

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Ecole Nationale Polytechnique



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique



Mémoire de projet de fin d'études pour l'obtention du diplôme d'ingénieur d'état en
Génie Industriel, option : Data Science et Intelligence Artificielle

Étude de l'impact du changement des prix sur les
performances de ventes par le recours à l'apprentissage
automatique.

Application : Decathlon El Djazair.

Présenté par :

Anes CHEMICHEM & Sofiane SADAT

Sous la direction de M. Iskander ZOUAGHI & de M. Ayoub ABBACI

Présenté et soutenu publiquement le (26/06/2023)

Composition du jury :

Président	M. Hakim FOURAR LAIDI	MCA	ENP
Examineur	M. Oussama ARKI	MCB	ENP
Promoteur	M. Iskander ZOUAGHI	MCA	ENP
Promoteur	M. Ayoub ABBACI	MCA	ENP
Invité	M Ayyoub OULD SETTI	SC. MANAGER	Decathlon

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Ecole Nationale Polytechnique



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique



Mémoire de projet de fin d'études pour l'obtention du diplôme d'ingénieur d'état en
Génie Industriel, option : Data Science et Intelligence Artificielle

Étude de l'impact du changement des prix sur les
performances de ventes par le recours à l'apprentissage
automatique.

Application : Decathlon El Djazair.

Présenté par :

Anes CHEMICHEM & Sofiane SADAT

Sous la direction de M. Iskander ZOUAGHI & de M. Ayoub ABBACI

Présenté et soutenu publiquement le (26/06/2023)

Composition du jury

Président	M. Hakim FOURAR LAIDI	MCA	ENP
Examineur	M. Oussama ARKI	MCB	ENP
Promoteur	M. Iskander ZOUAGHI	MCA	ENP
Promoteur	M. Ayoub ABBACI	MCA	ENP
Invité	M Ayyoub OULD SETTI	SC. MANAGER	Decathlon

ENP 2023

ملخص:

في سوق البيع بالتجزئة المتنافس في الوقت الحالي، تسعى الشركات باستمرار إلى إيجاد طرق لتحسين استراتيجيات التسعير الخاصة بها لزيادة الربحية والبقاء تنافسية. تدرك Decathlon، مثل الكثير من الشركات الأخرى، أهمية استراتيجيات التسعير وقامت بإجراء هذا المشروع لفهم أفضل لتأثير السعر على سلوك العملاء. يهدف هذا المشروع إلى إنشاء نموذج قادر على توقع المبيعات لكل سعر محتمل للمنتج، مما يتيح للشركة تحديد الأسعار التي تحافظ على المبيعات الثابتة وتزيد الأرباح. بالإضافة إلى ذلك، سيساعد النموذج في إدارة المخزون حيث يوفر توقعات تعتمد على الوقت.

الكلمات المفتاحية: التنبؤ، التسعير، السلاسل الزمنية، التعلم العميق، التعلم الآلي، زيادة البيانات، المرونة.

Abstract:

In today's competitive retail market, companies are constantly seeking ways to improve their pricing strategies to increase their profitability and remain competitive. Decathlon, like many others, recognizes the importance of pricing strategies and has undertaken this project to better understand the impact of price on customer behaviour. The goal of this project is to create a model that can predict sales for each potential price of a product, enabling the company to set prices that maintain stable sales and increase profit. Additionally, the model will aid in stock handling as it provides predictions based on time.

keywords: prediction, pricing, time series, deep learning, machine learning, data augmentation, elasticity

Résumé:

Dans le marché de la vente au détail compétitif d'aujourd'hui, les entreprises cherchent constamment des moyens d'améliorer leurs stratégies de tarification pour augmenter leur rentabilité et rester compétitives. Decathlon, comme beaucoup d'autres, reconnaît l'importance des stratégies de tarification et a entrepris ce projet pour mieux comprendre l'impact du prix sur le comportement des clients. Le but de ce projet est de créer un modèle capable de prédire les ventes pour chaque prix potentiel d'un produit, permettant à l'entreprise de fixer des prix qui maintiennent des ventes stables et augmentent les profits. De plus, le modèle aidera à la gestion des stocks, car il fournit des prévisions basées sur le temps.

Mots-clés : prédiction, tarification, séries temporelles, apprentissage profond, apprentissage automatique, augmentation de données, élasticité

Dédicaces

"À ma famille, mes parents, mes frères Yasser et Adem, et mes sœurs Insaf et Kaouter, je voudrais exprimer ma gratitude pour votre disponibilité constante et votre soutien inconditionnel tout au long de ces années.

Je tiens également à remercier mes amis Rayan, Rostom, Zinou, Anis et particulièrement Fayçal, pour les bons souvenirs partagés et pour m'avoir aidé lorsque j'en avais besoin.

Je n'oublierai jamais les moments passés avec les Daïta, vous allez me manquer. En particulier, Chaimaa, Insaf, Lina, Kimo, Moncef, Smail, Yousra, Sara7 et bien sur mon binôme Sofiane, j'aurais aimé continuer à réviser ensemble (surtout lilet r3ad). J'espère que nous pourrons nous retrouver le plus tôt possible dans de bonnes conditions, insha'Allah."

— ANES

" Je tiens à dédier ce travail à ma mère, qui s'est tant sacrifiée pour voir son fils réussir et atteindre ses objectifs, À mon père, qui a longtemps été garant de mon développement et mon évolution sur les bases qui ont fait de moi l'homme que je suis aujourd'hui. À mes chères sœurs Lydia, Lynda, qui ont été là depuis le début à suivre mon parcours de près, À mon cher frère Kaci a qui je leur souhaite tout le succès.

À mes grands-parents, que Dieu vous protège.

À la mémoire de ma tante, que Dieu t'accueille dans son vaste paradis. À mes cousins et cousines, tantes et oncles,

À mon binôme, mon frère et mon ami, Anes, qui a

Toujours était à côté de moi pour relever plusieurs défis et aujourd'hui dans la réalisation de ce projet.

À l'ensemble de mes camarades de la promotion les Daïtas "

— SOFIANE

Remerciement

Nous tenons à exprimer nos sincères remerciements à nos deux encadrants, M. ZOUAGHI Iskander et M. ABBACI Ayoub, pour leur encadrement de qualité tout au long de notre projet de fin d'études ainsi que pour l'année écoulée à l'École Nationale Polytechnique. Leur soutien, leur encouragement constant, leurs critiques constructives et leurs précieux conseils ont grandement contribué à notre évolution et à la réussite de ce travail.

Nous souhaitons également exprimer notre gratitude envers notre promoteur Decathlon, M. HADJADJ Ahmed, pour sa confiance en nous, sa grande disponibilité et ses conseils inestimables.

Nous remercions également le responsable IT M. FERDJOUKH Abderrahmane et le responsable supply Mr.OULD SETTI Ayyoub pour leur temps précieux, leur soutien et leur encouragement tout au long de nos recherches.

Nous tenons à remercier tous les leaders de Decathlon El Djazair, M. MEZINE Akim, M. BALI Redouane, Mr.SADI Mahdi, M. BELKACEMI Oualid, ainsi que les responsables juridique et RH Mme. BOUDOUR Dounia et M. DIFI Salan, ainsi que toute l'équipe e-commerce pour leur disponibilité, leur orientation et leur suivi pour que notre stage se déroule dans les meilleures conditions. Nous sommes fiers d'avoir travaillé avec vous et d'avoir réalisé ce travail ensemble.

Nous exprimons également notre gratitude à nos familles pour leur soutien continu et à nos amis pour leur présence et leur soutien dans tous les moments de notre vie.

Enfin, nous tenons à remercier toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce travail, ainsi que tous les enseignants et employés de l'ENP pour leur accompagnement tout au long de notre parcours pour devenir ingénieurs.

Table des matières

Liste des figures

Liste des tables

Liste des abréviations

Introduction Générale.....	12
Chapitre 1 : Aperçu générale sur l'entreprise Decathlon El Djazair, Son service Cloud et position de la problématique.	
1.1. Introduction	16
1.2. Présentation de Decathlon	16
1.2.1. Historique et évolution	16
1.2.2. Les valeurs et les missions de Decathlon	19
1.3. Présentation de Decathlon EL Djazair	23
1.3.1. Objectifs de l'entreprise Decathlon El Djazair.....	25
1.3.2. L'organisation de Decathlon EL Djazair.....	25
1.3.3. Le site E-commerce de Decathlon El-Djazair et ses fonctionnalités	29
1.4. Services cloud et la génération et exploitation des données	30
1.4.1. Génération et partage de données.....	31
1.4.2. Utilisation et exploitation des données.....	32
1.5. Stratégie de tarification et la difficulté de déterminer le juste prix	32
1.6. Conclusion.....	33
Chapitre 2 : Introduction à la tarification et à la prédiction des séries temporelle.	
2.1. Introduction	35
2.2. Élasticité-prix et les stratégies de tarification	35
2.2.1. Élasticité-prix	35
2.2.2. Stratégies de tarification.....	35
2.3. Les séries temporelles	37
2.3.1. Les composantes d'une série temporelle.....	38
2.3.2. Stationnarité d'une série temporelle.....	39
2.3.3. La transformation des séries chronologiques	41
2.4. Les modèles de prédiction des séries temporelles.....	42
2.4.1. Apprentissage automatique	42
2.4.2. ARIMA.....	43

2.4.3. Facebook Prophet	44
2.4.4. Random forest	45
2.4.5. Réseaux de neurones	45
2.4.6. Augmentation des données.....	54
2.5. Métriques d'évaluation.....	54
2.5.1. Erreur quadratique moyenne (RMSE).....	54
2.5.2. Erreur absolue moyenne en pourcentage (MAPE).....	55
2.5.3. R^2 et R^2 ajusté	55
2.6. Conclusion.....	55
Chapitre 3 : Collecte, nettoyage et exploration données de ventes.	
3.1. Introduction	57
3.2. Collecte et description des données	57
3.3. Nettoyage des données	61
3.4. Proposition d'une Solution IA.....	73
3.5. Choix d'un produit pour la modélisation prédictive	74
3.5.1. Choix de la marque	74
3.5.2. Choix du magasin.....	75
3.5.3. Choix d'un item.....	76
3.6. Conclusion.....	80
Chapitre 4 : Méthodologie de la résolution de la problématique et présentation des résultats.	
4.1. Introduction	82
4.2. Méthodologie de la résolution de la problématique	82
4.2.1. Évaluation de la stationnarité de la série temporelle.....	82
4.2.2. Sélection des paramètres pour le modèle de prédiction	84
4.2.3. Création des variables retardées	84
4.2.4. Division des données.....	85
4.2.5. Modélisation.....	86
4.3. Présentation et discussion des résultats :	88
4.4. Conclusion:.....	96
Conclusion générale et travaux futurs	97
Références bibliographiques	99

Liste des figures

Figure 1.1 : Graphique à barres de l'évolution du Chiffre d'affaires à l'international	19
Figure 1.2 : Courbe montrant l'évolution du nombre de collaborateurs	21
Figure 1.3 : Quelques statistiques sur Decathlon El Djazair	23
Figure 1.4 : L'organigramme de Decathlon El Djazair	25
Figure 1.5 : Documentation fonctionnelle et technique du stockage de la donnée	32
Figure 2.1 : La décomposition des séries temporelles de l'ensemble de données sur les passagers aériens (compagnies aériennes américaines de 1949 à 1960)	39
Figure 2.2 : Schéma d'un neurone artificiel	46
Figure 2.3 : Réseau de neurones à N couches cachées	47
Figure 2.4 : Fonction sigmoïde	48
Figure 2.5 : Fonction tanh	48
Figure 2.6 : Fonction ReLU	49
Figure 2.7 : Réseaux de neurones denses	50
Figure 3.1 : Schéma simplifié de la base de données	57
Figure 3.2 : Distribution de la variable quantité	62
Figure 3.3 : Matrice montrant la proportion de valeurs manquantes qu'il y a dans chaque colonne	67
Figure 3.4 : Histogramme de la distribution de l'élasticité calculée des produits	70
Figure 3.5 : Histogramme de la distribution de la moyenne globale de l'élasticité ajustée des produits	72
Figure 3.6 : Graphique en barres montrant la quantité vendue pour chaque marque de Decathlon	74
Figure 3.7 : Graphique en barres montrant la quantité vendue pour chaque	75

catégorie

Figure 3.8 : Graphique en barres montrant la quantité vendue pour chaque catégorie répartie par marque	75
Figure 3.9 : Analyse des ventes à Bab Ezzouar en fonction du nombre de magasins Decathlon ouverts durant la même période.	76
Figure 3.10 : Analyse du chiffre d'affaires généré au magasin de Bab Ezzouar en fonction du nombre de magasins de Decathlon durant la même période.	76
Figure 3.11 : Graphique des ventes mensuelles du stock pour la marque Kalenji au magasin de Bab Ezzouar.	77
Figure 3.12 : Répartition des quantités par genre et couleur (code modèle)	77
Figure 3.13 : Graphique circulaire montrant la répartition des ventes pour les 5 modèles	79
Figure 3.14 : Graphique circulaire montrant la répartition des ventes pour les 5 “items”	78
Figure 3.15 : Graphique en barres montrant les quantités vendues pour chaque prix de vente pour “item” 2005497	79
Figure 3.16 : Graphique en ligne montrant l'évolution de l'état du stock pour “item” 2005497 au magasin de Bab Ezzouar	79
Figure 3.17 : Décomposition de la série temporelle en tendance, saisonnalité et composantes irrégulières pour une période de 7 jours	79
Figure 4.1 : L'évolution des ventes après application des transformations	83
Figure 4.2 : Matrice des corrélations	84
Figure 4.3 : Illustration explicative de la création des variables retardées	85
Figure 4.4 : Graphiques linéaire de l'évolution des ventes séparées entre les ensembles d'entraînement et de test	86
Figure 4.5 : Schéma illustrant la génération des séquences	87
Figure 4.6 : Architecture du modèle LSTM utilisé dans l'étude	87
Figure 4.7 : La perte avant l'augmentation	
Figure 4.8 : La perte après l'augmentation	88

Figure 4.9 : La perte après la transformation Log	
Figure 4.10 : Exemple d'un surajustement	88
Figure 4.11 : Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle Random forest avant l'augmentation des données	90
Figure 4.12 : Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle ARIMA avant l'augmentation des données	90
Figure 4.13 : Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle Fb Prophet avant l'augmentation des données	91
Figure 4.14 : Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle LSTM avant l'augmentation des données	91
Figure 4.15 : Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle Random forest après l'augmentation des données	92
Figure 4.16 : Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle ARIMA après l'augmentation des données	92
Figure 4.17 : Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle Fb Prophet après l'augmentation des données	93
Figure 4.18 : Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle LSTM après l'augmentation des données	93
Figure 4.19 : Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle Random Forest après la transformation log	94
Figure 4.20 : Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle ARIMA après la transformation log	94
Figure 4.21 : Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle FB Prophet après la transformation log	95
Figure 4.22 : Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle LSTM après la transformation log	95
Figure 4.23 : L'évolution des ventes prédites en fonctions de plusieurs prix donnés aux modèles	96

Liste des tables

Table 1.1 : Fiche signalétique de Decathlon El-Djazair	24
Table 3.1 : Description des colonnes de la table de données transaction	58
Table 3.2 : Description des colonnes de la table de données des produits	59
Table 3.3 : Description des colonnes de la table de données du stock	60
Table 3.4 : État de l'ensemble de données avant le traitement.	61
Table 3.5 : État l'ensemble de données après l'élimination du B2B.	61
Table 3.6 : État de l'ensemble de données après l'élimination des transactions avec une quantité nulle.	62
Table 3.7 : État de l'ensemble de données après l'élimination des transaction "Seconde vie".	63
Table 3.8 : État de l'ensemble de données après l'élimination des produit peu vendu.	63
Table 3.9 : État de l'ensemble de données après l'élimination des produits sans code.	64
Table 3.10 : État de l'ensemble de données après le groupement par jour.	64
Table 3.11 : État de l'ensemble de données après le traitement des dates manquantes.	65
Table 3.12 : État de l'ensemble de données après la jointure avec le stock.	65
Table 3.13 : État de l'ensemble de données après l'ajout des variables.	66
Table 3.14 : Calcul des quantités moyennes vendues par jours et par magasin	69
Table 3.15 : Tableau représentatif du calcul de la moyenne supérieure et inférieure des prix unitaire et quantité	69
Table 3.16 : Tableau présentant les quantités vendues dans chaque magasin pour le mois de mars, pour plusieurs années différentes	71
Table 3.17 : Tableau présentant pour chaque magasin l'élasticité calculée	72
Table 4.1 : Tableau comparatif des résultats du R carré et R carré ajusté.	89

Liste des abréviations

IA	Intelligence artificielle
ARIMA	AutoRegressive Integrated Moving Average
ANN	Resaux de neurones artificiels
RNN	Réseaux de neurones récurrents
LSTM	Long-short term memory
MAPE	Erreur absolue moyenne en pourcentage
RMSE	Erreur quadratique moyenne
ReLu	La fonction d'activation rectified linear unit
ADF	Le test Dickey-fuller augmenté
B2B	Business to Business

Introduction Générale

Déterminer le juste prix d'un produit ou d'un service devient de plus en plus complexe et difficile pour les entreprises. Les décisions en matière de prix ne sont plus uniquement fondées sur le coût de production ou des matériaux, mais également sur des facteurs tels que la concurrence, la demande des clients et les tendances du marché. En outre, les entreprises sont censées apporter de la valeur à leurs clients tout en maximisant leurs profits et en garantissant leur satisfaction. Les stratégies de tarification doivent donc être élaborées sur la base d'une compréhension approfondie de ces facteurs.

La difficulté consiste à trouver l'équilibre entre la maximisation des profits et la fidélisation des clients. Fixer un prix trop élevé pour un produit peut entraîner une baisse des ventes et faire fuir les clients vers les concurrents, tandis qu'un prix trop bas peut entraîner une perte de bénéfices potentiels.

Pour relever ce défi, les entreprises peuvent s'appuyer sur des stratégies de tarification fondées sur des données qui prennent en compte des facteurs tels que l'offre et la demande, la concurrence et le comportement des clients. En mettant en œuvre des stratégies de tarification dynamique qui ajustent les prix en temps réel en fonction de l'évolution de la demande et d'autres facteurs du marché, les entreprises peuvent trouver le point de prix optimal qui maximise les bénéfices tout en maintenant la satisfaction des clients.

Les outils d'analyse avancée tels que ARIMA et les réseaux de neurones sont désormais couramment utilisés pour prévoir la demande et identifier des modèles dans le comportement des clients et les décisions d'achat. Cela peut aider les entreprises à mieux comprendre les facteurs qui influencent la volonté de payer des clients et à prendre des décisions plus éclairées en matière de tarification.

Notre projet de fin d'étude est centrée sur l'étude de l'effet du prix sur les ventes et de son impact sur la rentabilité globale de l'entreprise. Grâce à l'utilisation de techniques avancées d'analyse et d'apprentissage automatique, nous proposons un modèle de tarification qui peut aider l'entreprise à prendre des décisions plus éclairées en matière de tarification. En tirant parti de la puissance des données et de l'analyse. Les entreprises peuvent ainsi optimiser leurs stratégies de tarification et trouver le prix optimal pour leurs produits.

Ce travail est composé de quatre chapitres :

- Le chapitre 1 présente l'entreprise Decathlon, ses valeurs, sa mission, ainsi que sa filiale Decathlon El Djazair et son site e-commerce. Nous y aborderons également les différents services cloud utilisés pour l'analyse de données et leur stockage. Le chapitre aborde également la stratégie de tarification adoptée par Decathlon et les enjeux liés à la détermination du juste prix.
- Le chapitre 2 met l'accent sur les concepts clés liés à notre recherche, à savoir l'élasticité et les stratégies de tarification. Nous y définissons également les techniques d'apprentissage automatique que nous avons utilisées dans notre étude, afin de permettre au lecteur de mieux comprendre notre démarche.

- Le chapitre 3 présente les données utilisées, les différents traitements appliqués, l'évaluation de l'élasticité pour chaque produit, une introduction sur la solution d'apprentissage automatique, ainsi que la justification du choix du produit sur lequel nous avons appliqué le modèle prédictif.
- Le chapitre 4 décrit les étapes qui vont de l'évaluation de la stationnarité jusqu'à la modélisation des séries temporelles, ainsi que la présentation des résultats de l'évaluation des modèles. En plus de cela, une simulation de prédiction a été effectuée pour démontrer l'efficacité de la solution proposée pour la prédiction de nouveaux prix de vente.

