le modèle sera utilisée pour calculer le point suivant en utilisant le même modèle déjà entraîné (principe de récursivité).

III. Expérimentations et validation de la méthodologie

L'évaluation et la validation de notre approche ont été réalisées au sein de l'entreprise elm.leblanc spécialisée dans la fabrication des systèmes de chauffage résidentiels et industriels, sur des données réelles de ventes de plusieurs produits. À travers l'amélioration de ses prévisions de ventes, l'entreprise vise à réduire les coûts de stockage, améliorer la planification de la production et optimiser les performances globales de sa Supply Chain.

III.1. Expérimentations réalisées

Pour réaliser les différentes expérimentations, nous avons utilisé l'historique des données de ventes représentées par des séries chronologiques de plusieurs types de produits entre janvier 2015 et janvier 2021 inclus. L'évaluation est effectuée en calculant l'écart (erreur absolue) entre les prédictions des modèles et les valeurs réelles. Pour des raisons de confidentialité des données, nous avons « anonymisé » les produits étudiés.

Nous avons mené ces expérimentations sur l'horizon court (H1) et sur l'horizon moyen / long (H3). Pour ces expérimentations, nous avons utilisé plusieurs produits appartenant à différentes familles et ayant divers comportements de vente afin de discuter autour de la performance de notre méthodologie face à différentes situations. Nous avons comparé les résultats de chaque expérience avec les prévisions faites manuellement selon la méthodologie utilisée dans l'entreprise. Nous avons réalisé ces expérimentations sur les mois d'octobre, novembre et décembre de l'année 2018 et sur les mêmes mois de l'année 2019.

Pour comparer notre approche avec la méthode de prévision manuelle, nous avons retenu comme indicateur l'erreur moyenne absolue (MAE) sur l'ensemble des produits, calculé comme suit :

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

Où \hat{v}_i est la prévision des ventes du produit i , v_i est la vente réelle du produit i et n le nombre des produits pris en considération.

III.1.1. Données de l'année 2018

Dans ce paragraphe, nous allons évaluer la mise en application de notre méthodologie sur le dernier trimestre de l'année 2018. Durant ces expérimentations, nous avons travaillé sur les 300 références de produits les plus vendues dans l'entreprise représentant ainsi la plus grande partie du chiffre d'affaires. Ces expérimentations correspondent au calcul des prévisions des ventes pour différents horizons à travers la méthodologie proposée et la comparaison des résultats avec la méthode appliquée dans l'entreprise. Nous allons présenter les résultats pour l'horizon H1 du dernier trimestre de l'année 2018, c'est-à-dire les prédictions à une seule étape des mois d'octobre, novembre et décembre, et pour l'horizon H3 qui correspondent aux prédictions faites au mois d'octobre sur 3 mois. En adéquation avec notre méthodologie, les modèles peuvent varier d'un produit à l'autre et pour le même produit à l'acquisition de nouvelles données de ventes.

Dans l'histogramme de la figure 21, nous présentons quelques exemples des 20 produits les plus vendus avec les quantités prédites par notre approche, par rapport aux quantités prédites manuellement et les ventes réelles pour le dernier trimestre de 2018 et pour un horizon H3.

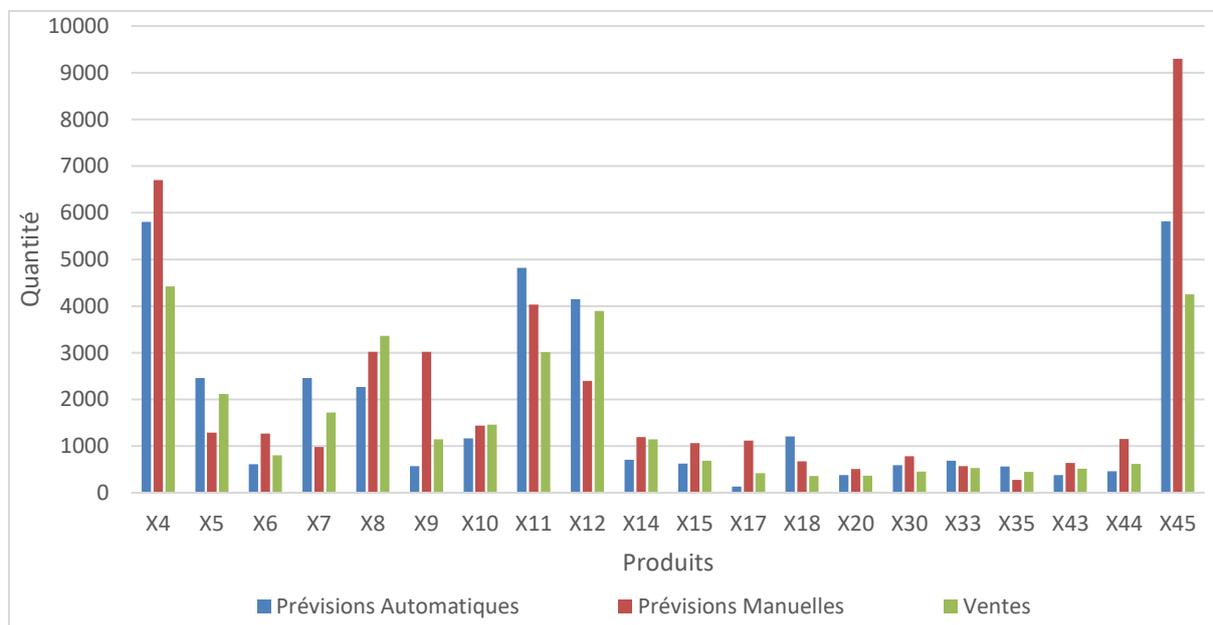


Figure 21 : Comparaisons des quantités prédites automatiquement et manuellement par rapport aux ventes réelles des exemples de produits pour l'horizon H3 de 2018

Dans ce graphique, une barre représente la somme des quantités prédites ou vendues sur les 3 mois pour lesquelles la prédiction a été faite. Comme nous pouvons le constater, pour plusieurs produits, notre approche donne des prévisions plus proches des quantités réelles

des ventes que les prévisions effectuées manuellement par l'entreprise, en particulier pour les produits avec des quantités importantes (sauf X8 et X11). Cependant, pour certains produits, notre méthodologie était moins précise, ce qui est peut-être expliqué par l'introduction des promotions sur ces produits ou encore car ils sont récemment lancés et donc avec des séries chronologiques courtes tel est le cas du produit X8.

Dans le tableau 4, nous présentons les erreurs absolues entre les ventes réalisées et les prévisions de notre approche, notées erreurs automatiques, ou les ventes réalisées et les prévisions manuelles faites selon l'expertise elm.leblanc, notées erreurs manuelles, pour les horizons H1 et H3 de 2018 pour plusieurs produits en vente. Nous présentons également, dans la colonne des pourcentages par rapport au total, le pourcentage d'erreur que représente chaque produit par rapport au total des erreurs de l'ensemble des produits dans le tableau. Dans la dernière ligne, nous présentons l'erreur moyenne absolue (MAE) sur toutes les références pour les deux horizons également.

Produit	H3				H1			
	Erreur Manuel	% par rapport au total	Erreur Automatique	% par rapport au total	Erreur Manuel	% par rapport au total	Erreur Automatique	% par rapport au total
X1	366	1,65%	158	0,97%	366	1,58%	153	0,93%
X2	42	0,19%	73	0,45%	42	0,18%	75	0,46%
X3	121	0,55%	118	0,72%	121	0,52%	143	0,87%
X4	2280	10,28%	1384	8,49%	2260	9,76%	1935	11,74%
X5	826	3,72%	341	2,09%	831	3,59%	318	1,93%
X6	472	2,13%	311	1,91%	467	2,02%	520	3,16%
X7	979	4,41%	1515	9,29%	979	4,23%	1426	8,65%
X8	341	1,54%	1096	6,72%	581	2,51%	955	5,80%
X9	1878	8,46%	572	3,51%	1638	7,07%	542	3,29%
X10	119	0,54%	329	2,02%	359	1,55%	410	2,49%
X11	1018	4,59%	1803	11,06%	1018	4,40%	999	6,06%
X12	1494	6,73%	992	6,09%	1494	6,45%	1377	8,36%
X13	321	1,45%	191	1,17%	321	1,39%	276	1,68%
X14	504	2,27%	439	2,69%	1066	4,60%	459	2,79%
X15	380	1,71%	102	0,63%	550	2,38%	143	0,87%
X16	207	0,93%	172	1,06%	159	0,69%	174	1,06%
X17	869	3,92%	288	1,77%	869	3,75%	268	1,63%

X18	317	1,43%	848	5,20%	334	1,44%	597	3,62%
X19	21	0,09%	38	0,23%	21	0,09%	38	0,23%
X20	145	0,65%	118	0,72%	145	0,63%	156	0,95%
X21	291	1,31%	102	0,63%	201	0,87%	107	0,65%
X22	92	0,41%	106	0,65%	83	0,36%	113	0,69%
X23	210	0,95%	70	0,43%	165	0,71%	59	0,36%
X24	194	0,87%	98	0,60%	199	0,86%	127	0,77%
X25	120	0,54%	135	0,83%	120	0,52%	148	0,90%
X26	136	0,61%	137	0,84%	136	0,59%	61	0,37%
X27	128	0,58%	72	0,44%	128	0,55%	91	0,55%
X28	121	0,55%	135	0,83%	121	0,52%	99	0,60%
X29	97	0,44%	141	0,86%	97	0,42%	123	0,75%
X30	323	1,46%	160	0,98%	388	1,68%	266	1,61%
X31	191	0,86%	114	0,70%	191	0,82%	123	0,75%
X32	92	0,41%	89	0,55%	92	0,40%	117	0,71%
X33	245	1,10%	217	1,33%	245	1,06%	252	1,53%
X34	272	1,23%	150	0,92%	272	1,17%	106	0,64%
X35	210	0,95%	113	0,69%	210	0,91%	135	0,82%
X36	38	0,17%	89	0,55%	38	0,16%	67	0,41%
X37	151	0,68%	76	0,47%	191	0,82%	122	0,74%
X38	149	0,67%	185	1,13%	149	0,64%	119	0,72%
X39	113	0,51%	169	1,04%	113	0,49%	100	0,61%
X40	183	0,82%	135	0,83%	157	0,68%	152	0,92%
X41	203	0,91%	106	0,65%	203	0,88%	167	1,01%
X42	166	0,75%	220	1,35%	266	1,15%	265	1,61%
X43	183	0,82%	154	0,94%	188	0,81%	104	0,63%
X44	530	2,39%	209	1,28%	530	2,29%	374	2,27%
X45	5051	22,76%	2231	13,69%	5051	21,81%	2116	12,84%
MAE	493,09	100%	362,24	100%	514,56	100%	366,16	100%

Tableau 4 : Erreurs de prévisions pour différents produits sur les horizons H1 et H3 de 2018 selon la méthode employée (manuelle ou automatique)

D'après le tableau 4, nous constatons que, pour plusieurs produits, notre approche donne de meilleurs résultats (erreur plus faible) que la méthode manuelle pour les horizons H1 et H3. C'est notamment le cas pour le moyen terme (H3) où nous remarquons une nette amélioration pour les produits X4, X5, X12, X45 qui sont considérés comme des « High Runners » c'est-à-dire des produits toujours vendus avec des grandes quantités. Néanmoins,

pour le produit X8 par exemple, notre approche a donné des résultats moins satisfaisants que la méthode manuelle sur cet horizon (H3) en étant moins optimiste sur les prédictions. Ceci est expliqué par des augmentations brusques et importantes des demandes clients justifiées par des chantiers ou des promotions. Le calcul des prévisions par les deux méthodes pour ce produit (manuelle et automatique) était ainsi impacté par cette augmentation comme nous pouvons le constater et a donc amené à des erreurs importantes selon les deux méthodes.

Pour des produits comme le X11 ou X18, notre approche était moins performante que l'approche manuelle également, en prévoyant des quantités plus importantes que les ventes réelles. C'est le cas aussi de la méthode manuelle où nous pouvons remarquer de grandes erreurs mais avec moins d'importance que la méthode automatique. Cet aspect peut s'expliquer par la rupture de stock de certains composants ou des défauts de qualité constatés lors de la réception des matières premières rendant les produits moins disponibles à la vente.

Nous pouvons constater également sur le tableau que, pour certains produits comme le X9, X15 ou le X17, les prévisions calculées par notre approche sont meilleures pour le moyen terme H3 que pour le court terme H1. Pour d'autres produits comme le X1 ou le X7, le phénomène inverse se produit pour les deux horizons. Cependant, au niveau de l'erreur absolue moyenne (MAE) sur tous les produits, notre approche était légèrement plus précise sur l'horizon H3 que sur l'horizon H1 avec une MAE respective de 362,24 et 366,16. La même remarque peut être faite également pour la méthode manuelle ayant respectivement une MAE de 493,09 pour H3 et 514,56 pour H1. La dégradation de la qualité des prévisions en passant de H3 à H1 à la fois pour la méthode automatique et manuelle, peut être expliqué par une rupture de stock des matières premières dans cette période, ce qui a impacté les ventes et donc des retards de livraison des produits chez les clients, impliquant ainsi des erreurs de prédiction plus importantes sur l'horizon H1.

Au niveau des performances globales, l'application de notre approche a permis un taux d'amélioration intéressant du MAE d'environ 26 % pour le moyen terme H3 et d'environ 29 % pour le court terme H1 par rapport à la méthode manuelle. Ce taux est calculé en faisant le rapport entre le gain en MAE (la différence) et la valeur du MAE manuelle (initiale). Ces améliorations touchent plusieurs produits dont les « High Runners » permettant ainsi d'avoir des profits économiques significatifs pour l'entreprise. Cependant, notre approche était moins bonne sur quelques produits qui présentaient parfois une énorme variabilité au fil des années

ou des courtes séries temporelles et donc très peu de données ou encore subissent des facteurs externes tels que des promotions et des événements spécifiques. Globalement, notre méthodologie apporte une amélioration importante par rapport à la méthode manuelle sur cette période étudiée.

Comme expliqué ci-dessus, notre méthodologie s'appuie sur différents modèles de prédiction des séries temporelles associés à chacune des références du catalogue produit. Nous présentons dans les tableaux suivants les résultats d'entraînement des modèles pour plusieurs produits. Les valeurs correspondent aux mesures d'erreurs (RMSE) calculées lors de la phase d'entraînement des modèles pour choisir la meilleure structure. Nous présentons également le meilleur modèle choisi pour chaque produit et les paramètres associés à ce dernier.

Le tableau 5 montre les erreurs d'entraînement des modèles pour différents produits pour un horizon H3 pour les prévisions du mois d'octobre 2018. Les valeurs marquées en vert correspondent aux erreurs minimales identifiées lors de la phase de compétition dans la méthodologie et donc la structure du modèle choisi pour chaque produit. Nous rappelons que pour le lissage exponentiel, nous pouvons avoir une tendance additive amortie ou non amortie et une saisonnalité de nature additive ou multiplicative. Un modèle de lissage exponentiel peut avoir une composante de tendance seulement sans saisonnalité (LE Double), une composante saisonnière sans tendance (LE Double) ou les deux composantes en même temps (LE Triple).

Produit	Lissage Exp (LE)	ARIMA	SARIMA	Prophet	Choix	Paramètres
X1	16	48	36	18	LE Triple	['add', False, 'add']
X2	13	31	-	18	LE Triple	['add', False, 'add']
X3	38	-	-	47	LE Double	['add', True, None]
X4	340	435	423	798	LE Triple	['add', False, 'add']
X5	725	915	-	392	Prophet	-
X6	70	44	-	144	ARIMA	[1, 0, 0]
X7	342	382	-	307	Prophet	-
X8	214	-	331	274	LE Double	['add', True, None]
X9	140	151	-	127	Prophet	-
X10	60	282	282	132	LE Triple	['add', True, 'mul']

X11	241	363	309	352	LE Triple	['add', False, 'add']
X12	243	305	-	181	Prophet	-
X13	59	47	47	65	ARIMA	[2, 0, 0]
X14	113	193	193	496	LE Triple	['add', False, 'add']
X15	67	-	93	136	LE Double	['add', False, None]
X16	33	-	-	52	LE Double	['add', True, None]
X17	30	199	157	128	LE Double	['add', False, None]
X18	154	158	158	182	LE Double	['add', True, None]
X19	13	13	-	12	Prophet	-
X20	27	-	32	36	LE Triple	['add', True, 'mul']
X21	16	25	-	20	LE Double	[None, False, 'add']
X22	6	20	-	13	LE Triple	['add', True, 'add']
X23	36	31	30	29	Prophet	-
X24	10	15	-	19	LE Double	['add', True, None]
X25	20	25	20	27	LE Double	[None, False, 'mul']
X26	24	37	-	27	Prophet	-
X27	22	36	24	39	LE Triple	['add', False, 'mul']
X28	21	23	38	38	LE Double	[None, False, 'mul']
X29	22	26	63	54	LE Double	[None, False, 'mul']
X30	50	34	-	152	ARIMA	[1, 0, 0]
X31	42	58	72	81	LE Triple	['add', True, 'mul']
X32	26	24	24	31	ARIMA	[1, 0, 0]
X33	101	-	-	115	LE Double	[None, False, 'add']
X34	26	-	55	56	LE Triple	['add', True, 'mul']
X35	35	75	72	70	LE Double	[None, False, 'mul']
X36	11	16	13	13	LE Triple	['add', True, 'mul']
X37	24	-	30	29	LE Triple	['add', False, 'mul']
X38	23	-	-	30	LE Double	['add', False, None]
X39	116	82	-	85	ARIMA	[2, 0, 1]
X40	75	47	-	61	ARIMA	[0, 0, 1]
X41	71	-	-	53	Prophet	-
X42	11	87	87	41	LE Double	['add', True, None]
X43	39	38	65	54	ARIMA	[1, 1, 1]

X44	45	35	75	74	ARIMA	[1, 0, 0]
X45	212	387	387	284	LE Triple	['add', False, 'mul']

Tableau 5 : Mesure d'erreur (RMSE) à l'entraînement pour un horizon H3 et choix du modèle par produit pour les prévisions au mois d'octobre 2018

Pour les paramètres des modèles de lissage exponentiel, le premier élément correspond à la tendance (additive, multiplicative ou sans composante), le deuxième à l'amortissement de la tendance (amortie ou pas) et le troisième à la saisonnalité (additive, multiplicative ou sans composante). Le tableau 6 présente également les erreurs d'entraînement des différents modèles pour plusieurs produits pour un horizon H1 pour les prévisions du mois de décembre 2018.

Produit	Lissage Exp (LE)	ARIMA	SARIMA	Prophet	Choix	Paramètres
X1	25	-	33	45	LE Double	['add', False, None]
X2	17	-	-	28	LE Triple	['add', False, 'mul']
X3	60	60	60	67	ARIMA	[0, 1, 1]
X4	261	417	335	322	LE Double	[None, False, 'mul']
X5	135	-	-	140	LE Double	['add', False, None]
X6	205	177	-	160	Prophet	-
X7	461	560	-	499	LE Double	['add', False, None]
X8	364	-	419	345	Prophet	-
X9	227	244	-	187	Prophet	-
X10	64	-	-	111	LE Double	[None, False, 'mul']
X11	370	284	340	449	ARIMA	[0, 0, 1]
X12	306	323	262	278	SARIMA	[(0, 0, 0), (1, 0, 0, 12)]
X13	64	71	71	72	LE Triple	['add', False, 'mul']
X14	196	160	160	529	ARIMA	[0, 1, 2]
X15	55	-	-	99	LE Double	['add', False, None]
X16	40	-	-	63	LE Double	['add', True, None]
X17	92	104	133	121	LE Triple	['add', False, 'mul']
X18	172	185	185	207	LE Double	['add', True, None]
X19	14	16	-	12	Prophet	-
X20	31	-	38	37	LE Triple	['add', False, 'mul']
X21	23	50	-	29	LE Triple	['add', False, 'mul']

X22	32	38	-	6	Prophet	-
X23	18	30	-	21	LE Triple	['add', False, 'mul']
X24	23	36	-	16	Prophet	-
X25	24	37	28	29	LE Triple	['add', True, 'add']
X26	10	-	39	40	LE Double	['add', True, None]
X27	22	31	21	53	SARIMA	[(0, 0, 1), (1, 0, 0, 12)]
X28	28	31	26	38	SARIMA	[(0, 0, 0), (2, 1, 1, 12)]
X29	18	36	27	27	LE Triple	['add', True, 'add']
X30	83	84	131	180	LE Double	['add', False, None]
X31	6	-	-	40	LE Double	['add', False, None]
X32	30	-	33	33	LE Double	[None, False, 'mul']
X33	57	63	63	54	Prophet	-
X34	18	-	27	50	LE Double	[None, False, 'add']
X35	39	70	54	53	LE Double	[None, False, 'add']
X36	20	16	16	22	ARIMA	[3, 0, 2]
X37	27	-	41	22	Prophet	-
X38	41	-	-	43	LE Double	['add', True, None]
X39	38	53	53	75	LE Double	['add', False, None]
X40	26	40	62	58	LE Double	[None, False, 'mul']
X41	45	-	-	48	LE Double	['add', True, None]
X42	57	73	73	51	Prophet	-
X43	57	78	63	51	Prophet	-
X44	48	87	90	54	LE Triple	['add', True, 'add']
X45	367	390	390	490	LE Double	[None, False, 'add']

Tableau 6 : Mesure d'erreur (RMSE) à l'entraînement pour un horizon H1 et choix du modèle par produit pour les prévisions au mois de décembre 2018

Dans les deux tableaux 5 et 6, certains produits ont des valeurs manquantes. Cela peut être dû au fait que la série temporelle soit trop courte pour entraîner un modèle comme SARIMA par exemple, ou encore au fait que le modèle ne fait pas partie de l'ensemble des modèles considérés (paragraphe II.4) lors de l'optimisation de ses paramètres, comme le modèle ARIMA(0,0,0). Nous pouvons remarquer que l'usage des modèles est très varié selon la nature du produit considéré, avec une dominance des modèles de lissage exponentiel grâce à la variété des combinaisons possibles des modèles selon leurs composantes, ce qui donne une

meilleure adaptation aux produits. Nous remarquons également pour certains produits comme le X5 ou le X17 au tableau 5 et le X4 ou le X11 au tableau 6, qu'il y a un modèle largement meilleur que les autres en compétition, et donc ce produit a un comportement de vente très spécifique et est bien modélisé par le modèle choisi. En revanche, pour d'autres produits comme le X9, le X18 ou le X23 au tableau 5 et le X3, le X13 ou le X33 au tableau 6, nous remarquons que, malgré la nature différente des modèles, les résultats de compétition étaient très proches et parfois avec des écarts minimes. Ceci peut remettre en question le choix d'un seul modèle par produit notamment lorsque les modèles sont de différentes natures et donc ce choix peut impacter les prévisions surtout sur le moyen et long terme. Par exemple, pour le produit X17 du tableau 6, le modèle ARIMA et le LE Tripple sont très proches en termes de performances, bien qu'ils aient deux natures différentes (modèle saisonnier et non saisonnier), alors que sur le moyen ou le long terme les résultats seront davantage différents. Cet aspect peut pousser la réflexion à l'usage d'une combinaison de modèles au lieu d'un seul pour exploiter les particularités de chaque modèle.

III.1.2. Données de l'année 2019

Pour s'assurer du bon fonctionnement de la méthodologie et sa validation, nous l'avons testée également sur une autre période du temps. Nous avons réalisé des mêmes expérimentations sur le dernier trimestre de l'année 2019 pour les 300 références de produits les plus vendues représentant la majorité du chiffre d'affaires de l'entreprise et pour les horizons H1 et H3. Les prévisions sont calculées à un horizon H3, c'est-à-dire en mois d'octobre sont calculées pour les 3 prochains mois, et à un horizon H1, c'est-à-dire des prédictions à une seule étape à chaque mois du trimestre. Pour chaque produit, nous avons employé notre procédure de choix de modèle et donc le meilleur modèle est identifié pour chaque produit. Notant que certains changements au niveau des produits sont introduits à travers le lancement des nouveaux produits remplaçant des anciens et l'arrêt d'autres produits.

L'histogramme de la figure 22 montre les quantités prédites par notre approche, par rapport aux quantités prédites manuellement et les ventes réelles pour le dernier trimestre de 2019 et pour un horizon H3, des 20 produits les plus vendus. Comparant avec les produits en 2018, nous remarquons l'introduction des nouveaux produits dans cette liste des produits les plus vendus, grâce à une augmentation importante dans leurs quantités de ventes sur l'année

2019, ou le fait qu'ils remplacent des anciens produits et donc récupèrent leurs parts de marchés, ou encore un lancement des produits très demandés sur le marché.

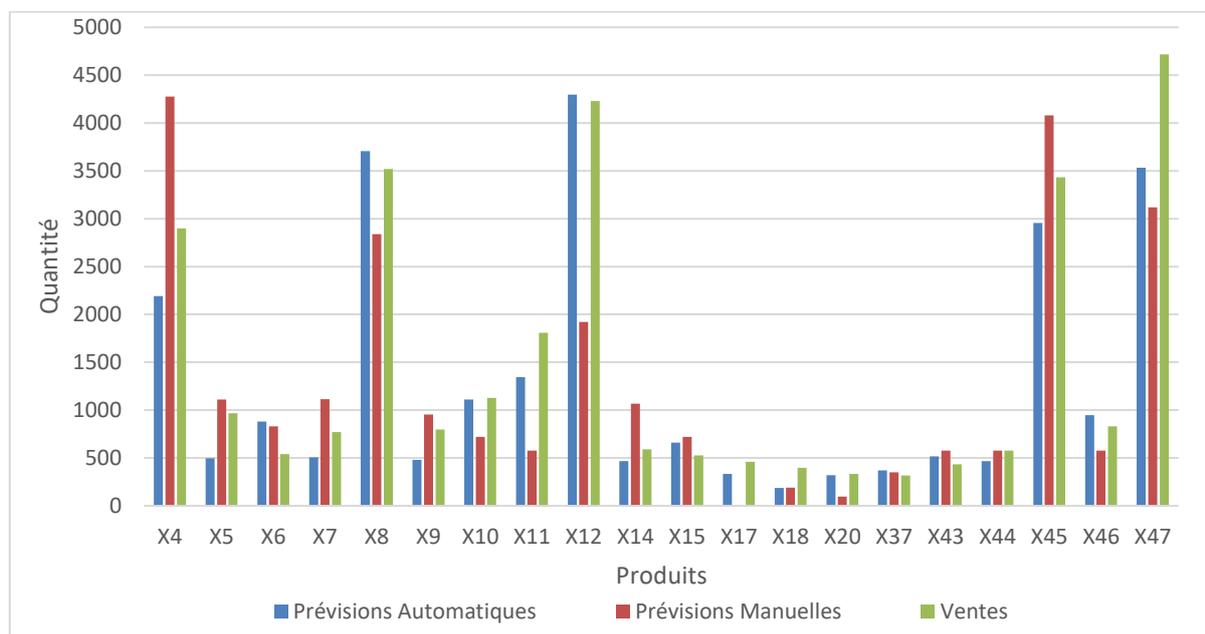


Figure 22 : Comparaisons des quantités prédites automatiquement et manuellement par rapport aux ventes réelles des exemples de produits pour l'horizon H3 de 2019

Comme l'année 2018, pour plusieurs produits, notre approche donne des prévisions plus précises que les prévisions effectuées manuellement par l'entreprise, en particulier pour les produits avec des quantités importantes comme le X4, X8 et X12. Cependant, pour certains produits impactés par des promotions, notre méthodologie était moins réussie et les prévisions étaient moins précises.

Dans le tableau 7, nous présentons les erreurs absolues entre les ventes réalisées et les prévisions de notre approche, notée erreur automatique, et entre les ventes réalisées et les prévisions manuelles faites selon l'expertise elm, notée erreur manuelle, pour les horizons H1 et H3 de 2019 pour plusieurs produits. Dans la dernière ligne, nous présentons l'erreur moyenne absolue (MAE) sur toutes les références pour les deux horizons également.

Produit	H3				H1			
	Erreur Manuelle	% par rapport au total	Erreur Automatique	% par rapport au total	Erreur Manuelle	% par rapport au total	Erreur Automatique	% par rapport au total
X2	73	0,50%	44	0,50%	78	0,59%	43	0,46%
X3	69	0,47%	21	0,24%	65	0,49%	14	0,15%

X4	1380	9,39%	736	8,42%	1489	11,18%	989	10,49%
X5	157	1,07%	471	5,39%	172	1,29%	365	3,87%
X6	292	1,99%	364	4,17%	313	2,35%	361	3,83%
X7	342	2,33%	264	3,02%	371	2,79%	168	1,78%
X8	765	5,21%	187	2,14%	460	3,45%	428	4,54%
X9	209	1,42%	317	3,63%	261	1,96%	396	4,20%
X10	864	5,88%	457	5,23%	1104	8,29%	564	5,98%
X11	1371	9,33%	461	5,28%	1371	10,30%	415	4,40%
X12	2309	15,72%	998	11,42%	2367	17,78%	990	10,50%
X13	125	0,85%	219	2,51%	61	0,46%	214	2,27%
X14	477	3,25%	123	1,41%	195	1,46%	99	1,05%
X15	194	1,32%	142	1,63%	134	1,01%	140	1,49%
X16	240	1,63%	24	0,27%	101	0,76%	46	0,49%
X17	459	3,12%	230	2,63%	646	4,85%	342	3,63%
X18	206	1,40%	210	2,40%	176	1,32%	145	1,54%
X19	308	2,10%	43	0,49%	303	2,28%	46	0,49%
X20	238	1,62%	23	0,26%	239	1,79%	32	0,34%
X21	62	0,42%	34	0,39%	35	0,26%	40	0,42%
X22	68	0,46%	26	0,30%	42	0,32%	25	0,27%
X23	57	0,39%	34	0,39%	58	0,44%	38	0,40%
X24	30	0,20%	34	0,39%	35	0,26%	25	0,27%
X30	134	0,91%	89	1,02%	299	2,25%	207	2,20%
X37	127	0,86%	101	1,16%	83	0,62%	95	1,01%
X38	141	0,96%	26	0,30%	73	0,55%	26	0,28%
X39	7	0,05%	110	1,26%	7	0,05%	45	0,48%
X40	72	0,49%	105	1,20%	72	0,54%	107	1,14%
X41	143	0,97%	5	0,06%	103	0,77%	6	0,06%
X42	169	1,15%	448	5,13%	7	0,05%	153	1,62%
X43	143	0,97%	115	1,32%	157	1,18%	125	1,33%
X44	95	0,65%	131	1,50%	217	1,63%	191	2,03%
X45	1408	9,59%	477	5,46%	1168	8,77%	849	9,01%
X46	359	2,44%	346	3,96%	359	2,70%	380	4,03%
X47	1596	10,87%	1322	15,13%	694	5,21%	1316	13,96%
MAE	419,69	100%	249,63	100%	380,43	100%	269,29	100%

Tableau 7 : Erreurs de prévisions pour différents produits sur les horizons H1 et H3 de 2019 selon la méthode employé (manuelle ou automatique)

D'après le tableau 7, pour plusieurs produits tels que le X4, le X12 et le X45 considérés comme des « High Runners », notre approche présente des erreurs plus faibles que la méthode manuelle pour les deux horizons H1 et H3. Ces gains en précision permettent à l'entreprise de limiter ces pertes liées au surstockage ou à la rupture des stocks sur ces produits. Nous pouvons également noter que pour quelques produits comme le X10 et le X17, la précision des prévisions s'est dégradée à la fois par la méthode manuelle et automatique en passant de l'horizon H3 à l'horizon H1. Ce phénomène s'explique par des ruptures en matières premières ou par la mise en place d'une campagne de promotion impliquant une augmentation brusque des ventes.

Pour d'autres produits comme le X47, la méthode manuelle a fourni des prévisions plus précises à l'horizon H1 qu'à l'horizon H3, pour lequel la méthode automatique était plus précise. Ceci est expliqué par une augmentation brusque et importante des demandes clients justifiée par des marchés de chantiers qui ont été validés récemment et auxquels il faut envoyer des quantités supplémentaires, donc ces quantités ont été prises en considération lors du calcul des prévisions manuellement, et n'ont pas été considérées par le modèle choisi. Néanmoins, pour les produits X5 et X9 par exemple, notre approche a donné des résultats moins satisfaisants que la méthode manuelle sur les deux horizons H1 et H3. Des facteurs externes comme la validation des marchés des chantiers ou des campagnes de promotion ou publicitaires peuvent être à l'origine de ces résultats, sans que nous n'ayons la moindre certitude.

Notons également que les réglementations gouvernementales instaurées peuvent être à l'origine des baisses des ventes de certains produits en mettant des fin-vies, et donc l'arrêt progressive de ces derniers. De telles informations ne peuvent pas être prises en considération par les modèles, d'où leurs incapacités à fournir des meilleures prédictions que la méthode manuelle dans ces cas.

Nous pouvons constater également sur le tableau 7, qu'au niveau de l'erreur absolue moyenne (MAE) sur tous les produits, notre approche était plus précise sur l'horizon H3 que sur l'horizon H1 avec une MAE respective de 249,63 et 269,29. Cependant avec la méthode manuelle, les résultats étaient plus précis sur H1 que H3 ayant respectivement une MAE de 380,43 pour H1 et 419,69 pour H3. La légère dégradation de la qualité de prédiction de la méthode automatique peut faire suite à une promotion réalisée sur cette période, qui pourra

être prise en considération dans la méthode manuelle, ce qui explique la non dégradation de la qualité des prévisions en passant de H3 au H1 de cette dernière. Cependant, les erreurs MAE se sont nettement réduites à travers l’approche automatique pour les deux horizons.

Au niveau des performances globales, l’application de notre approche a permis un taux d’amélioration important du MAE d’environ 40 % pour le moyen terme H3 et d’environ 29 % pour le court terme H1, par rapport à la méthode manuelle. Ce taux est calculé en faisant le rapport entre le gain en MAE (la différence) et la valeur du MAE manuelle (initiale). Ces améliorations touchent plusieurs produits dont les « High Runners » permettant ainsi d’avoir des profits économiques significatifs pour l’entreprise. Il est à noter que, de manière logique, notre approche était moins bonne sur quelques produits qui présentaient parfois une énorme variabilité suite à des facteurs externes tels que des promotions ou la validation des marchés des chantiers. Cependant, globalement, notre méthodologie apporte une amélioration importante par rapport à la méthode manuelle.

Etant donné que notre méthodologie se base sur plusieurs modèles de prédiction associés à chacune des références du catalogue produit, nous présentons dans les tableaux 8 et 9 les résultats de la mise en compétition des modèles lors de la phase d’entraînement pour différents produits. Les valeurs correspondent aux mesures d’erreurs (RMSE) calculées pour chaque structure ainsi que le meilleur modèle choisi pour chaque produit et les paramètres associés à ce dernier.

Le tableau 8 montre les erreurs d’entraînement des modèles pour différents produits pour un horizon H3 pour les prévisions au mois d’octobre 2019. Les valeurs marquées en vert correspondent aux erreurs minimales identifiées lors de la phase de compétition dans la méthodologie et donc la structure du modèle choisi pour chaque produit. Nous adoptons les mêmes notations présentées pour les données de l’année 2018.

Produit	Lissage Exp (LE)	ARIMA	SARIMA	Prophet	Choix	Paramètres
X2	26	-	-	38	LE Triple	['add', False, 'add']
X3	14	16	16	23	LE Double	['add', False, None]
X4	224	613	465	770	LE Double	['add', True, None]
X5	128	154	154	923	LE Double	['add', False, None]

X6	10	56	-	44	LE Triple	['add', True, 'mul']
X7	96	175	-	782	LE Double	['add', False, None]
X8	182	-	-	349	LE Triple	[None, False, 'mul']
X9	107	-	-	287	LE Double	['add', True, None]
X10	165	324	263	159	Prophet	-
X11	135	637	595	287	LE Double	[None, False, 'mul']
X12	283	522	280	223	Prophet	-
X13	68	-	-	73	LE Triple	['add', True, 'mul']
X14	148	137	137	382	ARIMA	[0, 1, 3]
X15	72	108	104	131	LE Triple	['add', False, 'mul']
X17	61	169	98	55	Prophet	-
X18	71	88	88	204	LE Double	['add', False, None]
X19	7	18	10	8	LE Triple	['add', False, 'mul']
X20	13	-	34	25	LE Triple	['add', False, 'mul']
X21	13	35	-	27	LE Double	['add', True, None]
X22	3	32	28	26	LE Triple	['add', True, 'mul']
X23	13	36	31	34	LE Triple	['add', False, 'mul']
X24	12	33	38	43	LE Triple	['add', False, 'mul']
X30	41	71	82	126	LE Double	['add', False, None]
X37	15	-	21	22	LE Triple	['add', False, 'add']
X38	37	-	-	43	LE Double	['add', False, None]
X39	38	54	54	46	LE Double	['add', True, None]
X40	41	39	39	15	Prophet	-
X41	56	47	-	41	Prophet	-
X42	44	69	69	38	LE Double	['add', True, None]
X43	36	60	57	51	LE Triple	['add', True, 'add']
X44	38	45	35	57	SARIMA	[(1, 0, 0), (0, 0, 1, 12)]
X45	322	731	731	1216	LE Double	[None, False, 'add']
X46	85	197	167	126	LE Triple	['add', True, 'mul']
X47	242	359	273	211	Prophet	-

Tableau 8 : Mesure d'erreur (RMSE) à l'entraînement pour un horizon H3 et choix du modèle par produit pour les prévisions au mois d'octobre 2019

Comme présenté pour les données de l'année 2018, nous considérons la même notation pour les paramètres des modèles de lissage exponentiel. Dans le tableau 9, nous présentons les

erreurs liées à la compétition des modèles pour les prévisions du mois de décembre 2019 à l'horizon H1.

Produit	Lissage Exp (LE)	ARIMA	SARIMA	Prophet	Choix	Paramètres
X2	23	36	29	32	LE Triple	['add', False, 'mul']
X3	5	13	13	27	LE Double	['add', True, None]
X4	183	353	267	286	LE Double	['add', False, None]
X5	261	175	170	535	SARIMA	[(1, 0, 1), (0, 0, 0, 12)]
X6	46	56	347	334	LE Triple	['add', False, 'add']
X7	194	33	33	861	ARIMA	[3, 0, 0]
X8	244	-	-	310	LE Double	[None, False, 'mul']
X9	160	-	-	298	LE Double	[None, False, 'mul']
X10	195	236	192	199	SARIMA	[(0, 1, 2), (2, 0, 0, 12)]
X11	156	566	545	477	LE Double	[None, False, 'mul']
X12	290	511	379	241	Prophet	-
X13	75	-	-	90	LE Double	[None, False, 'mul']
X14	146	139	139	325	ARIMA	[0, 1, 3]
X15	63	-	87	123	LE Triple	['add', False, 'mul']
X17	84	103	72	97	SARIMA	[(0, 0, 1), (2, 0, 0, 12)]
X18	83	92	92	190	LE Double	['add', False, None]
X19	12	15	16	16	LE Double	['add', False, None]
X20	7	-	21	25	LE Triple	['add', False, 'mul']
X21	22	27	-	33	LE Double	['add', True, None]
X22	7	20	26	32	LE Triple	['add', True, 'mul']
X23	11	34	34	33	LE Triple	['add', False, 'mul']
X24	8	24	32	43	LE Triple	['add', False, 'mul']
X30	51	-	-	160	LE Double	['add', True, None]
X37	24	-	19	32	SARIMA	[(0, 0, 0), (1, 0, 0, 12)]
X38	34	39	39	39	LE Double	['add', False, None]
X39	41	53	53	45	LE Double	[None, False, 'mul']
X40	29	34	34	23	Prophet	-
X41	5	46	-	41	LE Double	['add', True, None]
X42	47	36	38	41	ARIMA	[0, 1, 1]

X43	34	49	47	50	LE Double	[None, False, 'add']
X44	38	68	51	53	LE Triple	['add', True, 'mul']
X45	207	363	363	1049	LE Double	[None, False, 'add']
X46	98	98	99	146	ARIMA	[0, 1, 2]
X47	205	347	330	189	Prophet	-

Tableau 9 : Mesure d'erreur (RMSE) à l'entraînement pour un horizon H1 et choix du modèle par produit pour les prévisions au mois de décembre 2019

Pour les tableaux 8 et 9, nous remarquons une dominance des modèles de lissage exponentiel due à leur variété comme nous l'avons constaté pour les données de l'année 2018. Nous constatons également que, pour certains produits, l'un des modèles représente beaucoup mieux les ventes d'un produit et engendre de grands écarts entre les modèles. C'est le cas des produits X4 à l'horizon H3 (Tableau 8) et X11 à l'horizon H1 (Tableau 9). Pour d'autres produits comme X39 à l'horizon H3 (Tableau 8) et X10 à l'horizon H1 (Tableau 9), les modèles présentent des résultats très proches. Le premier constat justifie la stratégie du choix du meilleur modèle pour les produits du catalogue de l'entreprise pour prédire les valeurs futures, néanmoins, le deuxième constat amène la réflexion de pouvoir fusionner plusieurs modèles pour acquérir des meilleures prévisions notamment au long et moyen terme.

Nous remarquons que, pour certains produits comme le X2 et X12, la même structure de modèle a gagné la compétition pour l'horizon H3 (tableau 5 et tableau 8). Le même aspect est constaté pour les produits X20 et X45 pour l'horizon H1 (tableau 6 et tableau 9). Ceci peut s'expliquer par une stabilité dans le comportement de ventes de ces produits d'où le choix du même modèle pour la prédiction. En revanche, pour la plupart des produits le modèle choisi a changé pour s'adapter aux évolutions des produits en passant de 2018 à 2019 pour les deux horizons. Ceci permet à notre méthodologie d'avoir un caractère évolutif permettant de mieux modéliser les produits en ayant plus de données au fil du temps.

III.2. Analyses du choix des modèles par profil de produits

Dans le but de réaliser une analyse des modèles sélectionnés par profil de produits, nous avons étudié les résultats d'application de notre méthodologie sur environ 300 produits les plus vendus (y compris les produits de l'usine de Drancy) avec deux horizons de prédiction à moyen terme (H3) et à court terme (H1) sur le dernier trimestre de 2019. Nous analysons ainsi comment les produits sont regroupés par type de modèle afin de mieux comprendre et catégoriser nos séries temporelles en fonction de type de ventes et de produits. Cette

identification peut conduire à créer des groupes de produits, étudier les similitudes présentées dans ces séries chronologiques et obtenir une meilleure compréhension des ventes de produits afin d'améliorer les prévisions.

III.2.1. Analyse pour un horizon à moyen terme (H3)

Pour la prévision au moyen terme (H3), le tableau 10 présente la répartition des produits modélisés par chaque modèle.

Modèles	Produits
Prophet	20 produits
ARIMA/SARIMA	59 produits
Lissage Exponentiel	219 produits

Tableau 10 : Répartition des produits par modèle pour le moyen terme (H3)

Nous constatons que le modèle Prophet a été sélectionné pour 20 produits qui sont des séries chronologiques avec des variations et des comportements similaires. La figure 23 présente des exemples de ventes de ces produits.

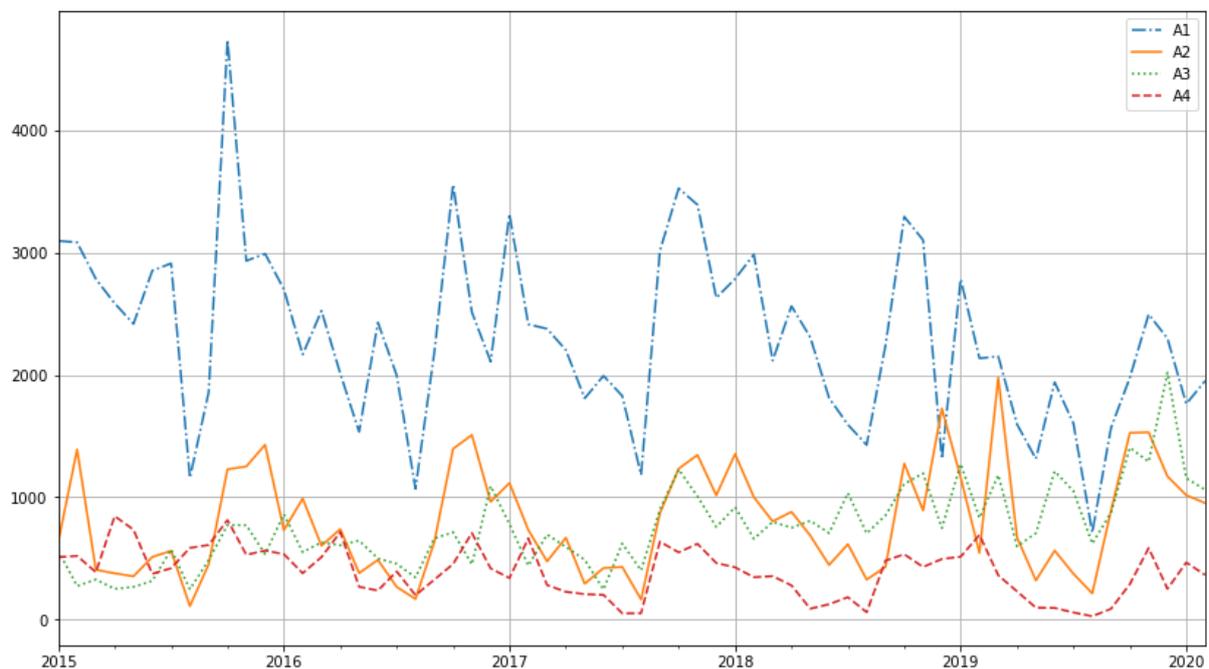


Figure 23 : Exemples de comportements de ventes réelles des produits modélisés par Prophet sur leurs cycles de vie représentant des variations semblables

Ces produits ont un comportement saisonnier reconnaissable avec une baisse des ventes en août de chaque année et une augmentation importante des quantités à la fin et au début de chaque année. Prophet a réussi à capturer la variabilité de ces produits mieux que d'autres

modèles. D'un point de vue industriel, ces produits représentent des « High Runners », c'est-à-dire des produits vendus avec des grandes quantités répondant ainsi à une forte demande client. Ce sont des produits vendus sur le marché des grossistes et des installateurs mais aussi sur le marché des chantiers avec des quantités parfois importantes, d'où les pics constatés sur les graphes des comportements. La saisonnalité présente dans les séries de vente de ces produits montre aussi une certaine régularité des ventes expliquée par des commandes régulières de plusieurs grossistes pour réalimenter leurs stocks.

Le modèle ARIMA a été utilisé pour modéliser 24 produits. Nous avons constaté que certains produits ayant un comportement particulier sont bien capturés par ARIMA comme présenté dans la figure 24.

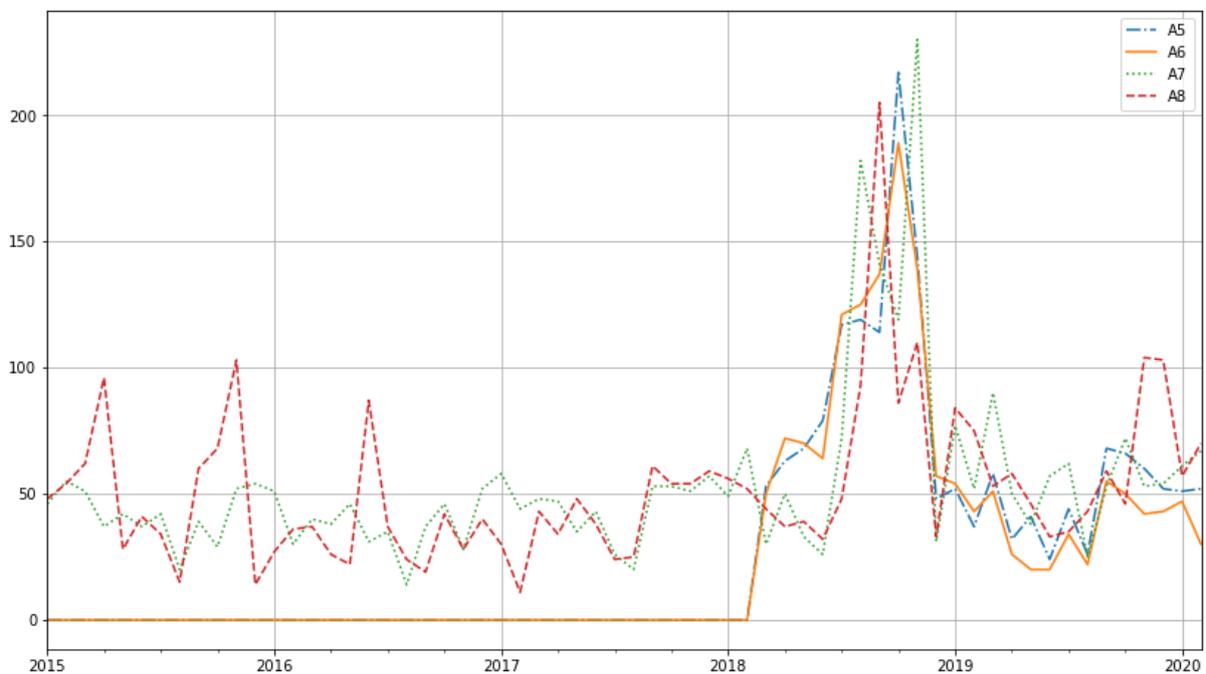


Figure 24 : Exemples de comportements de ventes réelles des produits modélisés par ARIMA sur leurs cycles de vie représentant des variations semblables

Nous constatons que ce type de produits, présentant une augmentation importante des quantités vendues en fin 2018, a été mieux interprété par ARIMA que par les autres modèles. De plus, ces séries chronologiques ne représentent pas de fortes tendances saisonnières qui justifient la performance d'ARIMA sur d'autres modèles saisonniers. D'un point de vue industriel, cette augmentation brusque en fin de 2018 est expliquée par la validation de quelques chantiers et donc une augmentation importante, temporaire et imprévisible dans la demande des clients, et qui s'est restabilisée rapidement après cette période.

Le type de modèle le plus utilisé était l'ensemble des lissages exponentiels (LE), elles ont été sélectionnées pour 219 séries chronologiques, avec différentes composantes. 12 structures de modèles de lissage exponentiel sont employées pour faire des prévisions avec de multiples combinaisons. Le tableau 11 résume la répartition des produits par modèle.

		Saisonnalité		
		<i>Aucune</i>	<i>Additive</i>	<i>Multiplicative</i>
Tendance	<i>Aucune</i>	-	34 produits	32 produits
	<i>Additive</i>	28 produits	19 produits	46 produits
	<i>Additive Amortie</i>	25 produits	15 produits	20 produits

Tableau 11 : Répartition des produits par type de lissage exponentiel pour H3

Les produits sont répartis sur les différents modèles avec des proportions proches. 66 produits sont représentés par des modèles n'ayant pas une composante de tendance, et donc leurs comportements de ventes ne présentent pas une orientation générale à la hausse ou à la baisse en passant d'une année à l'autre. C'est notamment le cas des produits qui sont vendus auprès des mêmes clients chaque année avec des quantités qui varient légèrement. Nous constatons également que 53 produits sont représentés par des modèles avec seulement une composante de tendance (amortie et non amortie), et sans composante saisonnière. Ces produits présentent alors un comportement avec une augmentation ou une baisse constante des chiffres des ventes sans avoir des motifs répétés chaque année. Ce type de série concerne les produits récemment lancés avec des ventes en augmentation chaque année ou des produits en fin de vie avec des baisses progressives. Le modèle le plus sélectionné pour le catalogue des produits de l'entreprise est celui avec une tendance non amortie additive et une saisonnalité multiplicative avec 46 produits modélisés. Ces produits sont caractérisés par des motifs répétés chaque année mais avec une augmentation ou une baisse de leur amplitude. La figure 25 présente quelques exemples de produits.

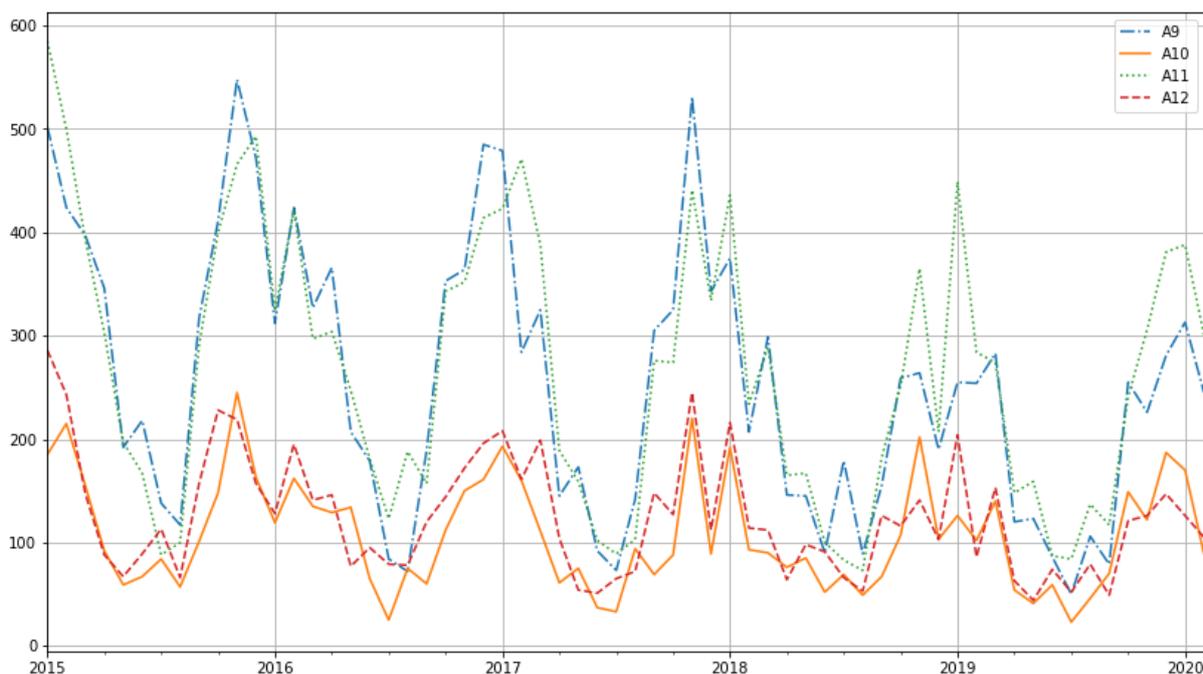


Figure 25 : Exemples de comportements de ventes réelles des produits modélisés par un lissage exponentiel avec une tendance additive non amortie et une saisonnalité multiplicative sur leurs cycles de vie représentant des variations semblables

Nous pouvons remarquer une baisse progressive de l'amplitude de la saisonnalité chaque année. Ce phénomène est plus répandu surtout pour le produit comme A9, tel que présenté sur la figure 25. Ce comportement s'explique par des produits qui commencent à perdre leurs parts de marché en faveur d'autres produits et donc se dirigent vers un arrêt progressif.

Au travers de ces exemples, nous constatons que les produits sont modélisés par différentes techniques dans le but de bien représenter leurs comportements de ventes. Ceci soutient notre méthodologie de prévision des ventes consistant à choisir un modèle approprié pour chaque produit et ne pas appliquer un seul modèle pour tous le catalogue de l'entreprise.

III.2.2. Analyse pour un horizon à court terme (H1)

Pour la prévision au court terme (H1), le tableau 12 présente la répartition des produits modélisés par chaque modèle.

Modèles	Produits
Prophet	34 produits
ARIMA/SARIMA	65 produits
Lissage Exponentiel	199 produits

Tableau 12 : Répartition des produits par modèle pour le court terme H1

Pour le court terme (H1), nous remarquons une augmentation des produits modélisés par Prophet et ARIMA/SARIMA impliquant logiquement une baisse pour les modèles de lissage exponentiel (LE) par rapport au moyen terme (H3). Nous présentons dans la figure 26 un exemple de comportement de vente des produits modélisés par Prophet.

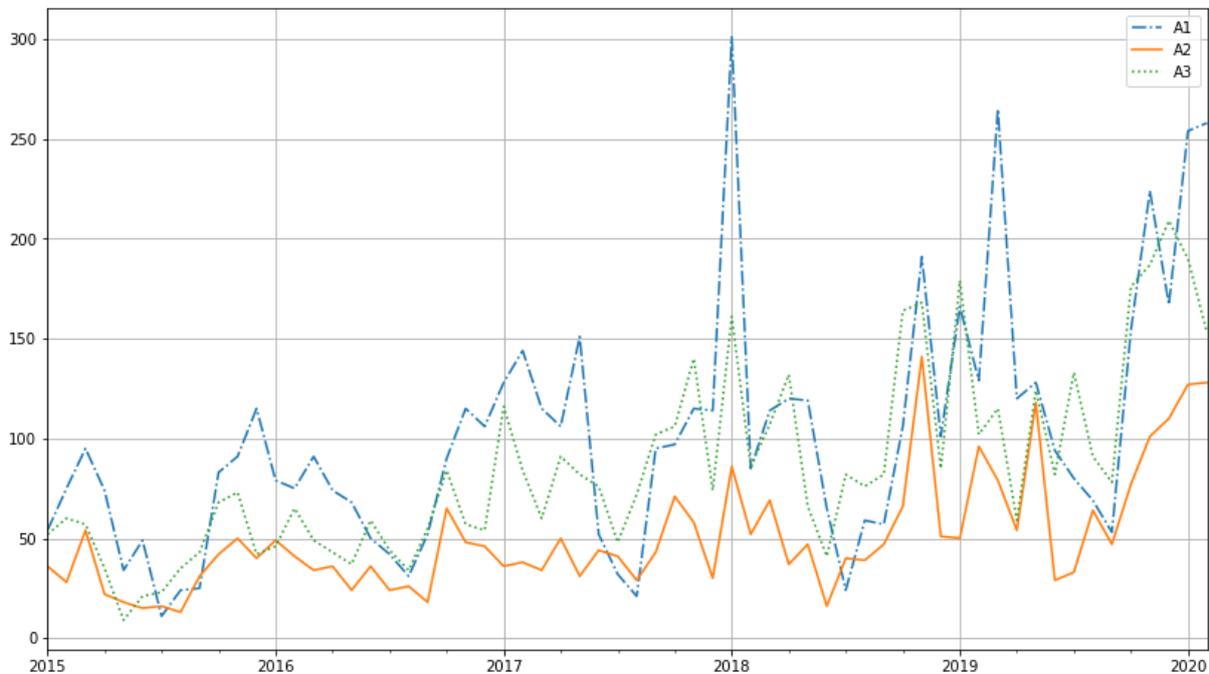


Figure 26 : Exemples de comportements de ventes réelles des produits modélisés par Prophet pour H1 sur leurs cycles de vie représentant des variations semblables

Le modèle Prophet était le plus à-même pour capturer le comportement de ventes des produits présentés dans la figure. Ces produits sont caractérisés par une certaine saisonnalité avec parfois des pics de ventes. Ces derniers sont expliqués par des volumes de ventes importants suite à des demandes clients supplémentaires durant la période de début 2018 et 2019.

Le modèle SARIMA a été utilisé pour modéliser 23 produits du catalogue. La figure 27 présente quelques exemples de comportement commun de ces produits.

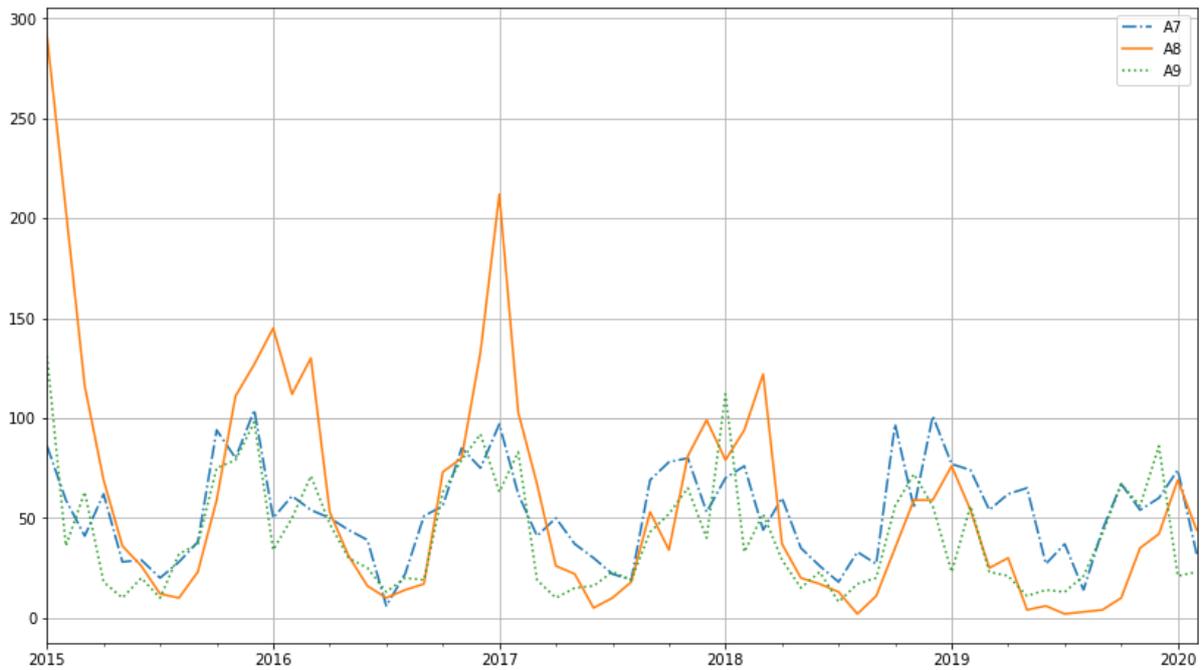


Figure 27 : Exemples de comportements de ventes réelles des produits modélisés par SARIMA pour H1 sur leurs cycles de vie représentant des variations semblables

On remarque l'existence d'une saisonnalité claire d'une année à l'autre pour ces produits, d'où la sélection du modèle SARIMA. Il s'agit des produits vendus régulièrement d'une année à l'autre avec une légère baisse pour le produit A8. La baisse des ventes à l'été de chaque année est due à la nature de l'activité de l'entreprise à travers la production des systèmes de chauffage, et aussi à la fermeture annuelle de l'usine.

Comme pour l'horizon H3, le type de modèle le plus utilisé est les techniques de lissage exponentiel (LE) vu qu'ils ont été sélectionnés pour 199 produits, avec différentes composantes. Les mêmes structures de modèles sont utilisées pour l'horizon H1 également. Le tableau 13 résume la répartition des produits par modèle.

		Saisonnalité		
		Aucune	Additive	Multiplicative
Tendance	Aucune	-	28 produits	46 produits
	Additive	12 produits	25 produits	35 produits
	Additive Amortie	21 produits	14 produits	18 produits

Tableau 13 : Répartition des produits par type de lissage exponentiel pour le court terme H1

D'après le tableau 13, les modèles avec une composante saisonnière multiplicative sont les plus utilisés reflétant une augmentation des ventes de ces produits d'une année à l'autre, avec

une domination du modèle sans composante de tendance. La figure 28 présente quelques comportements de ventes des produits représentés par ce modèle.

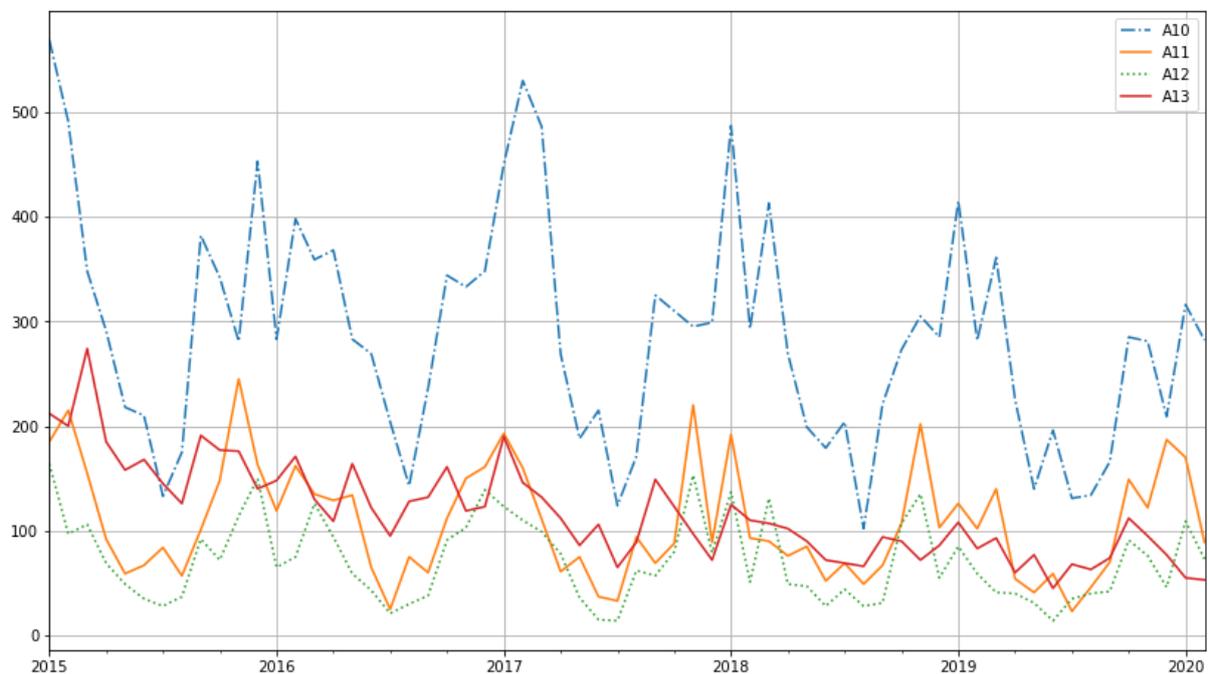


Figure 28 : Exemples de comportements de ventes réelles des produits modélisés par un lissage exponentiel sans tendance avec une saisonnalité multiplicative sur leurs cycles de vie représentant des variations semblables

Sur la figure 28, l'absence d'une tendance pour ces produits est remarquable, ils sont ainsi caractérisés par des ventes régulières d'une année sur l'autre avec une légère baisse de l'amplitude de la saisonnalité au fil du temps, ce qui justifie la sélection de ce modèle.

Nous constatons également dans le tableau 13, un changement léger de répartition des produits sur les différents modèles par rapport au moyen terme (H3), avec une diminution du nombre des produits représentés par des modèles sans composante saisonnière. L'augmentation du nombre des périodes utilisées pour l'entraînement des modèles peut être à l'origine de ce changement étant donné que les modèles ont plus d'informations et donc une meilleure modélisation des produits.

Un aspect intéressant également noté est que les produits appartenant à la même famille de produits qui peuvent avoir un comportement plus semblable sont modélisés avec des méthodes différentes, ce qui prouve leurs natures de ventes différentes et la nécessité d'être traités par des modèles séparés. La sélection du type de modèle diffère d'un horizon à l'autre et nous ne retrouvons pas la même taxonomie de modèle que pour les prédictions à moyen

terme (H3). Ce fait justifie également notre intérêt à utiliser différents modèles, car en évoluant dans le temps, nous obtenons plus de données et donc les modèles et les comportements de ventes des produits peuvent changer en fonction de ces nouvelles données. Par conséquent, notre méthodologie permet d'avoir un aspect évolutif pour bien modéliser les séries au fil du temps.

III.3. Performances des modèles par horizon de prédiction

Comme expliqué dans la méthodologie, notre approche emploie différentes techniques de prédiction dans le but de sélectionner la meilleure structure de modèles pour la prédiction à court terme (H1) et à moyen terme (H3). Ceci implique l'usage des modèles différents pour préparer les prévisions des ventes. Nous analysons dans ce paragraphe le choix des modèles selon l'horizon de prédiction. Nous présentons ainsi les résultats des analyses effectuées sur les produits les plus vendus de l'entreprise. Les prévisions sont effectuées pour l'horizon H1 et H3 du dernier trimestre de l'année 2019.

III.3.1. Usage des modèles par horizon

Nous avons observé l'importance de sélectionner des modèles pour chaque horizon et ainsi identifié ceux qui sont plus adaptés au court terme (H1) ou au moyen terme (H3). Le graphe de la figure 29 montre la répartition de l'usage des modèles pour le court terme.