Chapitre 1 :
Aperçu générale sur l'entreprise
Decathlon El Djazair, Son service
Cloud et position de la
problématique.

1.1. Introduction

Le chapitre 1 offre une présentation complète de l'entreprise Decathlon, en commençant par son histoire, ses valeurs et sa mission. Nous nous concentrons ensuite sur Decathlon El Djazair, en décrivant ses objectifs, son organisation et son site e-commerce. Nous abordons également le service cloud, en expliquant la génération et l'utilisation des données pour aider à soutenir les performances commerciales de l'entreprise. Enfin, nous examinons la stratégie de tarification adoptée par Decathlon et la problématique qu'elle soulève en termes de détermination du juste prix.

1.2. Présentation de Decathlon

L'entreprise multinationale Decathlon (ou Oxyane de 2008 à 2014) est une enseigne de droit français, qui est active dans la conception, la fabrication et la grande distribution d'articles de sports et de loisirs,

Decathlon est présente dans 72 pays différents, avec plus de 1 751 magasins et met à la disposition de ses clients et utilisateurs plusieurs articles et services liés aux activités sportives et de loisirs, notamment :

- Équipements de sports collectifs et individuels dont le cyclisme, aquatiques (natation, plongée et canoë), le ski, la randonnée et l'escalade,
- Machines de sport (fitness & cardio),
- Click & collect (retrait en magasin),
- Livraison à domicile,
- Trocathlon : reprise d'articles pour une seconde vie,
- Decathlon location d'équipements.

1.2.1. Historique et évolution

Decathlon, chaîne de magasins de vêtements de sport, a été fondée par Michel Leclercq le 27 juillet 1976. Sa vocation première a toujours été la conception, la fabrication et la vente d'articles de sport. Le siège social de l'entreprise est situé à Villeneuve d'Ascq, en France.

Au fil des années, Decathlon a connu d'importants changements et une forte expansion, en diversifiant ses activités, en créant de nouvelles filiales, en se développant à l'international et en renforçant constamment son réseau de distribution par l'ouverture de nouveaux magasins. Les principales étapes de son évolution sont résumées ci-dessous :

En 1986, l'entreprise crée Decathlon Production, chargée de la conception et de la fabrication des produits de la marque Decathlon. La même année, Decathlon ouvre son premier magasin hors de France, à Dortmund, en Allemagne.

En 1988, la société étend ses activités de production en ouvrant son premier site de production en Asie. Cela a conduit à la création de deux filiales en 1996 : Tribord pour les sports nautiques et Quechua pour les sports de montagne.

À partir de 1999, Decathlon commence à ouvrir des magasins aux États-Unis et au Royaume-Uni, respectivement à Boston et à Londres. En 2003, le développement international de Decathlon prend une nouvelle dimension avec l'ouverture de son premier magasin en Chine, à Shanghai. La même année, l'entreprise s'implante à Domancy, dans le Mont-Blanc, en France.

En 2004, Tribord, la marque spécialisée dans les sports nautiques, s'installe à Hendaye, au bord de l'océan Atlantique, pour être encore plus proche de ses utilisateurs et concevoir des produits adaptés à leurs besoins. Cette année-là, l'entreprise crée également deux nouvelles marques : Aptonia et Géonaute, dédiées aux produits et services pour le bien-être des sportifs aventuriers. En 2005, Kalenji (course à pied) et Fouganza (matériel d'équitation) viennent compléter la liste des marques de Decathlon.

En 2006, l'enseigne décide de fermer quatre magasins en raison d'une forte concurrence aux États-Unis, marché sur lequel Decathlon n'a pas su se différencier. Cependant, la même année, elle ouvre son premier magasin en Russie, à Moscou, et crée Artengo, une marque spécialisée dans l'équipement de tennis en Russie.

En 2008, le groupe Decathlon décide de changer de nom pour devenir Oxylane et adopte une nouvelle organisation plus verticale avec des filiales, le nom Decathlon ne reflétant plus la diversité des activités développées par l'entreprise. Cette évolution montre que, malgré les défis, Decathlon est resté fidèle à son objectif initial : proposer des articles de sport de qualité pour tous les types d'amateurs de sport.

En 2009, la marque de golf Inesis a installé son siège international à Marcq-en-Barœul dans le département du Nord. De même, la filiale Kipsta (sports collectifs) s'est installée dans la zone de l'Union à Tourcoing. Oxylane a installé son centre logistique international sur la plate-forme multimodale et logistique française Delta 3 à Dourges dans le Pas-de-Calais. C'est à la fois le siège international des équipes logistiques et un entrepôt de 51 000 m² où l'innovation dans les procédures de réception et d'expédition joue un rôle important. Enfin, cette année, le premier magasin Decathlon a été ouvert en Roumanie.

En 2010, une nouvelle structure, le B'Twin Village, a vu le jour à Lille. Ce site, entièrement dédié au vélo, regroupe les équipes de l'enseigne B'Twin, avec une usine d'assemblage et un magasin pour les passionnés de vélo. Decathlon s'est également implanté dans deux nouveaux pays : la République Tchèque et la Turquie.

En 2011, la marque FLX, spécialisée dans le cricket, a été créée pour l'Inde, ainsi que pour l'Australie et le Royaume-Uni. Le groupe a également pris une participation dans le magasin "Au vieux plongeur", spécialisé dans la vente de matériel neuf et d'occasion pour la plongée, la chasse sous-marine et la natation.

En 2014, le groupe reprend son ancienne dénomination "Decathlon". En 2015, Michel Aballea devient président-directeur général du groupe Decathlon. La même année, Decathlon ouvre 140 magasins à l'étranger, dont 51 en Chine.

En janvier 2016, la marque a testé un nouveau concept appelé "Osons" dans son magasin précurseur d'Englos. L'objectif était d'avoir zéro stock. Cette stratégie logistique et de production vise à minimiser la quantité de stock dans l'entrepôt afin de réduire les coûts d'exploitation et de n'avoir aucun produit immobilisé qui ne soit pas déjà réservé pour une commande.

En septembre 2019, Decathlon a acquis une partie du capital d'Alltricks, une société spécialisée dans le matériel de cyclisme, de course à pied et de triathlon sur internet, fondée par Gary Anssens en 2008. La même année, le magasin Decathlon d'Alger est ouvert.

Début février 2020, Decathlon a annoncé une augmentation de ses ventes mondiales de plus de 9%, portant son chiffre d'affaires à plus de 12,4 milliards d'euros en 2019. Le succès de Decathlon, qui opère dans 69 pays et couvre 87 sports différents, s'articule aujourd'hui autour d'une focalisation sur ses propres marques au détriment des grandes marques de sport, avec un ratio d'environ 80%/20%.

Lors de la pandémie de COVID-19 en 2019-2020, l'entreprise a retiré de la vente son masque de plongée Subea Easybreath pour en faire don au personnel soignant confronté à une pénurie de masques de protection et d'assistance respiratoire ; les professionnels de la santé en Italie et en France ont pu les adapter pour les utiliser avec des respirateurs.

En février 2021, Decathlon a lancé "Decathlon Experience", une plateforme de réservation de séjours sportifs. Cette plateforme permet de réserver l'hébergement, les activités sportives liées à la montagne (forfaits de ski), l'équipement sportif, le transport et l'assurance. Dans un premier temps, Decathlon Expérience se concentrera sur les sports de montagne, mais devrait à terme couvrir une soixantaine de sports différents.

Parallèlement, Decathlon a dévoilé le premier prototype d'un masque de protection spécialement conçu pour le sport. Ce masque a été conçu pour répondre aux besoins des athlètes et des sportifs qui souhaitent s'entraîner en toute sécurité dans les salles de sport qui ont rouvert après les fermetures liées à la pandémie de Covid-19.

En 2022, Decathlon a connu une croissance significative de son chiffre d'affaires, atteignant 15,4 milliards d'euros, soit une augmentation de 12% par rapport à l'année précédente. Dans le même temps, Decathlon a réduit ses émissions de carbone absolues de 1,5% et a augmenté la part de ses produits éco-conçus, qui représentent 23% de son chiffre d'affaires global en 2022. (voir figure 1.1)

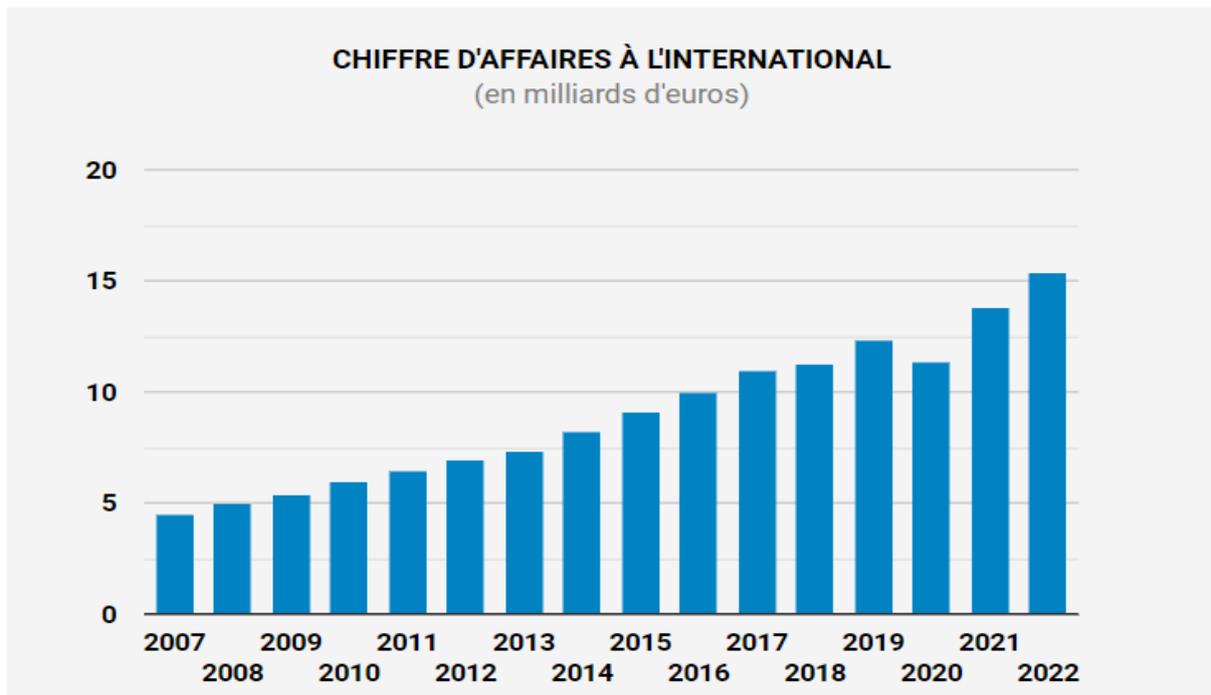


Figure 1.1 : Graphique à barres de l'évolution du Chiffre d'affaires à l'international.

1.2.2. Les valeurs et les missions de Decathlon

Le management de Decathlon se base sur le développement et l'enrichissement de l'ensemble de ses employés à travers ses valeurs fondamentales de vitalité, sincérité, responsabilité et générosité, créant ainsi une culture d'entreprise forte et distinctive, comme c'est le cas dans le monde entier.

- **Vitalité** : Autre synonyme, c'est avoir la pêche ! La vitalité, c'est la vie, l'activité intense, l'énergie, l'entrain, la vigueur. Les collaborateurs chez Decathlon sont pleins de vitalité, car ils sont avant tout positifs et pleins d'énergie. Ils sont enthousiastes, aiment l'innovation et la création, cherchent sans relâche à progresser et à faire évoluer les choses. Ils aiment l'action comme un sportif.
- **Sincérité** : C'est la transparence à travers leurs actes, entre collègues, avec leurs clients et leurs fournisseurs. Ils veulent beaucoup de cohérence entre ce qu'ils expriment et ce qu'ils réalisent.
- **Responsabilité** : Être responsable, cela consiste à prendre des décisions et à agir pour leur mise en œuvre. Les collaborateurs chez Decathlon assument le poids de leurs décisions vis-à-vis de leur équipe et de leurs clients. C'est également leur responsabilité d'anticiper les défis et enjeux sociétaux pour orienter leurs actions en faveur du développement durable, ainsi que de garantir la sécurité de leurs clients et de leurs équipes à travers le monde. Dans le management, la générosité des actions est associée à une exigence forte, qui est un signe de responsabilité. Leur principal objectif est de "créer l'envie et de rendre accessible au plus grand nombre le plaisir et les bienfaits du sport".

L'objectif de Decathlon est d'équilibrer progressivement la balance commerciale de chacun des pays où il est présent, tant au niveau de la production que de la distribution. Cette expansion internationale témoigne de leur volonté de globalisation, avec l'ouverture prochaine de nouveaux magasins pour les athlètes dans les grandes villes européennes et au-delà. La force de Décathlon, ce sont des hommes et des femmes de toutes nationalités qui partagent les mêmes objectifs. En tant que sportifs passionnés, ils partagent naturellement les mêmes objectifs. L'objectif premier de Decathlon n'est pas de vendre, mais de faire en sorte que le client "sportif satisfait" revienne.

- **Rendre le sport accessible au plus grand nombre :**

La mission première de Decathlon est de rendre les bienfaits et le plaisir du sport accessibles au plus grand nombre, durablement, et c'est pourquoi l'innovation est présente à chaque étape du processus de conception pour créer des produits exclusifs au service du bonheur des utilisateurs. Chez Decathlon, l'innovation est pragmatique : rendre le sport accessible à tous, quel que soit le niveau ou la discipline. Les équipes placent le sportif au centre du processus et se concentrent sur les innovations d'usage. Elles innovent par des produits opérationnels, astucieux, pratiques et uniques.

- **Placer l'homme au cœur du développement :**

Le bien-être des collaborateurs est au cœur des préoccupations de Decathlon. L'entreprise s'attache à déployer une politique sociale et humaine en adéquation avec ses valeurs. Decathlon est convaincu que des collaborateurs heureux et engagés contribuent à la performance de l'entreprise sur le long terme.

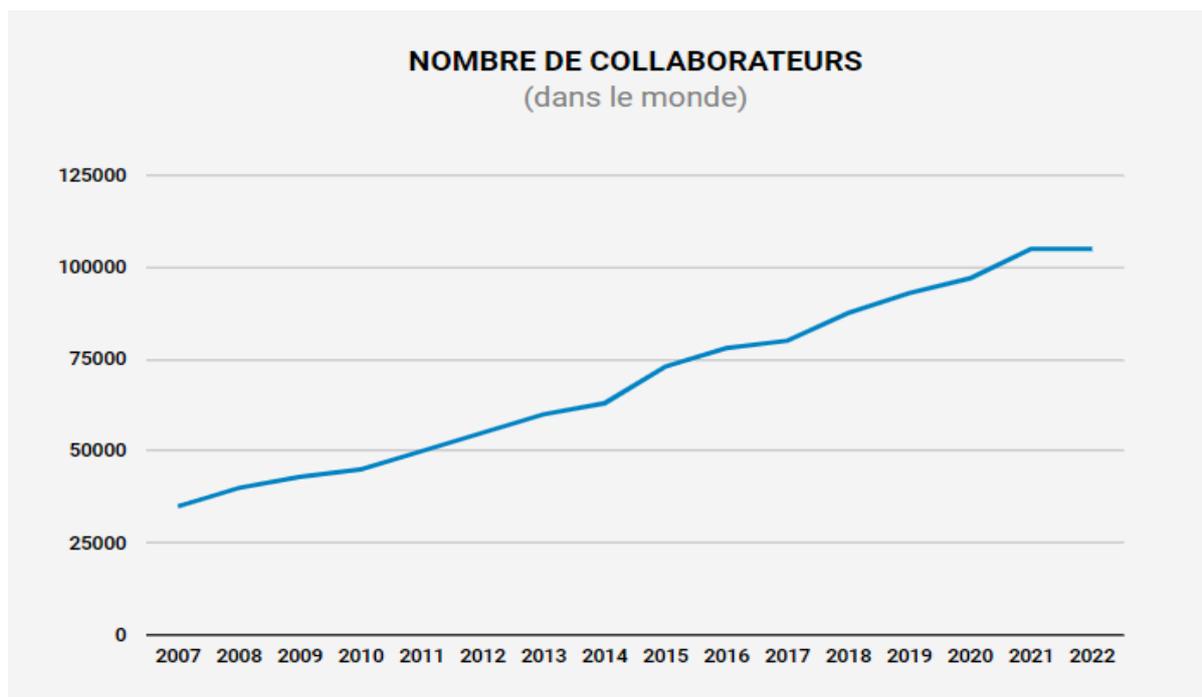


Figure 1.2: Courbe montrant l'évolution du nombre de collaborateurs.

- **Responsabiliser :**

Nous vivons dans un monde devenu " volatile, incertain, complexe et ambigu " (selon l'acronyme anglais VUCA), qui connaît la digitalisation, la mondialisation, le progrès d'une part, mais aussi l'épuisement des ressources naturelles et le creusement des inégalités entre les personnes d'autre part. Face à cela, les entreprises doivent être agiles, rapides et efficaces. Elles ont besoin de collaborateurs qui osent, entreprennent et prennent des décisions qui font sens localement avec leur environnement immédiat.

- **Développer les compétences et les talents :**

La formation continue des salariés est essentielle pour maintenir un niveau de compétences adapté au développement de toutes les stratégies de l'entreprise. Elle permet aux collaborateurs de poursuivre leur développement professionnel, tant sur le plan des compétences liées à leur métier que sur le plan du développement personnel, en les aidant à s'adapter aux nombreuses évolutions de leur environnement de travail.

Afin d'accompagner chaque collaborateur vers plus d'autonomie tout en lui proposant de nouveaux métiers et l'acquisition de nouvelles compétences, Decathlon veille à ce que chacun puisse trouver les ressources nécessaires à tout moment de son parcours professionnel.

Grâce à l'évolution du mode de transmission des savoirs, notamment en structurant le rôle du référent de l'entreprise, la transition vers une posture de partage de connaissances et d'expériences concrètes est encouragée, tout en mettant l'accent sur l'auto-évaluation et la transformation digitale. Cette méthode d'apprentissage favorise l'autonomie des

collègues en mettant l'accent sur l'importance du développement de leurs compétences et de leur épanouissement personnel, dans un cadre agile et accessible.

- **Intégrer l'innovation à tous les niveaux :**

La stratégie de Decathlon a toujours été de placer l'innovation au cœur de son activité pour offrir des solutions sportives adaptées à tous et se réinventer en permanence. En 2018, face aux bouleversements mondiaux qui remettent en question les modes de consommation, l'enjeu est encore plus important. Decathlon s'engage donc à intégrer l'innovation à tous les niveaux de l'entreprise pour continuer à répondre aux besoins des consommateurs.

- **Encourager chacun à agir à son niveau pour le développement durable :**

Chez Decathlon, il est important d'encourager chaque coéquipier à devenir acteur du développement durable en agissant chaque jour selon ses convictions. Des actions concrètes sont mises en place pour favoriser cette prise de conscience et permettre à chacun de contribuer à un avenir plus durable.

- **Proposer des offres utiles et écologiques :**

Le monde est confronté à des défis environnementaux majeurs qui résultent directement de l'exploitation des ressources naturelles de la planète. Dans ce contexte mondial instable, Decathlon s'engage dans une transition économique, écologique et sociétale visant à promouvoir de nouveaux modes de consommation.

Conscient de sa responsabilité face à ces enjeux, Decathlon explore de nouveaux modes de fonctionnement. En intégrant les concepts d'économie circulaire et de fonctionnalité dans son modèle économique, Decathlon valorise les efforts fournis par l'ensemble de ses collaborateurs impliqués dans plusieurs initiatives volontaires : plateformes favorisant le réemploi des produits sportifs, ateliers favorisant la réparabilité et la durabilité des produits, événements autour des produits d'occasion ou des dons.

Ces initiatives locales permettent à Decathlon de présenter de nouvelles options à ses clients sportifs : intégrer la réparabilité dès la conception, promouvoir l'apprentissage de l'entretien des produits sportifs, proposer des produits de seconde vie qui répondent aux besoins de clients, ou des articles de sport fabriqués à partir de matériaux recyclés.

1.3. Présentation de Decathlon EL Djazair

Déjà implanté au Maghreb, avec des magasins au Maroc et en Tunisie, il était naturel que Decathlon vienne servir les sportifs algériens. Samedi 29 juin 2019 marque en effet l'ouverture du premier magasin Decathlon à Alger, au centre commercial de Bab Ezzouar. Un premier lieu de contact et de découverte entre sportifs passionnés.

Les facteurs déterminants et constitutifs de Decathlon El-Djazair mais aussi Decathlon monde sont sa vision, ses valeurs et son cadre de jeu. De plus, l'entreprise se démarque de ses concurrents par ses marques « in-house », qu'ils qualifient de « marques passion ». Ces dernières, au nombre de 40, couvrent la quasi-totalité des sports comme Tribord pour les sports nautiques et aquatiques ou Wed'ze pour le ski et le snowboard, ces marques sont propres à Decathlon. Elles mettent à disposition 14 000 références pour 55 disciplines sportives comme le running, le cyclisme, les sports d'eau et les sports collectifs, tels que le football et le basketball.

À l'instar des services proposés, ces produits sont techniques, malins, innovants et aux meilleurs prix. Chaque pratiquant peut ainsi tester et s'équiper quel que soit son sport et son niveau tout en bénéficiant des conseils avisés d'un des collaborateurs, pratiquant et passionné qui ont un sens inné de la transmission et de l'accompagnement. Avec ce magasin, son ADN et la passion du sport, Decathlon El-Djazair entend accompagner celles et ceux qui souhaitent se mettre et se remettre au sport, dans la superbe et dynamique ville d'Alger.



Figure 1.3: Quelques statistiques sur Decathlon El Djazair.

Neuf mois après son ouverture, Decathlon El Djazair a lancé sa plateforme de commerce électronique le 23 mars 2020, en réponse à la fermeture de son magasin physique due aux mesures de confinement liées à la pandémie de COVID-19. L'entreprise a cherché à répondre aux besoins de ses clients en leur offrant un service à domicile grâce à cette plateforme, un service encore peu développé en Algérie.

Le 17 septembre 2020, Decathlon a inauguré son deuxième magasin à Oran, la deuxième ville la plus peuplée d'Algérie. Le choix de cette ville portuaire et cosmopolite est stratégique. Au premier semestre 2021, Decathlon se dote d'un entrepôt et met en place le projet de Centre d'Approvisionnement Régional (CAR), afin de mieux gérer son stock et de satisfaire ses clients. Le système informatisé permet une gestion automatisée des marchandises et un réapprovisionnement plus efficace.

Le 25 février 2021, Decathlon a poursuivi son expansion en ouvrant un magasin à El Mohamadia, dans le centre commercial Ardis à Alger. L'emplacement accessible et l'air frais ont attiré de nombreuses familles et sportifs. En avril 2021, Décathlon a lancé une première production locale en Algérie baptisée "MADE IN ALGERIA". Cette production s'inscrit dans la stratégie de développement de la production locale dans tous les pays où Decathlon est implanté. Ce projet a démontré l'engagement de Decathlon envers l'environnement et sa participation à l'économie nationale.

Le 1er juillet 2021, Decathlon a ouvert son quatrième magasin à Constantine, une ville remarquable par sa beauté et son histoire. Les sportifs de l'est du pays ont montré un engouement remarquable pour les produits depuis l'ouverture. Le 4 mars 2022, Decathlon a ouvert un magasin à Chéraga, dans l'incontournable Garden City Lifestyle Mall, afin que tous les habitants de la capitale et de sa banlieue aient accès aux magasins sans avoir à traverser la ville.

Le 20 avril 2022, Decathlon a ouvert son sixième magasin à Blida, grande ville commerciale et industrielle. L'Atlas blidéen, qui surplombe la ville, abrite l'une des premières pistes de ski découvertes en Algérie, attirant de nombreuses familles et sportifs. Aujourd'hui, Decathlon El Djazair dispose de six magasins, d'un entrepôt intelligent et d'un service de commerce électronique qui couvre l'ensemble du pays. Le site marchand et l'application mobile permettent le paiement électronique. Avec plus de 450 collaborateurs jeunes, dynamiques et passionnés, Decathlon El Djazair œuvre chaque jour pour rendre accessible au plus grand nombre le plaisir et les bienfaits du sport.

Table 1.1 : Fiche signalétique de Decathlon El-Djazair .

<i>Forme juridique</i>	Decathlon El-Djazair est une société à responsabilité limitée (SARL)
<i>Capital social</i>	350 200 000,00 DZD
<i>Chiffre d'affaire (Decathlon monde)</i>	15,4 milliards d'euros (2022)
<i>Dénomination sociale</i>	Decathlon El-Djazair
<i>Adresse du siège social</i>	Lot 43, N°, Quartier des affaires d'Alger, Centre commercial, 4ème étage, tour Nord N°420, Bab Ezzouar, Alger, Algérie

1.3.1. Objectifs de l'entreprise Decathlon El Djazair

Decathlon El Djazair comme toute entreprise, est à la poursuite d'objectifs tels que :

- Contribuer à la pratique sportive, dont le cyclisme, la randonnée et les activités aquatiques en Algérie (mettre les Algériens au sport).
- L'amélioration du choix de gamme.
- Être présent dans tout le pays.
- L'amélioration de la disponibilité des produits.
- La valorisation des ressources humaines.
- L'augmentation du volume de production.
- Diversification dans la production.
- Le renforcement de la sécurité des infrastructures.

1.3.2. L'organisation de Decathlon EL Djazair

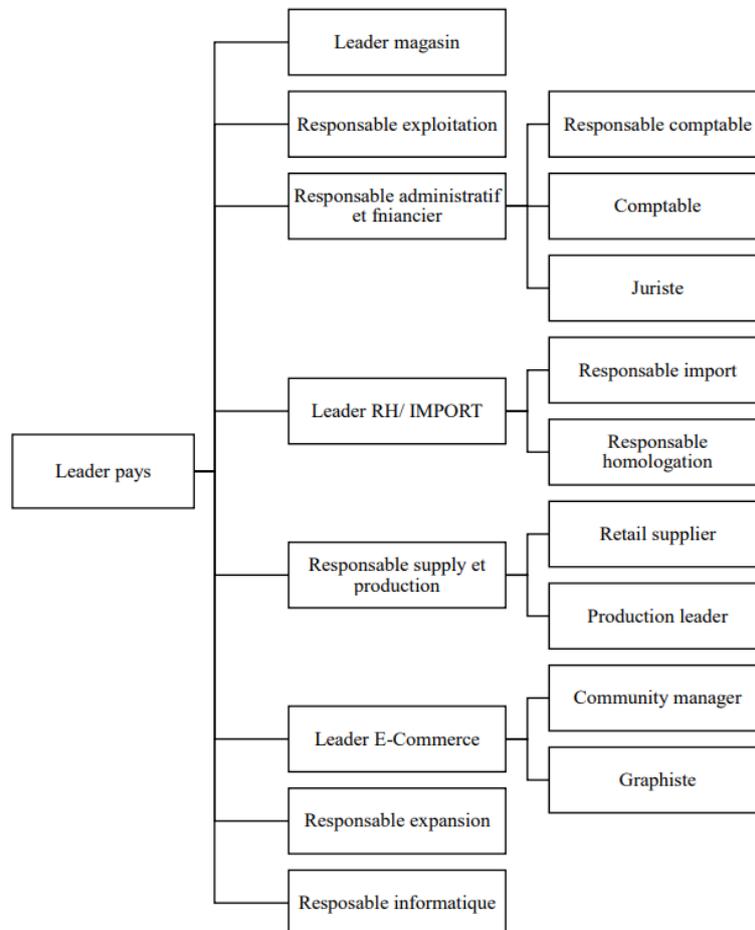


Figure 1.4: L'organigramme de Decathlon El-Djazair.

Decathlon El-Djazair comme illustrée dans l'organigramme cité ci-haut, est hiérarchisée suivant différentes structures dirigées par un leader pays qui veille à la sécurité et la gestion optimale de ses ressources

La décomposition de la structure organisationnelle de Decathlon El-Djazair nous donne les distinctions suivantes :

- **Leader pays** : Ses missions principales sont les suivantes :
 - Diriger l'entreprise conformément à la vision définie pour l'entreprise par le conseil d'administration.
 - Partenariat avec des dirigeants de haut niveau pour développer l'entreprise, la renforcer et assurer sa pérennité.
 - Travailler en étroite collaboration avec le directeur financier pour préparer les budgets annuels, compléter l'analyse des risques sur les investissements potentiels, et conseiller le conseil d'administration en ce qui concerne le risque d'investissement et de retour.
 - Travailler en étroite collaboration avec les Ressources humaines (RH) en ce qui concerne les pratiques d'embauche, la masse salariale et le déboursement des prestations.
 - Il se charge aussi de :
 - Déterminer les objectifs de l'entreprise
 - Affiner la stratégie pour les atteindre
 - Recruter
 - Licencier
 - Déléguer
 - Connaître et respecter la législation
 - Donner son avis à tous les niveaux
 - Prendre des décisions
 - Représenter l'image de l'entreprise

- **Leader magasin** : ses missions :
 - Élaborer et animer avec son équipe un projet à la fois commercial et humain pour son magasin.
 - Recruter, former et manager son équipe magasin constituée de 30 à 150 personnes
 - Cerner, mobiliser et développer les talents de chacun
 - Favoriser l'évolution de son équipe au sein du groupe Decathlon
 - Être le garant des performances économiques et humaines de son site
 - Prendre des décisions en totale autonomie

- Contribuer à ancrer le magasin dans son environnement local
- **Responsable exploitation:** Le responsable exploitation de Decathlon El-Djazair a pour mission de garantir la sécurité et la sûreté des personnes et des biens, tout en assurant une exploitation optimale de l'activité économique et humaine. Il est responsable de plusieurs tâches clés, telles que :
 - Rendre autonome les équipes sur l'exploitation (Sécurité-sûreté / exploitation journalière / maintenance...)
 - Piloter l'activité économique et humaine, animer la disponibilité et de la fiabilité des stocks.
 - Ses autres missions :
 - Assurer la maintenance, anticiper et garantir la tenue des frais .
 - Assurer une exploitation magasin et une image au service du commerce .
 - Contribuer aux résultats et aux ratios économiques des rayons et du magasin.
- **Directeur administratif et financier :** Le Directeur administratif et financier joue un rôle clé chez Decathlon El-Djazair, en assurant la gestion financière de l'ensemble des activités de l'entreprise. La finance est présente à chaque étape de la vie d'un produit, depuis la conception jusqu'à la vente en magasin, et les coéquipiers de Decathlon El-Djazair possèdent de diverses compétences pour soutenir cette fonction.

Le Directeur administratif et financier est responsable de la comptabilité, du contrôle de gestion, de la trésorerie, du juridique, de la fiscalité et de l'informatique financière. Ses missions principales sont les suivantes :

- Assurer la fiabilité des données comptables conformément aux réglementations fiscales, comptables et sociales.
- Gérer le financement des magasins.
- Veiller au respect des procédures internes.
- Analyser les flux de marchandises pour garantir la disponibilité des produits.
- Assurer la clôture mensuelle et la révision des comptes des magasins.
- Optimiser le traitement comptable des frais généraux en magasin.
- Contribuer à améliorer durablement les performances de l'activité et de l'organisation du service DAF.

- **Leader ressources humaines : ses missions :**
 - Assurer une gestion de la paie juste et fiable tout en respectant les réglementations en vigueur.
 - Mettre en place et accompagner la mise en œuvre de la politique RH interne.
 - Former, animer et accompagner les opérationnels des magasins pour qu'ils deviennent autonomes dans la gestion des ressources humaines.

- **Leader Import :** Le responsable import gère les commandes et participe à faire grandir le portefeuille de Decathlon El-Djazair. De ce fait, il étudie les besoins de l'entreprise, négocie les contrats, et réalise des propositions commerciales. Il est aussi amené à animer une équipe et à échanger régulièrement avec les partenaires et fournisseurs. Prospections, propositions commerciales, gestion de stocks, rédaction de cahiers des charges, négociation de contrats de ventes et d'achats, entretien des relations internationales clients et fournisseurs pour une fidélisation optimale, respect des règles du commerce international, gestion des commandes internes et externes et respect des délais, évaluation des produits et marchandises, gestion des contentieux, les missions ne manquent pas.

- **Responsable supply et production :** En ce qui concerne les missions de responsable supply :
 - Il est responsable de l'approvisionnement du magasin Decathlon El-Djazair. En toute autonomie, il choisit et mobilise son équipe pour garantir la réalisation de ses objectifs.
 - Il anime le fournisseur pour garantir les meilleures performances Coût/ Qualité / Délai des produits finis ou composants dont il a la responsabilité.
 - Il s'assure que les produits soient disponibles au bon moment, au bon endroit et au meilleur coût.
 - Achats : Il déploie la politique d'achats mondiale de son périmètre.
 - Amélioration continue : il soutient et anime les démarches d'amélioration continue des partenaires industriels de son périmètre.

- **Leader E-commerce :**ses missions :
 - Rendre le sport accessible au plus grand nombre en mettant à disposition des utilisateurs des applications informatiques de qualité.
 - Il fait évoluer ses solutions en collaboration avec ses utilisateurs et met en production les adaptations fonctionnelles et techniques.
 - Il a la charge d'animer des ressources pour garantir la qualité et la fiabilité de

ses applications et d'analyser les récurrences des incidents pour assurer la meilleure expérience utilisateur.

- Il communique sur des nouvelles solutions, accompagne et forme ses utilisateurs.
- Il gère le site E-commerce de Decathlon El-Djazair, ainsi que les commandes des clients et leurs réclamations.
- **Responsable expansion** : Le responsable expansion chez Decathlon El-Djazair favorise le développement commercial de l'entreprise. Pour atteindre cet objectif, il élabore plusieurs stratégies de développement et de dynamisation du réseau. Ce réseau, constitué de points de vente, est la sève nourricière de l'entreprise. Le chargé d'expansion analyse sa performance et corrige ses défaillances. Il conseille, donne des directives aux animateurs de réseau pour un regain de dynamisme. De plus, il recherche en permanence de nouveaux espaces pour l'installation de points de vente de l'entreprise.

Toutefois, il ne se contente pas uniquement de créer ces espaces de vente. Il se livre d'abord à une enquête approfondie sur le terrain grâce aux études de marché.

- **Responsable IT** : Le responsable de service informatique a la charge d'une unité informatique. Son rôle :
 - régler les problèmes à moindre coût et dans les plus brefs délais.
 - proposer des solutions informatiques améliorant la rentabilité de l'entreprise.
 - Il s'assure que son service atteigne les objectifs fixés par la direction des systèmes d'information. Pour cela, il planifie les activités, anime son équipe et met en place des projets d'évolution en fonction des besoins des utilisateurs.

Decathlon EL djazair est organisée en trois principales entités :

- **Magasins** : La principale activité de l'entreprise est centralisée dans les magasins.
- **Entrepôt (CAR)** : Stockage de toutes les références produits.
- **Administration** : gère l'aspect stratégique, juridique de l'entreprise et les différentes opérations de paiement, d'importation.

1.3.3. Le site E-commerce de Decathlon El-Djazair et ses fonctionnalités

Pour renforcer sa présence en ligne, Decathlon El-Djazair a lancé son site e-commerce, www.Decathlon.com.dz, ainsi que le service de livraison sur la wilaya d'Alger, le 23 mars 2020, en réponse à la situation sanitaire de 2019/2020. Le site offre une variété d'univers sportifs, avec des fiches-conseils et des guides pratiques rédigés en interne. Chaque article présenté sur le site dispose de sa propre "fiche produit", qui fournit toutes les informations nécessaires aux clients,

ainsi que des images pour compenser l'absence physique du produit pendant le processus de décision. Les clients peuvent vérifier la disponibilité des produits et leurs prix en ligne, tandis qu'un magazine en ligne fournit régulièrement des mises à jour sur les dernières innovations, produits, conseils et actualités du groupe.

Le site de Decathlon El-Djazair offre de nombreuses fonctionnalités pratiques pour les internautes, telles que :

- **Fonctionnalité site vitrine :**

Les utilisateurs peuvent découvrir tous les produits et leurs fiches d'informations, ainsi que vérifier en temps réel leur disponibilité et leur prix par rapport aux stocks du site e-commerce.

- **Possibilité de passer commande en ligne :**

Le site permet également aux clients de passer des commandes en ligne avec deux options de livraison :

- **Livraison à domicile :**

Les clients peuvent se faire livrer leur commande à l'adresse indiquée sur leur compte Decathlon El-Djazair.

- **Click and Collect :**

Les clients peuvent passer commande en ligne et récupérer leur colis dans le magasin.

1.4. Services cloud et la génération et exploitation des données

Decathlon reconnaît le potentiel transformateur de cloud computing en matière de stockage et de gestion efficaces de ses vastes quantités de données. Pour assurer un stockage et une gestion des données sans faille, Decathlon s'appuie sur la suite complète de services puissants offerts par Amazon Web Services (AWS). Les services AWS utilisés par Decathlon comprennent :

- **Amazon S3 :**

Amazon S3 (Simple Storage Service) est un service de stockage d'objets hautement évolutif, durable et sécurisé fourni par Amazon Web Services (AWS). Il permet aux utilisateurs de stocker et d'extraire n'importe quelle quantité de données, depuis n'importe quel endroit de l'internet, à l'aide d'une simple interface web ou d'API.

S3 offre un moyen simple de stocker et de récupérer des fichiers, des objets et des ensembles de données. Il peut être utilisé pour stocker un large éventail de données, notamment des images, des vidéos, des journaux, des sauvegardes et des archives. S3 est conçu pour offrir une durabilité et une disponibilité élevées, les données étant automatiquement répliquées sur plusieurs zones de disponibilité géographiquement dispersées au sein d'une région AWS.

S3 offre également des fonctionnalités telles que le versionnage, les politiques de cycle de vie et les mécanismes de contrôle d'accès pour aider les utilisateurs à gérer efficacement leurs données. S3 s'intègre à d'autres services AWS tels que Amazon EC2, Amazon Glacier, AWS Lambda et Amazon CloudFront, ce qui en fait un élément clé de nombreuses architectures AWS.

- **Redshift AWS :**

Amazon Redshift est un service de data warehousing dans le cloud fourni par Amazon Web Services (AWS). Il permet aux utilisateurs de créer rapidement et facilement un entrepôt de données scalable, sécurisé et économique.

Redshift est conçu pour traiter des charges de travail de data warehousing à grande échelle, jusqu'à plusieurs pétaoctets de données. Il est optimisé pour la performance, l'évolutivité et la facilité d'utilisation, en utilisant une architecture de stockage en colonnes qui permet une compression efficace des données et des performances d'interrogation. En outre, il offre des fonctionnalités telles que la sauvegarde automatique, le cryptage et la gestion des accès pour garantir la sécurité et la fiabilité des données.

Redshift permet également d'effectuer des analyses et des requêtes SQL sur les données stockées dans l'entrepôt de données, à condition que les données soient structurées de manière appropriée. Il peut être intégré à une variété d'autres services AWS, tels que Amazon S3 pour faciliter le chargement de données dans l'entrepôt de données et l'analyse de données.

Pour intégrer les données stockées sur Amazon S3 dans un entrepôt de données qui est Redshift à Decathlon, plusieurs étapes doivent être suivies, telles que le nettoyage des données, l'harmonisation des formats, la normalisation et la résolution des incohérences. Cette préparation est essentielle pour garantir la qualité et la cohérence des données stockées dans l'entrepôt. Ensuite, les données préparées peuvent être chargées dans Redshift.

1.4.1. Génération et partage de données

À travers ses multiples points stratégiques tels que les magasins, les entrepôts et les bureaux d'affaires, des données précieuses sont collectées à la fois en temps réel et hors temps réel.

En ce qui concerne les données collectées en temps réel, Decathlon est en mesure de suivre les achats de ses clients de manière instantanée. Chaque fois qu'un client effectue un achat, cette information est immédiatement enregistrée et intégrée dans le système de données de l'entreprise.

D'autre part, certaines données sont collectées hors temps réel, par exemple, le nombre de produits disponibles dans le stock est vérifié à intervalles réguliers plutôt qu'en temps réel.

1.4.2. Utilisation et exploitation des données

Decathlon utilise les données pour analyser la performance des produits, personnaliser l'expérience client, gérer les stocks et la chaîne d'approvisionnement, prendre des décisions stratégiques, améliorer la qualité des produits et mener des initiatives de marketing et de communication ciblées.



Figure 1.5 : Documentation fonctionnelle et technique du stockage de la donnée.

1.5. La stratégie de tarification de Decathlon

Decathlon a mis en place une stratégie de prix originale dans le but de rendre le sport accessible à tous. Au lieu de pratiquer des prix exorbitants que seul un certain segment de la communauté peut se permettre, Decathlon vise à pratiquer le juste prix qui reste abordable et génère un chiffre d'affaires raisonnable, qui peut ensuite être utilisé pour les investissements et l'expansion future.