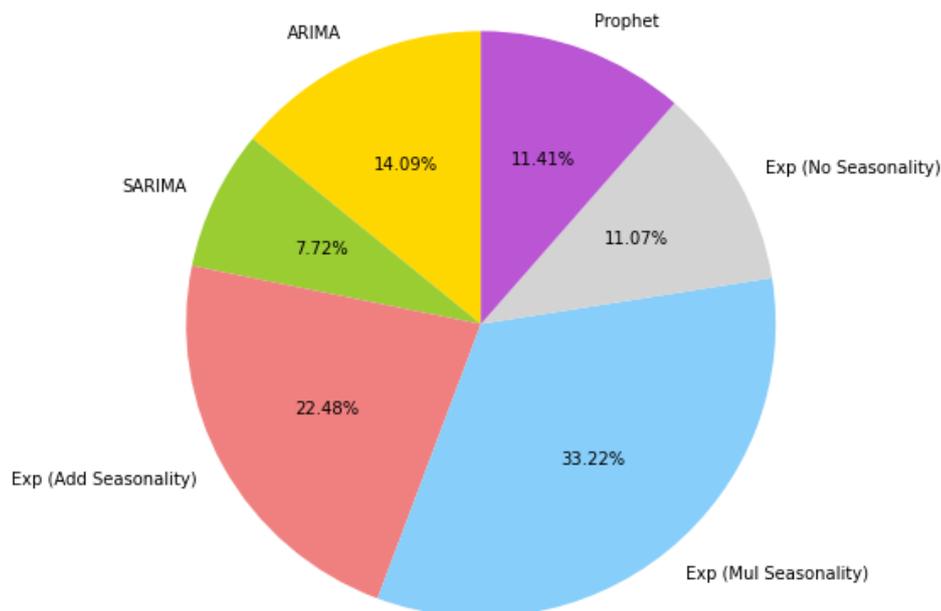


Figure 29 : Répartition des modèles pour l'horizon H1 de 2019

Nous avons trouvé que la famille des modèles de lissage exponentiel est la plus utilisée avec environ 66 % des produits qui ont été modélisés par ses différentes variétés. Sur le graphe, nous avons différencié l'usage de ce type de modèle selon la nature de la composante de saisonnalité. La majorité des produits (correspondant à 33 %) sont représentés par une composante de saisonnalité multiplicative. De plus, 22 % des produits sont modélisés par un modèle avec une saisonnalité additive et environ 11 % des produits sans composante saisonnière. Le modèle ARIMA vient en 2ème place avec près de 14 % de produits représentés par ce modèle suivi de Prophet et SARIMA avec respectivement 11 % et 7 %.

Cette répartition montre que pour le catalogue de produit de elm.leblanc, les modèles de lissage exponentiel avec une composante saisonnière multiplicative et additive sont les plus adéquats pour faire des prédictions à l'horizon H1. Ceci est expliqué par la nature saisonnière de l'activité de l'entreprise représentée par une saison de chauffe pendant l'hiver. Bien que pour plusieurs produits les ventes sont régulières d'une année à l'autre (pas de tendance), l'aspect de saisonnalité est bien présent à travers les fluctuations observées justifiant ainsi l'usage de ces modèles. Le modèle ARIMA est également intéressant pour cet horizon de prédiction. Vu que certains produits ont un comportement spécifique caractérisé par une augmentation durant une période du cycle de vie du produit (comme le cas pour l'année 2018), ARIMA est adéquat pour capturer ce comportement.

Pour les prévisions à l'horizon H3, le graphe de la figure 30 présente la répartition des modèles.

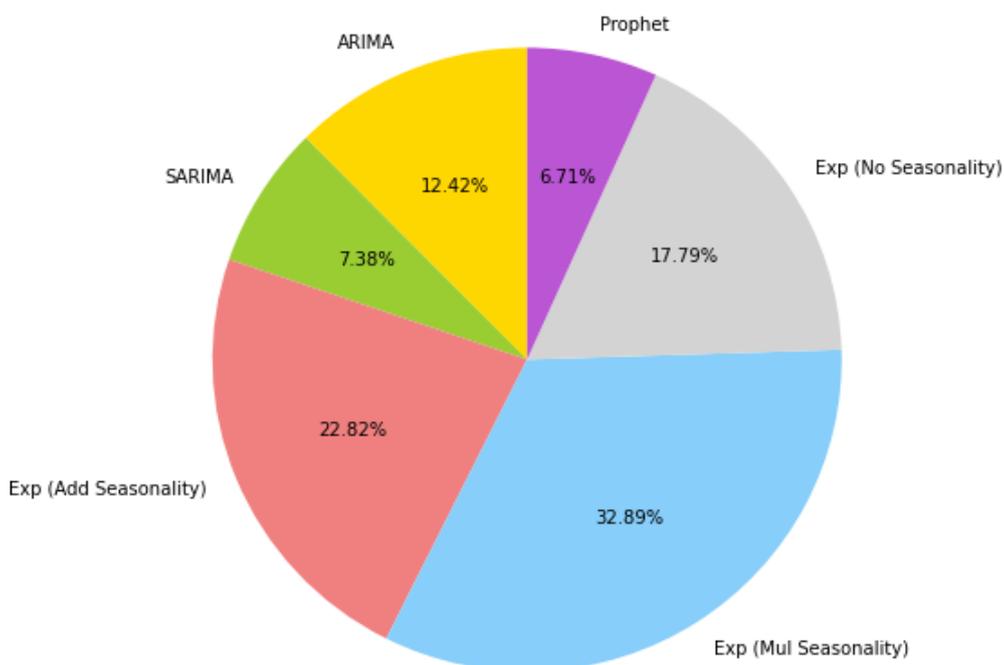


Figure 30 : Répartition des modèles pour l'horizon H3 pour 2019

Sur cette figure, nous constatons le même pourcentage d'usage des modèles de lissage exponentiel avec composante saisonnière (additive ou multiplicative) ce qui montre l'adéquation de ces modèles tant pour le moyen terme (H3) que pour le court terme (H1). En revanche, nous remarquons l'amplification de l'usage du modèle de lissage exponentiel sans composante saisonnière par rapport à H1 avec environ 17 %. Ce modèle s'est avéré plus adéquat pour prédire les ventes au moyen terme des produits moins saisonniers. Nous remarquons aussi une stabilité pour l'usage des modèles ARIMA et SARIMA justifiée par une stabilité dans le comportement de ventes des produits pour lesquels ces modèles ont été sélectionnés. Pour le catalogue des produits de l'entreprise, le modèle Prophet était plus adapté à la prédiction du court terme que le moyen terme puisque on remarque une diminution de son usage à l'horizon H3 par rapport à H1.

Selon les figures 29 et 30, l'évolution des produits au fil du temps et donc la fluctuation de leurs comportements implique un changement dans la nature du modèle utilisé pour la prédiction de l'horizon H1 et H3. Cet aspect montre le caractère évolutif de notre méthodologie et sa capacité à s'adapter aux évolutions du catalogue de produit de l'entreprise pour remettre en question le choix du modèle selon l'horizon de prédiction. Néanmoins, la plupart des produits ne présentent pas un changement drastique dans leurs comportements,

ainsi la répartition des modèles garde une certaine stabilité entre H3 et H1 notamment pour les modèles de lissage exponentiel avec composante saisonnière.

III.3.2. Produits à modèles communs pour différents horizons de prédiction

En croisant les modèles employés pour la prédiction de chaque produit sur le court terme (H1) avec ceux employés sur le moyen terme (H3), nous avons identifié quelques produits pour lesquels le même modèle a été sélectionné pour les deux horizons. Ces produits correspondent dans plusieurs cas où le modèle utilisé est Prophet, à des produits avec un comportement de ventes stables et saisonniers. Le graphe de la figure 31 montre quelques exemples de ces produits.

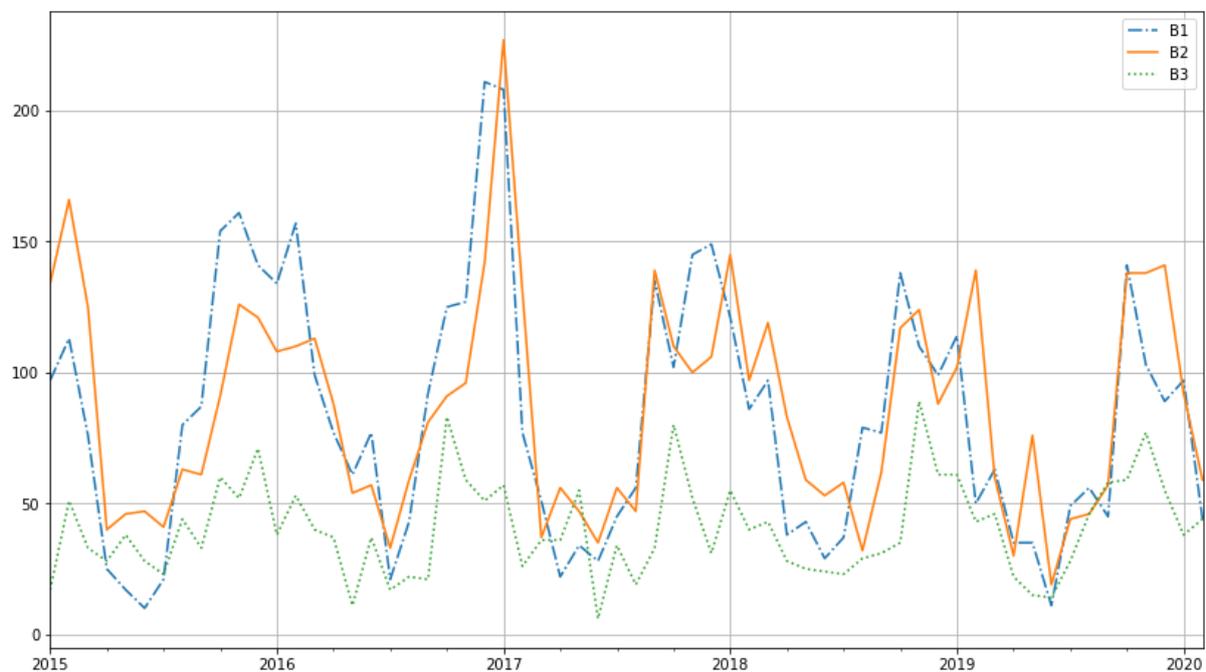


Figure 31 : Exemples de comportement de ventes des produits utilisant Prophet pour les prédictions à H1 et H3

Ceci est expliqué par le fait que durant la procédure de test, dans l'étape de choix du modèle dans la méthodologie, la stabilité des ventes de ces produits implique des erreurs faibles calculées lors de cette étape tant pour le court terme (H1) que pour le moyen terme (H3), impliquant ainsi le choix du même modèle de prédiction.

La figure 32 montre également quelques comportements de produits qui ont été modélisés par le même modèle pour le court et le moyen terme.

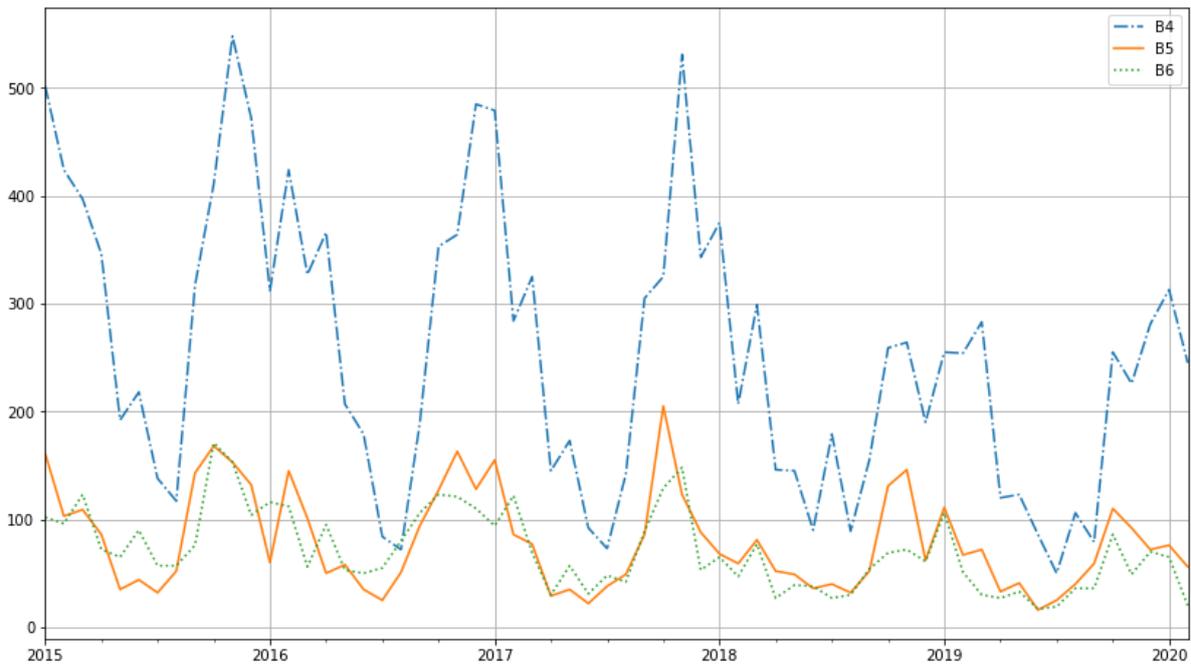


Figure 32 : Exemples de comportement de ventes des produits utilisant le lissage exponentiel pour les prédictions à H1 et H3

Le modèle employé pour la prédiction de ces produits est un lissage exponentiel avec une tendance additive non amortie et une saisonnalité multiplicative. Ce modèle a bien représenté ces séries montrant un motif identifiable et répétitif caractérisé par une baisse de l'amplitude de la saisonnalité au cours des années. Le comportement de vente n'a pas eu un grand changement en faisant des prédictions pour H1 ou H3 et donc le même modèle est sélectionné pour calculer les prédictions. Cet aspect implique une capacité de ce modèle à être performant pour faire des prédictions pour H1 et H3.

Pour la plupart des produits, nous ne retrouvons pas le même modèle pour les prédictions de H1 et H3. Ceci s'explique par une augmentation ou une baisse importante des ventes due à une augmentation du part du marché ou encore à d'autres facteurs externes, et donc un changement du comportement des produits. Cet aspect implique par la suite un changement du choix du modèle lors de la phase d'entraînement et de test de notre méthodologie.

IV. Vers une méthodologie générique

Dans le cadre de ce projet, notre méthodologie a été adaptée aux données d'elm.leblanc pour pouvoir tester et valider notre approche de choix des modèles employés pour la prédiction. Ce cas d'usage permet de souligner l'efficacité de cette méthodologie vis-à-vis de la modélisation et la prédiction de plusieurs produits du catalogue de l'entreprise. En effet, vue

la diversité des natures des séries temporelles présente dans les données d'elm, notre approche a réussi à modéliser ces comportements avec le modèle le plus adéquat prenant en considération l'évolution des séries temporelles au fil du temps.

IV.1.Généralisation

Bien que notre approche ait été appliquée à un cas d'usage spécifique qui concerne le secteur des systèmes de chauffage résidentiels et industriels, nous pouvons généraliser cette méthodologie à d'autres applications et domaines ayant un historique de données de séries temporelles varié. En effet, la multitude des modèles employés donne la possibilité de couvrir un large spectre de motifs de séries chronologiques de nature et de comportement différents qui peuvent exister dans d'autres champs d'étude.

Notons également que les modèles employés dans notre méthodologie peuvent être développés pour des données de différents échantillonnages allant de plusieurs années jusqu'aux millisecondes et avec différents horizons de prédiction selon l'application et le domaine d'étude. Ces aspects permettent à notre approche de répondre aux besoins de différentes problématiques de prédiction non seulement de prévision des ventes mais aussi d'autres problématiques de prévisions des séries temporelles.

Sachant que nous disposons d'un ensemble de modèles de prédiction M faisant chacun une prédiction d'une seule étape et noté $Model_{H_1}$, H_k représente l'horizon de prédiction avec $k \in [1, z]$ où z est le nombre de périodes maximales à prédire, et p est un paramètre d'un modèle décrivant le nombre d'observations utilisées par ce modèle pour prédire l'étape suivante, une formulation généraliste de la problématique peut s'écrire comme suit :

Si $k = 1$:

$$\hat{v}_i^{H_1}(j) = Model_{H_1}[v_i(j-k), \dots, v_i(j-p)] \quad ; \quad \forall i \in R \quad (1)$$

Si $1 < k \leq p$:

$$\hat{v}_i^{H_k}(j) = Model_{H_1} \left[\begin{array}{c} \hat{v}_i^{H_{k-1}}(j-1), \dots, \hat{v}_i^{H_{k-l}}(j-l), \dots, \hat{v}_i^{H_1}(j-k+1), \\ v_i(j-k), \dots, v_i(j-p) \end{array} \right] \quad (2)$$

Si $k > p$ et $k \neq 1$:

$$\hat{v}_i^{H_k}(j) = Model_{H_1} \left[\hat{v}_i^{H_{k-1}}(j-1), \dots, \hat{v}_i^{H_{k-p}}(j-p) \right] ; \forall i \in R \quad (3)$$

Où $\hat{v}_i^{H_k}(j)$ la prédiction des ventes du produit i à l'horizon H_k du mois j à prédire et $v_i(j)$ correspond aux ventes réelles du produit i au mois j . R représente l'ensemble des séries temporelles à prédire, N_i est l'ensemble des observations enregistrées pour la série i et donc $p \in [1, N_i - 1]$.

Lors de la procédure du choix de modèle, nous partageons les données de chaque produit en base d'entraînement et base de test. L'évaluation des modèles est réalisée sur les données de la base de test. Nous utilisons la formule suivante pour calculer l'erreur quadratique de prédiction de chaque modèle mis en compétition :

$$\varepsilon_{H_k} = \frac{1}{card(j)} \sum_{j \in \text{base de test}} |v_i(j) - \hat{v}_i^{H_k}(j)|^2 \quad (4)$$

Dans le pseudo-algorithme ci-dessous, nous détaillons les différentes étapes de construction de cette approche. Notons que f représente un élément de l'ensemble M des modèles disponibles.

- Pour chaque i dans R :

- Partager les données en entraînement (Train (i)) et test (Test (i))

- Pour chaque modèle f dans M_i :

- Pour p_i dans $[1 .. N_i - 1]$:

1. Apprendre (f_i) sur Train (i)

2. Tester (f_i) sur Test (i)

3. Calculer l'erreur ε_{p_i} selon (4)

- Enregistrer les paramètres du modèle ayant min (ε_{p_i})

- Choisir le meilleur modèle f_i ayant min (ε_{f_i})

- Associer chaque référence i avec le meilleur modèle (f_i)

- Enregistrer le paramètre p_i du modèle (f_i)

Procédure de choix de modèle pour chaque série

- Pour chaque i dans R :

Mise à jour de (f_i) sur la série i

- Si $k = 1$:

Procédur e de

Prédire les valeurs futures de i selon (1)

- Si $1 < k \leq p$:

Prédire les valeurs futures de i selon (2)

- Si $k > p$:

Prédire les valeurs futures de i selon (3)

- **Récupérer les prédictions pour tous les produits**

IV.2. Discussion autour de la méthodologie proposée

La préparation des prévisions des ventes est une tâche de grande importance pour toute entreprise quel que soit son secteur d'activité. Nous avons proposé dans ce chapitre une méthodologie qui permette la réalisation de cette tâche d'une façon automatisée en commençant par une première étape de préparation et nettoyage des données.

Cette étape vise à préparer une base de données fiable contenant un ensemble de séries temporelles de ventes à prédire suivant un catalogue de produits. D'après notre cas d'étude dans l'entreprise elm.leblanc, le rassemblement des données à partir des différentes bases et systèmes d'information, le croisement des données et le nettoyage de ces dernières, sont des tâches auxquelles il faut consacrer un temps significatif de traitement, mais aussi une grande rigueur vue la sensibilité des données et l'impact de leur qualité sur les résultats de prévision.

La deuxième étape est une procédure de choix du meilleur modèle pour chaque produit. Cette étape se déroule en deux phases : d'abord le choix du meilleur modèle de chaque structure mise en compétition et ensuite, le choix du meilleur modèle parmi ceux sélectionnés de chaque structure. Pour accomplir ces deux phases, nous partageant les données disponibles en base d'entraînement et de test pour choisir le modèle gagnant ayant l'erreur minimale sur l'ensemble de la base de test. Cette dernière représente généralement 20% des données disponibles mais cela peut varier selon le cas d'étude et la quantité des données à disposition. Dans notre cas et à la vue des quantités de données disponibles, nous avons réservé 10% pour le test.

Le choix des modèles ARIMA, SARIMA, Prophet et la famille des Lissages Exponentiels provient de leur capacité à modéliser plusieurs comportements et natures de séries temporelles. Comme nous l'avons déjà mentionné, plusieurs critères sont à prendre en considération lors de la sélection des modèles à utiliser dans la méthodologie permettant ainsi de répondre aux

besoins de chaque cas d'usage comme les ventes saisonnier, les ventes événementielles ou encore les ventes sans motifs précis. Nous avons essayé de répondre à ces critères dans l'approche proposée, mais ces choix peuvent être reconsidérés et changer avec l'apparition des modèles plus précis, ou encore en ayant à disposition une quantité de données plus importante permettant le réentraînement des modèles pour suivre l'évolution des produits ou l'exploitation des techniques nécessitant une grande quantité de données comme les approches d'apprentissage profond.

Une question importante à garder à l'esprit est le changement du profil de ventes d'un produit au fil du temps. En effet, dû à un changement des parts du marché, à l'imposition des réglementations impactant l'activité d'une entreprise ou à l'arrêt de certains produits, le comportement de ventes des produits peut varier et le modèle peut évoluer dans le temps. Avec la possibilité de réentraîner les modèles à chaque fois que de nouvelles données sont introduites, le choix du modèle gagnant pour chaque produit peut changer au fil du temps, donnant ainsi à notre méthodologie un critère adaptatif permettant de faire face au changement des profils de ventes. Nous pourrions également introduire un certain seuil d'erreur qui, lorsqu'il est dépassé, provoque le fait de réentraîner les modèles pour qu'ils modélisent mieux les produits.

Après avoir identifié le meilleur modèle pour chaque produit, la troisième étape sera de l'utiliser pour calculer les prévisions à l'horizon demandé. Tout dépend de la taille du catalogue produits et de la longueur des séries temporelles, le temps de calcul et les ressources à mobiliser peuvent être importants. Pour la préparation des prévisions des produits de l'entreprise elm.leblanc, le temps d'exécution moyen enregistré est de 12 heures avec un cluster de 64 cœurs et 64 giga-octets de RAM dans un environnement Linux. Plus l'entreprise dispose de produits et de données, plus le temps de traitement est important. Le choix de la stratégie récursive de calcul des prévisions permet un temps de calcul abordable et acceptable pour préparer les prévisions à temps. Une stratégie directe ou RECTIFY peut augmenter le temps de calcul puisqu'elle nécessite un réentraînement itératif d'un nouveau modèle pour chaque point à prédire, et donc une explosion du temps d'exécution. La mise à disposition de ressources de calcul puissantes peut réduire considérablement le temps d'exécution.

Bien que la méthodologie présentée puisse être mise en place dans plusieurs domaines d'application ayant le même type de données, nous constatons quelques limites qui peuvent affecter l'usage de cette approche.

Comme mentionné dans les hypothèses d'implémentations, nous avons filtré les produits ayant moins de 8 périodes d'historique. Ceci est dû à la difficulté rencontrée par les modèles pour prédire des séries temporelles courtes avec très peu d'historique, ce qui rend la méthodologie moins efficace pour les domaines ayant des produits à très courts cycles de vie et qui varient énormément, donc n'ayant pas assez de données. Dans la même idée, les nouveaux produits récemment lancés sur le marché ne pourront pas être prédits vu l'indisponibilité d'historique sauf si ces derniers vont remplacer des anciens produits arrêtés.

Le fait d'utiliser une approche qui se base seulement sur l'historique des séries temporelles, ne permet pas d'exploiter d'autres variables exogènes qui peuvent influencer les ventes. En effet, les promotions effectuées sur les produits, l'évolution des stocks ou encore les données météorologiques peuvent être employés comme variables externes à inclure dans le modèle de prédiction.

D'autres éléments pouvant également influencer les ventes et impliquent un changement complet du comportement des produits et donc la mise en place d'une nouvelle modalité de prévisions. En effet, l'évolution des parts de marché des concurrents ou des produits, a une incidence directe sur les quantités vendues. L'imposition de nouvelles réglementations ou le changement de la politique de l'entreprise peuvent causer l'arrêt ou la réduction des ventes de certains produits comme les réglementations environnementales des produits qui émettent des gaz toxiques. L'impact d'un événement externe imprévisible comme une pandémie (ici COVID 19) ne peut pas être pris en considération vu l'impossibilité de prévoir ce type d'évènement.

Pour faire face à certaines limites présentées ci-dessus, nous proposons de compléter cette méthodologie par un processus de planification de la Supply Chain comme le S&OP. Un tel processus permettra d'ajuster les prévisions selon les avis d'experts présents durant l'étape de planification de la demande où nous pourrions prendre en considération l'impact des événements externes ou l'arrêt de certains produits, et donc modifier le plan de la demande pour répondre à ces événements.

Comme nous l'avons constaté dans le paragraphe des expérimentations (tableaux 5, 6, 8 et 9), pour certains produits la sélection du meilleur modèle au cours de la phase d'entraînement présente quelques difficultés vu les écarts minimes entre les modèles entraînés. Ceci est constaté surtout lorsque nous trouvons avec des mesures d'erreurs très proches avec parfois deux chiffres après la virgule près. Cet aspect peut impacter la précision des prévisions surtout pour les longs et moyens termes lorsque les modèles ayant une mesure d'erreur très proche, ont des natures différentes (modèle saisonnier et modèle non saisonnier). Alors, la projection des ventes dans le futur peut varier énormément selon le choix du modèle.

Pour limiter l'impact de cet aspect sur la qualité des prévisions, nous allons exploiter les caractéristiques de différents modèles à travers la fusion de leurs résultats, permettant ainsi de réduire l'impact de ne pas avoir sélectionné le meilleur modèle pour faire les prédictions surtout au long terme et d'atténuer les biais qui peuvent être identifiés dans les prévisions du modèle choisi par la première approche. Ceci sera le sujet du chapitre suivant.

V. Exploration et ouverture

Dans le but de challenger notre méthodologie et la tester sur une période caractérisée par une grande variabilité, nous l'avons mise en application sur l'année de 2020 caractérisée par le facteur externe du COVID-19. Nous avons réalisé des expérimentations sur quelques produits de l'usine de Drancy en calculant les prévisions pour l'horizon H1 et H3 sur les mois de Novembre et Décembre 2020 et Janvier 2021. L'étude de cette période est importante pour évaluer l'efficacité de la méthodologie pour la période post-COVID.

Dans le but de limiter l'impact de la période du COVID sur les ventes, nous avons dû remplacer les données de ventes réelles des mois de Mars, Avril et Mai 2020 par des données de simulations créées à travers l'application de notre méthodologie. Nous avons donc calculé les ventes de ces mois en nous basant sur l'historique des ventes des produits. Cette rectification des données a été appliquée pour remplacer la période de fermeture de l'entreprise suite à la crise sanitaire et son influence sur les ventes réelles.

Dans le graphique de la figure 33, nous présentons la somme des quantités des ventes par rapport aux quantités prédites par notre méthodologie et les prévisions manuelles pour l'horizon H3 à partir de novembre 2020, pour quelques produits de l'usine de Drancy.

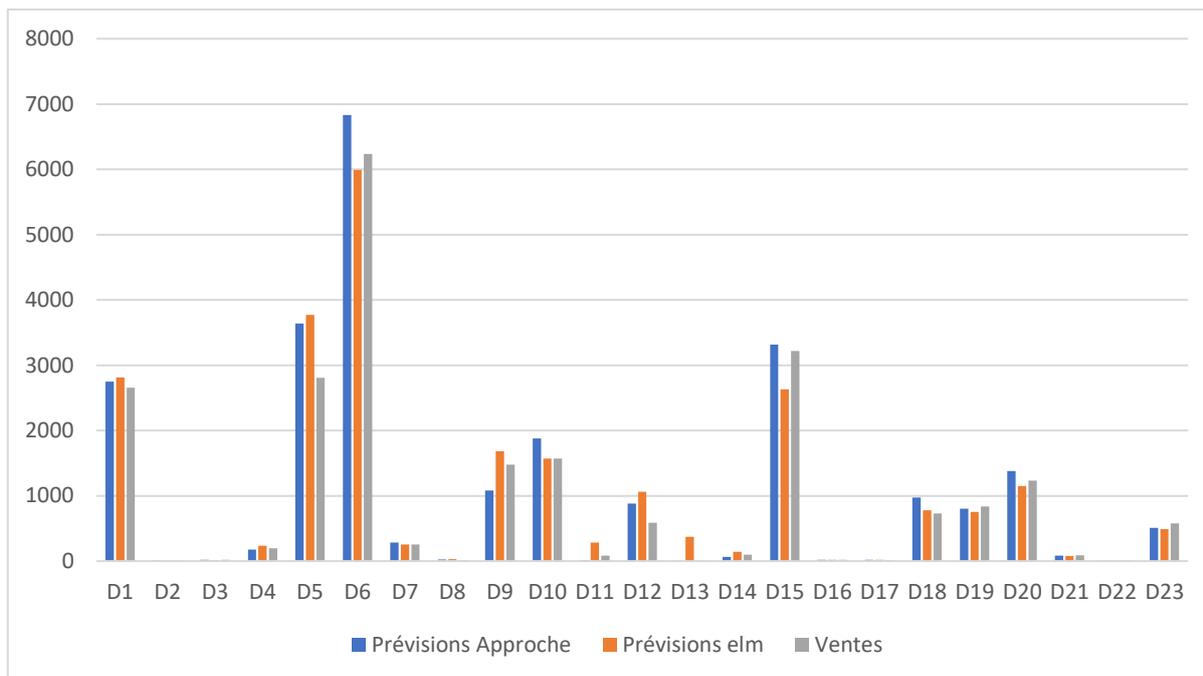


Figure 33 : : Comparaisons des quantités prédites par rapport aux ventes réelles des produits de Drancy pour l'horizon H3 de 2020/2021

Pour plusieurs produits ayant des quantités importantes, les quantités prédites par notre méthode s'approchent plus des ventes réelles que les prévisions manuelles, notamment pour les produits D1, D5 et D15. Pour le produit D6 par contre, la méthode manuelle était plus proche des ventes réelles.

Dans le tableau 14, nous présentons les erreurs calculées pour chaque produit de l'usine de Drancy ainsi que la MAE de notre approche et de la méthode manuelle pour l'horizon H3 et H1 de la fin de l'année 2020 et début de l'année 2021.

Produit	H3				H1			
	Erreur Manuel	% par rapport au total	Erreur Automatique	% par rapport au total	Erreur Manuel	% par rapport au total	Erreur Automatique	% par rapport au total
D1	487	8,27%	378	7,81%	511	8,65%	396	8,51%
D2	5	0,08%	5	0,10%	6	0,10%	5	0,11%
D3	8	0,14%	12	0,25%	4	0,07%	12	0,26%
D4	39	0,66%	20	0,41%	37	0,63%	26	0,56%
D5	967	16,41%	831	17,17%	880	14,89%	632	13,59%
D6	1070	18,16%	598	12,36%	1138	19,26%	861	18,51%
D7	54	0,92%	44	0,91%	56	0,95%	45	0,97%
D8	18	0,31%	10	0,21%	17	0,29%	10	0,21%

D9	253	4,29%	394	8,14%	269	4,55%	175	3,76%
D10	166	2,82%	305	6,30%	174	2,94%	166	3,57%
D11	198	3,36%	68	1,41%	183	3,10%	76	1,63%
D12	475	8,06%	292	6,03%	428	7,24%	207	4,45%
D13	367	6,23%	7	0,14%	356	6,02%	9	0,19%
D14	43	0,73%	32	0,66%	51	0,86%	46	0,99%
D15	772	13,10%	408	8,43%	719	12,17%	490	10,53%
D16	3	0,05%	7	0,14%	3	0,05%	7	0,15%
D17	5	0,08%	11	0,23%	7	0,12%	12	0,26%
D18	144	2,44%	410	8,47%	278	4,70%	410	8,81%
D19	247	4,19%	103	2,13%	170	2,88%	209	4,49%
D20	398	6,75%	775	16,02%	457	7,73%	734	15,78%
D21	13	0,22%	16	0,33%	13	0,22%	17	0,37%
D22	5	0,08%	4	0,08%	5	0,08%	5	0,11%
D23	155	2,63%	109	2,25%	148	2,50%	102	2,19%
MAE	256,17	100%	210,39	100%	256,96	100%	202,26	100%

Tableau 14 : Erreurs de prévisions pour différents produits sur les horizons H1 et H3 de 2020/2021 selon la méthode employé (manuelle ou automatique)

D'après le tableau 14, nous remarquons que, pour la méthode manuelle ou pour notre approche, il n'y a pas de règle générale qui peut s'appliquer en comparant les résultats pour H1 et H3. Nous obtenons pour certains produits des prévisions plus précises sur le court terme (H1) et pour d'autres produits la précision s'inverse. Nous constatons également que, pour la plupart des produits, notre méthodologie donne de meilleurs résultats (erreurs plus faibles) que ce soit pour le court terme ou le moyen terme (exemple pour les produits D1, D4, D6, D11 et D12).

Bien que la méthode manuelle ait donné de meilleurs résultats pour quelques produits (comme D15, D18 et D20), nous constatons globalement que notre approche montre de meilleurs résultats, illustré par une MAE de 210,39 pour H3 contre 256,17 pour la méthode manuelle, et 202,26 pour H1 contre 256,96 pour la méthode classique. Nous obtenons ainsi un taux d'amélioration d'environ 18 % pour le moyen terme et 21 % pour le court terme.

Comme expliqué ci-dessus, notre méthodologie s'appuie sur différents modèles de prédiction des séries temporelles associés à chacun des références du catalogue produits. Nous présentons, dans les tableaux suivants, les résultats d'entraînement des modèles pour les produits de l'usine de Drancy. Les valeurs correspondent aux mesures d'erreurs (RMSE) calculées lors de la phase d'entraînement des modèles pour choisir la meilleure structure.

Nous présentons également le meilleur modèle choisi pour chaque produit et les paramètres associés à ce dernier.

Le tableau 15 montre les erreurs d'entraînement des modèles pour différents produits pour un horizon H3 pour les prévisions du mois de novembre 2020. Les valeurs marquées en vert correspondent aux erreurs minimales identifiées lors de la phase de compétition dans la méthodologie et donc la structure du modèle choisi pour chaque produit.