Afin de déterminer ce prix, les responsables de Decathlon prennent en compte les coûts de revient liés à l'ensemble des charges directes (coûts de fabrication et d'approvisionnement) ainsi que les charges indirectes (abonnements internet et téléphoniques, frais d'assurance, loyer, frais postaux ou de publicité), auxquels ils ajoutent une petite marge. Cependant, suivre cette méthode peut conduire Decathlon à perdre du chiffre d'affaires potentiel. Au cours des années d'existence de Decathlon en Algérie, il a été remarqué qu'une augmentation de prix ne conduit pas nécessairement à une baisse des ventes, ni une diminution de prix à une augmentation des ventes.

Pour être plus clair, prenons l'exemple d'un produit vendu pendant un mois avec 140 unités vendues à 2 300 DZD. Le mois suivant, le leader a décidé de fixer un nouveau prix de 2 400 DZD et 145 unités ont été vendues. Malgré l'augmentation de prix, il y a eu une augmentation des ventes. Cela a entraîné un bénéfice supplémentaire de 100 DZD pour chaque unité vendue, ce qui est le meilleur scénario pour une entreprise de vente au détail comme Decathlon. Cependant, il est important de noter qu'un scénario opposé pourrait également se produire si le nouveau prix ne correspond pas à la volonté des clients de payer ce prix pour ce produit. Ce phénomène est expliqué par l'élasticité, qui classe les articles en fonction de la relation entre le prix et les ventes.

Si un niveau constant de ventes est maintenu malgré une augmentation ou une diminution de prix, le produit est appelé inélastique. En revanche, si une augmentation ou une diminution des ventes est observée avec un changement de prix, le produit est appelé élastique. Dans le cas des produits élastiques, trouver un prix qui maximise les bénéfices tout en maintenant les ventes constantes est difficile. C'est parce que si le prix fixé dépasse le prix psychologique qu'un client est prêt à payer pour l'article, les ventes chuteront considérablement. Il est donc essentiel de trouver la limite qui maintient la stabilité des ventes et maximise les profits.

1.6. Conclusion

Le chapitre présente l'entreprise Decathlon et sa filiale Decathlon El Djazair. Nous y avons également abordé le service cloud, en expliquant comment les données sont générées et utilisées pour améliorer les performances commerciales de l'entreprise. Enfin, nous avons mis en évidence la problématique traitée dans ce travail.

Chapitre 2 :
**Introduction à la tarification et à la
prédiction des séries temporelle.**

2.1. Introduction

Ce chapitre a pour objectif de présenter les connaissances nécessaires pour comprendre les concepts et la méthodologie de ce travail. Vous y trouverez une définition de l'élasticité-prix et des stratégies de tarification, ainsi qu'une définition des séries temporelles, des algorithmes utilisés et des concepts qui seront utilisés par la suite.

2.2. Élasticité-prix et les stratégies de tarification

La connaissance de l'élasticité-prix est cruciale pour les entreprises, car elle leur permet de déterminer la meilleure stratégie de tarification pour leur produit, dans cette partie, nous allons définir l'élasticité prix et voir quelques stratégies de tarifications

2.2.1. Élasticité-prix

L'élasticité-prix de la demande est une mesure économique permettant de quantifier la sensibilité de la demande d'un produit aux variations de son prix de vente. Soient X , Y et β : le prix de vente, la demande unitaire et l'élasticité-prix de la demande, respectivement. La formule générale de l'élasticité-prix de la demande est la suivante :

$$\beta = \frac{\Delta Y}{\Delta X}$$

Où ΔX et ΔY représentent la variation relative de la demande et du prix de vente en pourcentage. En règle générale, le prix influe sur la demande et $\beta \neq 0$.

- $\beta < 0$: lorsque la demande diminue à mesure que le prix augmente
- $\beta > 0$: Lorsque la demande augmente avec l'augmentation des prix, ou la demande diminue avec la diminution des prix.
- $\beta = 0$: Lorsque le prix d'un bien ou d'un service n'a aucun impact sur la demande de ce dernier, on qualifie cette dernière de parfaitement inélastique.

(Liu & Sustik, 2021)

2.2.2. Stratégies de tarification

Les stratégies de tarification sont des méthodes que les entreprises utilisent pour fixer les prix de leurs produits ou services. Elles ont pour but d'optimiser les revenus en maximisant les ventes, tout en prenant en compte les coûts de production et la demande du marché. Parmi les stratégies les plus utilisées auprès des entreprises, on cite :

2.2.2.1. Stratégies de prix élevés et de prix bas

Les stratégies de prix élevés sont basées sur la valeur perçue du produit par les clients, tandis que les stratégies de prix bas visent à attirer un grand nombre de clients grâce à des prix abordables. Par exemple, Apple a pu dominer le marché des lecteurs MP3 malgré le prix élevé de l'iPod en raison de sa qualité et de son esthétique. De même, les véhicules de la classe A de Mercedes-Benz ont utilisé l'image de marque pour compenser un prix plus élevé que l'analyse des coûts initiaux. D'autre part, des entreprises telles que Lidl, Aldi et Leader Price ont réussi à proposer des prix bas dans le secteur de la vente au détail de produits alimentaires. Ces stratégies peuvent être limitées par des attentes à long terme et une image fixe, ce qui peut réduire la flexibilité du système de prise de décision, tandis que les stratégies de prix bas nécessitent un volume de ventes élevé pour compenser les marges réduites.

2.2.2.2. Stratégies ajustables

Discrimination des prix

La discrimination par les prix, également connue sous le nom de différenciation ou de discrimination par les prix, est une stratégie de tarification dans laquelle une entreprise applique différents prix pour le même produit ou service à différents clients ou groupes de clients. L'objectif de la discrimination par les prix est d'extraire la valeur maximale de chaque segment de clientèle en fonction de sa volonté de payer. Il existe plusieurs types de stratégies de discrimination par les prix :

- **La tarification au premier degré ou personnalisée :**

Dans cette stratégie, les prix sont adaptés individuellement à chaque client en fonction de ses caractéristiques spécifiques, de ses préférences ou de son historique d'achat.

- **Le deuxième degré ou la tarification basée sur la quantité :**

Cette stratégie consiste à proposer des prix différents en fonction de la quantité ou du volume acheté. Les entreprises peuvent proposer des remises ou des prix de gros pour encourager les clients à acheter en plus grande quantité. Par exemple, les promotions du type "achetez-en un, obtenez-en un gratuitement".

- **La tarification au troisième degré ou basée sur le marché :** Cette stratégie consiste à appliquer des prix différents à différents segments du marché en fonction de leur sensibilité au prix ou de leur volonté de payer.

La mise en œuvre de la discrimination par les prix nécessite une segmentation minutieuse du marché, une analyse des données et l'élaboration d'une stratégie de tarification. Il peut s'agir d'un processus complexe, mais lorsqu'il est mis en œuvre de manière efficace, il peut conduire à une augmentation de la rentabilité et à une amélioration de la satisfaction de la clientèle. (Dolgui, 2017)

Stratégie de remise

Une vente à prix réduit consiste à vendre un ensemble spécifique d'articles à un prix inférieur pour une durée limitée. L'objectif de cette réduction est de générer des ventes supplémentaires qui compenserait la baisse des recettes. Toutefois, il est rare que cela se produise. De nombreuses entreprises ne reconnaissent pas le coût réel des remises. Lorsqu'une remise sur un produit est appliquée pour une période spécifique, elle est souvent appliquée à toutes les ventes, ce qui peut avoir des conséquences négatives.

Écrémage des prix

La stratégie d'écrémage des prix consiste à fixer initialement un prix relativement élevé pour un produit et à le réduire progressivement au fil du temps. Cette stratégie est similaire à la discrimination par les prix, mais en mettant l'accent sur le facteur temps. L'écrémage est souvent appliqué lorsque les clients sont moins sensibles au prix (comme dans l'industrie cosmétique) ou lorsqu'ils sont attirés par des produits innovants (notamment dans le secteur de l'électronique, comme les ordinateurs).

L'écrémage des prix est particulièrement utile pour récupérer les investissements substantiels réalisés dans la recherche et le développement. Si un prix élevé ne peut être maintenu indéfiniment, il peut l'être tant que l'entreprise jouit d'une position de monopole en raison de l'unicité du produit.

Prix de pénétration

Cette stratégie consiste à fixer des prix plus bas dans un premier temps afin de pénétrer ou de gagner une plus grande part de marché. Elle permet d'attirer des clients et de gagner rapidement des parts de marché. Au fil du temps, les prix peuvent être augmentés à mesure que l'entreprise renforce sa présence.

Gestion des revenus

La gestion des revenus vise à prévoir le comportement des clients et des concurrents afin de maximiser les recettes. Les entreprises qui utilisent cette stratégie, évaluent régulièrement les situations passées afin d'analyser l'impact des événements sur le comportement des clients et des concurrents (Agwu, 2014). Elles prennent également en compte les événements futurs pour ajuster leurs prix. La gestion des revenus est particulièrement efficace pour les produits soumis à des contraintes temporelles spécifiques, comme les billets d'avion, ou pour les produits périssables, comme les fruits et les aliments transformés.

2.3. Les séries temporelles

Une série temporelle (time series) est une série de points de données ou d'observations enregistrés à des intervalles de temps différents ou réguliers. En général, une série temporelle est une séquence de points de données pris à des intervalles de temps également espacés. La fréquence des points de données enregistrés peut être horaire, quotidienne, hebdomadaire,

mensuelle, trimestrielle ou annuelle. Si plusieurs variables sont mesurées dans le temps, la série temporelle est dite multivariée, sinon elle est univariée. Les séries temporelles sont utilisées dans les statistiques, la finance ou les applications commerciales. Un exemple très courant de données de séries temporelles est la valeur de clôture quotidienne d'un indice boursier comme le NASDAQ. D'autres applications courantes des séries temporelles sont la prévision des ventes et de la demande, les prévisions météorologiques, l'économétrie, le traitement des signaux, la reconnaissance des formes et la prévision des tremblements de terre.

2.3.1. Les composantes d'une série temporelle

La décomposition d'une série temporelle (ou chronologique) vise à distinguer différentes composantes de son évolution : une tendance générale, des variations saisonnières récurrentes, des variations cycliques et des variations imprévisibles. Voici une brève description de ces composantes :

- **Saisonnalité** : Cette composante englobe les fluctuations périodiques qui se produisent à l'intérieur d'une période donnée (par exemple, chaque mois, trimestre ou année). La saisonnalité a une période fixe et connue.
- **Tendance** : La tendance générale d'une série chronologique à augmenter, diminuer ou rester stable sur une longue période est appelée tendance séculaire ou simplement tendance. Elle représente un mouvement à long terme dans la série chronologique. Par exemple, les séries relatives à la croissance démographique ou au nombre de maisons dans une ville montrent une tendance à la hausse, tandis que les séries relatives aux taux de mortalité ou aux épidémies peuvent présenter une tendance à la baisse.
- **Cycle** : Les variations cycliques dans une série chronologique décrivent des changements à moyen terme causés par des circonstances qui se répètent de manière cyclique. La durée d'un cycle s'étend généralement sur deux ans ou plus.
- **Composante irrégulière** : Les variations irrégulières ou aléatoires d'une série chronologique sont causées par des influences imprévisibles qui ne suivent pas de modèle spécifique et ne se répètent pas régulièrement. Ces variations sont provoquées par des incidents tels que des guerres, des grèves, des tremblements de terre, des inondations, des révolutions, etc. Il n'existe pas de technique statistique définie pour mesurer ces fluctuations aléatoires dans une série chronologique. (*Components of Time Series Analysis: Trends and Mathematical Models*, n.d.)

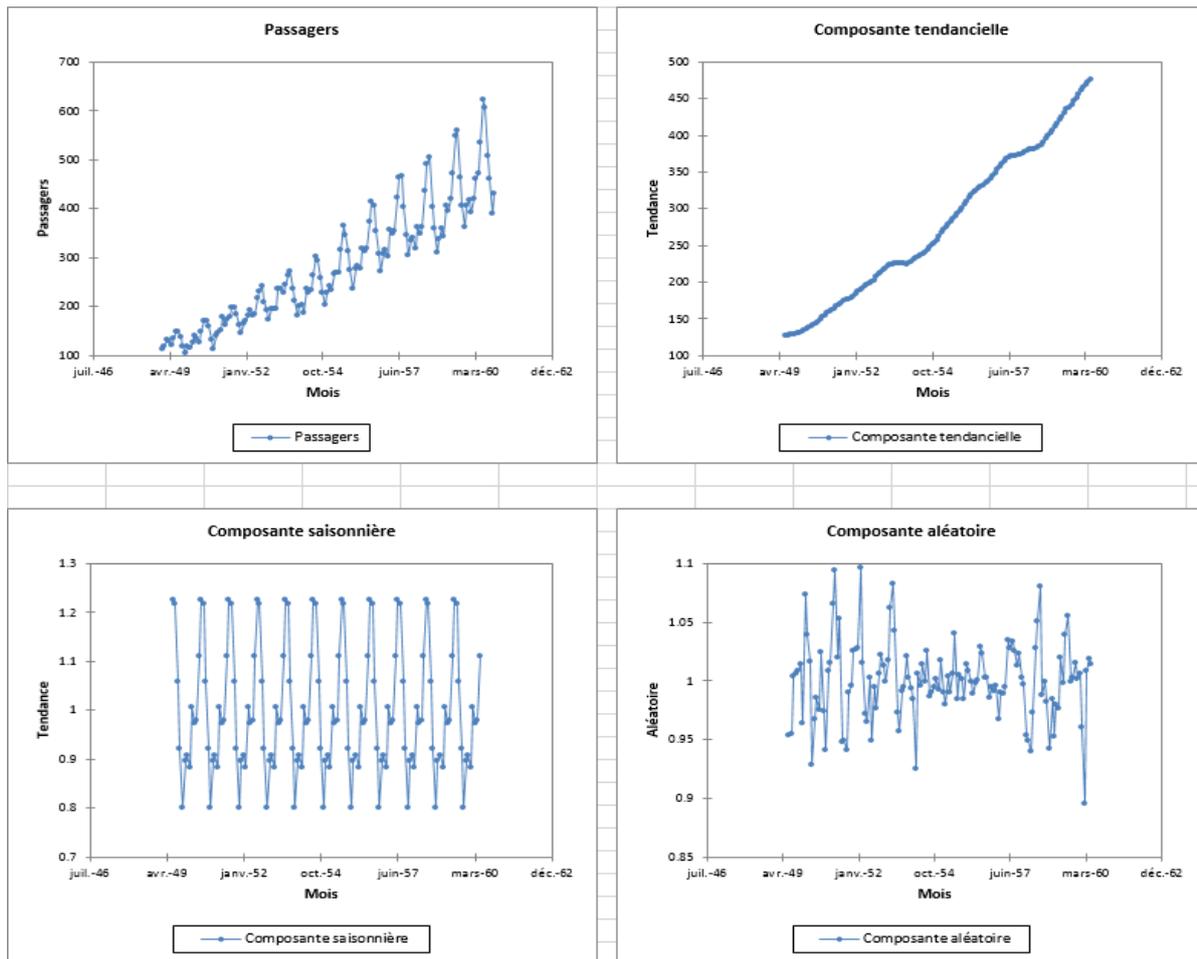


Figure 2.1: La décomposition des séries temporelles de l'ensemble de données sur les passagers aériens (compagnies aériennes américaines de 1949 à 1960).

2.3.2. Stationnarité d'une série temporelle

La stationnarité d'une série temporelle fait référence aux propriétés statistiques de la série qui restent constantes dans le temps. Plus précisément, cela signifie que la moyenne, la variance et la structure d'autocorrélation de la série temporelle sont constantes dans le temps. Une série temporelle stationnaire a une structure stable et les schémas observés dans le passé sont susceptibles de se poursuivre dans le futur.

La stationnarité est une propriété souhaitable dans les séries temporelles pour la modélisation et la prévision, car elle permet d'utiliser des méthodes statistiques qui supposent une structure constante dans le temps. Si une série temporelle est non stationnaire, ses propriétés statistiques changent au fil du temps, ce qui la rend difficile à modéliser et à prévoir avec précision.

Dans notre étude, nous avons choisi d'utiliser le test de Dickey-Fuller augmenté (ADF) pour évaluer la stationnarité de la série temporelle.

Augmented Dicky Fuller test (ADF) :

Le test ADF appartient à une catégorie de tests appelée "Test de racine unitaire", qui est une méthode appropriée pour tester la stationnarité d'une série temporelle.

- **La racine unitaire** : est une caractéristique d'une série temporelle qui la rend non stationnaire. Le test de racine unitaire est basé sur le modèle suivant :

$$y_t = \rho y_{t-1} + \varepsilon_t$$

où :

- y_t est la série temporelle au moment t
- ρ est le coefficient de la variable dépendante retardée y_{t-1}
- ε_t est le terme d'erreur

L'hypothèse nulle du test de racine unitaire est que si $|\rho| = 1$, cela signifie que la série temporelle a une racine unitaire et par conséquent elle est non stationnaire. L'hypothèse alternative c'est : si $|\rho| < 1$, la série temporelle est stationnaire.

- **Le test ADF** : c'est une extension du test de Dickey-Fuller qui inclut des termes supplémentaires dans l'équation de régression pour tenir compte de l'autocorrélation dans le terme d'erreur qui se produit lorsque ces termes sont corrélés entre eux dans le temps. Ceci viole l'hypothèse d'erreurs indépendantes et identiquement distribuées, qui est nécessaire à la validité des tests statistiques (Zivot & Wang, 2006, p. 120).
- ❖ Le test de Dickey-Fuller est basé sur le modèle suivant :

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \varepsilon_t$$

où :

- y_t sont les valeurs de la série temporelle $t = 0, 1, \dots$
- Δy_t est la différence entre y_t et y_{t-1}
- α est un terme constant
- β est le coefficient d'une tendance linéaire dans la série temporelle
- γ est le coefficient de la variable dépendante retardée y_{t-1}
- ε_t est le terme d'erreur.

- ❖ Le test ADF est basé sur le modèle suivant :

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum \delta_i \Delta y_{t-i} + \varepsilon_t$$

où :

- Δy_t la première différence de la série y_t
- α est un terme constant
- β est le coefficient d'une tendance linéaire dans la série temporelle
- γ est le coefficient de la variable dépendante retardée y_{t-1}
- $\sum \delta_i \Delta y_{t-i}$ est la somme des différences retardées de la série temporelle jusqu'à l'ordre p
- ε_t est le terme d'erreur

Le test Dickey-Fuller augmenté utilise les hypothèses suivantes:

- Hypothèse nulle: $\gamma = 0$
- Hypothèse alternative: $\gamma < 0$

L'hypothèse nulle dit qu'une racine unitaire se trouve dans la série chronologique, ce qui signifie qu'elle n'est pas stationnaire. Le rejet de l'hypothèse nulle indique que la série est stationnaire.

En pratique on calcule la valeur p qui représente une probabilité qui mesure la preuve par rapport à l'hypothèse nulle. Si cette dernière est inférieure à la valeur critique 0,05, l'Hypothèse nulle sera rejetée (Verma, 2021).

2.3.3. La transformation des séries chronologiques

Les séries chronologiques nécessitent souvent une transformation avant la prévision ou l'analyse. La transformation des données de séries chronologiques est souvent utile, car de nombreuses méthodes de prévision fonctionnent mieux avec des séries chronologiques qui respectent certaines hypothèses. La transformation des séries chronologiques peut également simplifier le problème de prévision en éliminant les composantes de tendance et saisonnières. Si une opération de transformation est appliquée aux données de la série temporelle, une opération inverse similaire doit être appliquée à la prévision. Identifier la meilleure méthode de transformation peut être difficile pour un problème de prévision donné, car il n'existe pas de règles standard pour le prétraitement des données chronologiques. Dans ce qui suit, nous présentons les opérations de transformation les plus courantes pour l'analyse et la prévision des séries chronologiques.

❖ **Différenciation:**

La différenciation (Differencing) est une technique utilisée dans l'analyse des séries temporelles pour transformer une série temporelle non stationnaire en une série stationnaire. La non-stationnarité rend en effet difficile la modélisation et la prévision précise des données. La différenciation consiste à calculer les différences entre les observations consécutives d'une série temporelle, ce qui supprime la tendance et la saisonnalité présente dans les données et permet d'obtenir une série temporelle stationnaire.

Deux types de différenciation sont couramment utilisés dans l'analyse des séries temporelles :

- La différenciation de premier ordre consiste à soustraire chaque observation de l'observation précédente $\Rightarrow Z(t) = Y(t) - Y(t - 1)$
- La différenciation saisonnière consiste à soustraire chaque observation de l'observation correspondante de la saison précédente $\Rightarrow Z(t) = Y(t) - Y(t - P)$, Avec P représente une période, qui peut être une semaine, un mois, une année...

La différenciation peut être appliquée plusieurs fois jusqu'à ce que la série temporelle résultante soit stationnaire.

❖ Transformation log :

La transformation log est un type de transformation de puissance dans lequel une fonction logarithmique est appliquée aux valeurs de série temporelle. Cette transformation est souvent utilisée dans les prévisions et analyses de séries temporelles pour stabiliser la variance d'une série. En d'autres termes, les prévisions basées sur $X_t = \log(Y_t)$ peuvent être bien meilleures que les prédictions directes de Y_t car X_t aurait une variance plus stable que Y_t .

❖ Normalisation :

La normalisation consiste à redimensionner la distribution des valeurs afin que la moyenne des valeurs observées soit égale à zéro et que l'écart-type en soit un. Ceci peut être réalisé en soustrayant la moyenne et en divisant le résultat par l'écart type de la série. Cette transformation améliore la convergence des méthodes de prévision appliquées à la série chronologique.

2.4. Les modèles de prédiction des séries temporelles

2.4.1. Apprentissage automatique

L'apprentissage automatique est un sous-ensemble de l'intelligence artificielle qui comprend des techniques mathématiques pour estimer les fonctions qui permettront de faire correspondre certaines données d'entrée à un résultat souhaité. Ces techniques mathématiques sont généralement utilisées lorsque la fonction $f(x)$ est difficile à spécifier explicitement. Elles peuvent être classées en deux grandes catégories, en fonction du type de données utilisées pour estimer les paramètres de la fonction. Les méthodes non supervisées ne disposent que d'un ensemble de données d'entrée possibles x et sont souvent utilisées pour trouver une structure ou des patterns intrinsèques à l'ensemble de données. En revanche, dans les modèles supervisés, les paramètres sont estimés à l'aide d'un ensemble de données contenant des exemples des entrées

x et des sorties souhaitées correspondantes y de f . L'apprentissage supervisé, qui fait l'objet de cette thèse, peut être divisé par tâche en deux sous-types : la classification ou la régression.

❖ **Classification :**

La classification est la tâche qui consiste à mettre en correspondance les entrées x à des sorties $y \in \{1, \dots, C\}$ où C est le nombre de classes. Les tâches de classification exigent que y soit une valeur catégorique ou numérique discrète dans un ensemble fini de classes possibles. Si $C = 2$, la tâche est appelée problème de classification binaire, tandis que pour tout $C > 2$, la tâche est appelée classification multi-classes.

Si les étiquettes y ne s'excluent pas mutuellement et que le modèle doit potentiellement produire plus d'une catégorie pour chaque entrée, la tâche est appelée problème de classification multi étiquettes. Le plus souvent, un modèle de classification fournit une seule sortie pour chaque entrée. Parmi les exemples de problèmes de classification, on peut citer la prédiction du contenu d'un courrier électronique en tant que spam ou non.

❖ **Régression :**

La régression consiste à prédire une fonction qui peut faire correspondre une entrée x à une sortie continue $y \in R$. Parmi les exemples de tâches de régression, citons la prédiction du prix de vente d'une maison à partir de détails sur sa taille, son état, etc.

2.4.2. ARIMA

ARIMA (Autoregressive Integrated Moving Average) est un modèle statistique utilisé pour la prévision de séries temporelles dans de nombreux domaines, notamment la finance, l'économie et le marketing.

Le modèle ARIMA se compose de trois éléments (HAYES, 2022): l'auto-régression(**AR**), l'intégration(**I**) et la moyenne mobile (**MA**). Ces trois composantes se combinent pour former le modèle ARIMA(p, d, q), où p , d et q sont respectivement l'ordre de chaque composante.

- **Autorégression (AR) :** modèle de régression capable de trouver la relation entre une observation et un certain nombre d'observations décalées. En d'autres termes, il suppose que la valeur actuelle de la série temporelle dépend de ses valeurs antérieures. L'ordre d'autorégression (p) détermine le nombre de valeurs décalées utilisées pour prédire la valeur actuelle (Ariton, 2021).

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \epsilon$$

Où Y_{t-1} est le lag_1 de la série, β_1 est le coefficient du lag_1 estimé par le modèle et α est le terme d'ordonnée à l'origine, également estimé par le modèle.

- **Intégration (I) :** La composante d'intégration modélise la différenciation des séries temporelles pour atteindre la stationnarité. L'ordre d'intégration (d) détermine le nombre de fois que la différenciation est nécessaire pour atteindre la stationnarité.

- **Moyenne mobile (MA) :** La composante moyenne mobile modélise la relation entre le terme d'erreur et un certain nombre de termes d'erreur retardés. Le terme d'erreur est la différence entre la valeur réelle de la série temporelle et la valeur prédite. L'ordre de la moyenne mobile (q) détermine le nombre de termes d'erreur retardés utilisés pour prédire la valeur actuelle.

$$Y_t = \alpha + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \dots + \phi_q \epsilon_{t-q}$$

où les termes d'erreur sont les erreurs du modèle autorégressif des retards respectifs.

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_0 Y_0 + \epsilon_t$$

$$Y_{t-1} = \beta_1 Y_{t-2} + \beta_2 Y_{t-3} + \dots + \beta_0 Y_0 + \epsilon_{t-1}$$

En d'autres termes, la méthode ARIMA n'est qu'une combinaison linéaire de ses 3 composants

$$Y_t \text{ prédit} = \text{Constante} + \text{Combinaison linéaire des lag de } Y \text{ (jusqu'à } p \text{ lags)} + \text{Combinaison linéaire des erreurs de prévision décalées (jusqu'à } q \text{ lags)}$$

Le modèle ARIMA nécessite une série temporelle stationnaire, de sorte que la première étape consiste souvent à appliquer une différenciation jusqu'à ce que la série soit stationnaire. Ensuite, le modèle est adapté à la série stationnaire et les paramètres du modèle (p, d, q) sont estimés à l'aide de techniques statistiques telles que l'estimation du maximum de vraisemblance (Fernandez, n.d.). Enfin, le modèle est utilisé pour prévoir les valeurs futures de la série temporelle.

2.4.3. Facebook Prophet

FBProphet est un algorithme développé par l'équipe de science des données de Facebook dans le but de fournir des prévisions évolutives, rapides et précises. Il a de nombreuses applications dans divers domaines, tels que la prévision des ventes de commerce électronique ou des modèles météorologiques. À la base, FBProphet fonctionne sur le principe de la décomposition des données de séries temporelles en composantes de tendance, de saisonnalité et de bruit afin de générer des prévisions précises. (FBProphet, n.d.)

La composante "tendance" caractérise la direction générale de la série temporelle, c'est-à-dire si elle présente une augmentation ou une diminution graduelle dans le temps. Pour modéliser la tendance, FBProphet utilise une approche de régression linéaire par morceaux qui permet de s'adapter aux données. La composante saisonnière, quant à elle, permet de capturer les schémas récurrents dans les données, tels que les tendances hebdomadaires ou mensuelles. FBProphet s'appuie sur la série de Fourier pour modéliser de manière flexible différents types de modèles saisonniers, en s'adaptant à diverses fluctuations dépendant du temps. Enfin, la composante bruit

tient compte des fluctuations aléatoires des données qui ne peuvent être attribuées à la tendance ou à la saisonnalité.

2.4.4. Random forest

Random Forest est un puissant algorithme d'apprentissage automatique utilisé pour diverses tâches, notamment la régression et la classification. C'est une méthode d'ensemble qui combine plusieurs arbres de décision, appelés estimateurs, pour créer un modèle de prédiction plus précis.

Contrairement aux arbres de décision individuels, qui sont sujets au surapprentissage (overfitting) ; un surapprentissage c'est quand le modèle essayer de trop s'adapter au bruit dans les données et par conséquent, il ne pourra pas généraliser sur de nouvelles données. Random Forest surmonte cette limitation en agrégeant les prédictions de plusieurs arbres. Cette conception d'ensemble permet au modèle de généraliser efficacement sur des données non vues, le rendant robuste et fiable. Il peut traiter différents types de données, tels que des caractéristiques catégorielles et numériques, et gérer efficacement de grands ensembles de données à haute dimensionnalité.

Les Random Forest excellent dans les tâches de classification, où ils offrent d'excellentes performances. Cependant, dans les problèmes de régression, ils peuvent présenter certaines limites par rapport aux modèles de régression linéaire. Les régresseurs Random Forest ne peuvent pas faire de prédictions en dehors de la plage des données d'entraînement, contrairement à la régression linéaire.

Un inconvénient des Random Forest est leur opacité. Contrairement à certains algorithmes d'apprentissage automatique traditionnels, il peut être difficile d'interpréter la logique derrière les décisions prises par un classificateur Random Forest. De plus, l'entraînement et l'exécution des modèles Random Forest peuvent être coûteux en termes de ressources informatiques.

Malgré ces limitations, les Random Forest offrent plusieurs avantages. Ils sont extrêmement robustes, faciles à mettre en œuvre et gèrent efficacement les types de données hétérogènes (numériques et catégoriques). De plus, ils ont relativement peu de paramètres à régler, ce qui en fait une option attrayante pour les scientifiques des données afin d'évaluer rapidement l'exactitude réalisable pour un problème donné. Même si la solution finale n'implique pas un Random Forest, il sert de point de départ précieux dans le processus d'apprentissage automatique.

2.4.5. Réseaux de neurones

Les réseaux de neurones artificiels, ou plus simplement les réseaux de neurones (ANN), sont un type de modèles d'apprentissage automatique qui s'inspirent des réseaux neuronaux humains. La structure des réseaux neuronaux est similaire à celle de notre cerveau, qui comporte des neurones et des connexions entre eux pour transmettre des impulsions électriques.

Chaque neurone est optimisé pour recevoir des informations d'autres neurones, traiter ces informations d'une manière unique et envoyer le résultat à d'autres cellules. En se basant sur le comportement fonctionnel des neurones biologiques, nous pouvons définir une unité de calcul

appelée neurone artificiel. Le neurone artificiel est l'unité de base de ce que l'on appelle les réseaux de neurones. Mathématiquement, il peut être représenté comme une fonction qui prend un certain nombre d'entrées x_1, x_2, \dots, x_m , chacune de ces entrées ayant un poids correspondant w_1, w_2, \dots, w_m . La somme pondérée des entrées plus le terme de biais forment la partie linéaire de la transformation produisant s , suivie de la partie non linéaire produisant la sortie du neurone y , qui est l'application de la fonction d'activation.

$$S = m \sum_{k=1} w_k x_k + b w_0$$

$$y = f(z)$$

La figure suivante illustre le concept de neurone unique. De nombreuses fonctions d'activation ont été proposées dans la littérature et seront discutées dans le titre suivant. Notez que dans certains cas, la fonction d'activation peut être linéaire (par exemple pour les problèmes de régression). En utilisant l'algèbre linéaire, nous pouvons réécrire les deux équations de façon simple.

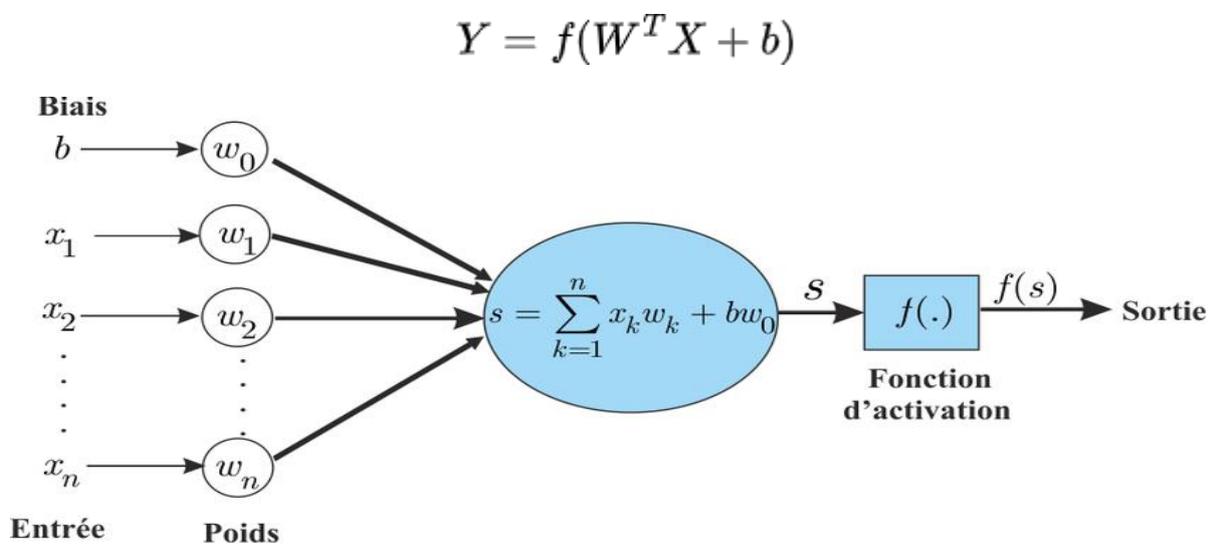


Figure 2.2: Schéma d'un neurone artificiel¹.

Nous pouvons maintenant passer à la compréhension de l'architecture des réseaux neuronaux, car l'idée d'un neurone est basique, mais essentielle. Une version simplifiée des réseaux neuronaux est représentée sous la forme d'une organisation en couches de neurones avec des connexions à d'autres neurones. La couche d'entrée est représentée par les données d'entrée, la couche de sortie est la dernière à traiter les données et à renvoyer les résultats de l'ensemble du modèle, tandis que toutes les couches intermédiaires sont appelées couches cachées. Ces couches cachées transforment les données d'une couche à l'autre et envoient les informations traitées à la couche suivante. C'est ce qui fait la profondeur du réseau. C'est cette caractéristique du réseau

¹ (Truong, 2017)

qui a popularisé le terme "Deep Learning" (apprentissage profond), qui est un sous-ensemble de l'apprentissage automatique dédié à l'étude des ANNs.

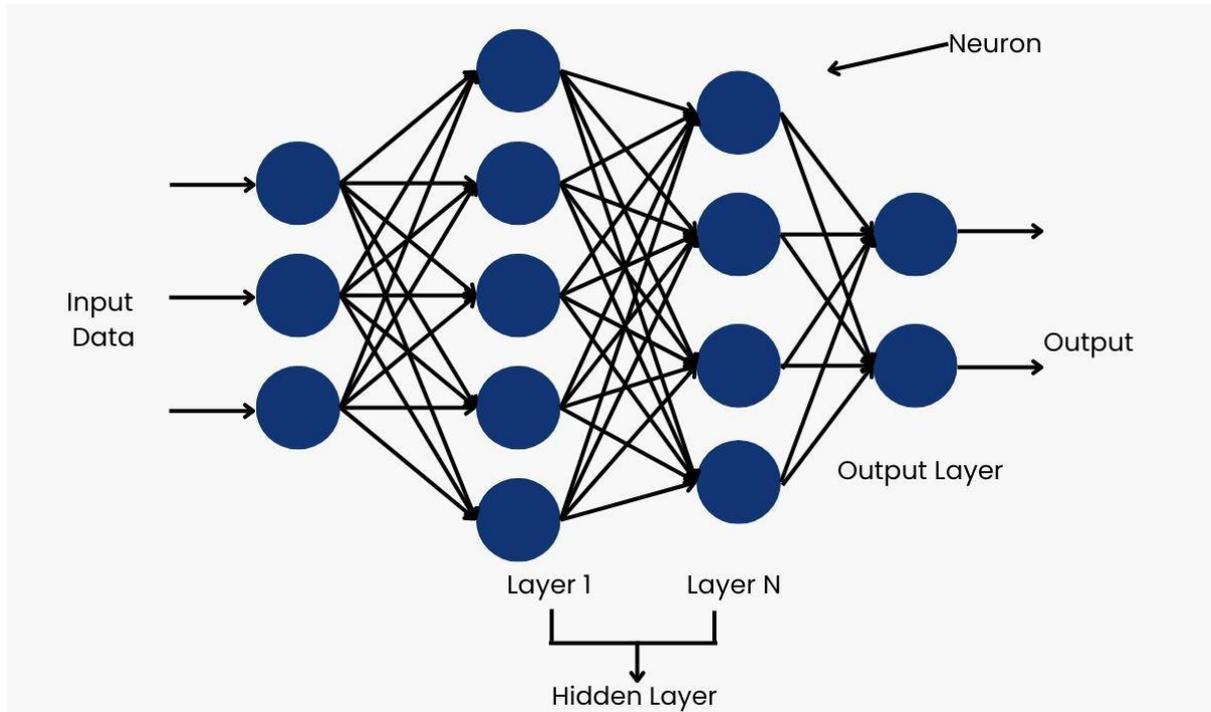


Figure 2.3: Réseau de neurones à N couches cachées.

2.4.5.1. Les fonctions d'activation :

Comme nous l'avons vu plus haut, la fonction d'activation est l'un des éléments constitutifs des réseaux neuronaux. Nous savons qu'il s'agit de la deuxième transformation de chaque neurone après la partie linéaire et qu'il s'agit d'une fonction non linéaire afin d'aider le modèle à apprendre des structures complexes. Si nous utilisons des fonctions d'activation linéaires dans un réseau neuronal profond, quelle que soit la profondeur de notre réseau, il sera équivalent à un simple réseau neuronal linéaire sans couches cachées, car ces fonctions d'activation linéaires peuvent être combinées pour former une autre fonction linéaire unique. En d'autres termes, un réseau neuronal sans fonction d'activation est essentiellement un modèle de régression linéaire. Nous utilisons donc une transformation non linéaire des entrées du neurone et cette non-linéarité dans le réseau est introduite par une fonction d'activation. Dans cette section, nous verrons les fonctions d'activation les plus courantes :

- **Fonction d'activation sigmoïde :** il s'agit de l'une des fonctions d'activation non linéaires les plus utilisées. Elle a été utilisée pour prédire des résultats basés sur des probabilités et a été appliquée avec succès à des problèmes de classification binaire. Elle est définie par :

$$f(x) = \frac{1}{1 + e^{-x}}$$

La fonction sigmoïde traduit l'entrée dans l'intervalle $[-\infty + \infty]$ par l'intervalle

$[0, 1]$. La fonction est monotone et continuellement différentiable. Pour les petites valeurs de x , elle est plus proche de 0. Inversement, plus x est grand, plus elle est proche de 1.

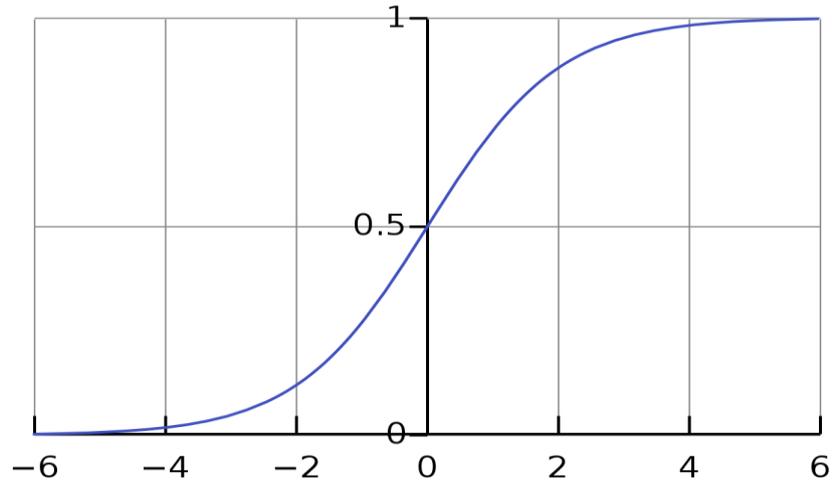


Figure 2.4: Fonction sigmoïde.