PRODUIT	ARIMA	LISSAGE EXP	PROPHET	SARIMA	CHOIX	PARAMETRES
D1	-	161	182	173	LE Triple	['add', False, 'add']
D2	4	4	4	-	ARIMA	[3, 0, 1]
D3	-	5	6	-	LE Double	['add', True, None]
D4	29	16	25	19	LE Double	[None, False, 'add']
D5	308	197	531	375	LE Triple	['add', False, 'mul']
D6	-	568	592	837	LE Triple	['add', True, 'mul']
D7	37	16	18	24	LE Double	[None, False, 'mul']
D8	8	6	12	8	LE Double	['add', False, None]
D9	247	168	214	247	LE Double	[None, False, 'mul']
D10	118	42	106	64	LE Triple	['add', True, 'add']
D11	94	134	146	108	ARIMA	[3, 0, 0]
D12	88	106	197	88	ARIMA	[3, 0, 0]
D13	-	2	0	-	Prophet	-
D14	1	2	7	-	ARIMA	[0, 1, 1]
D15	-	185	300	-	LE Triple	['add', False, 'add']
D16	-	1	11	-	LE Double	['add', False, None]
D17	4	4	48	4	ARIMA	[2, 1, 0]
D18	-	87	227	-	LE Double	[None, False, 'mul']
D19	96	113	153	-	ARIMA	[3, 0, 0]
D20	-	152	131	-	Prophet	-
D21	8	5	11	6	LE Triple	['add', False, 'add']
D22	-	1	1	-	Prophet	-
D23	51	25	48	51	LE Triple	['add', True, 'add']

Tableau 15 : Mesure d'erreur (RMSE) à l'entraînement pour un horizon H3 et choix du modèle par produit pour les prévisions du mois de novembre 2020

Le tableau 16 présente également les erreurs d'entraînement des différents modèles pour les produits de l'usine de Drancy pour un horizon H1 pour les prévisions du mois de novembre 2020.

PRODUIT	ARIMA	LISSAGE EXP	PROPHET	SARIMA	CHOIX	PARAMETRES
D1	-	164	189	230	LE Triple	['add', False, 'add']
D2	3	4	4	-	ARIMA	[3, 0, 1]
D3	-	6	6	-	LE Double	['add', True, None]
D4	26	17	25	19	LE Triple	['add', True, 'mul']
D5	339	212	542	296	LE Triple	['add', False, 'mul']
D6	-	553	494	826	Prophet	-
D7	31	18	20	25	LE Triple	['add', True, 'mul']
D8	9	6	11	9	LE Double	['add', False, None]
D9	270	74	122	270	LE Triple	['add', True, 'mul']
D10	155	65	143	94	LE Double	[None, False, 'add']
D11	86	128	173	96	ARIMA	[3, 0, 0]
D12	102	129	152	102	ARIMA	[3, 0, 0]
D13	-	2	3	-	LE Double	['add', True, None]
D14	6	5	5	-	Prophet	-
D15	-	195	259	-	LE Triple	['add', True, 'add']
D16	-	1	12	-	LE Double	[None, False, 'add']
D17	4	2	103	4	LE Double	['add', True, None]
D18	-	87	270	-	LE Double	[None, False, 'mul']
D19	69	114	47	-	Prophet	-
D20	-	99	64	-	Prophet	-
D21	-	7	11	6	SARIMA	[(0, 0, 0), (1, 0, 1, 12)]
D22	1	1	1	1	ARIMA	[1, 0, 0]
D23	52	33	54	48	LE Triple	['add', True, 'mul']

Tableau 16 : Mesure d'erreur (RMSE) à l'entraînement pour un horizon H1 et choix du modèle par produit pour les prévisions du mois de janvier 2020

Dans ces deux tableaux, certains produits ont des valeurs manquantes. Cela peut être dû au fait que la série temporelle soit trop courte pour entraîner un modèle SARIMA par exemple ou que le modèle utilisé corresponde à un modèle éliminé comme présenté dans le paragraphe II.4. Nous constatons également une domination de la famille des lissages exponentiels (LE) grâce à leur variété de modèles selon leurs composantes ce qui donne un

large choix de techniques. Comme cela peut être observé, les erreurs d'entraînement varient entre les modèles pour chaque produit. Ceci justifie notre choix de nous appuyer sur plusieurs techniques de modélisation et d'associer le modèle le plus adapté à chaque produit.

VI. Conclusion

Ce chapitre a été dédié à la présentation d'une première contribution autour d'une méthodologie employée pour la prévision des ventes de l'entreprise elm.leblanc, et basée sur une compétition entre modèles. Nous avons détaillé également les différents modèles utilisés dans cette approche ainsi que les hypothèses à prendre en considération lors de l'implémentation.

Les expérimentations sur le large catalogue de produits de l'entreprise a permis d'analyser la mise en application de cette approche. En effet, nous avons étudié les comportements et les spécificités des produits modéliser par la même technique afin de mieux comprendre le choix du modèle pour chaque série et identifier ses caractéristiques. Nous avons étudié également les performances des modèles par horizon de prédiction et nous avons identifié ceux qui sont plus adaptés au court terme et ceux qui sont plus performants sur le long et moyen terme de prédiction.

Vue la diversité des modèles employés, nous avons proposé de généraliser notre méthodologie pour d'autres champs d'étude et cadres d'applications. Le pseudo-algorithme décrit ci-dessus permet de répondre à plusieurs autres problématiques de prévisions ayant des séries temporelles univariées comme données. Nous avons souligné aussi quelques limites de cette approche et des cas ne permettant pas sa mise en place.

Dans le chapitre suivant, nous allons proposer une seconde contribution s'appuyant sur la proposition discutée en fin de ce chapitre, à savoir une extension des travaux actuels vers une méthode hybride pour la prévision des ventes.

Chapitre 4 : Méthode d'hybridation : Approfondissement

I. Introduction

Comme nous l'avons évoqué dans le chapitre 2 (état de l'art), l'évolution des techniques d'apprentissage automatique et d'apprentissage profond a permis de développer des nouvelles approches pour les prévisions des ventes, selon quatre approches prépondérantes dans la littérature : des modèles dédiés à la modélisation des séries chronologiques ; des méthodologies employant des approches de Data Mining et d'apprentissage automatique ; des approches basées sur l'apprentissage profond et enfin des méthodes de combinaison et des approches d'hybridation des modèles.

Nous avons vu également, dans notre méthodologie employée dans le chapitre 3, que lors de la mise en compétition des modèles pour identifier la meilleure structure liée à chaque produit, nous nous retrouvons avec deux types de résultats. Pour plusieurs produits, il y a un modèle qui se distingue des autres par des erreurs faibles et donc caractérise mieux le produit. Pour d'autres références, que ça soit pour le court, le moyen ou le long terme, les modèles étaient très proches en termes de performances. Le choix d'un modèle par rapport à un autre peut alors être délicat voire affecter la qualité des prévisions réalisées selon la nature du modèle choisi, notamment pour le moyen et le long terme.

Pour éviter ce biais identifié pour quelques produits, nous proposons de combiner les résultats afin d'exploiter au mieux la puissance de chacun des modèles. Cette combinaison peut être considérée comme une méthode d'hybridation de plusieurs modèles linéaires et non linéaires. Dans ce chapitre, nous allons étudier et tester une approche de combinaison qui permettra d'améliorer les résultats des prévisions dans plusieurs cas.

II. Contexte de l'approche d'hybridation

Comme évoqué lors du chapitre 2, les approches d'hybridation présentent plusieurs avantages par rapport aux autres catégories présentées ci-dessus, comme l'amélioration des prévisions grâce à une meilleure modélisation des motifs présents dans le « signal de vente », et la réduction du risque d'un mauvais choix de modèle grâce à la combinaison des résultats.

Lors de l'étude des résultats présentés au chapitre 3 (tableaux 5 et 6 P.92 ou tableaux 8 et 9 P.100), on remarque que malgré la nature différente des modèles utilisés dans la première approche, les résultats de compétition étaient très proches. Cela peut remettre en question le choix d'un seul modèle par produit notamment lorsque les modèles sont de différente nature et donc ce choix peut impacter les prévisions surtout sur le moyen et le long terme. Prenant l'exemple d'un modèle saisonnier et un autre non saisonnier qui sont très proches en termes de performances bien qu'ils aient deux natures différentes, sur le moyen ou le long terme l'usage de l'un ne donnera pas les mêmes résultats que l'autre modèle. Cet aspect a poussé notre réflexion à combiner les résultats des modèles au lieu d'utiliser un seul dans le but d'exploiter les particularités de chaque modèle. Il existe plusieurs approches et méthodes utilisées dans la littérature pour la combinaison des modèles. Nous détaillons, dans le paragraphe suivant, notre méthodologie proposée ainsi que les types de modèles et les données utilisées.

III. Méthodologie

III.1. Description de l'approche

Nous voulons ici exploiter les spécificités de chacun des modèles pris indépendamment tout en améliorant la qualité des prédictions. Ceci est possible en faisant des structures hybrides en série (dont les sorties d'un modèle seront les entrées d'un autre), en parallèle (pour combiner les résultats de plusieurs modèles avec une procédure donnée), ou encore avec une structure en série-parallèle en séparant les parties linéaires et non linéaires des séries temporelles. Pour la méthode en série, cette approche considère chaque série temporelle séparément et nécessite une grande quantité de données pour pouvoir entraîner un modèle à chaque fois. La méthode en série-parallèle permettra d'exploiter un nombre très limité de modèles puisque la partie non linéaire est modélisée par un réseau de neurone et la partie linéaire par un autre modèle, il sera ainsi nécessaire d'identifier quels modèles utiliser pour chaque produit.

Dans la première approche proposée au chapitre 3, nous avons utilisé des modèles non saisonniers comme ARIMA et certaines combinaisons de lissage exponentiel, et des modèles saisonniers comme SARIMA, Prophet et d'autres combinaisons de lissage exponentiel. Pour les produits du catalogue de l'entreprise, certains modèles ont excellé en fournissant de bons résultats, tandis que d'autres étaient moins réussis. Notre méthode vise alors à unifier les

résultats des modèles et trouver le bon équilibre pour réduire encore les erreurs de prédiction à l'échelle globale.

Dans cet objectif, nous avons employé une hybridation parallèle des modèles pour prendre en considération les caractéristiques de chacun. Ces modèles employés correspondent aux mêmes qui ont été présentés dans le chapitre 3 (paragraphe II.4), donnant ainsi la possibilité d'exploiter différentes structures et des méthodes de différentes natures. La figure 34 présente notre méthodologie de combinaison des modèles.

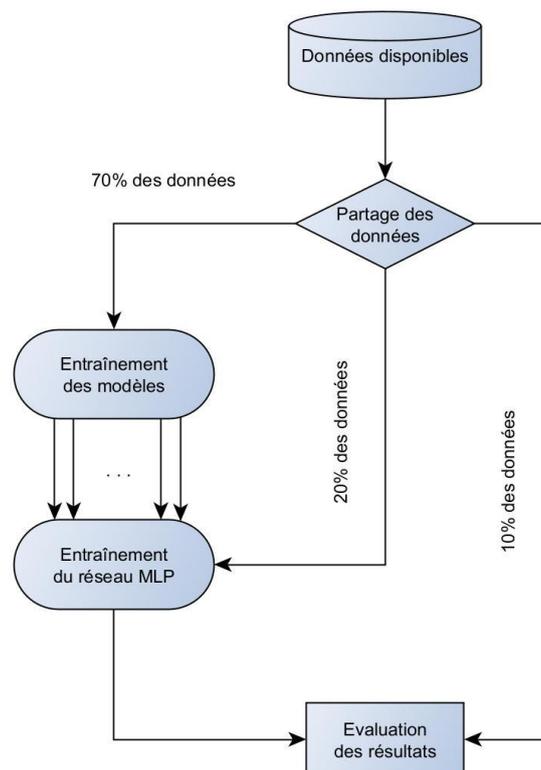


Figure 34 : Approche de combinaison des modèles

Dans notre approche, nous allons utiliser les prédictions des différents modèles pour trouver la meilleure fonction de combinaison linéaire ou non linéaire qui permettra de s'approcher au maximum des ventes réelles. Nous proposons alors l'usage d'un réseau de neurones de type Multilayer Perceptron (MLP) qui prendra en compte les possibles erreurs ou les réponses divergentes des modèles pour en créer un meilleur. Ce MLP permettra de multiplier les résultats des modèles par des coefficients spécifiques en donnant des poids plus ou moins importants aux résultats de chaque modèle, et donc repérer les endroits où les résultats de certains modèles sont moins intéressants pour les compenser par d'autres.

III.2. Description des données

Pour notre première approche de choix du meilleur modèle, nos données d'entrées étaient l'ensemble des séries temporelles représentant les historiques des ventes des différents produits. Nos données utilisées présentent des ventes mensuelles d'environ 6 ans d'historiques pour les produits les plus anciens et de quelques mois pour les plus récents. Étant donné la longueur des séries temporelles et leur nature, les modèles sont plus ou moins performants dans leurs modélisations des comportements de ventes des produits. De manière assez logique, certains ont été bien modélisés notamment ceux avec des longues séries temporelles, et d'autres moins bien surtout les produits récemment lancés.

Les résultats de prévision de la première approche proposée correspondent à des valeurs mensuelles selon l'horizon de prédiction introduit par l'utilisateur. L'idée de cette deuxième approche est alors de combiner les résultats des modèles utilisés dans la première approche.

A partir des données brutes qui correspondent à l'historique des ventes mensuelles des différents produits sous forme de séries temporelles, nous allons créer une base de données pour l'entraînement et la validation de cette approche. Nous allons partager notre base de départ en 3 parties. La première partie des données sera utilisée pour l'entraînement des modèles séparément pour chacun des produits. La deuxième partie sera réservée pour l'entraînement de la méthode de combinaison qui sera employée. Une dernière partie (base de test) pour évaluer les résultats des prédictions du modèle de combinaison (MLP) par rapport aux ventes réelles et donc tester sa convergence.

Comme nous pouvons le constater, les différents besoins en données pour l'entraînement à la fois des modèles et de la méthode de combinaison ainsi que pour l'évaluation, nécessite des séries temporelles de plus longue durée pour favoriser un bon apprentissage des modèles. Ceci peut restreindre l'usage de cette approche aux anciens produits ayant suffisamment de données historiques.

Vu la nécessité d'avoir des quantités de données importantes pour pouvoir entraîner les modèles et le MLP pour ensuite tester leurs performances, nous allons restreindre l'usage de cette approche aux produits éligibles c'est-à-dire ceux ayant plus de deux années d'historiques pour être en capacité d'entraîner les modèles saisonniers et avoir suffisamment de données pour tester le réseau de neurones.

III.3. Présentation du réseau MLP utilisée dans l'approche

Dans le domaine de l'apprentissage automatique, les perceptrons multicouches (MLP) représentent un type de réseau de neurones largement utilisé. La puissance de ce réseau de neurones vient de sa capacité à apprendre n'importe quel motif dans les données d'entraînement et à le relier au mieux à la variable de sortie qu'on souhaite prédire.

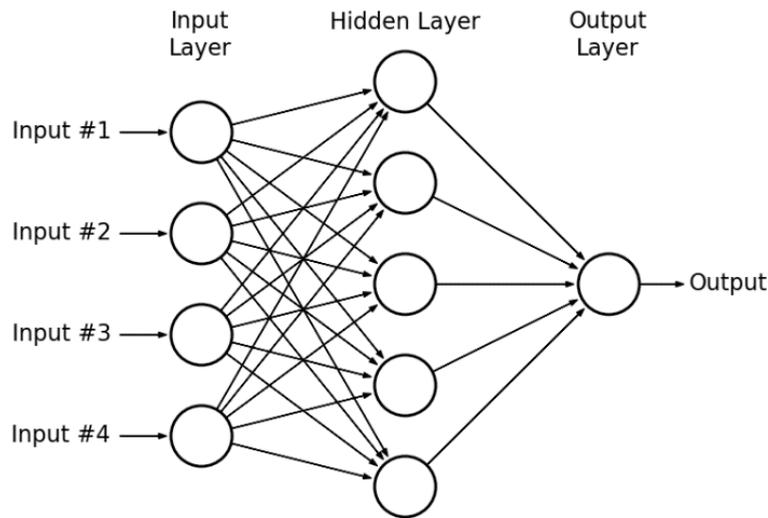


Figure 35 : Exemple d'un réseau de neurones de type MLP

Dans notre méthodologie, nous avons testé plusieurs architectures de réseaux de neurones (tableau 17) dans le but d'identifier celle qui modélise au mieux notre problématique de prévisions.

Architectures testées	Description
(13, 4, 1)	2 couches internes de 13 et 4 neurones et 1 couche de sortie
(13, 6, 1)	2 couches internes de 13 et 6 neurones et 1 couche de sortie
(13, 13, 1)	2 couches internes de 13 et 13 neurones et 1 couche de sortie
(13, 10, 5, 1)	3 couches internes de 13, 10 et 5 neurones et 1 couche de sortie
(13, 6, 3, 1)	3 couches internes de 13, 6 et 3 neurones et 1 couche de sortie
(13, 10, 6, 3, 1)	4 couches internes de 13, 10, 6 et 3 neurones et 1 couche de sortie

Tableau 17 : Description des architectures testées

Le diagramme de la figure 36 présente une des architectures testées dans notre méthodologie.

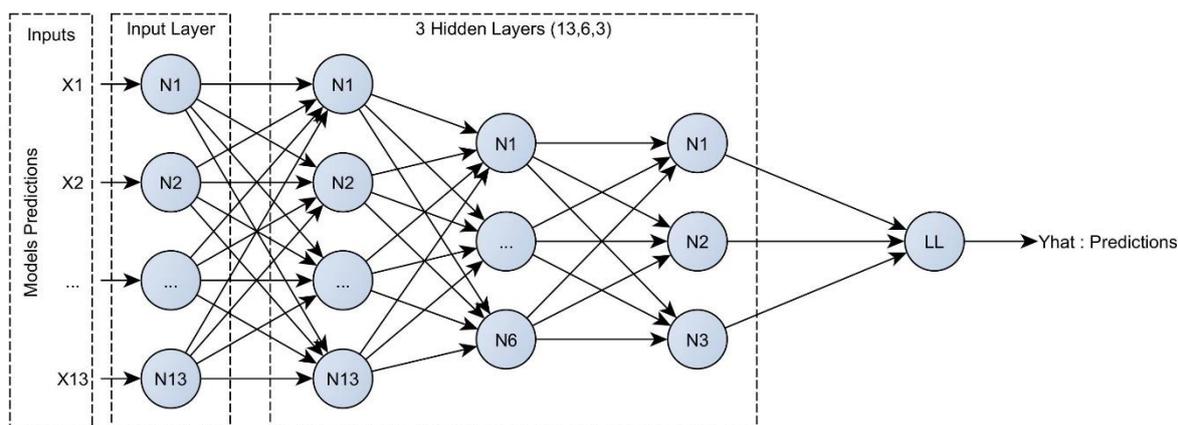


Figure 36 : Exemples d'architecture de réseau de neurones (13, 6, 3,1) testée dans notre méthodologie

Nous avons utilisé une couche d'entrée composée de 13 neurones qui correspondent aux nombres des données d'entrée fournies au réseau. Ces données correspondent aux prédictions des 13 modèles utilisées. Nous avons testé différentes architectures internes avec une ou plusieurs couches cachées contenant chacune plusieurs neurones. La fonction d'activation pour chaque neurone était de type ReLU (Rectified Linear Unit). La couche de sortie est composée d'un seul neurone pour calculer les prédictions finales.

III.4. Processus d'apprentissage et de test

Cette méthodologie d'hybridation comprend deux étapes. La première concerne l'entraînement et l'usage des modèles utilisés lors de la première approche (chapitre 3) pour préparer les données d'entrée du réseau. D'abord, l'historique des données de ventes composé des séries temporelles représentant chacune un produit du catalogue, sont partagées en trois parties pour l'entraînement des modèles de prédictions, entraînement du réseau MLP et pour tester la validité des résultats obtenus. Le schéma de la figure 37 présente la répartition des données dans notre méthodologie.

Entraînement des modèles (70%)	Entraînement du MLP (20%)	Test (10%)
--------------------------------	---------------------------	------------

Figure 37 : Répartition des données d'entraînement et de test

Les modèles sont entraînés sur 70% de l'ensemble des données disponibles. Ensuite, leurs prédictions sont réalisées sur 20% des données. Ces dernières constituent, avec les données réelles des ventes, les entrées du réseau de neurones (MLP) utilisées pour l'entraîner. L'objectif est d'apprendre au réseau à exploiter les résultats des meilleurs modèles et donc à compenser les faiblesses des autres. Ensuite, ses performances sont testées sur les 10% des

données restantes pour évaluer la qualité des prédictions obtenues. Dans cette approche, nous sommes en train d'exploiter le réseau comme une fonction de régression qui apprend à mapper les entrées (les prédictions des modèles) en sortie (les ventes réelles).

IV. Expérimentations

IV.1. Les données utilisées pour les expérimentations

Nous avons expérimenté l'efficacité de cette approche en utilisant des données de prédiction d'un seul horizon à chaque fois, et aussi en utilisant les données de plusieurs horizons en même temps pour construire un réseau multi horizon.

Les prédictions des modèles sont réalisées sur un ou plusieurs horizons de prédiction. Pour générer les données d'entrée du MLP, nous allons entraîner chacun des modèles sur 70% des données disponibles pour chaque produit, ensuite, nous allons réaliser des prédictions pour les 30 % des données restantes. 20 % de ces données seront utilisées pour entraîner le réseau et 10% pour évaluer ses performances. Dans le premier cas, les prédictions réalisées sont à un seul horizon à chaque fois, et donc les données d'entrée d'un réseau seront les prédictions des modèles à cet horizon. Ceci nous permettra de créer 6 modèles de réseaux de neurones pour les 6 horizons de prédiction testés. Dans le deuxième cas, les prédictions réalisées à plusieurs horizons seront les entrées d'un seul modèle de combinaison et donc nous allons construire un modèle multi horizons.

Le tableau 18 présente un exemple des données d'entrée fournies au réseau avec les prédictions de 13 modèles en entrée : Prophet, ARIMA, SARIMA, 10 modèles de lissage exponentiel. Nous rappelons que, pour les modèles de lissage exponentiel, la première composante représente la tendance avec une nature additive ou pas, le deuxième paramètre indique si la tendance est amortie, et la dernière composante représente la saisonnalité avec une nature additive, multiplicative ou l'absence de saisonnalité.

Produits	Prophet	ARIMA	SARIMA	(None, True, add)	(None, True, mul)	(None, False, add)	(None, False, mul)	(add, True, None)	(add, True, add)	(add, True, mul)	(add, False, None)	(add, False, add)	(add, False, mul)	Ventes
Produit 1	846	971	838	0	0	821	846	960	769	806	959	752	744	745
Produit 2	810	971	758	0	0	858	880	960	817	841	959	806	781	639
Produit 3	989	971	932	0	0	952	971	960	950	943	958	913	908	486
Produit 4	659	971	586	0	0	546	594	960	495	554	958	470	500	332
Produit 5	1011	971	1184	0	0	1102	1131	959	1100	1095	957	1050	1032	585

Tableau 18 : Exemples des données d'entrée du réseau avec 13 entrées

Chaque ligne du tableau constitue les valeurs prédites par chacun des modèles des produits, et la dernière colonne correspond aux ventes réelles enregistrées.

Pour le cas où notre méthodologie est testée pour un seul horizon, nous allons effectuer les prédictions à un seul horizon avec les différents modèles et ensuite les utiliser comme données d'entrée pour le réseau. La figure 38 illustre cette implémentation.

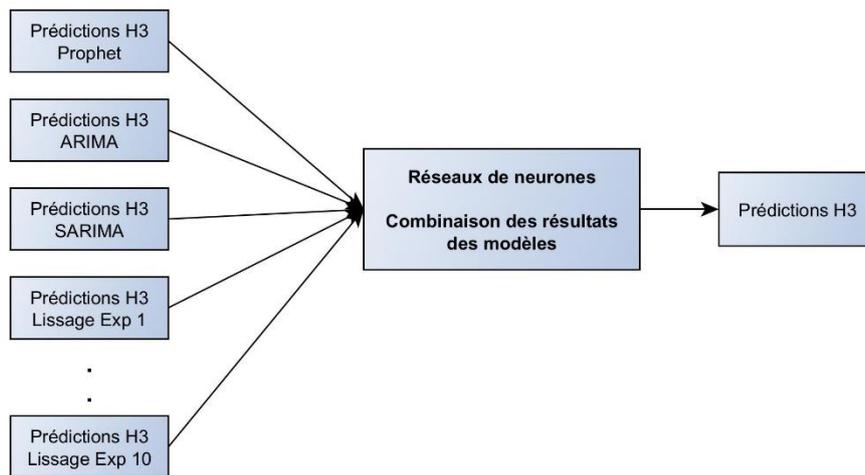


Figure 38 : Exemple d'un modèle à un seul horizon H3

Dans le cas où notre méthodologie est testée en utilisant les données de prédiction des modèles pour plusieurs horizons, nous allons introduire au réseau les prédictions des modèles aux différents horizons en même temps. Ces données supplémentaires apporteront plus d'informations au réseau, ce que constituera de ce dernier un modèle multi horizons. La figure 39 illustre cette implémentation.

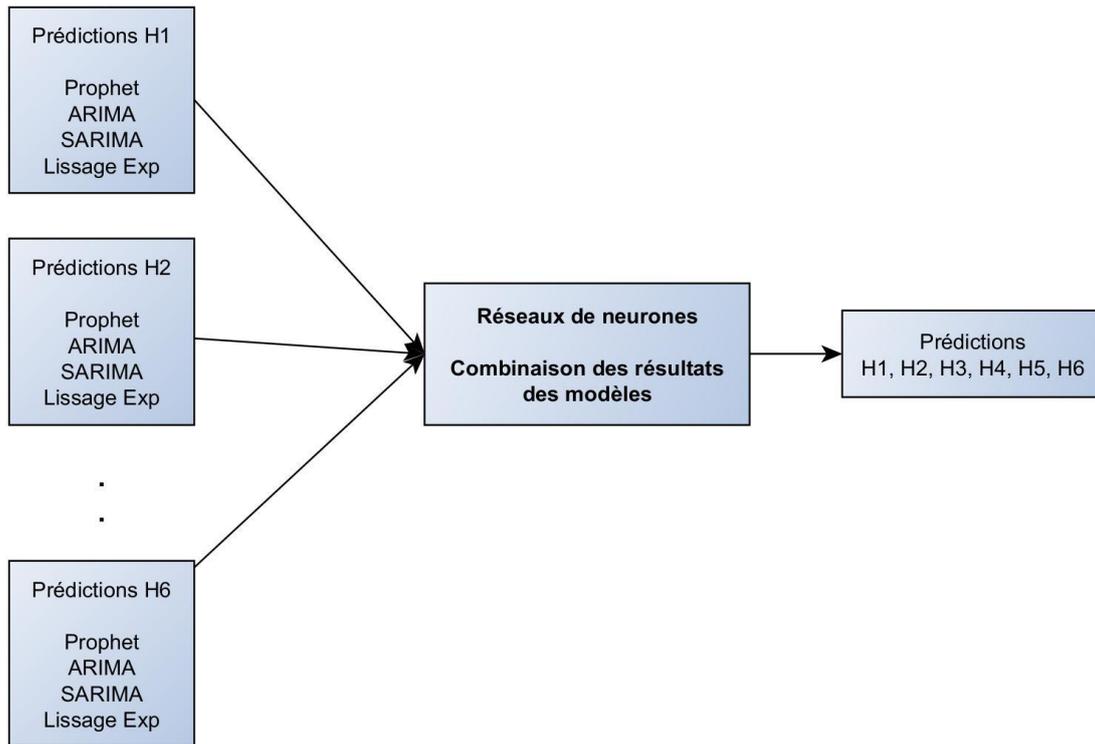


Figure 39 : Exemple d'un modèle multi horizons

IV.2.Choix de configuration

Ce choix consiste en tester plusieurs configurations possibles en manipulant les paramètres du réseau. Ceci peut être réalisé en modifiant le nombre des couches cachées, le nombre des neurones dans chaque couche et les fonctions d'activation dans les neurones. Nous avons testé dans notre méthodologie plusieurs configurations possibles dans le but d'identifier celle qui fournit des meilleurs résultats finaux en termes de prédiction des ventes. Dans le tableau 19 nous présentons les configurations testées avec les mesures d'erreurs enregistrées pour chaque configuration, en utilisant un modèle pour chaque horizon et un modèle multi-horizons. Nous avons utilisé la fonction ReLU comme fonction d'activation dans chacun des neurones et le RMSE comme mesure d'erreur. Le nombre d'epochs est fixé à 300 avec une taille de batch égale à 5. Ce choix est justifié par différentes expérimentations réalisées pour s'assurer de ne pas avoir un sous/sur entraînement du réseau, tout en gardant un temps d'exécution raisonnable.

Architecture	Modèle à un seul horizon						Modèle multi horizons
	H1	H2	H3	H4	H5	H6	
(13, 4, 1)	304	295	297	310	335	316	278

(13, 6, 1)	297	294	291	309	332	309	250
(13, 13, 1)	304	303	289	319	351	325	275
(13, 10, 5, 1)	316	301	306	311	315	309	291
(13, 6, 3, 1)	303	300	321	316	344	306	284
(13, 10, 6, 3, 1)	295	308	300	309	340	315	292

Tableau 19 : Erreurs des architectures des réseaux testées selon l'horizon de prédiction

Les résultats d'erreur correspondent aux erreurs calculées sur les mêmes données de test pour les différentes configurations testées. Le nombre d'épochs et la taille des batch constituent également des paramètres sur lesquelles nous pouvons agir pour améliorer l'efficacité d'entraînement du réseau utilisé, mais nous avons gardé les mêmes paramètres pour toutes les architectures testées.

Au cours des expérimentations et comme nous pouvons le constater sur le tableau 19, le modèle multi horizons permet de fournir des meilleurs résultats par rapport aux autres modèles quelle que soit l'architecture utilisée (couleur verte). Cela s'explique par la quantité des données plus importante fournie à ce modèle puisque nous lui avons fourni en entrée les résultats des modèles pour tous les horizons de différents produits. Ceci lui fournira davantage d'informations que les autres modèles. En plus, les modèles à un seul horizon nécessitent un temps d'exécution important étant donnée la nécessité de construire un modèle pour chaque horizon à prédire. D'après le tableau, les architectures utilisées fournissent des résultats proches entre elles pour les différents horizons ainsi que pour le modèle multi-horizons. Nous pouvons pourtant constater que l'architecture ayant deux couches cachées avec respectivement 13 et 6 neurones avait de meilleurs résultats de prédictions par rapport aux autres architectures, notamment pour le modèle multi-horizons (couleur bleue).

D'après ces résultats, nous avons retenu le modèle multi-horizons ainsi que l'architecture ayant deux couches cachées de 13 et 6 neurones pour implémenter et analyser ses résultats. La figure 40 présente l'architecture finale retenue pour le modèle multi horizons.

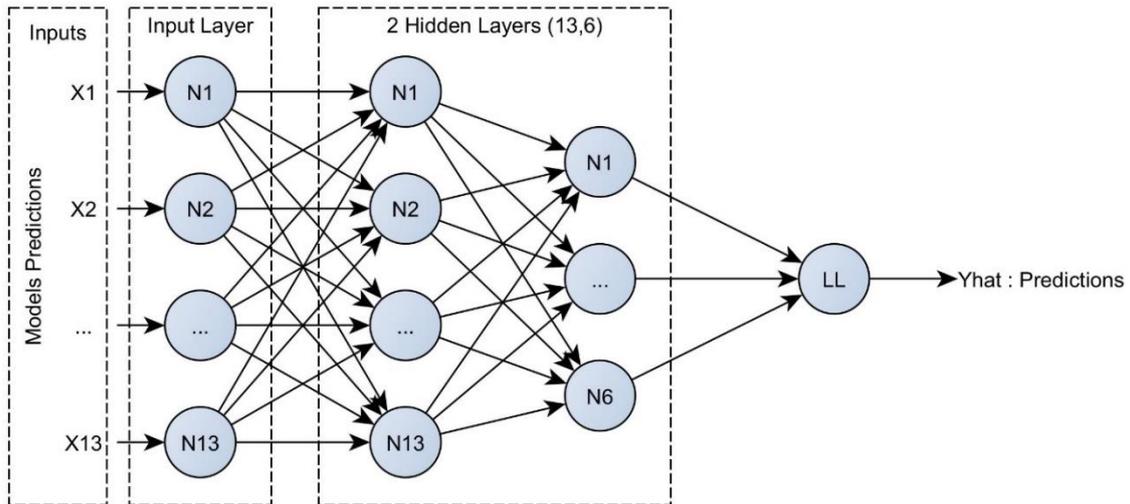


Figure 40 : Architecture retenue du réseau utilisée pour la combinaison (13, 6, 1)

Comme indiqué auparavant, la couche d'entrée reçoit les prédictions des modèles pour plusieurs produits à différents horizons et les envoie vers la première couche cachée à 13 neurones. Les neurones des deux couches cachées du réseau ainsi que de la dernière couche utilisent la fonction ReLU. Pour chaque ligne dans le tableau d'entrée, le réseau combine ces données des différents modèles pour fournir une prévision finale, et donc pour chaque ligne en entrée, nous aurons une prédiction en sortie.

IV.3. Résultats obtenus après implémentation

Dans le but de mesurer l'efficacité de la méthodologie développée, nous avons réalisé plusieurs expérimentations de l'architecture choisie. Dans le tableau 19, nous avons évalué les différentes configurations sur les mêmes données représentées par le dernier trimestre de l'année de 2019. Les erreurs calculées sont par rapport aux ventes réelles réalisées durant cette période. Pour cette dernière étape d'évaluation, nous allons tester l'architecture retenue sur différentes données par validation croisée.

Dans ces expérimentations, nous avons utilisé 30% des données pour entraîner et tester cette architecture. En utilisant la validation croisée avec différentes valeurs de K, nous avons évalué notre méthodologie sur des ensembles de données différents à chaque fois, et par la suite sa capacité de généralisation sur l'ensemble des données d'entrée. K désigne le paramètre de la validation croisée c'est-à-dire, le nombre de partitions de données créées à chaque fois. Pour le cas où $k=5$, les données sont partagées en 5 partitions et ensuite, à chaque fois le modèle est entraîné sur 4 partitions et est testé sur la partie restante. Ceci est effectué 5 fois pour que

la méthode soit évaluée sur l'ensemble des données disponibles, et finalement l'erreur moyenne est calculée pour ce modèle. Dans le tableau 20, nous présentons les erreurs calculées pour différentes valeurs de k, et pour les différents modèles séparés ainsi que pour le modèle de combinaison de l'architecture choisie.