- **Fonction d'activation tangente hyperbolique (Tanh):** utilise un type de non-linéarité en forme de S similaire à la fonction sigmoïde, mais au lieu de s'étendre de 0 à 1, la sortie de \tanh s'étend de -1 à 1. Elle est définie par l'équation suivante :

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

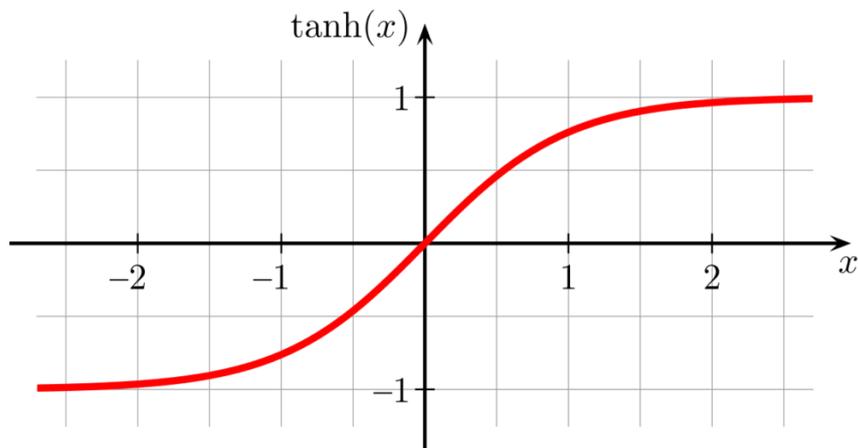


Figure 2.5 : Fonction tanh.

- **Fonction d'activation Rectified Linear Unit (ReLU):** effectue une opération de seuil sur chaque élément d'entrée où les valeurs inférieures à zéro sont mises à zéro et renvoie la sortie identité dans le cas contraire.

$$f(x) = \max(0, x)$$

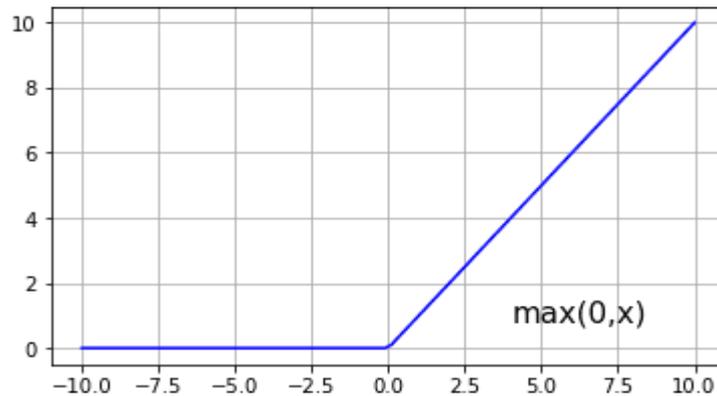


Figure 2.6: Fonction ReLU.

- **Fonction d'activation Softmax:** produit un vecteur représentant une distribution de probabilité sur un ensemble d'étiquettes mutuellement exclusives. Elle diffère des autres fonctions d'activation que nous avons vues précédemment, car la sortie de chaque neurone de la couche qui lui est associée dépend de la sortie des autres neurones de la même couche. Cette dépendance est le résultat de la loi de probabilité qui exige que la somme de toutes les sorties soit égale à 1. On y parvient en normalisant z et en fixant la sortie du neurone à :

$$y_{oi} = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

Pour une bonne prédiction, l'une des valeurs du vecteur de sortie est proche de 1, tandis que les autres entrées sont proches de 0. À l'inverse, pour une mauvaise prédiction, plusieurs étiquettes possibles sont à peu près égales à la même valeur. Par conséquent, Softmax est utilisé spécifiquement pour la couche de sortie du réseau.

2.4.5.2. Fonctions de perte et de coût

Les fonctions de perte (Loss function) et de coût (Cost function) indiquent dans quelle mesure le modèle est performant dans la tâche qu'il est censé accomplir. Dans les problèmes d'apprentissage supervisé, il s'agit souvent d'une sorte de mesure de distance entre les étiquettes des données d'apprentissage y et la valeur prédite par le modèle $\hat{y} = f(x)$.

La différence entre ces deux fonctions est que la fonction de perte capture la différence entre les valeurs réelles et prédites pour une seule observation alors que les fonctions de coût agrègent la différence pour l'ensemble des données d'apprentissage.

2.4.5.3. Réseaux neuronaux de type feed-forward

Les réseaux feed-forward profonds, également appelés réseaux neuronaux feed-forward ou réseaux neuronaux denses, sont au cœur des modèles d'apprentissage profond. L'objectif d'un réseau feed-forward est d'approximer une fonction f . Il définit une correspondance

$y = f(x ; \omega)$ et apprend la valeur des paramètres ω qui donnent la meilleure approximation de la fonction. Ils sont appelés "feed-forward" parce que l'information circule uniquement vers l'avant du réseau (pas de boucles), d'abord à travers les nœuds d'entrée, puis à travers les nœuds cachés (s'il y en a), et enfin à travers les nœuds de sortie. Le réseau est également appelé entièrement connecté parce que la sortie de chaque neurone d'une couche est envoyée à l'entrée de chaque neurone de la couche suivante (Goodfellow,2016).

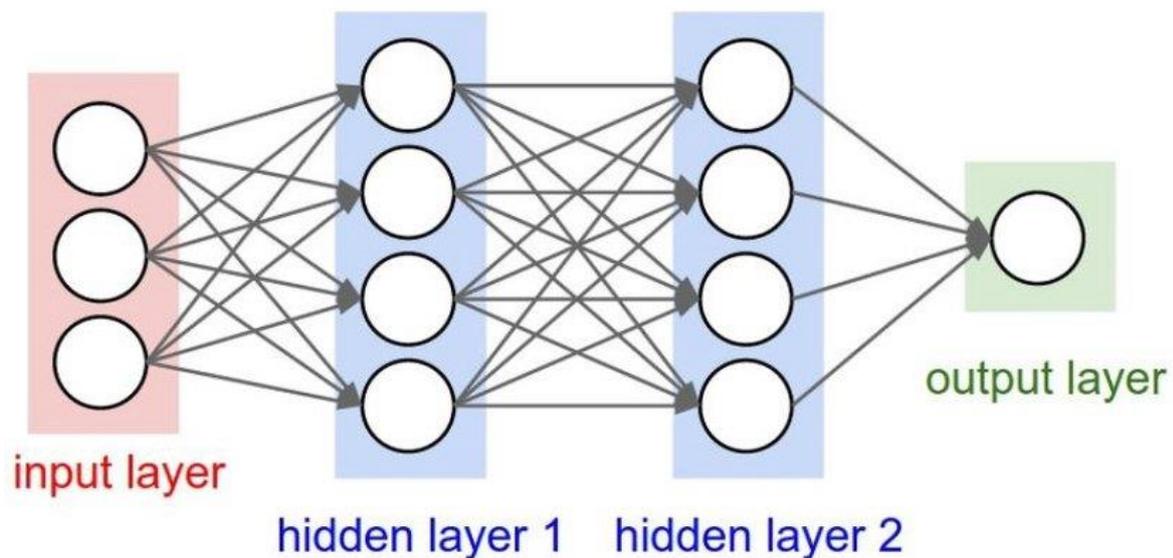


Figure 2.7: Réseaux de neurones dense².

2.4.5.4. Apprentissage des réseaux de neurones

Au cours de cette phase d'apprentissage, le réseau apprend en ajustant les paramètres de manière à pouvoir prédire la valeur ou l'étiquette correcte des échantillons d'entrée selon le type de problème (régression ou classification). Les paramètres du modèle, dans ce cas, font référence aux poids et aux biais de chaque couche. Tout d'abord, les paramètres sont initialisés, dont il existe de nombreux schémas d'initialisation dans la littérature. Ensuite, le réseau feed-forward est utilisé pour accepter une entrée x et produire une sortie \hat{y} , l'information circule dans le réseau comme expliqué avant. L'entrée x fournit l'information initiale qui se propage ensuite vers les unités cachées de chaque couche et produit finalement \hat{y} . C'est ce qu'on appelle la propagation vers l'avant. Avec ces sorties \hat{y} , et connaissant la valeur réelle de la sortie y , la fonction de coût

² De Heatmap-based Object Detection and Tracking with a Fully Convolutional Neural Network - Scientific Figure on ResearchGate- figure 6

peut être calculée. Ensuite, l'erreur mesurée par la fonction de coût est propagée à travers chaque couche dans l'ordre inverse, jusqu'à ce qu'elle atteigne la couche d'entrée. Cette étape permet à l'information du coût de remonter dans le réseau pour calculer le gradient. Cette phase s'appelle l'algorithme de rétropropagation. Les paramètres du modèle sont ensuite mis à jour par un algorithme d'optimisation tel que Adam, en utilisant les gradients.

L'entraînement d'un réseau neuronal par rétropropagation implique le choix d'un certain nombre de composants et d'hyper paramètres:

- **Fonction de coût** : c'est une fonction utilisée pour estimer les performances d'un modèle avec un ensemble spécifique de poids sur les exemples de l'ensemble de données d'apprentissage.
- **Initialisation des poids** : une procédure par laquelle de petites valeurs aléatoires initiales sont attribuées aux poids du modèle au début de l'apprentissage.
- **Taille du lot (Batch size)**: Nombre d'exemples utilisés pour estimer le gradient d'erreur avant de mettre à jour les paramètres du modèle.
- **Taux d'apprentissage (Learning rate)** : Le montant que chaque paramètre du modèle est mis à jour par cycle de l'algorithme d'apprentissage.
- **Époques (Epochs)** : Nombre de passages complets dans l'ensemble de données d'apprentissage avant que le processus de formation ne soit interrompu.

2.4.5.5. Réseaux de neurones récurrents (RNN)

Les réseaux neuronaux récurrents (RNN) sont un type de réseaux neuronaux destinés au traitement de données séquentielles. Ils sont appelés récurrents et non pas feed-forward parce que les sorties des étapes temporelles précédentes sont introduites dans l'étape temporelle actuelle.

Les RNN ont l'avantage de pouvoir traiter des séquences de longueur variable. De nombreux problèmes complexes ont été résolus à l'aide de RNN, tels que la traduction automatique, la reconnaissance vocale, le sous-titrage d'images et la génération de musique. Les RNN sont très efficaces pour la classification des séries temporelles et d'autres tâches séquentielles parce qu'ils ont une "mémoire". Ils peuvent lire les entrées X_t une à la fois et se souvenir de certaines informations/contextes par le biais des activations de la couche cachée qui sont transmises d'un pas de temps à l'autre.

Un RNN unidirectionnel, également connu sous le nom de RNN à sens unique, traite les séquences en ne considérant que l'information du passé. Il utilise les informations contextuelles précédentes pour traiter les entrées ultérieures, ce qui est approprié pour des tâches où l'information future n'est pas pertinente ou n'est pas disponible lors de la prédiction, comme la modélisation du langage ou l'analyse de sentiment.

D'autre part, un RNN bidirectionnel traite les séquences en utilisant à la fois l'information du passé et du futur. Il combine deux RNN distincts, l'un traitant la séquence de début à fin et l'autre de fin à début. Cette approche permet au modèle de capturer le contexte des éléments passés et futurs, ce qui le rend approprié pour des tâches où l'information des deux directions temporelles est importante, telles que la reconnaissance vocale ou la traduction automatique.

2.4.5.6. Limitations des RNN

Lors de l'entraînement de réseaux de neurones récurrents, on rencontre couramment des défis spécifiques, parmi lesquels se trouvent les problèmes suivants :

❖ Évanouissement du gradient :

Le problème de l'évanouissement du gradient (gradient vanishing) est un problème qui peut survenir lors de la formation de réseaux neuronaux profonds, en particulier ceux dotés d'architectures de réseaux neuronaux récurrents. Il s'agit du problème où les gradients, qui sont utilisés pour mettre à jour les paramètres du réseau pendant l'algorithme de rétropropagation, diminuent ou disparaissent au fur et à mesure qu'ils sont propagés vers l'arrière à travers les couches.

Lorsqu'un gradient devient très faible, cela signifie que l'impact des paramètres des couches précédentes sur la fonction de perte globale est négligeable. Par conséquent, les paramètres de ces couches reçoivent de très petites mises à jour, ce qui peut ralentir considérablement le processus d'apprentissage, voire l'entraver. Le problème du gradient disparaissant est particulièrement prononcé dans les réseaux profonds comportant de nombreuses couches, tels que les RNN, où les gradients doivent passer par de multiples itérations récurrentes ou étapes temporelles.

Le problème de l'évanouissement du gradient est souvent dû aux fonctions d'activation utilisées dans le réseau, en particulier celles qui écrasent leurs entrées dans une plage limitée, comme les fonctions sigmoïde ou tangente hyperbolique. Ces fonctions d'activation ont des gradients qui peuvent devenir très faibles pour des valeurs d'entrée extrêmes, ce qui fait que les gradients disparaissent lorsqu'ils sont rétropropagés à travers plusieurs couches. Par conséquent, le réseau ne parvient pas à capturer et à propager des informations importantes des couches précédentes aux couches suivantes. (Pascanu et al., 2013)

❖ Explosion du gradient :

Le problème de l'explosion du gradient (gradient exploding) est un autre problème qui peut survenir pendant la formation des réseaux neuronaux récurrents. Il se produit lorsque les gradients du réseau deviennent extrêmement importants au fur et à mesure qu'ils sont rétropropagés à travers les couches, ce qui entraîne une dynamique d'entraînement instable.

Lorsque les gradients deviennent trop importants, ils peuvent entraîner la mise à jour des poids du réseau par incréments importants, ce qui se traduit par des mises à jour de paramètres instables. Cela peut empêcher le réseau de converger vers une solution optimale et peut entraîner une oscillation ou une divergence du réseau pendant l'entraînement.

Le problème de l'explosion du gradient est plus fréquent dans les RNN en raison de la nature des connexions récurrentes. Dans les RNN, les gradients doivent être rétropropagés sur plusieurs pas de temps et si les gradients sont multipliés de manière répétée, ils peuvent croître de manière exponentielle.

L'une des causes courantes de l'explosion du gradient est la présence d'initialisations de poids instables ou de taux d'apprentissage trop élevés. Lorsque les poids sont initialisés à de grandes valeurs ou que le taux d'apprentissage est trop élevé, les gradients peuvent rapidement augmenter et devenir ingérables pendant la rétropropagation (Chung, 2014).

2.4.5.7. Long-short term memory -LSTM

Le modèle RNN standard présente des limitations comme mentionné en dessus. Tout d'abord, lorsqu'il est entraîné sur de longues séquences, il peut rencontrer des problèmes d'évanouissement ou explosion du gradient(mais surtout l'évanouissement du gradient), ce qui affecte la convergence de l'entraînement. De plus, le RNN fonctionne mieux lorsqu'il peut utiliser un contexte local pour estimer chaque sortie \hat{y} . Cela signifie que le modèle ne prend en compte que les informations proches de l'instant de prédiction t . Cependant, cette approche à court terme n'est pas optimale lorsque nous devons prédire des événements basés sur de longues séquences temporelles. Afin de remédier à ces problèmes, nous introduisons un modèle plus complexe appelé Long Short-Term Memory (LSTM) qui est un type d'architecture de RNN conçu pour capturer efficacement les dépendances à long terme et traiter le problème du gradient qui s'évanouit (Schmidhuber, 1997).

Les composants clés d'un LSTM sont les suivants :

- 1. Cellule mémoire :** La cellule mémoire est responsable de stocker et de maintenir l'information pertinente sur de longues séquences. Elle agit comme la mémoire à long terme du LSTM.
- 2. Porte d'entrée (input gate) :** La porte d'entrée détermine quelles informations de l'entrée actuelle et de l'état caché précédent doivent être prises en compte et ajoutées à la cellule mémoire.
- 3. Porte d'oubli (forget gate) :** La porte d'oubli décide quelles informations de l'état de la cellule mémoire précédente doivent être oubliées ou supprimées. Cela permet de se débarrasser des informations non pertinentes ou obsolètes.
- 4. Porte de sortie (output gate) :** La porte de sortie contrôle la quantité d'information de l'état de la cellule mémoire qui est utilisée pour générer la sortie du LSTM et l'état caché actuel.

Lors de son fonctionnement, un LSTM reçoit une entrée à chaque pas de temps, ainsi que l'état caché et l'état de la cellule précédente. Il calcule trois portes : la porte d'entrée, la porte d'oubli et

la porte de sortie. La porte d'entrée détermine quelles informations sont pertinentes à partir de l'entrée actuelle et de l'état caché précédent, et doivent être stockées dans l'état de la cellule. La porte d'oubli décide quelles informations doivent être supprimées de l'état de la cellule précédente. La porte de sortie contrôle le flux d'informations de l'état de la cellule vers l'état caché et la sortie du LSTM. L'état de la cellule est mis à jour en combinant l'état de la cellule précédente avec les informations de la porte d'entrée et de la porte d'oubli. L'état caché est calculé à l'aide de l'état de la cellule mis à jour et de la porte de sortie. En mettant à jour dynamiquement l'état de la cellule et en mettant à jour sélectivement l'état caché à l'aide de ces portes, les LSTMs peuvent capturer efficacement les dépendances à long terme et traiter des tâches d'analyse de données séquentielles.

2.4.6. Augmentation des données

L'augmentation des données est une technique largement utilisée dans l'apprentissage automatique pour augmenter artificiellement la taille d'un ensemble de données en créant de nouvelles versions modifiées de points de données existants. Elle consiste à appliquer un ensemble de transformations ou de modifications aux données d'origine, ce qui permet d'obtenir des exemples synthétiques qui conservent les caractéristiques essentielles des informations d'origine. Bien qu'elle soit couramment utilisée dans le traitement des images et des textes, l'augmentation des données peut également être utilisée efficacement dans le contexte des données de séries chronologiques (Wen, 2023).

Dans le domaine de l'analyse des séries temporelles, les techniques d'augmentation des données peuvent être utilisées pour générer des instances supplémentaires de séquences temporelles. En appliquant diverses transformations à la série temporelle originale, telles que l'interpolation et l'injection de bruit (Dilmegani, 2021), il est possible de créer des points de données supplémentaires. Ces séquences transformées fournissent au modèle un ensemble d'exemples plus diversifiés et plus robustes à partir desquels il peut apprendre, ce qui améliore sa capacité à généraliser et à faire des prédictions précises.

2.5. Métriques d'évaluation

Trois mesures différentes sont utilisées pour évaluer les modèles à l'aide des données de test. Supposons que le nombre d'instances de données est m , que la valeur réelle au i ème indice est y et que la valeur prédite est \hat{y} .

2.5.1. Erreur quadratique moyenne (RMSE)

Il s'agit de l'écart-type des résidus (erreurs de prédiction) ou bien la racine carrée de la moyenne de la différence carrée entre la valeur réelle et la valeur prédite et elle est obtenue par :

$$RMSE = \sqrt{\frac{\sum_i^m (\hat{y}_i - y_i)^2}{m}}$$

2.5.2. Erreur absolue moyenne en pourcentage (MAPE)

Elle indique l'écart moyen de la prédiction par rapport à la valeur réelle en pourcentage. Il est donné par la formule suivante,

$$MAPE = \frac{100}{m} \sum_i^m \left| \frac{\hat{y}_i - y_i}{y_i} \right|$$

2.5.3. R² et R² ajusté

Le coefficient de détermination R² et le coefficient de détermination ajusté R² ajusté sont des mesures statistiques utilisées pour évaluer la qualité d'ajustement d'un modèle de régression. Ils fournissent des informations sur la façon dont le modèle s'adapte aux points de données observés.

Le coefficient de détermination R² représente la proportion de la variance de la variable dépendante (la variable à prédire) qui est expliquée par les variables indépendantes (les "predictors") dans le modèle de régression. Il varie de 0 à 1, où 0 indique que le modèle n'explique aucune variance et 1 indique un ajustement parfait. Cependant, le R² seul peut ne pas fournir une évaluation précise des performances du modèle, en particulier lorsque des "predictors" supplémentaires sont ajoutés au modèle.

Le coefficient de détermination ajusté prend en compte le nombre de "predictors" dans le modèle et ajuste en conséquence le R². Il pénalise l'ajout de "predictors" non pertinents qui ne contribuent pas de manière significative à expliquer la variable dépendante. Contrairement au R², qui a tendance à augmenter avec l'ajout de "predictors", le R² ajusté peut diminuer si les "predictors" ajoutés n'améliorent pas significativement le modèle. Par conséquent, le R² ajusté fournit une estimation plus conservatrice de la qualité d'ajustement du modèle. (Li, 2022)

2.6. Conclusion

En conclusion, ce chapitre a introduit les définitions de l'élasticité prix et de différentes stratégies de tarification pour mieux comprendre l'importance du choix du prix. Nous avons également abordé les séries temporelles, étant donné que notre étude se concentre sur une analyse temporelle des données. De plus, nous avons défini les modèles de prédiction qui seront appliqués dans ce travail.