	RMSE k= 1	RMSE k= 2	RMSE k= 3	RMSE k= 4	RMSE k= 5
prophet	392	384	374	354	359
arima	465	443	416	441	429
sarima	401	370	362	365	372
(None, True, 'add')	1138	1193	1150	1187	1107
(None, True, 'mul')	1138	1193	1150	1187	1107
(None, False, 'add')	355	333	337	328	334
(None, False, 'mul')	433	416	435	417	415
('add', True, None)	431	427	395	442	419
('add', True, 'add')	372	338	348	340	341
('add', True, 'mul')	430	418	433	403	409
('add', False, None)	439	435	405	458	437
('add', False, 'add')	375	344	351	353	353
('add', False, 'mul')	432	418	435	409	413
Models combination	308	300	290	292	286

Tableau 20 : Erreurs de prédictions des modèles par rapport à la combinaison

D'après le tableau 20, nous remarquons que les erreurs minimales correspondent aux erreurs réalisées par l'approche de combinaison, en comparant avec les prédictions réalisées par les modèles séparément pour les différentes valeurs de k testées. Les erreurs correspondent à la moyenne de la RMSE calculée pour chaque partition pour chaque valeur de k. Les valeurs marquées en vert représentent les erreurs minimales constatées. Les deux modèles n'ayant pas une composante de tendance amortie présentent les erreurs les plus importantes. On remarque aussi que plusieurs modèles se rapprochent beaucoup en termes d'erreurs malgré leurs natures différentes ce qui renforce l'intérêt d'une combinaison de ces modèles.

Nous avons effectué également une analyse de sensibilité pour cette approche dans le but d'identifier le ou les modèle(s) ayant le plus d'importance dans le calcul des prédictions finales par le réseau. Cette étude de sensibilité est réalisée pour savoir quelle entrée du réseau i.e. quel modèle utilisé est le plus important et qui affecte le plus le résultat final de prédiction.

Le processus se réalise en mélangeant les valeurs d'une seule variable d'entrée i.e. les données de prédiction d'un modèle, à chaque fois et en calculant les erreurs des prédictions finales après cette modification. Cette technique est répétée pour toutes les variables d'entrée du réseau. Les erreurs maximales correspondent aux variables les plus importantes, c'est-à-dire que leur modification affecte le plus les performances du modèle de combinaison final, et donc les variables ayant des erreurs les plus importantes correspondent aux modèles i.e. variables d'entrée les plus pertinents. Dans les deux graphes ci-dessous, nous présentons l'importance de chacun des modèles utilisés en entrée du réseau dont ces résultats sont présentés au tableau 20.

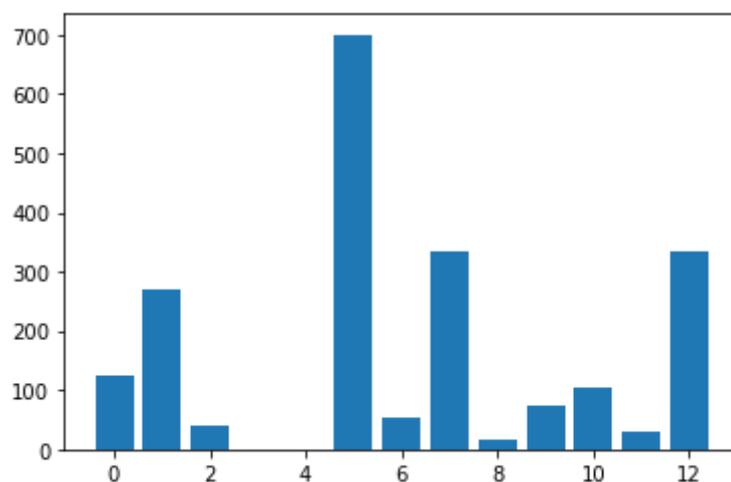


Figure 41 : Erreurs réalisées par le modèle final en modifiant à chaque une variable d'entrée

La figure 41 présente les erreurs (RMSE) constatées dans les résultats finaux en modifiant à chaque fois les variables concernées. L'axe des ordonnées correspondent aux mesures d'erreur et l'axe des abscisses aux modèles utilisées en entrée : 0 - Prophet, 1 - ARIMA, 2 - SARIMA, 3 - (None, True, add), 4 - (None, True, mul), 5 - (None, False, add), 6 - (None, False, mul), 7 - (add, True, None), 8 - (add, True, add), 9 - (add, True, mul), 10 - (add, False, None), 11 - (add, False, add), 12 - (add, False, mul). Les paramètres des modèles de lissage exponentielle correspondent à la nature de la tendance et si elle est amortie, ainsi que la nature de la saisonnalité. Les mêmes résultats peuvent être constatés dans la figure 42 également.

Weight	Feature
712.9983 ± 15.4815	(None, False, 'add')
341.2252 ± 10.1338	('add', True, None)
332.8036 ± 10.5281	('add', False, 'mul')
263.8570 ± 16.0376	arima
124.1046 ± 2.8420	prophet
104.5681 ± 5.7520	('add', False, None)
74.1631 ± 5.7244	('add', True, 'mul')
51.3464 ± 12.2820	(None, False, 'mul')
40.8099 ± 5.5123	sarima
30.4884 ± 3.7624	('add', False, 'add')
17.7594 ± 4.1696	('add', True, 'add')
0 ± 0.0000	(None, True, 'mul')
0 ± 0.0000	(None, True, 'add')

Figure 42 : Erreurs réalisées par le modèle final en modifiant à chaque fois une variable d'entrée

Nous remarquons que, pour ces résultats, le modèle de lissage exponentiel sans tendance et avec une saisonnalité additive représente la variable d'entrée la plus importante du réseau et sa modification a impliqué le plus d'erreurs pour les résultats finaux. Les trois modèles les plus pertinents sont ensuite, un modèle de lissage exponentiel avec tendance additive et sans composante saisonnière, un modèle de lissage exponentiel avec une tendance additive et une saisonnalité multiplicative et le modèle ARIMA. On constate que les quatre premiers modèles les plus importants sont de différente nature avec deux modèles saisonniers et deux modèles non saisonniers. Ceci montre qu'un seul modèle n'est pas suffisant pour modéliser le catalogue de produit de l'entreprise et que l'exploitation des spécificités de plusieurs modèles par la combinaison peut améliorer les performances des prédictions.

V. Conclusion

Nous avons étudié dans ce chapitre une nouvelle approche qui permettra de surmonter quelques limites constatées dans la première approche proposée au chapitre 3. La méthodologie proposée consiste en la combinaison des prédictions des différents modèles pour obtenir de meilleures prédictions en exploitant les caractéristiques de chaque modèle. Les natures différentes des modèles utilisés permettent de mieux modéliser les comportements de ventes des produits. L'usage de plusieurs modèles réduit le risque des grandes erreurs surtout pour les prévisions au long et au moyen terme en se ne basant pas que sur un seul modèle. En revanche, cette approche présente quelques limites notamment, la nécessité de beaucoup de données pour assurer les étapes d'entraînement des modèles de séries temporelles et du réseau de neurones, ce qui rend l'approche applicable plus aux

anciens produits. L'entraînement des modèles et du réseau est également coûteux en termes de temps d'exécution ce qui pourra limiter leur usage pour la préparation des prévisions rapidement.

En comparaison avec la méthodologie présentée au chapitre 3 où nous sélectionnant le meilleur modèle pour chaque produit, cette méthodologie de combinaison permet de réduire le risque d'un mauvais choix de modèle mais en rajoutant une complexité supplémentaire qui se traduit par une quantité de données importante et par l'usage d'un réseau de neurone. Pour ces raisons, la première méthodologie demeure plus efficace pour modéliser le catalogue produit de l'entreprise et pour répondre à ces besoins de prédiction des ventes. En plus, vu le grand nombre de produits avec des courtes séries temporelles, la deuxième méthodologie ne sera pas applicable pour calculer leurs prévisions. Pour conclure, le meilleur compromis sera de mettre en place la première méthodologie proposée tout en gardant à l'esprit le risque qui peut exister. Cette implémentation sera le sujet du chapitre 5.

Chapitre 5 : Industrialisation des travaux et mise en production

I. Introduction

Nous avons présenté tout au long du troisième chapitre la méthodologie proposée pour améliorer les prévisions des ventes pour le catalogue de produits de l'entreprise elm.leblanc. Cette méthodologie repose sur un ensemble de modèles de prédiction des séries temporelles et permet de modéliser les produits avec la meilleure estimation possible en mettant en compétition plusieurs modèles. Cette méthodologie a été développée pour répondre au besoin de l'entreprise d'avoir un outil de prévision automatisé et plus précis que la méthode actuelle employée. Cette solution doit être accessible aux différents acteurs concernés et sera intégrée dans le processus S&OP pour faciliter la planification de la production selon la demande. Le principal enjeu sera de rendre cet outil utilisable par les membres des équipes avec une facilité et aisance grâce à une implémentation dans un environnement industriel de production pour assurer son utilité, son accessibilité et sa pérennité.

Dans le cadre de cette thèse, les travaux exploratoires ont été développés et testés dans un environnement Python en exploitant plusieurs librairies de ce langage, et implémentés sur une machine locale au sein de l'entreprise. Néanmoins, plusieurs problématiques interviennent lors de l'implémentation dans un environnement local. Les durées de vie prévues des machines sont généralement courtes et peuvent être raccourcies suite à des accidents ou des défauts matériels. Les données et les packages installés sont en risque constant de perte ou de suppression accidentelles ou inattendues par l'utilisateur. En plus, les packages installés sont toujours en amélioration continue suite à des mises à jour ce qui augmente le risque de non compatibilité des versions des librairies et donc nécessite une maintenance régulière.

Pour éviter ces risques et assurer la pérennité de la solution et le bon fonctionnement et l'exploitation de cet outil, nous nous sommes dirigés vers l'implémentation des travaux sur un service cloud vu les différents avantages que présente cet environnement par rapport à un environnement physique local. Parmi ces avantages on peut citer :

- **Haute disponibilité** : les applications cloud peuvent fournir une expérience utilisateur continue, en limitant les temps d'arrêt et d'indisponibilité en cas de problème.

- **Scalabilité & Élasticité** : les applications cloud peuvent être configurées pour utiliser la mise à l'échelle automatique, ce qui garantit que les applications disposent toujours des ressources dont elles ont besoin.
- **Agilité** : les ressources cloud peuvent être déployées et configurées rapidement, au fur et à mesure que les besoins évoluent.
- **Géodistribution** : les applications et les données peuvent être déployées dans des centres de données à travers le monde entier, garantissant ainsi toujours des meilleures performances pour chaque site.
- **Reprise d'activité après sinistre** : grâce aux services de sauvegarde dans le cloud, à la réplication des données et à la géodistribution, les applications déployées et les données sont sécurisées en cas de sinistre.

Dans la première partie de ce chapitre, nous allons discuter de l'intégration de la solution dans le processus S&OP et la planification des processus de la Supply Chain et de son rôle d'équilibrage de l'offre et la demande. Ensuite, nous présentons la valeur ajoutée de cet outil dans le processus de prévisions des ventes et comment il impacte des modifications dans le processus S&OP existant. La deuxième partie sera consacrée à la présentation des services cloud utilisés et nous décrivons les étapes d'implémentation des travaux réalisés sur une de ces plateformes. Dans la dernière partie de ce chapitre, nous présentons la mise en production de l'outil développé et son intégration dans les processus de l'entreprise, ainsi que son exploitation par les collaborateurs.

II. Intégration de l'outil dans le processus S&OP

Comme présenté dans le premier chapitre, le processus S&OP est considéré comme un processus pilote pour la planification de la Supply Chain. En ayant comme but l'équilibrage de l'offre et la demande, ce processus repose généralement sur cinq étapes permettant la mise en place et la cohérence des différents plans de la demande, de production et de réapprovisionnement de la matière première (voir figure 1 dans le chapitre 1).

Les travaux réalisés sur l'amélioration des prévisions des ventes s'intègrent dans la première étape et vise à fournir une base solide pour le reste des étapes du processus. L'outil développé vise à fournir un plan de prévision pour l'ensemble des références de produits du catalogue

de l'entreprise constituant ainsi le sujet de validation lors de la deuxième étape. Concrètement, l'outil développé doit être utilisé en début de chaque mois pour calculer des prévisions à court terme et à moyen terme pour chaque produit. Ensuite, ces prévisions sont discutées et validées par les différents responsables des départements de vente, de production, de marketing et de logistique durant la deuxième réunion de la planification de la demande comme le montre la figure 43.

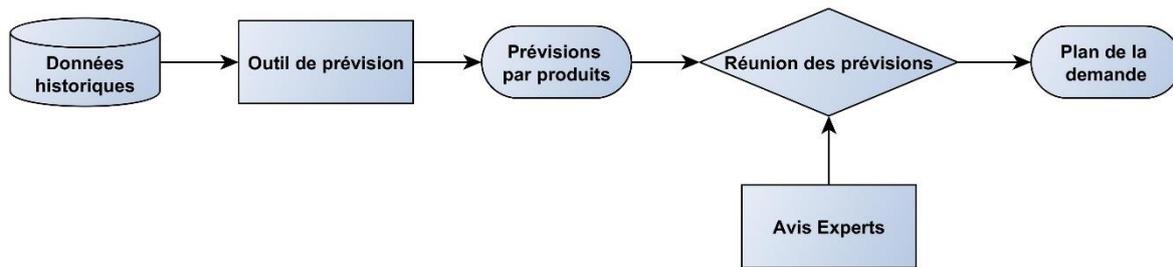


Figure 43 : Nouveau processus de prévisions des ventes

Ces discussions peuvent apporter des nouveaux éléments qui ne sont pas pris en considération par les modèles lors de calcul des prévisions comme l'arrêt de certains produits ou la mise en place des campagnes publicitaires sur certaines gammes, ou encore l'obtention d'un ou plusieurs marchés de type chantier nécessitant ainsi des quantités supplémentaires des produits. Cette réunion permet également de prendre en considération les retours du terrain à travers les informations renvoyées par les directeurs commerciaux régionaux qui sont en contact direct avec les clients et les distributeurs. Tous ces éléments sont assimilés pour rectifier si besoin le plan de prévision déjà préparé par l'outil, d'où l'importance d'intégrer les meilleures structures de modèles dans le cadre de ce processus S&OP. Une fois le plan de la demande est validé, il sera renvoyé au service de production pour poursuivre le cycle du processus.

L'importance de l'outil s'avère non seulement en préparant des prévisions plus précises, rapidement et d'une façon automatique, mais aussi en traitant un grand nombre de références de produits. En effet, les discussions menées lors de la réunion de plan de la demande concernent plus les familles de produits et les références les plus vendues qui subissent des changements, et donc beaucoup de produits avec des faibles quantités ne sont discutés qu'à l'échelle de leur famille et par la suite les prédictions des produits individuels peuvent être moins précises. Cette capacité qu'a l'outil développé de permettre de calculer

des prévisions pour tout type de référence, quel que soit le volume de vente réalisé, représente dans ce contexte un apport significatif. De fait, cet outil joue un rôle important pour renforcer et améliorer les prévisions de ventes et soutenir le processus S&OP dans l'entreprise, en ayant un impact indirect sur le reste des étapes.

Dans le paragraphe suivant, nous présentons le cloud computing et ses spécificités, quelques solutions disponibles sur le marché ainsi que la solution choisie et sur laquelle nos travaux ont été implémentés, en détaillant plusieurs éléments importants liés à une intégration dans cet environnement ainsi que l'architecture mise en œuvre pour l'implémentation.

III. Le Cloud Computing

Le cloud computing est le terme qui désigne l'ensemble des services informatiques fournis sur internet par plusieurs entreprises du marché, et qui incluent des services de calcul et de stockages de données, des logiciels, des outils analytiques et d'intelligence artificielle, etc... Dans le cloud computing, il existe trois différents modèles de service cloud, connus sous les noms « Infrastructure As A Service (IaaS) », « Platform As A Service (PaaS) » et « Software As A Service (SaaS) ». Ces modèles définissent les différents niveaux de responsabilité partagée entre un fournisseur et un locataire du Cloud.

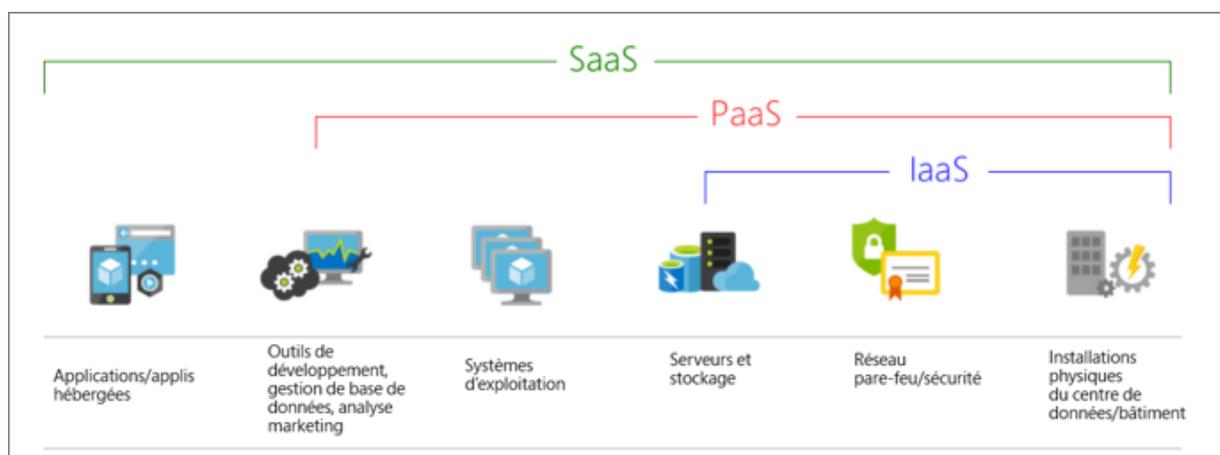


Figure 44 : Modèles des services Cloud

Infrastructure As A Service : Ce modèle de service cloud est le plus proche de la gestion de serveurs physiques : un fournisseur de cloud maintient le matériel à jour, mais la maintenance du système d'exploitation et la configuration du réseau sont laissées au locataire du cloud.

Platform As A Service : Ce modèle de service cloud est un environnement d'hébergement managé, où le fournisseur de cloud gère les machines virtuelles et les ressources réseau et

dans lequel le locataire du cloud déploie ses applications dans l'environnement d'hébergement managé.

Software As A Service : Dans ce modèle de service cloud, le fournisseur de cloud gère tous les aspects de l'environnement d'application (machines virtuelles, ressources réseau, stockage des données, applications, etc.), et le locataire du cloud doit seulement fournir ses données à l'application gérée par le fournisseur.

Pour un service cloud, il existe trois types de déploiement possibles : *cloud public*, *cloud privé* et *cloud hybride*. Chaque modèle de déploiement a différents aspects à prendre en compte lors d'une migration vers le cloud.

- **Cloud public** : Les services sont fournis via internet et sont accessibles à toutes personnes physiques et morales. Les ressources cloud, comme les serveurs et l'espace de stockage, sont détenues et gérées par un fournisseur de services cloud tiers et sont mises à disposition via Internet.
- **Cloud privé** : Un cloud privé se compose de ressources de calcul utilisées exclusivement par des utilisateurs d'une entreprise ou d'une organisation. Un cloud privé peut être situé physiquement dans le centre de données sur site (local), ou être hébergé par un fournisseur de services tiers.
- **Cloud hybride** : Un cloud hybride est un environnement de calcul qui combine un cloud public et un cloud privé en autorisant le partage de données et d'applications entre ceux-ci.

Dans le cas de l'entreprise elm.leblanc et du groupe Bosch, un type hybride est employé permettant ainsi de garder le fonctionnement de quelques applications et le stockage des données sur des serveurs locaux, tout en bénéficiant de certaines fonctionnalités et services du cloud public.

Pour bénéficier des services du cloud computing, il existe sur le marché plusieurs solutions appartenant à différents fournisseurs. Parmi les environnements dominants et les plus connus et utilisés, on distingue Amazon Web Services (AWS) (*Services et Produits de Cloud Amazon | AWS, 2021*), Google Cloud Platform (GCP) (*Premiers Pas – Google Cloud Platform, 2021*) et Microsoft Azure (*Services de Cloud Computing | Microsoft Azure, 2021*). Ces solutions contiennent différents services d'hébergement des applications, de stockage et gestion des

données, d'intelligence artificielle, des ressources de calcul, etc.... Ces trois géants proposent des services très similaires mais se distinguent entre eux en termes de maturité sur le marché et des coûts d'exploitation. GCP s'avère la solution la moins chère entre eux, néanmoins, AWS est la plus mature et propose le plus large éventail de produits. Quant à Microsoft Azure, il propose une large gamme de services également et est plus demandé au sein des organismes qui s'appuient déjà sur d'autres solutions Microsoft. C'est dans ce dernier cas que l'entreprise elm.leblanc se situe. En effet, l'entreprise emploie plusieurs technologies de Microsoft pour le développement de ses applications, d'où le choix d'Azure comme solution de cloud computing pour l'implémentation de l'outil de prévision des ventes.

III.1. Microsoft Azure

Microsoft Azure est une plateforme du cloud computing qui fournit un ensemble de services aux entreprises pour les aider à créer des solutions couvrant une large variété d'objectifs. Les services Azure sont très divers, de l'ordre de 200. La figure 45 ci-dessous présente les principaux services proposés par Azure :

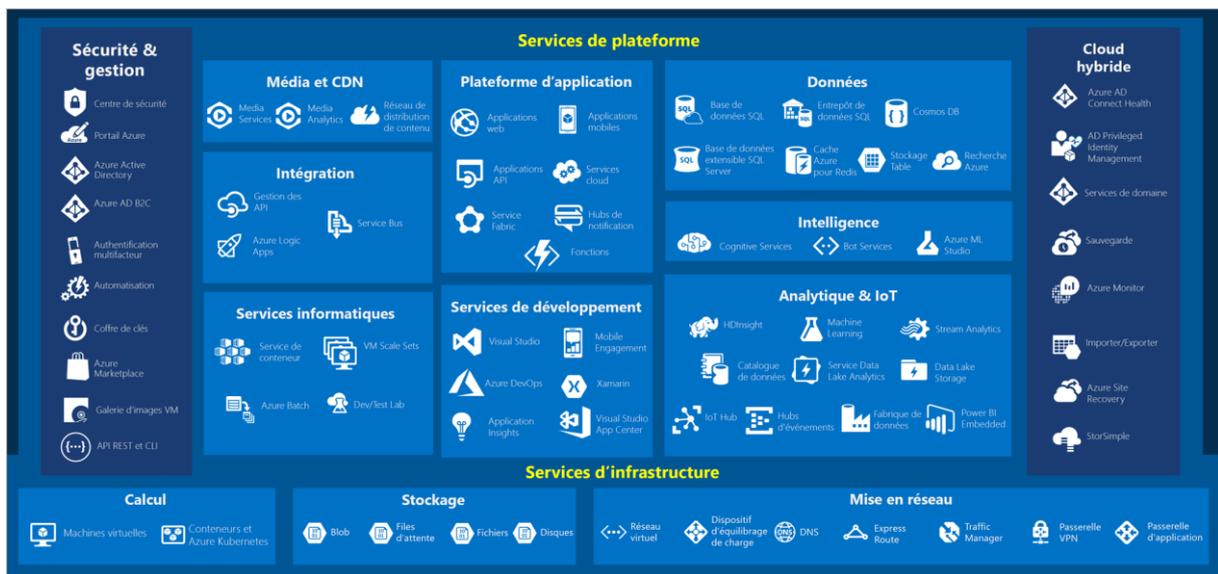


Figure 45 : Exemples des services Cloud d'Azure

Ils vont de simples services web utilisés pour héberger les activités dans le cloud, à des serveurs entièrement virtualisés sur lesquels les solutions logicielles personnalisées sont exécutées. Azure fournit plusieurs services cloud, parmi lesquels le stockage étendu, l'hébergement de bases de données, la gestion centralisée des comptes et l'accès à des nouvelles fonctionnalités liées au calcul, à l'apprentissage automatique et l'intelligence

artificielle et l'Internet des objets (IoT). Azure permet de créer, de gérer et de déployer des applications sur un réseau de dimension mondiale reposant sur une infrastructure massivement distribuée.

Pour assurer le bon fonctionnement de l'outil de prévisions, nous employons principalement deux services de Microsoft Azure à savoir le Blob Storage (*Introduction Au Stockage (d'objets) Blob - Azure Storage | Microsoft Docs, 2021*) et Azure Machine Learning (*Qu'est-Ce Que Azure Machine Learning ? - Azure Machine Learning | Microsoft Docs, 2021*) comme étant des services PaaS. Le premier sera consacré au stockage des données d'entrée et de sortie de l'outil, tandis que le deuxième a pour objectif d'implémenter et de faire apprendre les modèles de prédiction ainsi que les différentes fonctions de nettoyage et de préparation des données. Nous présentons dans les deux paragraphes suivants ces deux services ainsi que ses différents composants.

III.2. Présentation du Blob Storage

Le Blob Storage Azure est un service qui permet de stocker des quantités massives de données structurées et non structurées, ou *objets blob*, dans le cloud. Le Blob Storage peut gérer des milliers de chargements simultanés, des quantités énormes de données vidéo, des fichiers journaux à croissance constante, et est accessible depuis n'importe quel emplacement disposant d'une connexion Internet.

Les bases de données qui correspondent aux historiques des ventes, des références des produits à prédire, des références de remplacement des anciens produits et l'ensemble des fonctions Python développées en local, sont placées sur la solution Blob Storage d'Azure pour les connecter directement à Azure ML. Les résultats des prévisions ainsi que les données de ventes brutes téléchargées chaque mois, sont stockées dans le Blob Storage pour être consommé par l'outil de prévision.

III.3. Azure Machine Learning

Microsoft Azure fournit le service Azure Machine Learning qui correspond à une plateforme cloud qui permet de créer, gérer et publier des modèles d'apprentissage automatique. Elle permet la création des projets end-to-end d'apprentissage automatique. La figure 46 ci-dessous présente les étapes de déroulement d'un projet ML sur cette plateforme.

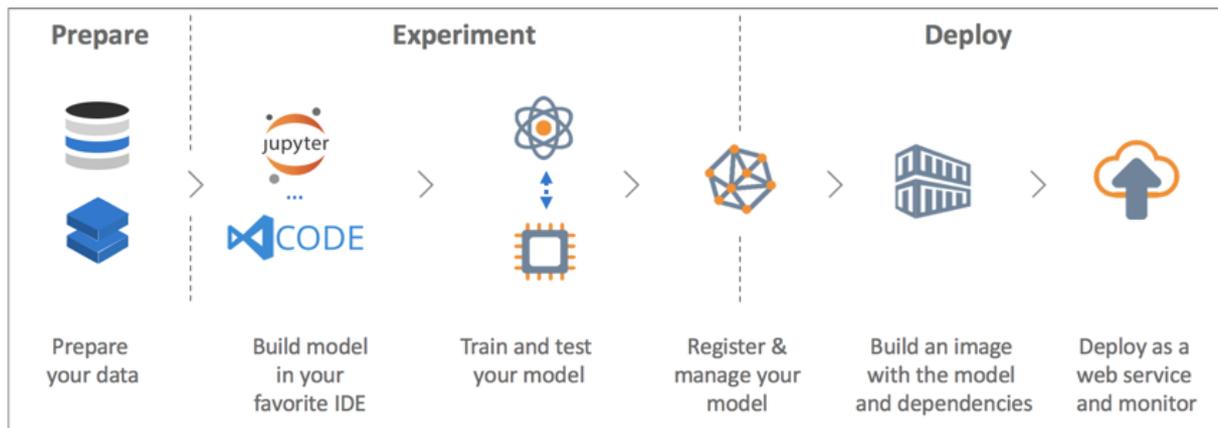


Figure 46 : Etapes de création d'un projet ML sur Azure ML

Azure Machine Learning offre les fonctionnalités et capacités suivantes :

- **Machine Learning automatisée :** Cette fonctionnalité permet de tester plusieurs modèles de Machine Learning présélectionnés selon le cas d'usage.
- **Concepteur Azure Machine Learning :** Une interface graphique permet le développement avec et sans code de solutions Machine Learning.
- **Gestion des données et des calculs :** Un stockage de données dans le cloud et des ressources de calcul que les utilisateurs peuvent employer pour exécuter du code d'expérimentations à grande échelle.
- **Pipelines :** Une fonctionnalité qui permet aux utilisateurs de définir un ensemble d'étapes pour orchestrer les tâches d'entraînement, de déploiement et de gestion des modèles à travers des Kits de développement logiciel (SDK) en langage Python et R.

L'outil concepteur d'Azure ML (*Qu'est-Ce Que Le Studio Azure Machine Learning ? - Azure Machine Learning | Microsoft Docs, 2021*) permet de créer et d'exécuter des pipelines en utilisant une interface visuelle glisser-déposer pour connecter des modules qui définissent les étapes et le flux de données du pipeline. Il est également possible d'utiliser des « blocs » de code qui permettent le développement des fonctionnalités spécifiques à la problématique étudiée à travers le langage Python ou R. Il est possible aussi d'installer des packages qui n'existent pas par défaut sur Azure ML, ou importer des blocs de code pour les utiliser directement dans le pipeline. La figure 47 ci-dessous illustre un exemple d'implémentation d'un pipeline dans le concepteur d'Azure ML.

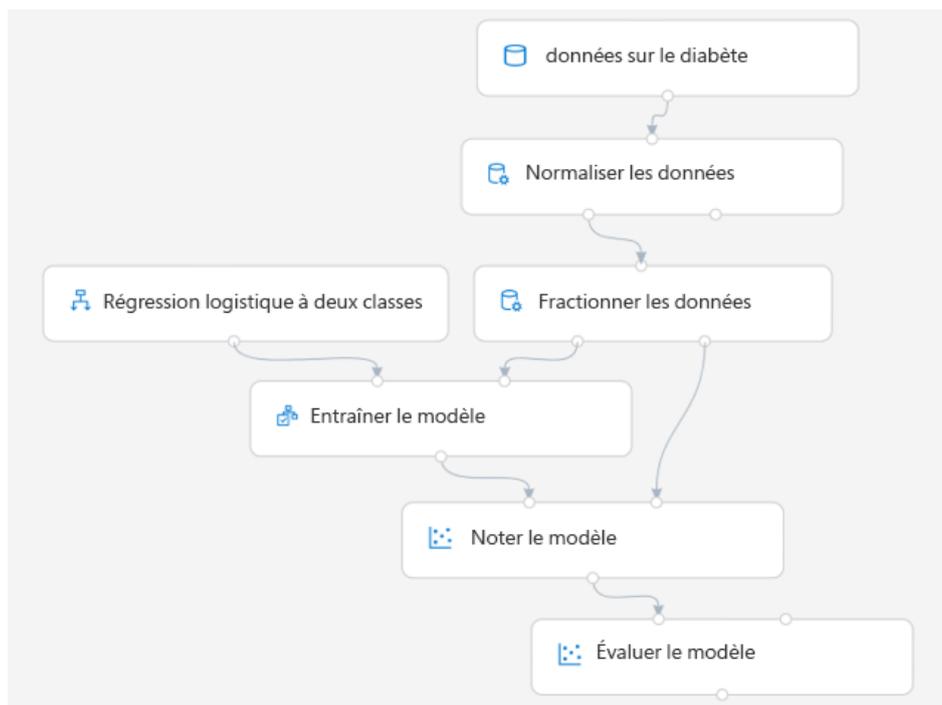


Figure 47 : Exemple d'un pipeline sur Azure ML

Dans le cadre de la mise en application des travaux réalisés, nous avons employé cet outil pour la création d'un pipeline personnalisé pour le traitement des données de l'entreprise elm.leblanc, l'entraînement des modèles de prédiction et la préparation des prévisions des ventes. Dans le paragraphe suivant, nous présentons les différentes étapes du pipeline conçu dans le concepteur d'Azure ML.

IV. Description du pipeline retenu pour le portage de la solution

Dans le cadre de l'implémentation de nos travaux dans l'environnement Azure ML et dans le but de faciliter l'usage des modèles développés pour la prévision des ventes aux services concernés, nous avons utilisé le concepteur d'Azure ML pour mettre en place tout le processus depuis le nettoyage et la préparation des données, jusqu'à la visualisation des résultats de prédiction à l'utilisateur en passant par l'entraînement des modèles.

Ce processus peut être partagé en deux phases : la première concerne la préparation et le nettoyage des données, tandis que la deuxième vise à la construction et l'entraînement des modèles, et le calcul des prévisions selon la méthodologie proposée dans ces travaux.

IV.1. Etapes du pipeline

Le pipeline ou encore le processus de stockage et de traitement des données, d'entraînement et de prédiction, est constitué de deux phases faisant l'objet de cet outil. Pour la phase de

préparation des données, la première étape consiste en la lecture et le nettoyage des données brutes historiques des ventes à partir du Blob Storage, avec une mise à jour des données téléchargées du mois écoulé. Un travail de suppression des aléas, des valeurs invalides et de sélection des données pertinentes est effectué durant cette étape. La deuxième étape concerne la sélection des produits éligibles selon les conditions discutées au chapitre 3 (paragraphe II.2). Ensuite, il convient d'ajouter une étape de vérification s'il y a un changement au niveau des références de produits pour ajouter de nouvelles références ou supprimer des anciennes. Cette première phase vise à construire une base de données propre et complète des historiques des ventes de l'ensemble des produits qu'on souhaite prédire.

La deuxième phase de ce pipeline vise à préparer les prévisions et les enregistrer dans le Blob Storage. D'abord, l'ensemble des fonctions développées en machine locale auparavant sont importées depuis le Blob. Ces fonctions représentent une implémentation de la méthodologie présentée dans le chapitre 3. En utilisant les données préparées à l'issue de la première phase, le processus d'entraînement et de validation des modèles est initié dans Azure ML. La dernière étape consiste à calculer les prévisions des ventes de chaque produit selon l'horizon souhaité par l'utilisateur. L'horizon de choix constitue une entrée par l'utilisateur et est pris en considération dans l'étape de calcul des prédictions. Les résultats finaux sont exportés vers le Blob Storage pour être téléchargés et exploités par la suite.

Dans le paragraphe suivant, nous décrivons l'implémentation de ces étapes sur le concepteur d'Azure ML et les liaisons entre les différentes étapes du pipeline.

IV.2. Implémentation

Dans le concepteur d'Azure ML, les différentes étapes décrites dans le paragraphe précédent sont représentées par un ou plusieurs blocs de code. Chaque bloc prend en entrées des bases de données, des blocs de code ou encore les résultats d'autres blocs. La figure ci-dessous présente l'implémentation du pipeline de l'outil de prévision des ventes. Nous décrivons dans cette partie les différents blocs du pipeline ainsi que les liaisons entre ces derniers et les données employées dans cet outil. La figure 48 ci-dessous présente le pipeline implémenté sur Azure ML pour les prévisions des ventes.

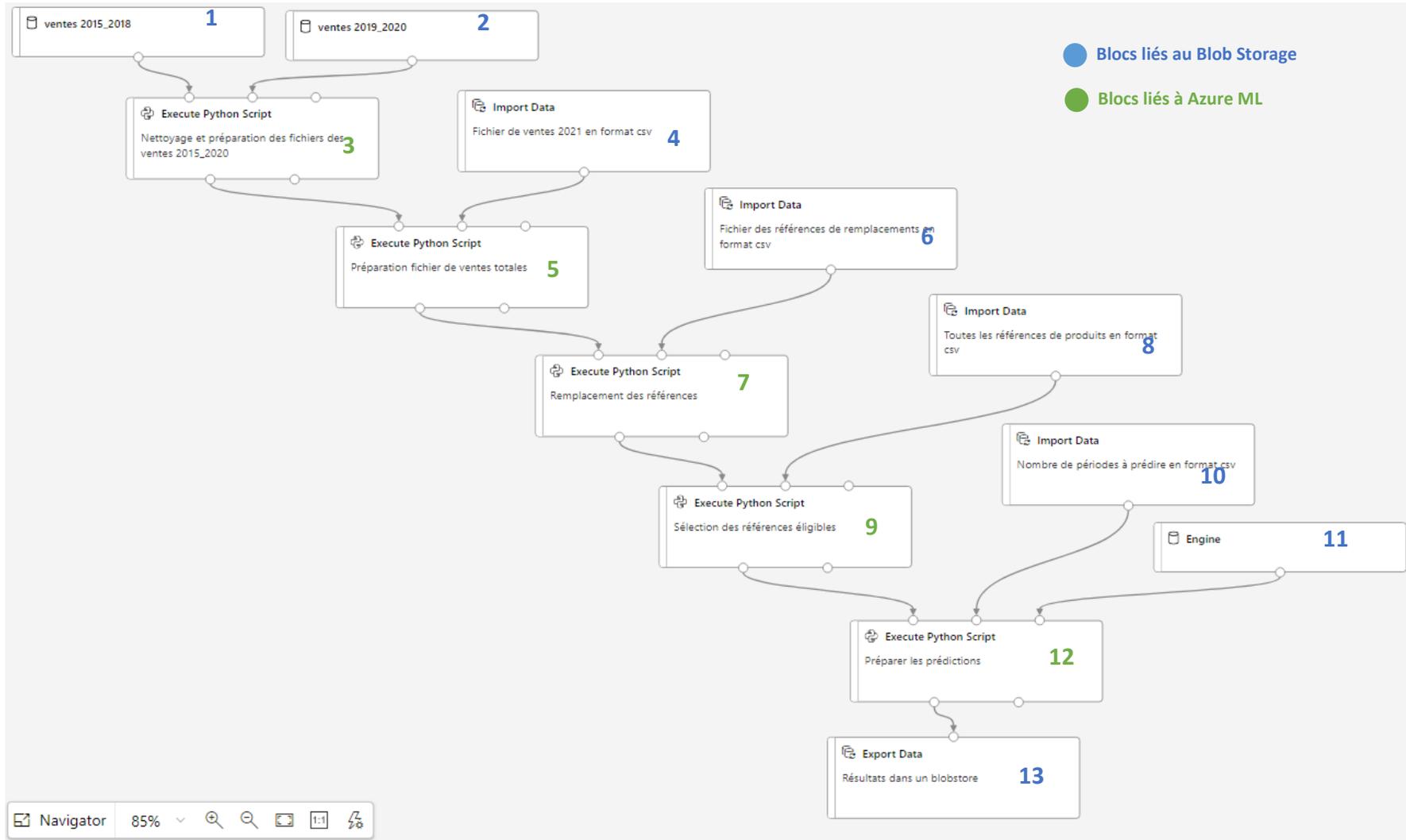


Figure 48 : Capture du pipeline implémenté sur Azure ML pour la prédiction des ventes

En partant du haut à gauche, les deux premiers blocs correspondent aux données historiques des ventes entre 2015 et 2018 et entre 2019 et 2020. Cette répartition sur deux fichiers est due au fait qu'on ne dispose pas des mêmes colonnes de données suite à un changement de format en 2019, et donc ces deux fichiers sont traités et nettoyés puis concaténés ensemble pour former la première sortie du bloc 3. Le bloc 4 représente les données de ventes de l'année 2021 mises à jour à chaque mois écoulé, et importées depuis le Blob Storage. Il constitue avec le résultat du bloc 3, les entrées pour le bloc 5 nécessaire pour consolider les données et les remettre au même format pour toutes les années. Le bloc 6 représente la liste des nouveaux produits qui remplacent les anciens, et donc ces données sont croisées avec les historiques des ventes dans le bloc 7 pour créer une base de données à jour (sortie du bloc 7). A cette étape, nous avons construit une base de données propre de l'ensemble des produits vendus dans l'entreprise avec un historique complet depuis 2015. Le bloc 8 correspond à la liste des références de produits que l'utilisateur souhaite prédire et qui est importée du Blob Storage avec la possibilité de la mettre à jour régulièrement. Ce bloc 8 constitue avec la sortie du bloc 7 les entrées du bloc 9 qui sera consacré à la sélection des produits éligibles à la prédiction en respectant les conditions discutées lors de l'explication de la méthodologie au chapitre 3 (paragraphe I.2). A la sortie du bloc 9, nos données sont prêtes pour être exploitées et utilisées par les structures de modèles dont l'estimation et la mise à jour se font sur Azure ML. Le bloc 11 correspond aux fonctions développées en machine locale avec quelques adaptations nécessaires, et qui sont téléchargés sous format zip dans le Blob Storage et ensuite importés au pipeline. Dans le bloc 12, nous faisons appel aux différentes fonctions développés (bloc 11), pour l'estimation et la mise à jour des structures des modèles en utilisant les données à la sortie du bloc 9. Après leur entraînement, les modèles sont mis en compétition selon notre méthodologie développée, et ensuite, le calcul des prédictions de chaque produit est réalisé selon le modèle choisi et selon l'horizon souhaité par l'utilisateur introduit par l'intermédiaire du bloc 10. Les sorties du bloc 12 correspondent alors à des prévisions des ventes pour les produits demandés, exportées ensuite au Blob Storage sous format csv pour être exploitées et téléchargées.

Ces blocs de code sont exécutés grâce à des ressources de calcul qui sont choisies selon la puissance souhaitée par l'utilisateur selon son abonnement. L'offre d'Azure ML permet de mettre en veille automatiquement les ressources lorsqu'elles ne sont pas utilisées et donc ne

pas engendrer des coûts supplémentaires à l'utilisateur. Grâce à une autre fonctionnalité intéressante d'Azure ML, les blocs délivrant les mêmes résultats à chaque lancement comme le bloc 3, ne seront pas dans l'obligation de réexécuter le code contenu dans ces blocs et vont exploiter directement les résultats sauvegardés en cache depuis le dernier lancement. Cet aspect accélère l'exécution de l'ensemble du pipeline et réduit la consommation des ressources de calcul et donc réduit les coûts d'exploitation.

Dans le paragraphe suivant, nous présentons les résultats d'implémentation des travaux présentés auparavant sur Azure ML et son usage au sein de l'entreprise.

V. Mise en œuvre et Résultats

Après le déploiement de l'approche proposée dans le chapitre 3 sur Azure ML, cet outil est désormais employé dans l'entreprise pour préparer les prévisions des ventes. L'outil est fonctionnel depuis le début de l'année de 2021 et est utilisé régulièrement chaque mois. L'outil est intégré dans un processus S&OP pour préparer les prévisions de ventes des produits avant la réunion de la planification de la demande. Le responsable des prévisions se sert de cet outil chaque début de mois pour lancer les calculs la nuit et retrouver les résultats le lendemain. La prise en main de l'outil s'est réalisée à travers une démonstration de l'usage de l'outil pour les personnes concernées des différents départements. Ensuite, un guide d'utilisation a été réalisé pour servir d'assistance en cas de besoin. Les premières utilisations de l'outil étaient accompagnées pour assurer la bonne maîtrise des différentes étapes de la mise en marche, et avoir les retours d'expériences des utilisateurs. Ensuite, les utilisations ont été réalisées en toute autonomie renvoyant ainsi une satisfaction des utilisateurs concernant la facilité d'usage et l'ergonomie de la solution.

Nous présentons dans le tableau 21 ci-dessous les résultats de prévisions de l'outil par rapport aux ventes réelles de quelques produits sur le premier semestre 2021.

Produit	Janvier-21		Février-21		Mars-21		Avril-21		Mai-21		Juin-21	
	Ventes	Prévisions	Ventes	Prévisions	Ventes	Prévisions	Ventes	Prévisions	Ventes	Prévisions	Ventes	Prévisions
X2	64	61	47	51	75	54	100	32	81	62	74	65
X3	32	18	35	24	60	28	10	44	45	28	50	32
X4	1103	935	887	752	864	810	966	800	700	928	1099	863
X5	593	538	503	491	734	411	588	756	712	585	814	677
X6	9	29	53	21	43	34	10	37	339	26	208	148

X7	222	224	213	304	202	259	221	222	199	202	169	189
X8	1034	1071	671	1050	330	897	950	358	1052	589	1036	887
X9	149	351	514	146	328	341	505	361	217	327	216	304
X10	477	506	245	450	306	281	103	374	238	170	179	222
X11	573	1097	933	682	1110	682	788	632	815	510	271	823
X12	1156	1133	1451	1002	1077	1053	827	934	524	599	877	787
X13	8	17	12	27	18	42	0	0	2	15	10	34
X14	100	189	119	78	223	178	113	159	132	84	228	98
X15	24	249	12	132	7	68	21	75	10	72	28	65
X16	0	0	1	2	2	1	0	0	2	1	2	1
X17	195	111	49	138	79	5	59	79	14	59	27	14
X18	77	162	119	118	114	107	97	127	88	135	46	78
X46	307	321	311	274	119	298	221	196	225	224	438	235
X19	27	20	28	20	10	17	28	15	14	16	16	16
X20	79	74	79	74	57	73	57	70	80	68	83	68
X21	66	20	32	18	23	38	20	16	24	30	47	34
X22	47	26	29	29	16	32	21	22	17	20	30	28
X23	43	34	38	35	33	35	20	32	53	28	40	28
X24	52	26	24	37	33	31	19	31	27	26	48	26
X30	48	17	45	40	41	43	22	40	30	27	69	22
X37	73	84	79	83	41	82	74	80	74	79	56	79
X38	8	14	8	12	25	5	1	20	17	1	31	17
X39	24	15	2	22	22	14	2	10	2	6	2	4
X40	7	13	90	9	11	34	6	23	9	31	5	16
X41	2	4	3	3	8	1	0	0	4	2	8	3
X42	2	2	2	2	5	2	3	2	3	4	4	2
X43	203	188	158	119	110	154	80	129	180	115	229	119
X44	110	90	119	68	130	97	101	99	130	113	195	116
X45	966	1139	907	724	1372	930	878	1119	1872	1015	1602	1535

Tableau 21 : Ventes et prévisions réalisées sur le S1 2021

Ces prévisions sont ensuite discutées durant la première réunion du processus S&OP qui est la planification de la demande pour les ajuster en intégrant les avis d'experts et les informations externes qui peuvent influencer ces prévisions.

Nous présentons dans la figure ci-dessous les erreurs absolues des prévisions fournies par l'outil par rapport aux ventes réalisées pour chaque mois du premier semestre de 2021.

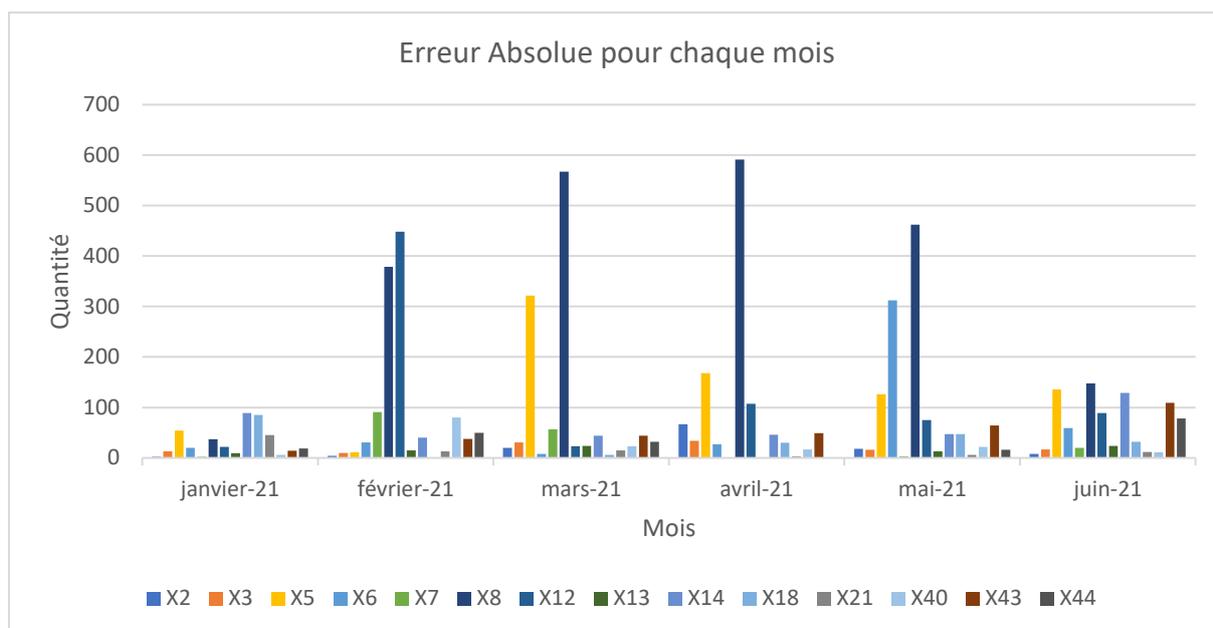


Figure 49 : Exemples pour quelques produits des erreurs absolues enregistrées pour chaque mois

On remarque que notre approche était assez proche de la réalité des ventes sur la plupart des produits, néanmoins pour certains produits comme le X8, notre méthode a fourni des erreurs élevées. Ceci peut être justifié par le fait que ce produit a subi une rupture de stock sur certains composants pour le mois de février et mars, et ensuite une augmentation brusque sur les mois d'après pour compenser les ventes non honorées sur les mois précédents. Ces effets sont pris en considération dans la réunion de préparation du plan de la demande lors du processus S&OP dans l'entreprise, ce qui permet de réduire l'impact de ces causes sur la précision des prévisions. On remarque également que pour le produit X12, nous avons une erreur importante sur le mois de février qui n'apparaît pas sur les autres mois. Ceci peut être expliqué par la validation d'un chantier et donc une demande importante brusque notée sur ce produit pour ce mois. Pour des produits comme le X5 le X6, les erreurs constatées restent modérées et non élevées par rapport aux autres produits puisqu'il s'agit plutôt des produits réguliers ne subissant pas de perturbation particulière. Dans le tableau 22 suivant, nous présentons ces résultats pour un plus grand nombre de produits.

Référence	Janvier-21	Février-21	Mars-21	Avril-21	Mai-21	Juin-21
X2	2	4	20	67	18	8
X3	13	10	31	34	16	17
X4	167	134	53	165	228	235
X5	54	11	322	168	126	136
X6	20	31	8	27	312	59
X7	2	91	57	1	3	20

X8	37	379	567	591	462	148
X9	202	367	13	143	110	88
X10	29	205	24	271	67	43
X11	524	250	427	155	304	552
X12	22	448	23	107	75	89
X13	9	15	24	0	13	24
X14	89	40	44	46	47	129
X15	225	120	61	54	62	37
X16	0	1	0	0	0	0
X17	83	89	73	20	45	13
X18	85	0	6	30	47	32
X46	14	36	179	24	0	202
X19	6	7	7	12	2	0
X20	4	4	16	13	11	14
X21	45	13	15	3	6	12
X22	20	0	16	1	3	1
X23	8	2	2	12	24	11
X24	25	13	1	12	0	21
X30	31	4	2	18	2	46
X37	11	4	41	6	5	23
X38	6	4	19	19	15	13
X39	8	20	7	8	4	2
X40	6	80	23	17	22	11
X41	2	0	6	0	1	4
X42	0	0	2	0	1	1
X43	14	38	44	49	64	109
X44	19	50	32	1	16	78
X45	173	182	441	241	856	66

Tableau 22 : Erreurs absolues enregistrées sur le premier semestre de 2021

Comme nous pouvons le constater à partir du tableau 22, pour plusieurs produits notre approche a fourni des bons résultats en termes de précision. Nous rappelons que l'ensemble des prédictions sont ajustées lors de la réunion de préparation du plan de la demande pour prendre en considération les facteurs externes liés à la nature de l'activité de l'entreprise comme les promotions, les campagnes publicitaires ou encore les chantiers.

VI. Conclusion

Ce dernier chapitre a été consacré pour la phase d'industrialisation et la mise en place de l'outil développé dans un environnement d'application réel. Pour assurer cet objectif, nous avons employé la plateforme Cloud de Microsoft Azure Machine Learning qui permet le développement et le déploiement rapide des applications en ML. Azure Blob Storage était

utilisé pour télécharger et sauvegarder les bases de données. La mise en place de l'outil passe à travers l'implémentation d'un pipeline permettant la préparation des données ainsi que l'entraînement des modèles et le calcul des prévisions.

Ce travail est aujourd'hui mis en œuvre au sein du département des ventes dans l'entreprise et est exploité pour préparer les prévisions des ventes dans un cycle mensuel du processus S&OP. Cet outil constitue le livrable demandé par l'entreprise au début de ces travaux, et consiste en une solution clé en main pour le service des ventes dans un environnement industriel. L'implémentation sur un service cloud comme Azure ML permet une pérennité de la solution évitant ainsi les inconvénients liés à l'infrastructure locale. Cette solution offre également une automatisation de calcul des prévisions remplaçant ainsi la méthode manuelle utilisée auparavant chez l'entreprise. Cet aspect facilite l'usage de la solution et est le sujet d'appréciation des collaborateurs qui utilisent cet outil. L'intégration de l'application dans les processus de l'entreprise s'est effectuée progressivement, en formant d'abord les collaborateurs concernés à son utilisation, ensuite, un usage avec accompagnement pour avoir les retours des utilisateurs et identifier les difficultés rencontrées, et puis un usage autonome complété par un guide d'utilisation en cas de besoin. Les retours des collaborateurs sont plutôt positifs autant pour l'aisance d'usage que pour les résultats de prédictions fournis par l'outil. Cette appréciation est justifiée par l'utilisation de la solution jusqu'au aujourd'hui en toute autonomie au sein de l'entreprise, tout en permettant des améliorations futures, tant sur la mise à jour des modèles utilisés, l'introduction de nouvelles fonctionnalités ou l'évolution du catalogue produits elm.leblanc.

Conclusion Générale

L'objectif global pour chaque entreprise est de maximiser ses gains et réduire ses coûts en améliorant ses processus et sa Supply Chain. Dans ce manuscrit, nous avons traité une problématique visant à atteindre cet objectif à travers la mise en place d'un outil d'aide à la décision pour l'amélioration du processus de planification de la demande. Les travaux menés dans cette thèse ont deux objectifs : d'un point de vue recherche, proposer une nouvelle méthodologie scientifique pour la prévision des ventes des produits de différentes natures, et d'un point de vue industriel, mettre en place un outil pour aider les entreprises à mieux planifier leur plan de demande. L'entreprise elm.leblanc était le cadre d'application industrielle de ces travaux en utilisant ses données pour réaliser les différentes expérimentations.

Dans ce projet, nous avons essayé de mettre en application les travaux de recherche réalisés dans un contexte industriel. L'évaluation de l'état existant de l'entreprise nous a révélé plusieurs points d'amélioration. En se basant sur les étapes du processus S&OP et en évaluant le besoin de l'entreprise, nous nous sommes focalisés sur l'amélioration du processus de prévision des ventes. Cette problématique représente la première étape de planification de la Supply Chain et son amélioration affecte les étapes suivantes telles que la planification de la production et le réapprovisionnement des matières premières, d'où l'intérêt accordé par l'entreprise pour l'améliorer. Nous nous sommes focalisés alors sur l'étude des techniques existantes dans la littérature, dans le but de proposer une méthodologie qui à la fois répond à ce besoin, et contribue au progrès scientifique de ce domaine d'étude.

Nos recherches se sont concentrées sur les méthodes utilisées dans différents domaines d'applications pour faire les prévisions des ventes. Cette étude s'avère intéressante pour comprendre comment les spécificités et les particularités de chaque domaine influent le choix des méthodes de prévisions. Elle révèle que plusieurs facteurs entrent en jeu et doivent être pris en considération pour chaque application. L'approche utilisée doit correspondre aux données existantes et à la nature de l'activité de chaque entreprise.

Dans cette étude bibliographique, nous nous sommes intéressés aussi à la description mathématique des approches et modèles de la littérature. Nous avons étudié les principales

techniques utilisées pour les prévisions des ventes et nous avons identifié quatre catégories principales. Nous trouvons tout d'abord des techniques traitant cette problématique par une approche de séries temporelles uniquement, employant ainsi des modèles dédiés pour leur modélisation. D'autres approches considèrent davantage de dépendances statistiques des variables et des facteurs externes et donc emploient des techniques d'apprentissage automatique. Une autre catégorie emploie des techniques plus avancées d'apprentissage profond, notamment en utilisant les réseaux de neurones. Enfin, des méthodologies présentent des méthodes hybrides et de combinaison, utilisant plusieurs types de modèles en même temps. Cette investigation nous a permis de construire une connaissance sur les méthodes et les approches adéquates pour chaque cas d'étude, et nous a aidé à développer notre méthodologie appropriée à notre problématique.

Etant donnée la variété et la diversité du catalogue des produits de l'entreprise elm.leblanc, nous avons proposé dans ces travaux une première méthodologie pour fournir des meilleurs prévisions. Notre approche repose sur une structure multi-modèle qui permet d'identifier celui le plus adéquat pour chaque produit du catalogue et l'utiliser pour calculer les prévisions à différents horizons. Nous avons employé plusieurs types de modèles pour couvrir l'ensemble des produits de l'entreprise. Cette méthodologie a été évaluée sur des données réelles de l'entreprise pour valider son efficacité. Nous nous sommes intéressés également à la répartition des produits par type de modèle utilisé dans le but de mieux comprendre le comportement des ventes de ces produits et justifier le choix des modèles employés. Comme nous avons étudié plusieurs domaines d'application des techniques de prévisions des ventes, nous avons généralisé notre approche pour qu'elle soit adaptable à d'autres domaines. Ainsi, elle peut être appliquée dans d'autres cas d'études. Nous avons également testé notre approche sur des données réelles irrégulières, notamment pour la prédiction des ventes pour la période post-COVID, où nous avons eu des résultats satisfaisants malgré l'impact de la situation sanitaire sur les entreprises.

Comme pour toute méthodologie, il n'existe pas d'approche parfaite et chacune peut avoir un biais. Notre méthode ne fait pas l'exception, et pour faire face au biais identifié, nous avons proposé une extension de cette approche. L'aspect compétitif de la première méthodologie sera remplacé par un aspect collaboratif. Nous avons donc proposé de combiner les résultats des différents modèles utilisés auparavant dans le but d'améliorer les prévisions finales. Cette

combinaison est réalisée à travers un réseau de neurones qui prendra la décision après avoir affecté des poids à chaque modèle. La nécessité d'une quantité importante de données pour l'entraînement des modèles et l'usage du réseau, rend cette approche applicable d'avantage aux produits ayant suffisamment de données d'historique de ventes. Cette deuxième méthodologie a été évaluée également sur les données d'elm.leblanc, montrant ainsi des résultats satisfaisants.

Comme le succès d'un projet industriel est conditionné par une mise en place réussite ainsi qu'une facilité d'usage, les travaux menés durant la thèse étaient l'objet d'une implémentation et mise en œuvre dans l'entreprise elm.leblanc. Nous avons réussi à implémenter cette approche de prédiction sur la plateforme Cloud de Microsoft Azure ML. Cette mise en œuvre permet la création d'un outil d'aide à la décision pour la prévision des ventes du catalogue des produits de l'entreprise, et qui est utilisé aujourd'hui avec aisance par le département des ventes. D'après ce constat, il nous semble que ce projet est couronné de succès et que l'objectif fixé au départ a été atteint.

Comme chaque méthodologie développée dans la littérature, il existe toujours des limites et des conditions d'utilisation. Pour notre première approche proposée, nous avons spécifié quelques critères de choix des produits qui peuvent être prédits. Le choix des modèles utilisés peut être reconsidéré avec l'apparition des nouveaux modèles plus précis, ou avec le cumul d'une grande quantité de données, et donc l'emploi d'autres approches d'apprentissage profond. Avec l'intervention des facteurs externes influant les ventes comme le changement des réglementations ou encore la rupture d'approvisionnement, notre approche actuelle ne pourra pas prévoir la baisse ou l'augmentation des ventes puisqu'elle ne prend pas en compte ces facteurs. Cette limite est levée aujourd'hui par le processus S&OP où les prévisions sont ajustées dans l'étape de planification de la demande pour prendre en considération ces aspects. Néanmoins, elle peut être éliminée aussi avec l'usage des modèles qui permettent de prendre en considération ces facteurs.

Grâce aux différentes natures des modèles utilisés dans la méthodologie, cette dernière peut s'adapter au fur et à mesure aux évolutions des comportements de ventes des produits. Le choix du modèle utilisé pour chaque produit n'est pas fixe et peut être mis à jour. Néanmoins, notre approche présente une autre limite caractérisée par les prévisions des nouveaux produits ou les produits récemment lancés. En effet, avec l'absence des données historiques,

notre méthodologie ne peut être appliquée puisque les modèles ne peuvent pas être entraînés. Une dernière limite constatée sur notre approche concerne le choix du meilleur modèle. En effet, pour certains produits les modèles peuvent présenter des performances très proches même en étant de natures différentes, et donc ceci peut impacter les prévisions surtout au long et moyen terme.

Cette dernière limite constitue l'origine de la deuxième approche proposée, visant à combiner les prédictions des différents modèles, et donc réduire le risque des grandes erreurs. Bien que cette méthodologie puisse être une solution à cet aspect, elle ne peut pas être appliquée à tous les produits vu la quantité des données importantes nécessaires pour la mettre en place. Le temps d'exécution de cette deuxième approche peut être constaté comme limite aussi, étant donnée la durée d'entraînement nécessaire pour les modèles de prédictions pour tous les produits et le réseau de neurones utilisé, notamment pour le modèle multi-horizons.

La mise en place de l'outil de prévisions des ventes sur la plateforme d'Azure ML permet une pérennité d'usage même avec l'évolution des environnements de développement. Le critère automatique ou semi-automatique de cet outil donne une aisance d'usage sans besoin de maintenance régulière. L'exécution de l'outil est faite en temps caché (la nuit) ce qui évite des temps d'attente pour l'obtention des résultats.

L'outil proposé dans ces travaux peut représenter l'objet de plusieurs améliorations futures. L'usage des nouveaux modèles ayant la capacité de prendre en considération des facteurs externes influant les ventes est une piste intéressante de recherche, permettant ainsi d'atteindre une certaine autonomie sans avoir recours au processus S&OP pour l'ajustement des prévisions. Nous pouvons intégrer également des méthodes d'intelligence artificielle, comme l'apprentissage par transfert, qui permettent de faire les prévisions pour les nouveaux produits en se basant sur les produits similaires du catalogue. Cette piste permettra de dépasser la limite constatée auparavant.

Etant donnée la nature des modèles utilisés, cet outil peut être aussi employé dans d'autres applications indépendamment des prévisions des ventes, puisqu'il peut gérer les données de plusieurs séries temporelles. Ceci peut être appliqué pour le calcul des prévisions d'achat des produits non fabriqués sur site, ou encore pour la prédiction des appels des clients pour le service après-vente.

D'un point de vue industriel, l'automatisation des flux des données d'entrée à travers les systèmes d'informations de l'entreprise permettra d'automatiser le lancement de l'outil et donc la préparation des prévisions sans avoir besoin d'intervention humaine. Une fois couplé avec un outil de planification de la production, l'ensemble sera un système efficace capable de préparer les plans de la demande, de la production, de réapprovisionnement et de gestion des stocks d'une façon automatique. Cet outil peut aussi faire l'objet d'une application web dédiée à la prévision pour faciliter encore plus l'envoi et la récupération des données sans avoir besoin de passer par la plateforme d'Azure ML.

Références

- Abdellatief, M., Shaaban, E. M., & Abu-Raya, K. A. (2019). Egyptian Case Study-Sales forecasting model for automotive section. *2019 International Conference on Smart Applications, Communications and Networking, SmartNets 2019*. <https://doi.org/10.1109/SMARTNETS48225.2019.9069751>
- Adithya Ganesan, V., Divi, S., Moudhgalya, N. B., Sriharsha, U., & Vijayaraghavan, V. (2019). Forecasting Food Sales in a Multiplex Using Dynamic Artificial Neural Networks. *Advances in Intelligent Systems and Computing, 944*, 69–80. https://doi.org/10.1007/978-3-030-17798-0_8
- Aghabozorgi, S., Seyed Shirkhorshidi, A., & Ying Wah, T. (2015). Time-series clustering – A decade review. *Information Systems, 53*, 16–38. <https://doi.org/10.1016/j.is.2015.04.007>
- Ahmad, T., Chen, H., Guo, Y., & Wang, J. (2018). A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. In *Energy and Buildings* (Vol. 165, pp. 301–320). Elsevier Ltd. <https://doi.org/10.1016/j.enbuild.2018.01.017>
- Al-Musaylh, M. S., Deo, R. C., Adamowski, J. F., & Li, Y. (2018). Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. *Advanced Engineering Informatics, 35*, 1–16. <https://doi.org/10.1016/j.aei.2017.11.002>
- Armando, E., & Craparotta, G. (2017). A Meta-Model for Fashion Retail Category Sales Forecasting. *Lecture Notes in Electrical Engineering, 525*, 79–93. https://doi.org/10.1007/978-3-319-98038-6_7
- Arunraj, N. S., & Ahrens, D. (2015). A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. *International Journal of Production Economics, 170*, 321–335. <https://doi.org/10.1016/j.ijpe.2015.09.039>
- Aye, G. C., Balcilar, M., Gupta, R., & Majumdar, A. (2015). Forecasting aggregate retail sales: The case of South Africa. *International Journal of Production Economics, 160*, 66–79. <https://doi.org/10.1016/j.ijpe.2014.09.033>
- Baecke, P., De Baets, S., & Vanderheyden, K. (2017). Investigating the added value of integrating human judgement into statistical demand forecasting systems. *International Journal of Production Economics, 191*, 85–96. <https://doi.org/10.1016/j.ijpe.2017.05.016>
- Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q., & Seaman, B. (2019). Sales demand forecast in E-commerce using a long short-term memory neural network methodology. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11955 LNCS*, 462–474. https://doi.org/10.1007/978-3-030-36718-3_39
- Bee Dagum, E., & Bianconcini, S. (2016). *Seasonal Adjustment Methods and Real Time Trend-Cycle Estimation*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-31822-6>

- Behera, G., & Nain, N. (Eds.). (2019). A Comparative Study of Big Mart Sales Prediction. In *4th International Conference on Computer Vision and Image Processing*. https://www.researchgate.net/publication/336530068_A_Comparative_Study_of_Big_Mart_Sales_Prediction
- Beheshti-Kashi, S., Karimi, H. R., Thoben, K.-D., Lütjen, M., & Teucke, M. (2015). A survey on retail sales forecasting and prediction in fashion markets. *Systems Science & Control Engineering*, 3(1), 154–161. <https://doi.org/10.1080/21642583.2014.999389>
- Ben Taieb, S., Bontempi, G., Atiya, A. F., & Sorjamaa, A. (2012). A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Systems with Applications*, 39(8), 7067–7083. <https://doi.org/10.1016/j.eswa.2012.01.039>
- Bourdeau, M., Zhai, X. qiang, Nefzaoui, E., Guo, X., & Chatellier, P. (2019). Modeling and forecasting building energy consumption: A review of data-driven techniques. In *Sustainable Cities and Society* (Vol. 48, p. 101533). Elsevier Ltd. <https://doi.org/10.1016/j.scs.2019.101533>
- Cadavid, J. P. U., Lamouri, S., & Grabot, B. (2018). *Trends in Machine Learning Applied to Demand & Sales Forecasting: A Review*.
- Cao, G., & Wu, L. (2016). Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting. *Energy*, 115, 734–745. <https://doi.org/10.1016/J.ENERGY.2016.09.065>
- CATAL, C., ECE, K., ARSLAN, B., & AKBULUT, A. (2019). Benchmarking of Regression Algorithms and Time Series Analysis Techniques for Sales Forecasting. *Balkan Journal of Electrical and Computer Engineering*, 7(1), 20–26. <https://doi.org/10.17694/BAJECE.494920>
- Cecere, L., Barrett, J., & Mooraj, H. (2009). Sales and Operations Planning : Transformation From Tradition. *AMR Research*, May.
- Chen, C. Y., Lee, W. I., Kuo, H. M., Chen, C. W., & Chen, K. H. (2010). The study of a forecasting sales model for fresh food. *Expert Systems with Applications*, 37(12), 7696–7702. <https://doi.org/10.1016/j.eswa.2010.04.072>
- Chen, F. L., & Ou, T. Y. (2009). Gray relation analysis and multilayer functional link network sales forecasting model for perishable food in convenience store. *Expert Systems with Applications*, 36(3 PART 2), 7054–7063. <https://doi.org/10.1016/j.eswa.2008.08.034>
- Chen, R., & Liu, J. (2017). A Novel Framework for Online Sales Burst Prediction. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10536 LNAI, 3–14. https://doi.org/10.1007/978-3-319-71273-4_1
- Cheriyana, S., Ibrahim, S., Mohanan, S., & Treesa, S. (2019). Intelligent Sales Prediction Using Machine Learning Techniques. *Proceedings - 2018 International Conference on Computing, Electronics and Communications Engineering, ICCECE 2018*, 53–58. <https://doi.org/10.1109/icCECOME.2018.8659115>
- Ching-Chin, C., Ka Ieng, A. I., Ling-Ling, W., & Ling-Chieh, K. (2010). Designing a decision-support system for new product sales forecasting. *Expert Systems with Applications*,