Chapitre 3 :
Collecte, nettoyage et exploration
données de ventes

3.1. Introduction

Ce chapitre résume l'étape de collecte et de traitement des données, en détaillant les étapes suivies. Il présente également les résultats du calcul de l'élasticité prix et notre solution basée sur l'intelligence artificielle.

3.2. Collecte et description des données

La collecte des données a été réalisée à partir du datawarehouse Redshift en utilisant des requêtes PostgreSQL. Pour notre étude, nous nous sommes concentrés sur les ventes en Algérie, en excluant les ventes e-commerce, et la période s'étend de l'ouverture du premier magasin Decathlon le 29 juin 2019 jusqu'au 6 mai 2023.

Nous avons collecté les données à partir de trois tables distinctes :

Voici un schéma simplifié résumant la relation entre ces trois tables

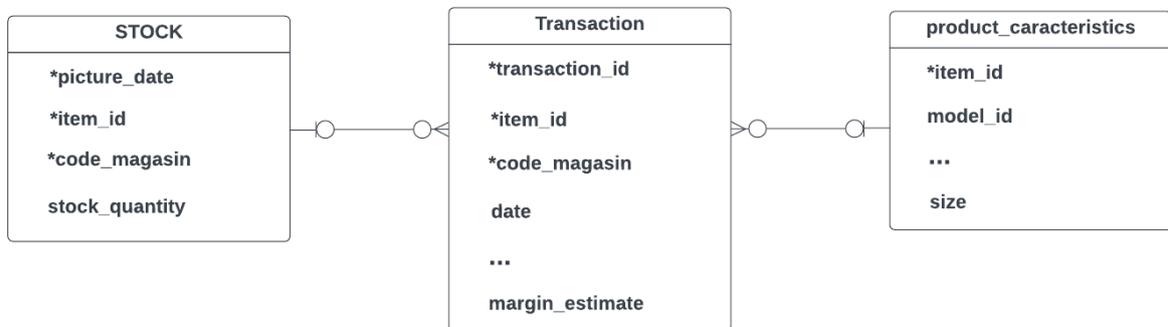


Figure 3.1: Schéma simplifié de la base de données.

Table des transactions : Cette table contient des informations détaillées sur chaque transaction effectuée en magasin. Elle inclut des éléments tels que l'identifiant de la transaction, la date et l'heure de la vente, le montant total de la transaction, les produits achetés et d'autres détails pertinents.

Table 3.1: Description des colonnes de la table des transactions.

Table des transactions	
Nom de la colonne	Description de la colonne
transaction_id	Identifiant du ticket d'achat du client
item_id	L'identifiant du produit acheté avec une granulation la plus petite au niveau de la taille du produit.
Date	L'instant de la transaction
code_magasin	le magasin de l'achat du produit
sale_price_unit	le prix unitaire de chaque produit
quantity	Combien d'instance du produit est acheté
line_num	Le numéro de la ligne du produit acheté sur le ticket de caisse
sale_price	Le prix a payé pour le produit multiplié par la quantité acheté
margin_estimate	La marge estimée sur le produit acheté multiplié par la quantité
discount_amount	La réduction sur le produit acheté multiplié par la quantité

Table des produits : Cette table contient les informations spécifiques sur les produits vendus chez Decathlon. Chaque produit est associé à un identifiant unique, une description, une catégorie, une marque, une taille, un prix et d'autres attributs pertinents.

Table 3.2: Description des colonnes de la table des caractéristiques des produits.

Table des caractéristiques des produits	
Nom de la colonne	Description de la colonne
item_id	L'identifiant du produit acheté avec une distinction au niveau de la taille du produit.
model_id	L'identifiant du produit acheté avec une distinction entre les différentes couleurs des produits
label	le nom du produit
brand	le nom de la marque du produit
univers	l'univers auquel appartient le produit
product_nature	le type de produit
family	la famille à laquelle le produit appartient
category	la catégorie à laquelle le produit appartient
size	la taille du produit acheté

Table du stock : La table du stock contient des informations telles que la date de l'inventaire, la quantité disponible et le code d'identification de chaque article. Cette table permet de suivre et de gérer les niveaux de stock des produits chez Decathlon.

En enregistrant la date de l'inventaire, nous sommes en mesure de connaître le moment précis où les informations sur le stock ont été recueillies. La quantité indiquée représente la quantité disponible de chaque article à cette date spécifique. Le code d'identification de l'article permet de faire correspondre les données du stock à d'autres informations pertinentes dans le système.

Cette table est essentielle pour la gestion du stock chez Decathlon, permettant de suivre les niveaux de stock, de détecter les éventuelles ruptures de stock, de planifier les réapprovisionnements et d'optimiser la disponibilité des produits pour répondre à la demande des clients.

Table 3.3 : Description des colonnes de la table du stock.

Table de l'état du stock	
Nom de la colonne	Description de la colonne
picture_date	L'identifiant du produit acheté avec une distinction au niveau de la taille du produit.
item_id	L'identifiant du produit acheté avec une distinction entre les différentes couleurs des produits
stock_quantity	le nom du produit
code_magasin	le nom de la marque du produit
Code	Indicateur de la performance des ventes du produit

3.3. Nettoyage des données

État de la table avant le preprocessing :

Table 3.4 : État de l'ensemble de données avant le traitement.

État de l'ensemble de données avant le nettoyage des données	
La Taille de la table	(5 268 684, 18)
Le nom des colonnes	transaction_id, item_id, date, code_magasin, sale_price_unit, quantity, line_num, sale_price, margin_estimate, discount_amount, model_id, label, brand, univers, product_nature, family, category, size

Après avoir centralisé les données de la table des transactions et de la table des caractéristiques des produits dans une même table, des étapes de nettoyage de données sont nécessaires afin de pouvoir utiliser ces données :

Éliminer les ventes B2B :

Pour quelques transactions la quantité inscrite pour un même produit est grande ce qui peut se traduire par une transaction B2B, on considérera dans notre étude que pour une transaction et pour un même produit qui a une quantité d'au-delà de cinq comme une transaction B2B qui faut éliminer.

Table 3.5: État l'ensemble de données après l'élimination du B2B.

État de l'ensemble de données après avoir appliqué l'étape ci-dessus	
La Taille de la table	(5 258 038, 18)
Nom des colonnes	les mêmes que l'étape précédente

Éliminer des transactions avec une quantité nulle :

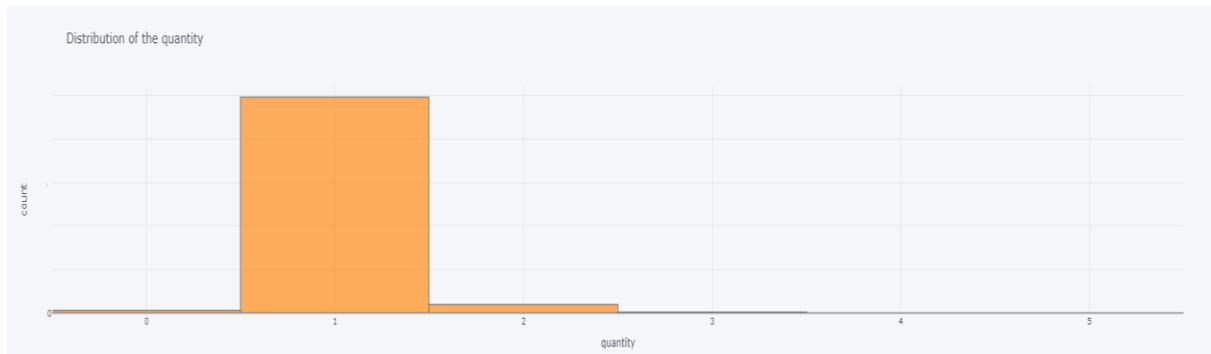


Figure 3.2: Distribution de la variable quantité .

Certaines transactions présentent une quantité nulle. Ces transactions sont défectueuses et doivent être supprimées.

Table 3.6 : État de l'ensemble de données après l'élimination des transactions avec une quantité nulle.

État de l'ensemble de données après avoir appliqué l'étape ci-dessus	
La Taille de la table	(5 193 245, 18)
Nom des colonnes	les mêmes que l'étape précédente

Éliminer les transactions avec comme label Seconde vie :

À Decathlon, un service permet aux particuliers de vendre leurs produits dans les magasins Decathlon, ces transactions sont inscrites avec le label Seconde vie et il n'entre pas dans le cadre de notre étude. Il faut donc éliminer toutes les transactions avec le label "SECONDE VIE".

Table 3.7: État de l'ensemble de données après l'élimination des transactions "Seconde vie".

État de l'ensemble de données après avoir appliqué l'étape ci-dessus	
La Taille de la table	(5 186 479, 18)
Nom des colonnes	les mêmes que l'étape précédente

Éliminer les transactions avec des produits pas assez vendus :

Après avoir effectué le calcul de la quantité totale vendue pour chaque produit, il a été constaté que certains produits n'ont pas atteint un niveau de vente suffisant pour pouvoir fournir assez de données pour les étudier. Cela peut être attribué à des problèmes de réapprovisionnement, qui ont entraîné des ventes insuffisantes pour ces produits.

Ces produits dont le total des ventes est inférieur à 100 au cours des quatre années ne seront pas pris en compte dans notre étude. Ainsi, pour qu'un produit soit inclus, il doit avoir une moyenne de vente d'au moins deux unités par mois.

Table 3.8 : État de l'ensemble de données après l'élimination des produits peu vendus.

État de l'ensemble de données après avoir appliqué l'étape ci-dessus	
La Taille de la table	(4 701 890, 18)
Nom des colonnes	les mêmes que l'étape précédente

Éliminer les produits avec une valeur nulle dans la colonne code :

La colonne "code" définit la fréquence d'achat de chaque article, allant de SA à F. Les produits SA correspondent aux articles les plus vendus, tandis que les produits F sont les moins vendus. Ces codes ont été déterminés par l'équipe d'approvisionnement de Decathlon El Djazair. Ainsi, les produits qui ne disposent pas de code ne sont pas considérés comme des éléments intéressants pour notre étude.

Table 3.9 : État de l'ensemble de données après l'élimination des produits sans code.

État de l'ensemble de données après avoir appliqué l'étape ci-dessus	
La Taille de la table	(3 405 609, 18)
Nom des colonnes	les mêmes que l'étape précédente

Grouper les données par jour :

Pour passer de la représentation de chaque ligne comme quantité vendue dans une transaction pour chaque produit à la quantité vendus pour chaque article et dans chaque magasin par jour, nous devons regrouper les données par "date", "item" et "code_magasin", puis agréger les autres variables de la manière suivante :

"sale_price_unit": moyenne

"margin_estimate": somme

"quantity": somme

"sale_price": somme

Table 3.10 : État de l'ensemble de données après le groupement par jour.

État de l'ensemble de données après avoir appliqué l'étape ci-dessus	
La Taille de la table	(1 602 736, 16)
Nom des colonnes	item_id, date, code_magasin, sale_price_unit, quantity, sale_price, margin_estimate, discount_amount, model_id, label, brand, univers, product_nature, family, category, size

Traitement des dates manquantes :

Après avoir regroupé les données et obtenu, pour chaque magasin, jour et date, la quantité vendue, nous avons remarqué qu'il y avait plusieurs dates manquantes correspondant à des jours où la quantité vendue était nulle. Il est donc nécessaire d'ajouter les dates manquantes pour chaque produit et code magasin.

Table 3.11 : État de l'ensemble de données après le traitement des dates manquantes .

État de l'ensemble de données après avoir appliqué l'étape ci- dessus	
La Taille de la table	(12 683 592, 16)
Nom des colonnes	les mêmes que l'étape précédente

Faire la jointure avec la table du stock :

maintenant que nous avons pour chaque jour et magasin et produit une valeur, il devient facile de faire la jointure avec la table du stock en se basant sur les 3 colonnes qui sont communes entre les deux tables :

“Item”, “code_magasin” et “date”

Nous aurons donc dans notre table une nouvelle variable “stock_quantity” qui représente le nombre d'unités en stock pour chaque produit et pour chaque date et magasin.

Table 3.12 : État de l'ensemble de données après la jointure avec le stock .

État de l'ensemble de données après avoir appliqué l'étape ci- dessus	
La Taille de la table	(12 683 592, 17)
Nom des colonnes	les mêmes que l'étape précédente + stock_quantity

Ajout des variable :

L'ajout de variables peut être crucial pour un modèle de prédiction, car cela permet d'ajouter d'autres facettes d'analyse. Chaque variable ajoutée peut fournir des informations supplémentaires qui peuvent aider à mieux comprendre les relations entre les entrées et les sorties qui sont souvent complexes, et ainsi améliorer la précision du modèle.

Nous avons créé deux nouvelles variables pour capturer les fluctuations de ventes en fonction du jour de la semaine : "day_of_week" et "weekend"

"day_of_week" est une variable numérique allant de 0 à 6, où 0 correspond à dimanche et 6 correspond à samedi. Cette variable nous permettra de comprendre comment les ventes varient en fonction du jour de la semaine.

"weekend" nous avons créé une variable booléenne qui prend la valeur 0 pour les jours du weekend et 1 pour les jours de la semaine. Cette variable nous permettra de comprendre comment les prix varient entre les jours de la semaine et les jours du weekend.

"holiday" cette variable est un indicateur de la nature de la journée en question. Elle prend trois valeurs distinctes : 0, 1, ou 2. La valeur 0 correspond à une journée où les magasins sont fermés, par exemple lors de la fête de l'Aïd. La valeur 1 correspond à une journée normale sans fête particulière. Enfin, la valeur 2 correspond à une journée spéciale où les magasins sont ouverts, telle que la fête des mères ou la fête du travail.

"season" : Est une variable catégorique qui indique la saison à laquelle appartient la date en question. Elle est divisée en quatre catégories distinctes : Hiver, Printemps, Été et Automne. Cette variable est utile pour comprendre la variation saisonnière des données.

"num_magasin": Elle représente le nombre de magasins de Decathlon ouverts en Algérie à une date donnée. Cette variable est importante pour étudier l'effet de la cannibalisation entre les magasins, c'est-à-dire comment la présence d'un magasin peut affecter les ventes d'un autre magasin de la même chaîne.

Table 3.13 : État de l'ensemble de données après l'ajout des variables.

État de l'ensemble de données après avoir appliqué l'étape ci- dessus	
La Taille de la table	(12 683 592, 22)
Nom des colonnes	item_id, date, code_magasin, sale_price_unit, quantity, sale_price, margin_estimate, discount_amount, stock_quantity, model_id, label, brand, univers, product_nature, family, category, size, day_of_week, weekend, holiday, season, num_magasin

Traitement des valeurs manquantes :

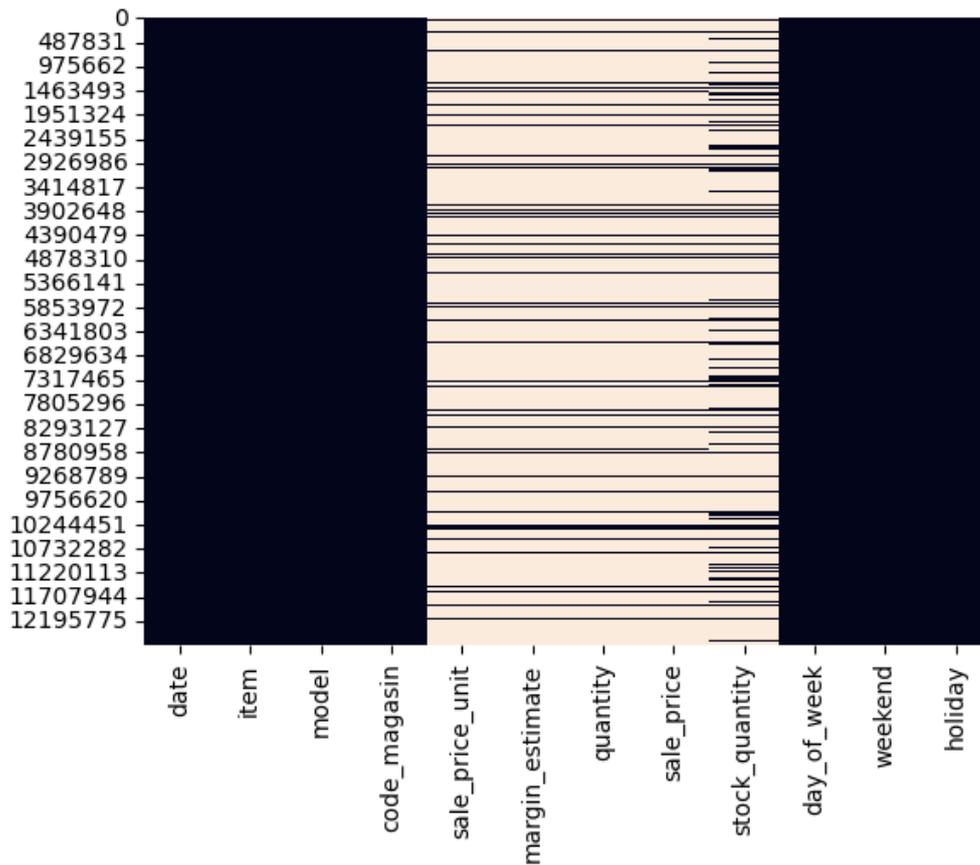


Figure 3.3: Matrice montrant la proportion de valeurs manquantes qu'il y a dans chaque colonne.

Dans les colonnes "sale_price", "quantity" et "margin_estimate" :

Les colonnes "sale_price", "quantity" et "margin_estimate" contiennent des valeurs manquantes qui correspondent aux jours où il n'y a eu aucune activité commerciale. Ces jours ont été ajoutés pour combler les vides entre les différentes dates que nous avons. Par conséquent, nous avons rempli ces valeurs manquantes avec la valeur 0 pour refléter l'absence d'activité durant ces jours.

Dans les jours ajoutés pour combler les vides, la variable "sale_price_unit" est restée vide. Pour pallier ce manque, nous avons tout d'abord ordonné le dataset selon les colonnes "item", "code_magasin" et "date". Nous avons ensuite appliqué la méthode de remplissage vers l'avant (forward filling), qui consiste à remplir les valeurs manquantes avec la dernière valeur disponible pour la variable "sale_price_unit" dans l'historique. Cela nous permet d'obtenir la valeur de prix unitaire pour la période où il n'y a pas eu de transactions, et ainsi avoir un historique complet des prix appliqués pour chaque produit.

Certains jours manquent des enregistrements pour la variable d'état du stock. Pour pallier ce manque, nous avons utilisé une méthode de remplissage vers l'avant de la même

manière qu’avec la colonne de “sale_price_unit”. Concrètement, nous avons rempli les valeurs manquantes en utilisant la dernière valeur disponible pour la variable d’état du stock dans l'historique. Cette approche nous a permis d'obtenir des estimations fiables de l’état du stock pour les jours où les données manquent.

Calcul d’élasticité pour chaque produit :

Dans cette section, nous allons explorer et calculer le coefficient d’élasticité pour chaque produit afin de mieux comprendre les dynamiques de marché auxquelles ils sont soumis. Il est important de noter que cette approche repose sur l'hypothèse que la demande fluctue linéairement par rapport au prix, ce qui peut ne pas toujours refléter la réalité.

En effet, la demande est souvent influencée par une multitude de facteurs tels que la présence de produits concurrents, les préférences des consommateurs, les évolutions économiques, et bien d'autres variables.

Toutefois, malgré ces limitations, l’évaluation du coefficient d’élasticité se révèle être un outil précieux pour estimer l'impact des variations de prix sur les ventes de chaque produit. Cela étant dit, il peut s'avérer complexe de déterminer l’élasticité pour un produit donné, car la performance des ventes dépend de nombreuses variables autres que le prix. Parmi ces variables, on peut citer les conditions météorologiques, l'état des routes, la disponibilité des stocks, les campagnes publicitaires, ou encore la notoriété de la marque auprès des consommateurs.

Pour faciliter notre analyse, nous ferons l'hypothèse que les différents prix pratiqués pour un produit sont soumis aux mêmes conditions externes. Cette simplification nous permet d'isoler l'effet du prix sur les ventes. Bien que cette approche puisse sembler quelque peu simpliste, elle nous offre néanmoins un aperçu précieux de l'influence du prix sur les performances de vente de chaque produit.