- 37(2), 1654–1665. <https://doi.org/10.1016/j.eswa.2009.06.087>
- Choi, T. M., Hui, C. L., Ng, S. F., & Yu, Y. (2012). Color trend forecasting of fashionable products with very few historical data. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 42(6), 1003–1010. <https://doi.org/10.1109/TSMCC.2011.2176725>
- Christopher Martin. (2011). *LOGISTICS & SUPPLY CHAIN MANAGEMENT* (PEARSON EDUCATION LIMITED (Ed.); FOURTH EDI). https://www.icesi.edu.co/blogs/supplychain0714/files/2014/07/Martin_Christopher_Logistics_and_Supply_Chain_Management_4th_Edition____2011-1.pdf
- Craparotta, G., Thomasse, S., & Biolatti, A. (2019). A Siamese Neural Network Application for Sales Forecasting of New Fashion Products Using Heterogeneous Data. *International Journal of Computational Intelligence Systems*, 12(2), 1537–1546. <https://doi.org/10.2991/IJCIS.D.191122.002>
- Da Veiga, C. P., Da Veiga, C. R. P., Puchalski, W., Coelho, L. dos S., & Tortato, U. (2016). Demand forecasting based on natural computing approaches applied to the foodstuff retail segment. *Journal of Retailing and Consumer Services*, 31, 174–181. <https://doi.org/10.1016/j.jretconser.2016.03.008>
- Dai, W., Chuang, Y. Y., & Lu, C. J. (2015). A Clustering-based Sales Forecasting Scheme Using Support Vector Regression for Computer Server. *Procedia Manufacturing*, 2, 82–86. <https://doi.org/10.1016/j.promfg.2015.07.014>
- De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, 22(3), 443–473. <https://doi.org/10.1016/J.IJFORECAST.2006.01.001>
- Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. In *Renewable and Sustainable Energy Reviews* (Vol. 74, pp. 902–924). Elsevier Ltd. <https://doi.org/10.1016/j.rser.2017.02.085>
- Doganis, P., Alexandridis, A., Patrinos, P., & Sarimveis, H. (2006). Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. *Journal of Food Engineering*, 75(2), 196–204. <https://doi.org/10.1016/J.JFOODENG.2005.03.056>
- Dwivedi, A., Niranjana, M., Computer, K. S.-I. J. of, & 2013, U. (2013). A business intelligence technique for forecasting the automobile sales using Adaptive Intelligent Systems (ANFIS and ANN). *Academia.Edu*. <http://www.academia.edu/download/46663697/pxc3889383.pdf>
- Elamin, N., & Fukushige, M. (2018). Modeling and forecasting hourly electricity demand by SARIMAX with interactions. *Energy*, 165, 257–268. <https://doi.org/10.1016/j.energy.2018.09.157>
- Elmasdotter, A., & Nyströmer, C. (2018). A comparative study between LSTM and ARIMA for sales forecasting in retail. *DEGREE PROJECT TECHNOLOGY*. <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-229747>
- Fan, S., & Hyndman, R. J. (2012). Short-term load forecasting based on a semi-parametric

- additive model. *IEEE Transactions on Power Systems*, 27(1), 134–141. <https://doi.org/10.1109/TPWRS.2011.2162082>
- Fan, Z. P., Che, Y. J., & Chen, Z. Y. (2017). Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis. *Journal of Business Research*, 74, 90–100. <https://doi.org/10.1016/j.jbusres.2017.01.010>
- Franses, P. H., & Legerstee, R. (2013). Do statistical forecasting models for SKU-level data benefit from including past expert knowledge? *International Journal of Forecasting*, 29(1), 80–87. <https://doi.org/10.1016/j.ijforecast.2012.05.008>
- Gao, J., Xie, Y., Cui, X., Yu, H., & Gu, F. (2018). Chinese automobile sales forecasting using economic indicators and typical domestic brand automobile sales data: A method based on econometric model. *Advances in Mechanical Engineering*, 10(2), 168781401774932. <https://doi.org/10.1177/1687814017749325>
- Grimson, J. A., & Pyke, D. F. (2007). Sales and operations planning : an exploratory study and framework. *The International Journal of Logistics Management*. <https://doi.org/10.1108/09574090710835093>
- Gurnani, M., Korke, Y., Shah, P., Udmale, S., Sambhe, V., & Bhirud, S. (2017). Forecasting of sales by using fusion of machine learning techniques. *2017 International Conference on Data Management, Analytics and Innovation, ICDMAI 2017*, 93–101. <https://doi.org/10.1109/ICDMAI.2017.8073492>
- Hajirahimi, Z., & Khashei, M. (2019). Hybrid structures in time series modeling and forecasting: A review. *Engineering Applications of Artificial Intelligence*, 86, 83–106. <https://doi.org/10.1016/j.engappai.2019.08.018>
- Hastie, T., & Tibshirani, R. (1987). Generalized additive models: Some applications. *Journal of the American Statistical Association*, 82(398), 371–386. <https://doi.org/10.1080/01621459.1987.10478440>
- Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, 154, 72–80. <https://doi.org/10.1016/j.IJPE.2014.04.018>
- Hsu, D. (2017). Time Series Forecasting Based on Augmented Long Short-Term Memory. *Undefined*.
- Huang, T., Fildes, R., & Soopramanien, D. (2014). The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. *European Journal of Operational Research*, 237(2), 738–748. <https://doi.org/10.1016/j.ejor.2014.02.022>
- Hyndman, R.J. Athanasopoulos, G. (2019). *Forecasting: principles and practice, 3rd edition*. OTexts: Melbourne, Australia. <https://otexts.com/fpp3/>
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 1–22. <https://doi.org/10.18637/jss.v027.i03>

- Hyndman, R. J., & Kostenko, A. V. (2007). *Minimum sample size requirements for seasonal forecasting models*. <https://robjhyndman.com/publications/minimum-sample-size-requirements-for-seasonal-forecasting-models/>
- Introduction au Stockage (d'objets) Blob - Azure Storage | Microsoft Docs*. (2021). <https://docs.microsoft.com/fr-fr/azure/storage/blobs/storage-blobs-introduction>
- Jeong, D.-B. (2017). Prediction of Sales on Some Large-Scale Retailing Types in South Korea. *Asian Journal of Business Environment*, 7(4), 35–41. <https://doi.org/10.13106/EAJBM.2017.VOL7.NO4.35>
- Ji, S., Wang, X., Zhao, W., & Guo, D. (2019). An application of a three-stage XGboost-based model to sales forecasting of a cross-border e-commerce enterprise. *Mathematical Problems in Engineering*, 2019. <https://doi.org/10.1155/2019/8503252>
- Ju-Long, D. (1982). Control problems of grey systems. *Systems and Control Letters*, 1(5), 288–294. [https://doi.org/10.1016/S0167-6911\(82\)80025-X](https://doi.org/10.1016/S0167-6911(82)80025-X)
- Kache, F., Seuring, S., Kache, F., & Seuring, S. (2017). Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management. *International Journal of Operations & Production Management*. <https://doi.org/10.1108/IJOPM-02-2015-0078>
- Kathuria, R., Joshi, M. P., & Porth, S. J. (2007). Organizational alignment and performance: Past, present and future. *Management Decision*, 45(3), 503–517. <https://doi.org/10.1108/00251740710745106>
- Kong, J. H. L., & Martin, G. P. M. D. (1995). Backpropagation neural network for sales forecasting. *IEEE International Conference on Neural Networks - Conference Proceedings*, 2, 1007–1012. <https://doi.org/10.1109/icnn.1995.487558>
- Krishna, A., Akhilesh, V., Aich, A., & Hegde, C. (2018). Sales-forecasting of Retail Stores using Machine Learning Techniques. *Proceedings 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions, CSITSS 2018*, 160–166. <https://doi.org/10.1109/CSITSS.2018.8768765>
- Kristensen, J., & Jonsson, P. (2018). Context-based sales and operations planning (S&OP) research : A literature review and future agenda. *International Journal of Physical Distribution & Logistics Management*. <https://doi.org/10.1108/IJPDLM-11-2017-0352>
- Lapide, B. L. (2004a). SALES AND OPERATIONS PLANNING. *The Journal of Business Forecasting*, 17–19.
- Lapide, B. L. (2004b). SALES AND OPERATIONS PLANNING PART II: TECHNOLOGY ARCHITECTURE. *The Journal of Business Forecasting*, 8–10.
- Lapide, B. L. (2005). SALES AND OPERATIONS PLANNING. *The Journal of Business Forecasting*, 13–16.
- Lapide, L. (2011). S&OP: The Linchpin Planning Process. *Supply Chain Management Review*. www.scmr.com
- LARSEN, K. (2015). *GAM: The Predictive Modeling Silver Bullet*. <https://multithreaded.stitchfix.com/blog/2015/07/30/gam/>

- Li, K., Li, J., Tang, J., Chen, S., Wu, J., Xie, Y., & Xiao, K. (2020). Electricity Sales Forecasting Based on Model Fusion and Prophet Model. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12653 LNCS, 299–315. https://doi.org/10.1007/978-3-030-73671-2_26
- Li, M., Ji, S., & Liu, G. (2018). Forecasting of Chinese E-Commerce Sales: An Empirical Comparison of ARIMA, Nonlinear Autoregressive Neural Network, and a Combined ARIMA-NARNN Model. *Mathematical Problems in Engineering*, 2018, 1–12. <https://doi.org/10.1155/2018/6924960>
- Liu, N., Ren, S., Choi, T. M., Hui, C. L., & Ng, S. F. (2013). Sales forecasting for fashion retailing service industry: A review. In *Mathematical Problems in Engineering* (Vol. 2013). <https://doi.org/10.1155/2013/738675>
- Liu, T., Tan, Z., Xu, C., Chen, H., & Li, Z. (2020). Study on deep reinforcement learning techniques for building energy consumption forecasting. *Energy and Buildings*, 208, 109675. <https://doi.org/10.1016/j.enbuild.2019.109675>
- Liu, X., & Ichise, R. (2017). *Food Sales Prediction with Meteorological Data — A Case Study of a Japanese Chain Supermarket* (pp. 93–104). Springer, Cham. https://doi.org/10.1007/978-3-319-61845-6_10
- Loureiro, A. L. D., Miguéis, V. L., & da Silva, L. F. M. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems*, 114, 81–93. <https://doi.org/10.1016/j.dss.2018.08.010>
- Marino, D. L., Amarasinghe, K., & Manic, M. (2016). Building energy load forecasting using Deep Neural Networks. *IECON Proceedings (Industrial Electronics Conference)*, 7046–7051. <https://doi.org/10.1109/IECON.2016.7793413>
- Mccarthy, T. M., Davis, D. F., Golicic, S. L., & Mentzer, J. T. (2006). The evolution of sales forecasting management: A 20-year longitudinal study of forecasting practices. *Journal of Forecasting*, 25(5), 303–324. <https://doi.org/10.1002/for.989>
- Nagbe, K., Cugliari, J., & Jacques, J. (2018). Short-Term Electricity Demand Forecasting Using a Functional State Space Model. *Energies*, 11(5), 1120. <https://doi.org/10.3390/en11051120>
- Panigrahi, S., & Behera, H. S. (2017). A hybrid ETS–ANN model for time series forecasting. *Engineering Applications of Artificial Intelligence*, 66, 49–59. <https://doi.org/10.1016/j.engappai.2017.07.007>
- Pankaj Malhotra, Lovekesh Vig, Gautam Shroff, & Puneet Agarwal. (2015). Long Short Term Memory Networks for Anomaly Detection in Time Series. *European Symposium on Artificial Neural Networks*. <http://www.i6doc.com/en/>
- Pankratz, A. (1991). *Forecasting with Dynamic Regression Models*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118150528>
- Pavlyshenko, B. (2019). Machine-Learning Models for Sales Time Series Forecasting. *Data*, 4(1), 15. <https://doi.org/10.3390/data4010015>
- Pinho, J. M., Oliveira, J. M., & Ramos, P. (2016). Sales forecasting in retail industry based on

- dynamic regression models. *Advances in Transdisciplinary Engineering*, 3, 483–488. <https://doi.org/10.3233/978-1-61499-668-2-483>
- Polikar, R. (2012). Ensemble Learning. In *Ensemble Machine Learning* (pp. 1–34). Springer US. https://doi.org/10.1007/978-1-4419-9326-7_1
- Pongdatu, G. A. N., & Putra, Y. H. (2018). Seasonal Time Series Forecasting using SARIMA and Holt Winter's Exponential Smoothing. *IOP Conference Series: Materials Science and Engineering*, 407(1), 012153. <https://doi.org/10.1088/1757-899X/407/1/012153>
- Premiers pas – Google Cloud Platform*. (2021). <https://console.cloud.google.com/getting-started?hl=fr&pli=1>
- Punam, K., Pamula, R., & Jain, P. K. (2019). A two-level statistical model for big mart sales prediction. *2018 International Conference on Computing, Power and Communication Technologies, GUCON 2018*, 617–620. <https://doi.org/10.1109/GUCON.2018.8675060>
- Qi, Y., Li, C., Deng, H., Cai, M., Qi, Y., & Deng, Y. (2019). A deep neural framework for sales forecasting in e-commerce. *International Conference on Information and Knowledge Management, Proceedings*, 299–308. <https://doi.org/10.1145/3357384.3357883>
- Qu'est-ce que Azure Machine Learning ? - Azure Machine Learning | Microsoft Docs*. (2021). <https://docs.microsoft.com/fr-fr/azure/machine-learning/overview-what-is-azure-machine-learning>
- Qu'est-ce que le studio Azure Machine Learning ? - Azure Machine Learning | Microsoft Docs*. (2021). <https://docs.microsoft.com/fr-fr/azure/machine-learning/overview-what-is-machine-learning-studio>
- Ramos, P., Santos, N., & Rebelo, R. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing*, 34, 151–163. <https://doi.org/10.1016/j.rcim.2014.12.015>
- Ratnayaka, R. M. K. T., Seneviratne, D. M. K. N., Jianguo, W., & Arumawadu, H. I. (2015). A hybrid statistical approach for stock market forecasting based on Artificial Neural Network and ARIMA time series models. *2015 International Conference on Behavioral, Economic and Socio-Cultural Computing, BESC 2015*, 54–60. <https://doi.org/10.1109/BESC.2015.7365958>
- Robin Hyndman, Anne B. Koehler, J. Keith Ord, & Ralph D. Snyder. (2008). *Forecasting with Exponential Smoothing - The State Space Approach* (1st ed.). Springer-Verlag Berlin Heidelberg. <https://www.springer.com/gp/book/9783540719168>
- Sa-Ngasoongsong, A., Bukkapatnam, S. T. S., Kim, J., Iyer, P. S., & Suresh, R. P. (2012). Multi-step sales forecasting in automotive industry based on structural relationship identification. *International Journal of Production Economics*, 140(2), 875–887. <https://doi.org/10.1016/j.ijpe.2012.07.009>
- Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3), 1181–1191. <https://doi.org/10.1016/j.ijforecast.2019.07.001>
- Schoenherr, T., & Speier-pero, C. (2015). Data Science , Predictive Analytics , and Big Data in

- Supply Chain Management : Current State and Future Potential. *Journal of Business Logistics*, 1–13. <https://doi.org/10.1111/jbl.12082>
- Services de cloud computing | Microsoft Azure*. (2021). <https://azure.microsoft.com/fr-fr/Services-et-produits-de-cloud/Amazon-AWS/>
- Shahabuddin, S. (2009). Forecasting automobile sales. *Management Research News*, 32(7), 670–682. <https://doi.org/10.1108/01409170910965260>
- Shao, Y., & Tsai, Y.-S. (2018). Electricity Sales Forecasting Using Hybrid Autoregressive Integrated Moving Average and Soft Computing Approaches in the Absence of Explanatory Variables. *Energies*, 11(7), 1848. <https://doi.org/10.3390/en11071848>
- Sharma, S. K., Chakraborti, S., & Jha, T. (2019). Analysis of book sales prediction at Amazon marketplace in India: a machine learning approach. *Information Systems and E-Business Management 2019 17:2*, 17(2), 261–284. <https://doi.org/10.1007/S10257-019-00438-3>
- Singh, P. K., Gupta, Y., Jha, N., & Rajan, A. (2019). Fashion Retail: Forecasting Demand for New Items. *ArXiv*. <http://arxiv.org/abs/1907.01960>
- Somu, N., M R, G. R., & Ramamritham, K. (2020). A hybrid model for building energy consumption forecasting using long short term memory networks. *Applied Energy*, 261, 114131. <https://doi.org/10.1016/j.apenergy.2019.114131>
- Subrmanian, K., Othman, M. Bin, Sokkalingam, R., Thangarasu, G., & Subramanian, K. (2020). *A Hybrid Model of Holt-Winter and Neural Network Methods for Automobile Sales Forecasting*. 242–245. <https://doi.org/10.1109/icci51257.2020.9247838>
- Sun, Z. L., Choi, T. M., Au, K. F., & Yu, Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems*, 46(1), 411–419. <https://doi.org/10.1016/j.dss.2008.07.009>
- Tanaka, K. (2010). A sales forecasting model for new-released and nonlinear sales trend products. *Expert Systems with Applications*, 37(11), 7387–7393. <https://doi.org/10.1016/j.eswa.2010.04.032>
- Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *American Statistician*, 72(1), 37–45. <https://doi.org/10.1080/00031305.2017.1380080>
- Thomassey, S., & Fiordaliso, A. (2006). A hybrid sales forecasting system based on clustering and decision trees. *Decision Support Systems*, 42(1), 408–421. <https://doi.org/10.1016/j.dss.2005.01.008>
- Thome, A. M. T., Scavarda, L. F., Fernandez, N. S., & Scavarda, A. J. (2012a). Sales and operations planning : A research synthesis. *Int . J . Production Economics*, 138, 1–13. <https://doi.org/10.1016/j.ijpe.2011.11.027>
- Thome, A. M. T., Scavarda, L. F., Fernandez, N. S., & Scavarda, A. J. (2012b). Sales and operations planning and the firm performance. *International Journal of Productivity and Performance Management*. <https://doi.org/10.1108/17410401211212643>
- Thouvenot, V., Pichavant, A., Goude, Y., Antoniadis, A., & Poggi, J. M. (2016). Electricity forecasting using multi-stage estimators of nonlinear additive models. *IEEE Transactions*

- on *Power Systems*, 31(5), 3665–3673. <https://doi.org/10.1109/TPWRS.2015.2504921>
- Tiwari, S., Wee, H. M., & Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016 : Insights to industries. *Computers & Industrial Engineering*, 115(October 2017), 319–330. <https://doi.org/10.1016/j.cie.2017.11.017>
- Tuomikangas, N., & Kaipia, R. (2014). A coordination framework for sales and operations planning (S & OP): Synthesis from the literature. *Intern. Journal of Production Economics*, 154, 243–262. <https://doi.org/10.1016/j.ijpe.2014.04.026>
- Vahabi, A., Seyyedi, S., & Alborzi, M. (2016). A Sales Forecasting Model in Automotive Industry using Adaptive Neuro-Fuzzy Inference System(Anfis) and Genetic Algorithm(GA). *International Journal of Advanced Computer Science and Applications*, 7(11). <https://doi.org/10.14569/IJACSA.2016.071104>
- Vandeput, N. (2018). *Data Science for Supply Chain Forecast* (1st editio). <https://www.amazon.fr/Data-Science-Supply-Chain-Forecast/dp/1730969437>
- Wagner, S. M., Ullrich, K. K. R., & Transchel, S. (2014). The game plan for aligning the organization. *Business Horizons*, 57(2), 189–201. <https://doi.org/10.1016/j.bushor.2013.11.002>
- Wallace, B. T., & Stahl, B. (2008). The Demand Planning Process in Executive S&OP. *The Journal of Business Forecasting*, 19–24.
- Waller, M. A., & Fawcett, S. E. (2013). Data Science , Predictive Analytics , and Big Data : A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, 34(2), 77–84.
- Wang, G., Gunasekaran, A., Ngai, E. W. T., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management : Certain investigations for research and applications. *Intern. Journal of Production Economics*, 176, 98–110. <https://doi.org/10.1016/j.ijpe.2016.03.014>
- Wang, L., Wang, Z., Qu, H., & Liu, S. (2018). Optimal Forecast Combination Based on Neural Networks for Time Series Forecasting. *Applied Soft Computing*, 66, 1–17. <https://doi.org/10.1016/J.ASOC.2018.02.004>
- Weng, T., Liu, W., & Xiao, J. (2019). Supply chain sales forecasting based on lightGBM and LSTM combination model. *Industrial Management and Data Systems*, 120(2), 265–279. <https://doi.org/10.1108/IMDS-03-2019-0170>
- Xu, S., Chan, H. K., & Zhang, T. (2019). Forecasting the demand of the aviation industry using hybrid time series SARIMA-SVR approach. *Transportation Research Part E: Logistics and Transportation Review*, 122, 169–180. <https://doi.org/10.1016/j.tre.2018.12.005>
- Zhang, C., Tian, Y. X., Fan, Z. P., Liu, Y., & Fan, L. W. (2020). Product sales forecasting using macroeconomic indicators and online reviews: a method combining prospect theory and sentiment analysis. *Soft Computing*, 24(9), 6213–6226. <https://doi.org/10.1007/s00500-018-03742-1>
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)

- Zhao, J., Tang, W., Fang, X., Wang, J., Liu, J., Ouyang, H., Fang, H., Liu, Y., Wang, S., Lu, Y., & Qiang, J. (2015). A novel electricity sales forecasting method based on clustering, regression and time-series analysis. *Proceedings of Science*, 12-13-Sept, 015. <https://doi.org/10.22323/1.259.0015>
- Zunic, E., Korjenic, K., Hodzic, K., & Donko, D. (2020). Application of Facebook's Prophet Algorithm for Successful Sales Forecasting Based on Real-world Data. *ArXiv*, 12(2). <https://doi.org/10.5121/ijcsit.2020.12203>

Annexe A : Secteur de la vente au détail

Une description plus détaillée de quelques travaux de la littérature pour le secteur de la vente au détail est donnée dans les paragraphes suivants.

Dans leur article, (Aye et al., 2015) étudient le problème de la prévision des ventes au détail du marché Sud-Africain. Leurs travaux visent à améliorer la capacité des investisseurs à prévoir les mouvements des cours des actions des chaînes de ventes au détail. Pour cela, les chercheurs exploitent un historique des données de ventes mensuelles sur une période de 42 ans pour entraîner et évaluer leurs approches. Ces données sont caractérisées par une tendance globale croissante et d'une saisonnalité identifiable chaque année avec une hausse des ventes en décembre et une baisse importante en janvier. Ceci est expliqué par les achats intensifs durant la période de la fin d'année grâce aux fêtes de Noël et du Nouvel An et aux primes de fin d'année. Ces dépenses sont moins importantes en janvier étant donné que les gens ont repris leurs activités productives (travail, étude...), et aussi pour économiser ou rembourser des emprunts suite à d'importantes dépenses en décembre. Les données présentent également quelques volatilités et des pics plus ou moins importants. Pour répondre à cette problématique de prévision, les auteurs mettent en compétition 23 modèles ainsi que 3 méthodes de combinaison pour différents horizons de prédiction. Les résultats montrent qu'il n'y a pas un modèle qui performe le mieux dans toutes les situations et que les méthodes de combinaison ont tendance à donner des résultats plus précis.

(Krishna et al., 2018) étudient la problématique de prévision des ventes au détail pour 10 différents magasins ayant une offre très diversifiée. L'objectif est de créer un modèle de prédiction capable de déterminer les quantités de vente pour chaque produit référencé. Les auteurs se basent sur les historiques de vente mais aussi sur les caractéristiques de chaque produit et des magasins. Ces produits sont marqués par leur diversité avec 16 catégories présentes dans les données. Un processus de préparation des données a été employé pour identifier les variables d'entrée les plus importantes et les consolider avant l'implémentation des modèles (en prenant au final 12 variables en entrée). Les auteurs emploient 4 techniques de régression et 2 techniques de « Boosting » pour identifier l'approche la plus intéressante à utiliser. Les résultats montrent que les algorithmes de « Boosting », notamment le

« GradientBoost », performant mieux au niveau des différentes métriques d'erreur utilisées pour la comparaison des méthodes.

D'autres travaux comme (Ramos et al., 2015) traitent aussi la problématique de prévision des ventes au détail pour une entreprise portugaise spécialisée dans la fabrication des chaussures ayant à peu près 70 magasins de ventes. Cette activité est fortement influencée par les saisons notamment hivernale et estivale. Les travaux s'intéressent à la prédiction des ventes mensuelles de 5 catégories de chaussures en se basant sur un historique de ventes de 5 ans. Certains produits sont vendus plus en hiver tandis que d'autres plutôt en été avec des tendances parfois croissante, décroissante et constante selon la catégorie du produit. Les auteurs emploient 80 % des données pour l'entraînement des approches et 20 % des données pour leur évaluation. Ils comparent l'usage des techniques de Lissage Exponentiel et des modèles ARIMA pour le calcul des prévisions. Leurs résultats montrent que l'usage de ces deux techniques donne des résultats similaires pour le court terme (un mois) et le long terme (plusieurs mois) de prédiction selon différentes mesures d'erreurs utilisées.

Annexe B : Secteur de la mode

Une description plus détaillée des travaux évoqués pour caractériser les données et les techniques utilisées pour le secteur de la mode est donnée dans les paragraphes suivants.

Les travaux de (Loureiro et al., 2018) étudient la problématique de prévision des ventes d'une entreprise dans le domaine de ventes des produits de mode, dans le but d'adapter leurs stratégies pour satisfaire les préférences et les attentes des consommateurs. Ces produits correspondent à 684 types des sacs pour les femmes, vendus principalement sur le marché espagnol et portugais. Ils sont caractérisés généralement par un cycle de vie court mais ils diffèrent dans leurs propriétés physiques. Les auteurs proposent une approche basée sur les réseaux de neurones profonds pour calculer les prévisions. Les variables d'entrée considérées dans cette étude sont les caractéristiques physiques des produits (couleur, famille, prix, taille, etc.) avec leurs historiques des ventes ainsi que les avis des experts du domaine. Dans cette étude, les auteurs comparent les performances de leur approche par rapport à d'autres techniques comme les arbres de décision, forêts aléatoires, Machine à vecteurs de support (SVM), la Régression linéaire et un simple réseau de neurones (figure 50).

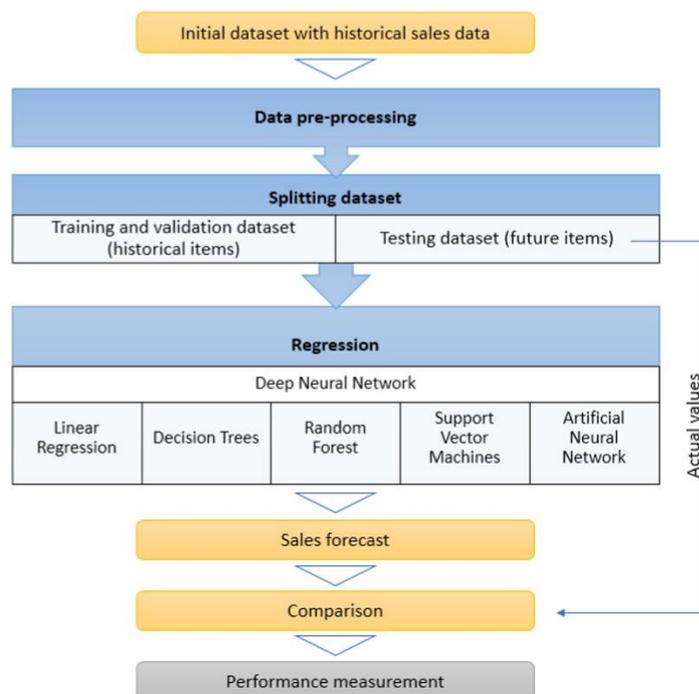


Figure 50 : Méthodologie proposée par (Loureiro et al., 2018)

L'évaluation est faite sur 5 métriques d'erreur. Leur approche présente de meilleurs résultats par rapport aux autres techniques pour deux des métriques employées. Les résultats montrent également que le modèle de forêts aléatoires présente de bons résultats notamment pour les trois autres métriques de comparaison utilisées.

Les travaux de (Singh et al., 2019) constituent un cas d'étude de la prévision des ventes pour une entreprise indienne dans le but de déterminer les quantités à prévoir pour de nouveaux articles lancés sur le marché. Ces nouveaux produits présentent un risque de ne pas être à la hauteur des attentes des consommateurs et donc le risque de rester en stock invendu posant ainsi des pertes financières pour l'entreprise. Les données utilisées dans cette étude correspondent à un historique de ventes hebdomadaires variant de 12 à 182 observations pour 5 types d'articles, ainsi que des caractéristiques physiques des produits en question. Les ventes de ce type de produits sont influencées par les événements nationaux et religieux, et par les promotions proposées par les magasins. Pour répondre à la problématique de prévision, les auteurs entraînent et évaluent différentes techniques basées sur les arbres de décision et sur les réseaux de neurones. Leurs résultats montrent que, après une transformation des données, le modèle « XGBoost » donne de meilleurs résultats pour la prédiction des nouveaux produits.

(Sun et al., 2008) étudient la problématique de prévisions des ventes pour une entreprise de ventes au détail des articles de la mode basée à Hong Kong. Les données utilisées dans les expérimentations correspondent aux facteurs affectant les ventes d'un article ainsi que son historique de ventes. Ils proposent l'usage d'une technique basée sur les réseaux de neurones à propagation du type « Extreme Learning Machine » (ELM). Avant d'utiliser leur approche, les données d'entrée sont normalisées et ensuite dénormalisées pour retrouver les quantités prédites. L'application de leur approche sur des données réelles montre de meilleures performances par rapport à d'autres techniques basées sur les architectures classiques de réseaux de neurones à rétropropagation.

Les travaux de (Beheshti-Kashi et al., 2015) représentent une enquête sur les méthodes utilisées dans l'état de l'art pour les prévisions des ventes dans plusieurs domaines avec un focus sur les techniques employées dans le secteur de la mode, et surtout pour les produits récemment lancés. D'après leur étude, les méthodes traditionnelles rencontrent des difficultés à produire des prévisions précises pour les articles récemment lancés sur le marché

et les produits orientés par les goûts des consommateurs, notamment, lorsque la demande est incertaine avec une grande variabilité des ventes et lorsque quelques données historiques sont manquantes. Les approches les plus récentes, notamment les techniques hybrides combinant des modèles linéaires et non linéaires donnent des résultats plus précis.

Annexe C : Secteur de l'agroalimentaire

Une description plus détaillée des travaux évoqués pour caractériser les données et les techniques utilisées pour le secteur de l'agroalimentaire est donnée dans les paragraphes suivants.

(F. L. Chen & Ou, 2009) étudient une chaîne logistique « froide » destinée à la distribution des produits alimentaires frais nécessitant des conditions de stockage spécifique. Ce type de Supply Chain nécessite de prendre en considération des précautions nécessaires pour maintenir la qualité et la stabilité des produits pour permettre aux consommateurs de les recevoir avec une fraîcheur garantie. Ces précautions constituent des coûts supplémentaires pour l'entreprise qu'il faut réduire à travers une connaissance au préalable de la demande des consommateurs. Dans cete optique, les auteurs exploitent les données historiques des ventes journalières d'un magasin alimentaire à Taiwan et de ses magasins voisins ainsi que des données métrologiques pour calculer ses prévisions de ventes de lait frais. Ils reposent sur les données de 11 mois de ventes. Les auteurs appliquent d'abord une méthode d'identification des facteurs influenceurs des ventes appelée « Gray Relation Analysis (GRA) » (Ju-Long, 1982), ensuite entraînent et évaluent différents modèles de séries temporelles pour choisir le meilleur. Ils proposent une approche de séries temporelles basée sur les réseaux de neurones appelée « Multilayer Function Link Network (MFLN) » pour la prévision des ventes (voir figure 51).

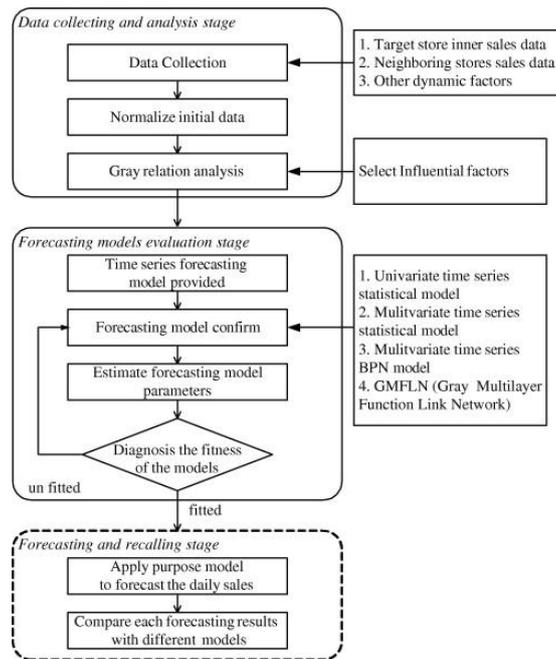


Figure 51 : Approche proposée par (F. L. Chen & Ou, 2009)

Ils mettent en compétition les modèles Moving Average (MA), ARIMAX, Réseau de neurones à rétropropagation multivarié (Multivariate BPN) et leur modèle MFLN. L'application de ces modèles sur des données des kiosques dépanneurs montre l'efficacité de leur approche et donne des meilleures prévisions que les autres techniques employées.

Les travaux (Arunraj & Ahrens, 2015) étudient le problème de prévision des ventes dans le cadre de prédiction des quantités de ventes journalières des bananes pour une entreprise allemande. L'objectif est de préserver le gaspillage et de réduire les ruptures de stocks des produits périssables comme les bananes, caractérisés par des ventes volatiles. Les auteurs proposent d'abord une taxonomie des techniques employées dans la littérature en les décomposant en trois types : des méthodes qualitatives basées sur les jugements et les avis des experts employés généralement lorsqu'on ne dispose pas des données suffisantes, des méthodes quantitatives et la combinaison des deux. Les méthodes quantitatives consistent au développement des modèles mathématiques avec les variables qui influencent les ventes, ou par l'extrapolation des données historiques des ventes à travers les modèles des séries temporelles, ou encore des modèles hybrides qui combinent les deux approches. Dans leurs travaux, les auteurs développent deux approches basées sur le modèle SARIMA avec des variables externes (les facteurs qui influent les ventes), à savoir SARIMA avec le modèle « Multiple Linear Regression » (SARIMA-MLR) et SARIMA avec le modèle « Quantile Regression » (SARIMA-QR). Les résultats montrent que ces techniques donnent de meilleurs

résultats comparant avec d'autres modèles comme SARIMA classique et le « Multilayer Perceptron (MLP) ».

Les travaux de (Doganis et al., 2006) constituent un cadre d'application des techniques de prédiction pour une entreprise opérant dans l'industrie laitière. Les données utilisées correspondent seulement à l'historique des volumes des ventes journalières du lait frais. Dans le but de réduire les produits invendus et retournés, ils proposent une approche non linéaire de prédiction, pour les séries temporelles avec un cycle de vie court, basée sur les réseaux de neurones avec un algorithme génétique spécifique. La comparaison des résultats avec des techniques classiques et récentes montre que leur méthodologie présente des meilleures performances et a été employée avec succès.

(C. Y. Chen et al., 2010) adressent le problème de management des commandes et des stocks des produits alimentaires chez les petites épicereries en prenant en considération les spécificités des lieux urbains, semi-urbains et montagneux. Dans le but de mieux gérer les stocks et éviter les gaspillages, les auteurs exploitent les données de vente des épicereries pour 35 jours de 10 types de produits périssables. Leur approche propose l'usage d'un réseau de neurones à rétropropagation qui donne des résultats meilleurs qu'une régression logistique ou encore la moyenne mobile.

Les travaux de (Da Veiga et al., 2016) étudient l'impact de la précision des prédictions des ventes d'une entreprise opérant dans l'industrie laitière, sur le taux de satisfaction des demandes clients ainsi que sur la totalité de l'activité de l'entreprise. Ils comparent les résultats de deux modèles de prédiction linéaires (ARIMA et Holt-Winters) avec les résultats de deux autres modèles non linéaires (Réseaux de neurones à ondelettes et Takagi-Sugeno Fuzzy System). Ces techniques sont employées sur les données de trois groupes de produits périssables mesurées entre 2005 et 2013. Leurs résultats montrent que les réseaux de neurones à ondelettes (WNN) performant beaucoup mieux que d'autres modèles et le taux de satisfaction de demande s'approche des 100 % pour les trois types de produits. Leur approche répond alors mieux au cas d'étude pour tous les groupes de produits à prédire et donc, leur proposition permet de faire des prévisions précises en employant un seul modèle à différents produits. Néanmoins, la mise en place de cette technique nécessite un temps de calcul relativement important et une grande expertise pour déterminer les paramètres du réseau.

Comme les secteurs étudiés précédemment, les approches de prédiction pour ce domaine d'activités reposent sur les modèles classiques tels que SARIMA et Lissage Exponentiel mais aussi sur des techniques basées sur les réseaux de neurones lorsqu'on dispose de suffisamment de données.

Annexe D : Secteur de la vente en ligne

Une description plus détaillée des travaux évoqués pour caractériser les données et les techniques utilisées pour le secteur de la vente en ligne est donnée dans les paragraphes suivants.

Les travaux de (Qi et al., 2019) proposent un modèle de prédiction des séries temporelles prenant en considération les campagnes des promotions faites sur différents produits. Leur modèle appelé « Deep Sales Forecasting Framework (DSF) » est basé sur un type spécifique de réseau de neurones récurrent appelé « Gated Recurrent Unit (GRU) » permettant l'apprentissage de la dépendance à long terme. Cette approche propose de faire les prédictions en deux parties : des ventes basiques sont estimées par le modèle selon l'historique récent, et en deuxième partie les prévisions de ventes sont ajustées prenant en compte les promotions affectant des produits concurrents ou similaires. Ces travaux ont été appliqués à deux bases de données de ventes de la plateforme Taobao (site de vente en ligne en Chine), et les résultats montrent de meilleures performances comparant avec des modèles de séries temporelles classiques tels que ARIMA et d'autres modèles d'apprentissage profond.

(Ji et al., 2019) étudient la problématique des prévisions des ventes dans le cadre d'une application au sein de l'entreprise Jollychic spécialisée dans le e-commerce. Cette entreprise vise à améliorer son service logistique et à optimiser le management de ses entrepôts à travers l'amélioration de ses prévisions de vente. La préparation des données à utiliser passe d'abord par un nettoyage, un clustering et une sélection des variables les plus intéressantes, vu la grande variété des données disponibles, de leurs sources et de leurs formats (texte, audio, vidéo, tweet, fichiers de logs, ...). Cet aspect de grande variété et un volume exceptionnel des données caractérisent surtout ce secteur de ventes en ligne. Les données employées dans cette étude correspondent à des données journalières collectées. Les auteurs proposent une méthodologie basée sur le modèle « XGBoost » à 3 niveaux permettant de faire la prédiction des ventes de cette plateforme de ventes en ligne (voir figure 52).

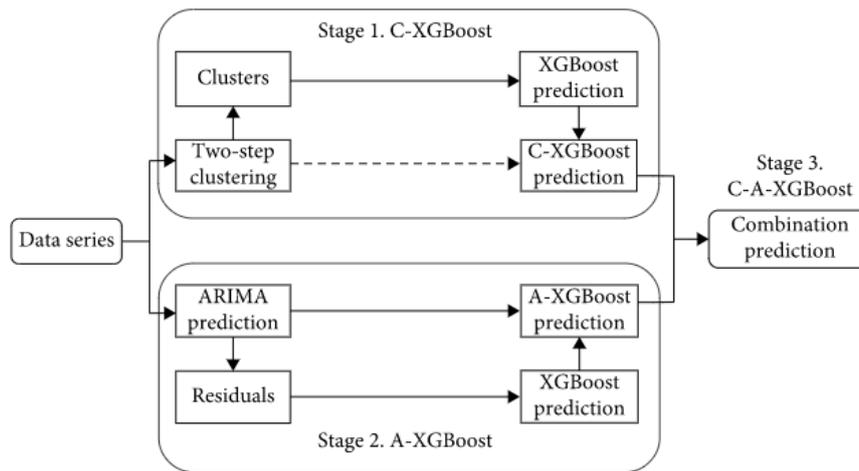


Figure 52 : Approche proposée par(Ji et al., 2019)

Le premier niveau consiste à regrouper les séries de données grâce un algorithme de clustering à 2 étapes afin de préparer les variables d'entrée, ensuite des prédictions sont calculées par un modèle XGBoost. Le deuxième niveau consiste à modéliser la tendance dans les données par un modèle hybride d'ARIMA pour la partie linéaire et de XGBoost pour la partie non linéaire et finalement, les résultats des deux premiers niveaux sont combinés en accordant des poids à chacun pour obtenir les prévisions finales. Les expérimentations montrent que le modèle final résultant de cette combinaison « C-A-XGBoost » donne des meilleurs résultats que les modèles séparés.

Dans les travaux de (M. Li et al., 2018), les auteurs étudient la problématique des prévisions des ventes de la plateforme de ventes en ligne Jingdong en Chine visant la mise en place des stratégies d'achat et de stockage reposant sur les prévisions des ventes. Les produits vendus sur la plateforme sont de différentes natures (bricolage, jouets, vêtements, aliments, ...) subissant parfois des promotions et des changements des prix. Les données employées dans cette étude correspondent à des historiques de ventes hebdomadaires de 60 types de produits différents se présentant sous la forme de séries temporelles de 167 observations couvrant le période entre janvier 2014 et mars 2017. Les auteurs proposent des modèles hybrides d'ARIMA avec des modèles non linéaires comme le réseau de neurones autorégressif non linéaire (NARNN). La figure 53 présente l'approche proposée par (M. Li et al., 2018).

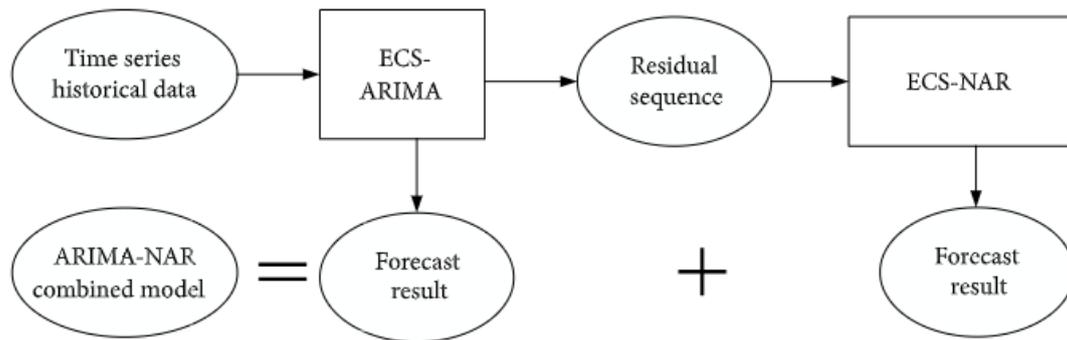


Figure 53 : Approche d'hybridation proposée par(M. Li et al., 2018)

Cette approche est appliquée aux 60 produits en question. Cette technique considère que les prévisions finales seront la somme des prédictions du modèle ARIMA et les prédictions du réseau de neurones. Les données sont d'abord modélisées par ARIMA pour identifier la partie linéaire de la série, ensuite, les résidus seront les entrées pour les réseaux de neurones qui modéliseront la partie non linéaire de la série. Pour entraîner et évaluer cette approche, ils utilisent une répartition de 85 % et 15 % respectivement pour l'entraînement et le test. Les résultats montrent que cette méthodologie d'hybridation donne de meilleurs résultats que l'usage des modèles séparément.

Annexe E : Secteur automobile

Une description plus détaillée des travaux évoqués pour caractériser les données et les techniques utilisées pour le secteur automobile est donnée dans les paragraphes suivants.

Les travaux de (Subrmanian et al., 2020) étudient la problématique de prévisions des ventes dans le cadre de l'industrie d'automobile indienne. Cette tâche facilite la prise de décision concernant la planification de la production et de la logistique en fournissant un guide stratégique à long terme. Les données utilisées dans cette étude correspondent aux données de ventes de 20 constructeurs automobiles indiens entre 2010 et 2019. Les auteurs proposent un modèle hybride composé d'un modèle « Adaptive Holt-Winters (AHW) » et un réseau de neurones à rétropropagation (BPNNs) comme le montre la figure 54.

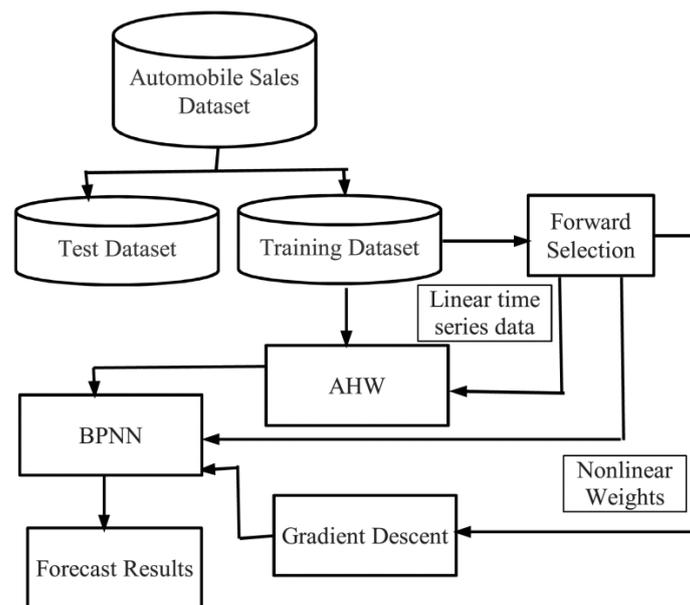


Figure 54 : Méthodologie proposée par (Subrmanian et al., 2020)

Le lissage exponentiel modélise la partie linéaire de la série temporelle, ensuite, ces prédictions seront des inputs pour un réseau de neurones (BPNN) avec la partie non linéaire de la série. Un algorithme de descente de gradient est employé pour calculer les biais dans le réseau de neurones. Le BPNN combine toutes les entrées et calcule les prédictions finales. Cette approche montre des résultats suffisants sur les données de test par rapport aux méthodologies existantes.

Les travaux de (Sa-Ngasoongsong et al., 2012) et (Gao et al., 2018) constituent deux cadres d'application de la problématique de prévision des ventes employant des variables économiques externes pour les calculer dans le secteur de l'industrie automobile. L'objectif étant d'améliorer les prévisions sur un long horizon (6 – 24 mois), permettant ainsi d'être plus compétitif sur ce marché. Les données employées dans les travaux de (Sa-Ngasoongsong et al., 2012) correspondent au marché américain en exploitant les historiques mensuels de ventes entre 1975 et 2010, ainsi que d'autres facteurs économiques externes tels que le prix de l'essence, l'indice du prix du consommateur et le taux de chômage dans le pays. Dans la même idée, (Gao et al., 2018) emploient les données de ventes de l'entreprise d'automobile chinoise Chery entre 2007 et 2016, avec des facteurs externes tels que le taux de production de l'acier, l'indice de confiance du consommateur et le prix de l'essence. Vu les différentes variables économiques qui peuvent influencer le marché, ces deux travaux emploient une méthodologie d'identification des relations structurelles basée sur plusieurs tests statistiques (Unit Root Test, Weak Exogeneity Test, Granger-causality Test, Cointegration Test) pour identifier les relations entre l'évolution des ventes et les indicateurs économiques, et les exploiter ensuite à travers des modèles de prédiction des séries temporelles. La figure 55 présente les différentes étapes de leur méthodologie.

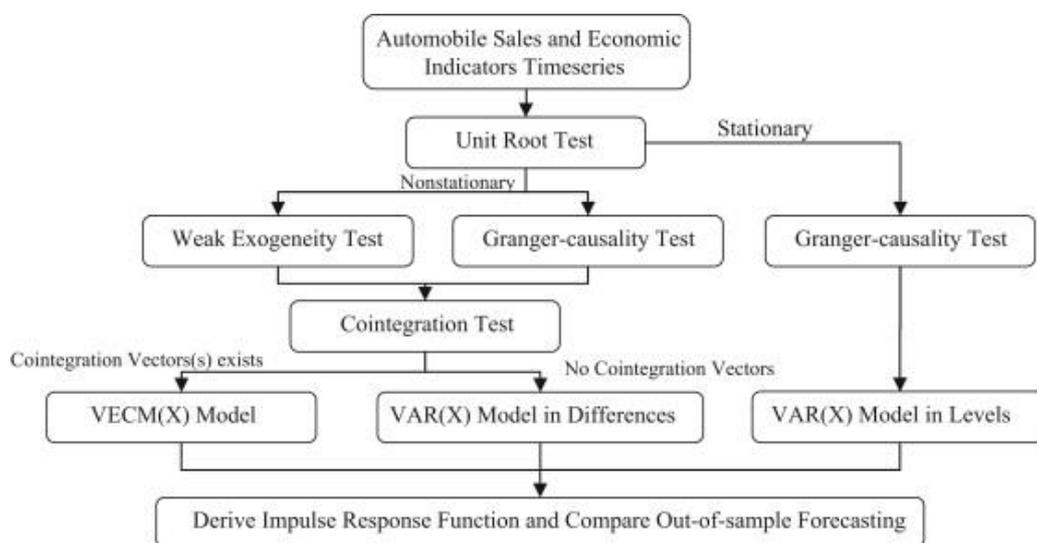


Figure 55 : Approche proposée par (Sa-Ngasoongsong et al., 2012)

Selon le résultat de chaque test, les prédictions vont être déterminées soit par un modèle Vector Autoregressive (VAR) ou Vector Autoregressive eXogenous (VARX), soit par un modèle Vector Error Correction Model (VECM) ou Vector Error Correction Model With eXogenous Variables (VECMX). Dans les deux cas d'études (Gao et al., 2018; Sa-Ngasoongsong et al.,

2012), les résultats montrent de meilleures performances de cette approche en comparaison avec d'autres techniques classiques de modélisation des séries temporelles.

Les travaux de (Z. P. Fan et al., 2017) étudient la problématique de prévision des ventes dans le secteur automobile en exploitant les avis des utilisateurs en ligne. Vu que ces derniers fournissent des informations riches pour les consommateurs lors de leurs achats, ces travaux exploitent leur influence pour améliorer les prévisions des ventes du modèle de voiture « Elantra » de la marque Hyundai en Chine. Les données correspondent aux historiques de ventes mensuelles de 3 générations de ce modèle avec plusieurs avis des utilisateurs entre juillet 2007 et mars 2015 collectés à partir de la plateforme Bitauto. Les auteurs emploient le modèle Naïf Bayésien pour l'analyse des sentiments afin d'exploiter les informations des avis des consommateurs. Ils proposent ensuite, l'usage d'un modèle Bass/Norton pour calculer les prédictions. Les résultats montrent que leur approche obtient de meilleures performances par rapport au modèle Bass/Norton standard et par rapport au modèle « Log-linear ».

De même, (C. Zhang et al., 2020) étudient les prévisions des ventes dans le secteur de l'automobile en s'appuyant sur des indicateurs macroéconomiques et les avis des utilisateurs pour améliorer les prédictions dans le secteur de l'automobile. Les données employées correspondent à l'historique des ventes trimestrielles de 3 types de voitures collectées de la part de l'entreprise Wind Client, les avis des consommateurs à partir du site Bitauto et 242 indicateurs macroéconomiques de la Chine sur la période entre 2008 et 2018. Les auteurs proposent l'usage d'un modèle autorégressif logarithmique en lui fournissant comme entrées les indicateurs macroéconomiques choisis par un algorithme de sélection, un indice de sentiment calculé à travers les avis des consommateurs en ligne et les historiques des ventes. Leur méthodologie réussit à fournir des prévisions plus précises que des modèles plus classiques existants.

Les travaux de (Dwivedi et al., 2013) étudient les prévisions des ventes pour l'entreprise Maruti Car dans le but de les améliorer et de renforcer sa compétitivité pour le marché indien. Les données employées dans cette étude correspondent à des historiques de ventes mensuelles entre 2008 et 2012. Ils développent une approche hybride basée sur un modèle « Adaptive Neuro Fuzzy Inference System (ANFIS) ». La méthodologie consiste à faire des prédictions par la méthode des moyennes mobiles (MA) et le Lissage Exponentiel, ensuite, les résultats seront des inputs pour le modèle ANFIS. L'application de cette approche dans leur

cas d'étude, montre de meilleurs résultats par rapport à l'usage d'un réseau de neurones classique (ANN) et la Régression Linéaire.

D'après les travaux présentés, on remarque l'usage des techniques basées sur des réseaux de neurones ainsi que le modèle ANFIS pour surpasser les limites qui peuvent être rencontrées avec les méthodes classiques (Lissage Exponentiel, Régression Linéaire, ARIMA), ainsi que l'exploitation des données liées aux variables externes et pas seulement les historiques de ventes.

Annexe F : Secteur automobile

Une description plus détaillée des travaux évoqués pour caractériser les données et les techniques utilisées pour le secteur de l'énergie est donnée dans les paragraphes suivants.

(Al-Musaylh et al., 2018) traitent la problématique de prévisions des ventes dans le cadre de l'étude de la consommation électrique dans la région de Queensland en Australie. Ceci est un sujet critique puisqu'ils aident les ingénieurs en énergie renouvelable et conventionnelle, les fournisseurs d'électricité, les utilisateurs finaux et les entités gouvernementales à relever les défis de durabilité énergétique pour le marché national de l'électricité en Australie. Les données employées sont collectées à partir de l'opérateur du marché australien correspondant à la consommation entre janvier 2012 et décembre 2015 avec des enregistrements chaque demi-heure donc 48 observations par jour. Dans cette étude, les auteurs proposent l'usage des techniques de régression basées sur les modèles « Multivariate Adaptive Regression Spline » (MARS), « Support Vector Regression » (SVR), et « Autoregressive Integrated Moving Average » (ARIMA). Les résultats montrent que le modèle MARS présente de meilleurs résultats sur les horizons 0,5 h et 1h, tandis que SVR était plus précis sur un horizon de 24 h. Le modèle ARIMA était moins performant sur les différents horizons.

(Shao & Tsai, 2018) étudient différents modèles de prévision dans le cadre de l'absence des variables explicatives pour prédire les ventes de l'électricité industrielle, résidentielle et commerciale sur le marché taiwanais. Les données employées concernent les 3 types des marchés cités avec des ventes mensuelles de 2006 à juillet 2016 et donc 127 observations pour l'entraînement et le test des modèles. Les auteurs proposent l'usage des modèles hybrides pour la prédiction des ventes de l'électricité comme présenté dans la figure 56.

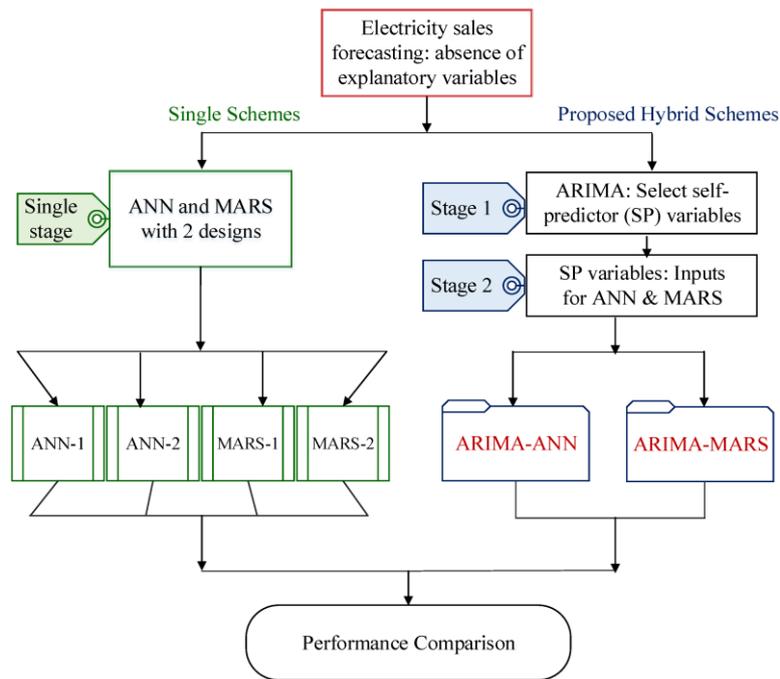


Figure 56 : Approche proposée par (Shao & Tsai, 2018)

Pour commencer, un modèle ARIMA est employé pour préparer les variables explicatives qui seront des entrées pour un réseau de neurones (ANN) et pour un modèle « Multivariate Adaptive Regression Splines (MARS) ». Ces modèles hybrides (ARIMA-ANN et ARIMA-MARS) sont employés pour prédire les trois types de marché (industriel, résidentiel et commercial). Les résultats montrent que l'emploi des méthodes hybrides présente des améliorations considérables par rapport à l'usage des modèles ANN et MARS séparément.

D'autres travaux proposent différentes techniques pour répondre à la problématique de la prédiction de la demande de l'énergie électrique. (Nagbe et al., 2018) étudient cette problématique dans le cadre d'une application sur les données des entreprises Enercoop et RTE afin de répondre à la demande croissante de l'électricité. Ces données correspondent à la consommation globale du marché de l'entreprise représentées par des séries temporelles de janvier 2012 à décembre 2017 avec des observations toutes les 30 minutes et donc 48 observations par jour. Pour l'entreprise RTE, les données correspondent à des observations chaque 15 minutes de 2012 jusqu'au 2018. Les auteurs proposent un modèle « Functional State Space Model » (FSSM) qui prend en considération principalement les valeurs de consommation passée pour prédire la future demande permettant ainsi de modéliser le comportement dynamique de la série temporelle. Pour l'estimation des paramètres du modèle, les auteurs emploient plusieurs techniques comme la maximisation de la

vraisemblance et le filtre de Kalman. La mise en application avec des données des entreprises Enercoop permet la validation de leur approche, et leur méthodologie donne des meilleurs résultats par rapport à la méthode de régression employée chez l'entreprise RTE.

Les travaux de (Elamin & Fukushige, 2018) sont développées dans le cadre de prédiction de la demande horaire en électricité pour une région spécifique au Japon. Le but étant d'anticiper la consommation de la région pour éviter les ruptures d'électricité. Les données employées dans cette étude correspondent aux quantités horaires d'électricité générées entre janvier 2012 et décembre 2015 du distributeur d'électricité japonais TEPCO, ainsi que des facteurs externes affectant la consommation tels que l'impact de la météo (température et humidité), les différentes saisonnalités (journalière, hebdomadaire, annuelle) ainsi que d'autres variables pouvant influencer la consommation. Les auteurs proposent l'usage d'un modèle SARIMA avec variables exogènes (SARIMAX) pour répondre à cette problématique. Leurs résultats montrent de meilleurs résultats par rapport à au modèle « Multiple Linear Regression » (MLR).

(Zhao et al., 2015) constituent un cadre de développement d'une méthodologie de prédiction d'électricité pour le réseau de distribution en Chine, suite au développement d'un réseau de distribution intelligent qui utilise la technologie des réseaux de communication pour détecter et réagir aux changements locaux d'utilisation. Les auteurs utilisent les données historiques mensuelles de consommation de 27 compagnies productrices d'électricité sous forme de séries temporelles entre janvier 2010 et décembre 2014. Ils proposent une approche de prédiction des ventes de l'électricité basée sur le clustering, la régression et les techniques des séries temporelles. La figure 57 présente leur méthodologie.

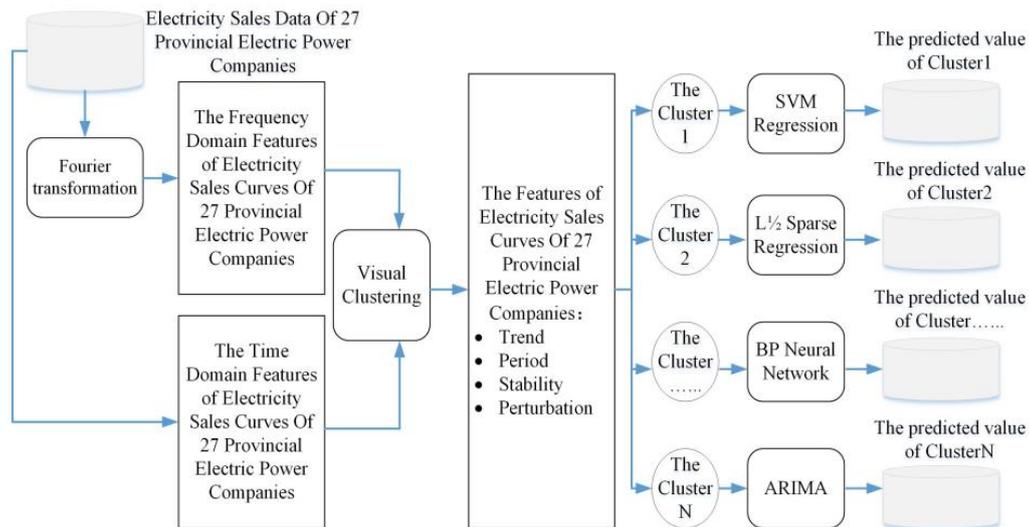


Figure 57 : Approche proposée par (Zhao et al., 2015)

La première étape consiste à créer des groupes de séries temporelles à travers la technique de clustering visuel selon différents critères comme la périodicité, la stabilité, la tendance, etc. Ces critères seront des entrées pour les modèles de prédiction. La deuxième étape consiste à calculer les prévisions pour chaque groupe selon le modèle approprié. Le choix de modèle à utiliser est défini selon un tableau de correspondance qui affecte chaque technique selon les critères de chaque cluster. Les données mises en application représentent les ventes entre 2010 et 2014 de ces compagnies. Les résultats des tests montrent que cette approche a un taux de précision intéressant.

Les travaux de (Ahmad et al., 2018) proposent une revue des approches employées dans la littérature pour prédire la demande d'énergie pour les bâtiments. Ils proposent quatre catégories : des approches guidées par les données, les approches basées sur la physique, les approches de prévisions énergétiques des bâtiments à grande échelle et des approches hybrides. Leurs travaux se focalisent plutôt sur les approches guidées par les données et les approches à grande échelle. Les techniques basées sur les données sont ainsi classées en quatre groupes également avec des méthodes basées sur les réseaux de neurones représentant plus que 40 % des applications dans la littérature, des méthodes statistiques et d'apprentissage automatique avec plus de 25 % des applications, des méthodes basées sur la technique de Machine à Vecteurs de Support (SVM) dans 12% des applications, et des méthodes basées sur le regroupement (clustering).

Les travaux de (Bourdeau et al., 2019) présentent une revue des méthodes guidées par les données proposées dans la littérature pour ce domaine. Ils se focalisent sur les caractéristiques des données d'entrée et les méthodes de prétraitement de ces données. La figure 58 présente un résumé des types des approches et des techniques appartenant à chaque catégorie.

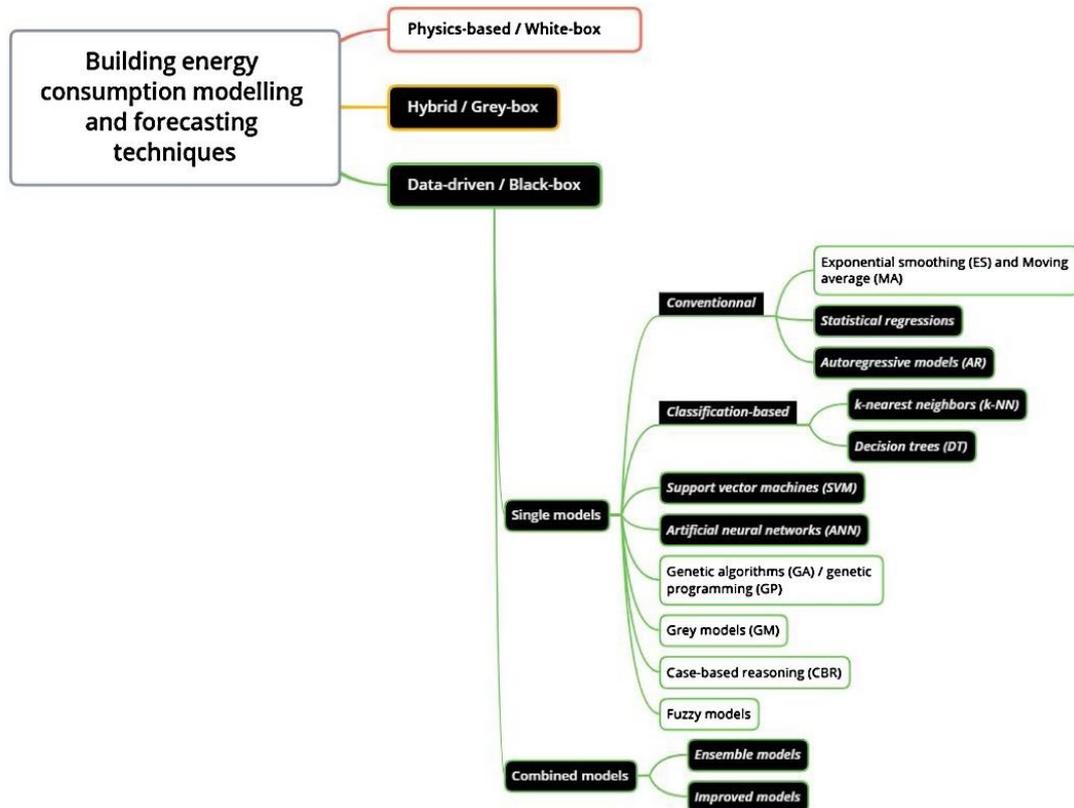


Figure 58 : Techniques par catégorie (Bourdeau et al., 2019)

D'autres revues dans la littérature se sont intéressées à ce sujet également. (Deb et al., 2017) propose de s'intéresser principalement aux approches basées sur les techniques de modélisation des séries temporelles pour la prédiction de la consommation de l'énergie des bâtiments. Ils identifient 9 techniques principales à savoir le réseau de neurones (ANN), ARIMA, Support Vector Machine (SVM), les méthodes floues, l'approche Grey, le lissage exponentiel et la moyenne mobile, réseaux de neurones profonds, Case-Based Reasoning (CBR) et les méthodes hybrides.

Résumé

Quelques soient le secteur d'activités des entreprises industrielles aujourd'hui, ces dernières doivent faire face à plusieurs enjeux et challenges économiques pour garantir leur pérennité et compétitivité sur leur marché. L'amélioration continue des processus de production et la gestion efficiente de la Supply Chain sont des pistes incontournables pour une bonne maîtrise des coûts et des risques associés à l'activité de chaque entreprise. Avec l'avancement techniques et technologiques, plusieurs pistes à fort potentiel telle que l'intelligence artificielle, se présentent comme des opportunités intéressantes et innovantes permettant de répondre à des problématiques industrielles complexes. Parmi ces dernières, les tâches de planification de la demande, de la production ou encore la gestion des stocks représentent des éléments clés pour une bonne maîtrise de la Supply Chain. C'est dans ce cadre que les travaux menés dans cette thèse s'intègrent, en ayant comme objectif de mettre en place un outil d'aide à la décision permettant de fiabiliser, faciliter et automatiser la préparation des prévisions des ventes. Nous proposons donc une approche guidée par les données permettant de répondre à cette problématique dans un cadre d'application industrielle chez l'entreprise elm.leblanc du groupe Bosch. Notre méthodologie proposée est complétée par une mise en œuvre à travers un développement sur une plateforme de service cloud dans l'entreprise.

Abstract

Whatever the sector of activity of industrial companies today, they must face several economic challenges to ensure their sustainability and competitiveness on their market. The continuous improvement of the production processes and the efficient management of the Supply Chain are essential tracks for a good control of the costs and the risks associated with the activity of each company. With the advancement of techniques and technologies, several high-potential avenues such as artificial intelligence, present themselves as interesting and innovative opportunities to respond to complex industrial problems. Among these, the tasks of demand planning, production or inventory management are key elements for a good control of the Supply Chain. It is within this framework that the work carried out in this thesis is integrated, with the objective of setting up a decision support tool allowing to make the preparation of sales forecasts more reliable, easier and more automated. We therefore propose a data-driven approach to address this issue in an industrial application framework at the elm.leblanc company of the Bosch group. Our proposed methodology is completed by an implementation through a development on a cloud service platform in the company.