Pour calculer l’élasticité pour chaque produit il faut procéder à une transformation des données comme suit :

Pour assurer l'intégrité des données relatives à la performance des différents prix d'un produit, il est essentiel de ne pas inclure les jours où le stock était indisponible. Il est donc recommandé de filtrer les données pour ne conserver que les jours où le stock du produit choisi était disponible. Cela permettra d'obtenir une analyse plus précise et fiable de la performance des différents prix du produit, en éliminant les biais potentiels liés à la disponibilité du stock.

Grouper les jours par “sale_price_unit” et faire les agrégation suivantes :

“sale_price” : moyenne

“quantity” : moyenne

“margin_estimate” : moyenne

De cette manière, nous pourrions calculer la moyenne quotidienne de la quantité vendue, du revenu généré et de la marge bénéficiaire pour chaque prix, pour chaque magasin.

Exemple d'un résultat pour un certain produit :

Table 3.14 : Calcul des quantité moyenne vendue par jours et par magasin

Prix unitaire (DA)	quantité moyenne vendue par jours et par magasin
7000	36,58
7400	50,44
7800	9,32
8000	39,04
9000	78,54
11600	3,56

En calculant la moyenne supérieure des prix et la moyenne inférieure des prix ainsi que la moyenne supérieure de la quantité ainsi qu'inférieure du tableau classé dans un ordre croissant vis-à-vis du prix, nous obtenons ces valeurs suivantes :

Table 3.15 : Tableau représentatif du calcul de la moyenne supérieure et inférieure des prix unitaire et quantité

	Prix unitaire		Quantité
Prix 1 (p1)	7400	Quantité 1 (q1)	32,11
Prix 2 (p2)	9533,33	Quantité 2 (q2)	40,38

En utilisant la formule suivante de l'élasticité-prix moyenne :

$$e = \frac{\frac{q_2 - q_1}{q_1}}{\frac{p_2 - p_1}{p_1}}$$

L'élasticité calculée pour ce produit est de 0,85, ce qui signifie qu'en moyenne, une augmentation de 1% des prix entraîne une augmentation de 0,85% de la demande. Cependant, il convient de noter que ce résultat peut sembler inhabituel et peut être dû à des variations dans la période à laquelle chaque prix a été appliqué, ainsi qu'aux

caractéristiques propres à chaque période. Par conséquent, il est important de prendre en compte ces facteurs lors de l'interprétation de l'élasticité-prix et de ses implications pour la demande du produit.

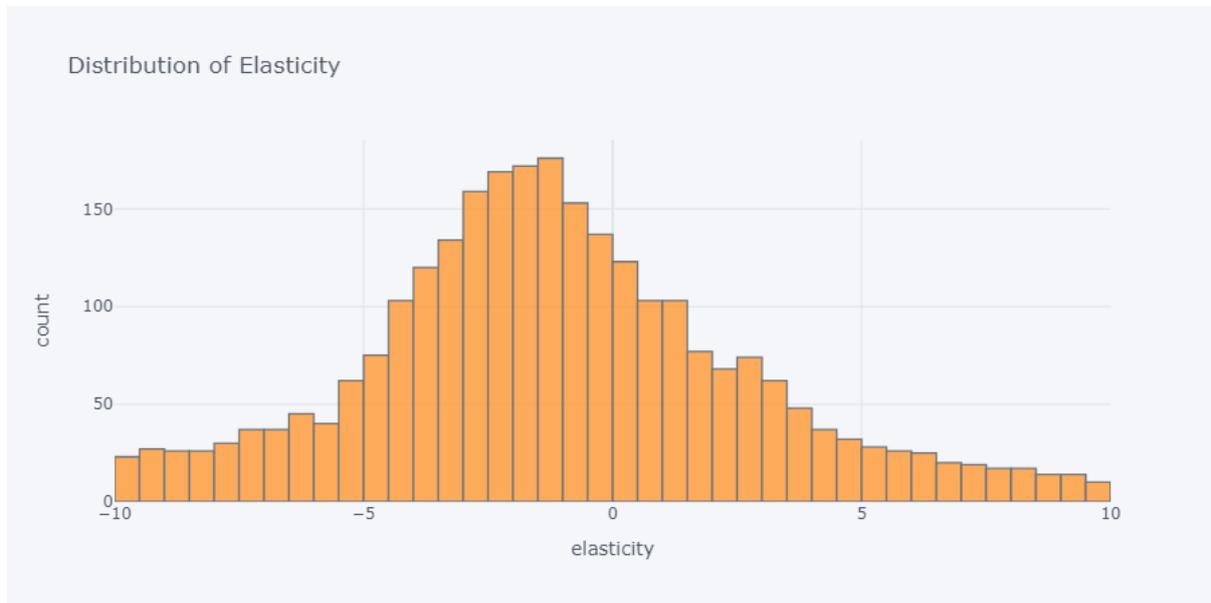


Figure 3.4: Histogramme de la distribution de l'élasticité calculée des produits.

Élasticité ajustée : Pour ajuster l'élasticité-prix et prendre en compte les effets de saisonnalité et les performances de vente différentes dans chaque magasin, une méthode courante consiste à calculer l'élasticité-prix entre des prix appliqués dans un seul magasin sur une même période de l'année. En utilisant cette approche, les variations dues aux différences de performance des magasins et aux fluctuations saisonnières sont minimisées, offrant ainsi une mesure plus fiable de l'élasticité-prix pour le produit. Il convient de noter que le calcul de l'élasticité-prix entre des prix appliqués dans un seul magasin sur une même période de l'année n'est pas toujours possible pour tous les produits. En effet, certains produits peuvent ne pas avoir de prix appliqué dans la même période en raison de leur disponibilité limitée. En laissant que les prix qui ont duré durant tout un mois et en groupant par mois la somme de la quantité vendus, on obtient le tableau suivant qui concerne le mois de Mars pour différentes années :

Table 3.16 : Tableau présentant les quantités vendues dans chaque magasin pour le mois de mars, pour plusieurs années différentes

Magasin	Année	Prix unitaire	quantité
1	2022	7000	838
1	2021	7400	1692
1	2023	11600	132
2	2022	7000	564
2	2021	7400	682
2	2023	11600	82
3	2022	7000	682
3	2021	7400	1888
3	2023	11600	114
4	2022	7000	410
4	2023	11600	66

En calculant l'élasticité de la même façon pour le produit dans chaque magasin, on obtient:

Table 3.17 : Tableau présentant pour chaque magasin l'élasticité calculée

Magasin	Moyenne de la tranche supérieure des prix	Moyenne de la quantité pour les prix de la tranche supérieure par mois	Moyenne de la tranche inférieure des prix	Moyenne de la quantité pour les prix de la tranche inférieure	Élasticité
1	7200	1265	9500	912	-0,87
2	7200	623	9500	382	-1,21
3	7200	1285	9500	1001	-0,69
4	7000	410	11600	66	-1,28

Nous remarquons que les coefficients d'élasticité pour chaque magasin variant de -1,28 jusqu'à -0,61 avec une moyenne globale de -1,01.

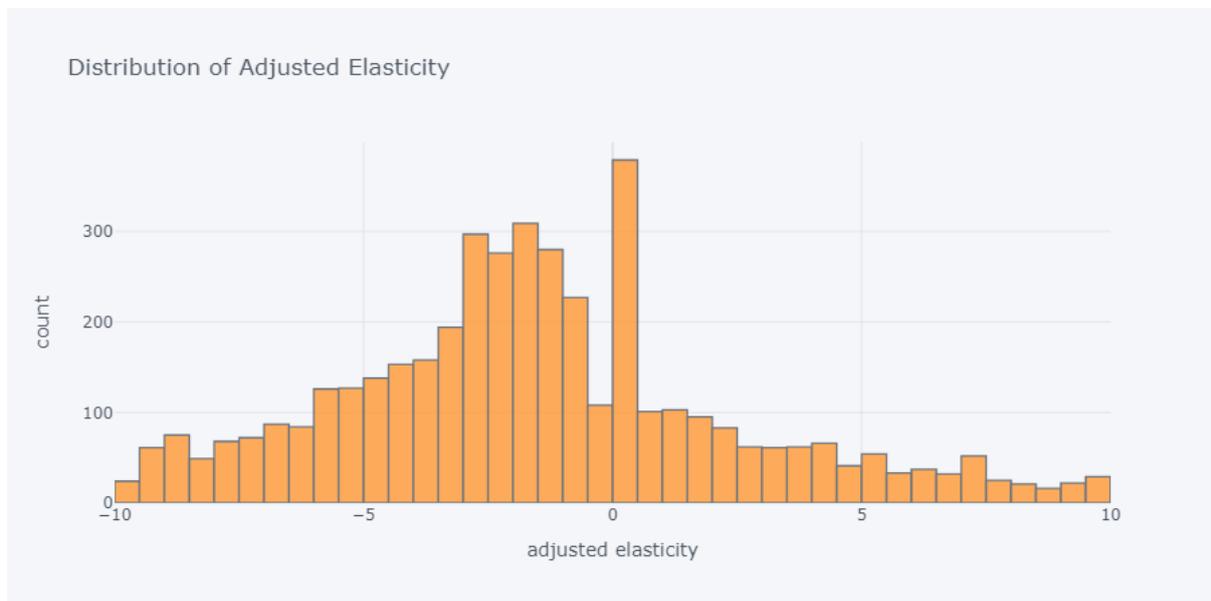


Figure 3.5: Histogramme de la distribution de la moyenne globale de l'élasticité ajustée des produits.

En conclusion, nous avons calculé deux coefficients d'élasticité pour les produits de Decathlon, en utilisant deux méthodes distinctes. La première méthode consistait à agréger les données en fonction du prix, tandis que la deuxième méthode prenait en compte la performance des ventes à différents prix pour une même période de l'année et pour un même magasin. Cette approche a permis d'éliminer l'effet de la saisonnalité, ainsi que l'effet de performance de vente propre à chaque magasin.

Ces deux coefficients d'élasticité fournissent des informations précieuses sur la sensibilité des ventes de chaque produit par rapport à leur prix. En comparant les résultats obtenus avec les deux méthodes, nous pouvons identifier les produits qui présentent une élasticité-prix particulièrement élevée ou faible et ainsi, mieux comprendre l'impact des variations de prix sur les ventes.

Ces connaissances permettront à Decathlon de mieux adapter sa stratégie de tarification, en tenant compte de l'élasticité-prix de chaque produit, afin d'optimiser les ventes et la rentabilité. À long terme, cette analyse peut aider l'entreprise à être plus réactive face aux fluctuations du marché et à améliorer sa position concurrentielle.

Cependant, il convient de souligner que cette analyse a ses limites. En effet, bien que les coefficients d'élasticité calculés fournissent des indications précieuses sur la sensibilité des ventes aux variations de prix, ils ne permettent pas de prédire les performances des prix jamais appliqués. De plus, l'analyse se concentre sur les données de ventes passées et ne prend pas en compte les changements potentiels dans le comportement des consommateurs ou les tendances du marché.

Il convient également de souligner que les deux méthodes utilisées pour calculer les coefficients d'élasticité ont leurs propres limites. La première méthode, qui agrège les données en fonction du prix, peut être influencée par des facteurs tels que les écarts de prix entre les magasins ou les variations saisonnières. La deuxième méthode, qui prend en compte la performance des ventes à différents prix pour un même magasin et une même période, peut être limitée par la disponibilité des données et la difficulté de trouver des périodes comparables.

Pour surmonter ces limites et fournir des prévisions plus précises et plus dynamiques de l'élasticité-prix, une solution d'intelligence artificielle pourrait être envisagée.

3.4. Proposition d'une Solution IA

Notre deuxième solution consiste à créer un simulateur basé sur l'intelligence artificielle (IA) qui utilise l'historique des ventes comme entrée pour fournir des prévisions de performances de vente pour n'importe quel prix. Nous utilisons les changements de prix présents dans les données historiques pour permettre au modèle d'apprentissage automatique de déterminer la relation entre les prix et les ventes. Notre hypothèse est que si notre modèle peut prédire avec précision les ventes pour les données comportant des changements de prix, il devrait être en mesure de prédire les ventes pour un prix qu'il n'a jamais rencontré. En effet, notre modèle aura déjà identifié la

formule qui relie les ventes au prix, lui permettant d'extrapoler ces connaissances à de nouvelles situations.

Cependant, nous sommes conscients que la précision de nos prédictions dépendra de la qualité et de la taille des données utilisées pour entraîner notre modèle. Nous avons donc adopté une approche proactive en augmentant la taille de notre base de données grâce à l'ajout de données bruitées, tout en nous assurant que notre ensemble de données est complet et représentatif. Nous prenons également en compte d'autres facteurs pertinents qui peuvent influencer les ventes, tels que la saisonnalité, le niveau de stock et les jours fériés.

Une fois que notre modèle aura atteint des résultats satisfaisants, nous le testerons avec plusieurs nouveaux prix pour déterminer s'il a pu trouver la bonne relation entre le prix et les ventes.

En raison de la taille limitée de notre base de données et de la variabilité des comportements de vente d'un produit à l'autre. Ainsi, pour simplifier notre étude et améliorer la précision de notre modèle de prédiction, on a décidé de travailler sur un seul produit.

3.5. Choix d'un produit pour la modélisation prédictive

L'étape d'analyse exploratoire consiste à sélectionner un produit spécifique comme objet d'étude, en raison de la taille limitée de notre base de données et de la variabilité des comportements de vente d'un produit à l'autre. Ainsi, pour simplifier notre étude et améliorer la précision de notre modèle de prédiction des ventes, nous avons opté pour le produit qui présente les meilleures performances de vente.

3.5.1. Choix de la marque

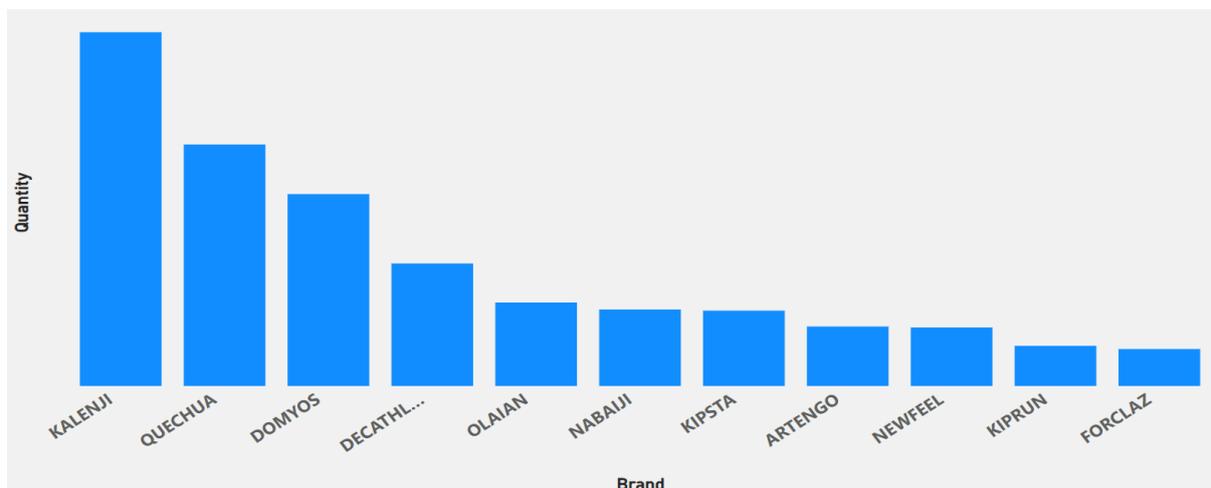


Figure 3.6 : Graphique en barres montrant la quantité vendue pour chaque marque de Decathlon.

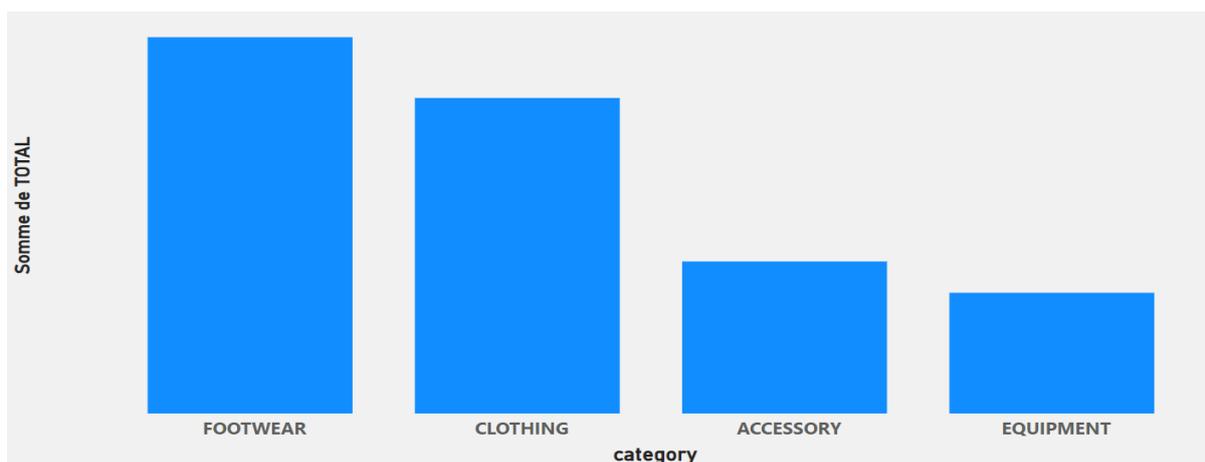


Figure 3.7: Graphique en barres montrant la quantité vendue pour chaque catégorie.

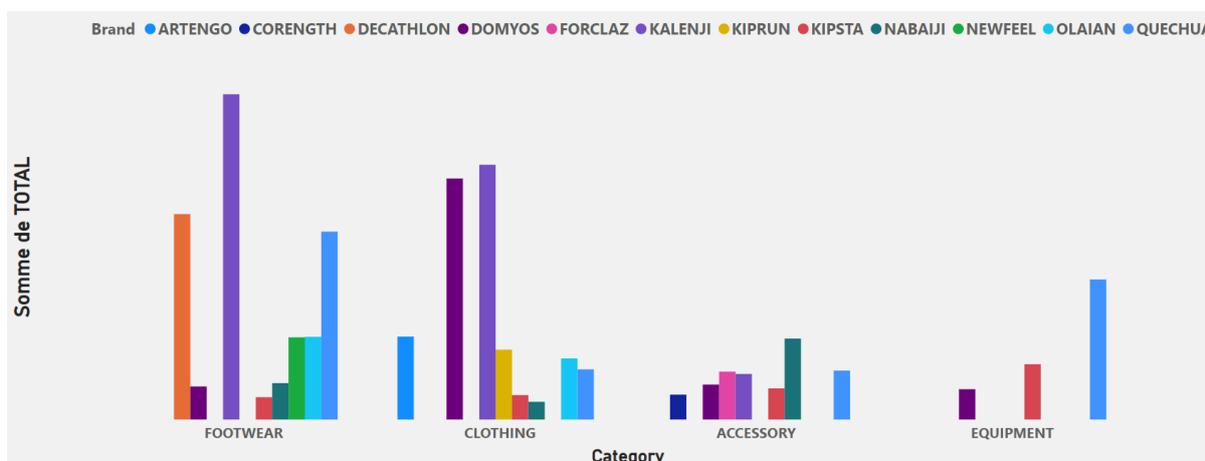


Figure 3.8: Graphique en barres montrant la quantité vendue pour chaque catégorie répartie par marque.

Les histogrammes présentés auparavant démontrent que les chaussures sont les produits les plus vendus chez Decathlon. Kalenji est la marque la plus populaire parmi toutes les marques de chaussures et toutes les autres catégories (vêtements, accessoires, équipement) depuis l'ouverture de Decathlon en Algérie. Cette popularité s'explique par les prix abordables proposés par Kalenji, qui attirent particulièrement les clients de la classe moyenne. Il est également important de souligner que les produits de Kalenji sont polyvalents et peuvent être utilisés pour d'autres activités que leur fonction initiale, on appelle ça "l'extension d'usage". Un exemple courant, une chaussure de course peut être utilisée pour la randonnée ou le travail.

3.5.2. Choix du magasin

Les performances de ventes varient d'un magasin à un autre. En effet, nous avons choisi de limiter notre étude au magasin de Bab Ezzouar, car il s'agit du premier magasin ouvert en Algérie, ce qui nous permet d'avoir un maximum de données par rapport à un autre magasin. De plus, le stock de produits de la marque Kalenji, est majoritairement disponible à Bab Ezzouar. Par

ailleurs, il convient de souligner que le magasin de Bab Ezzouar ne semble pas être impacté par la concurrence des autres magasins. Les ventes et le chiffre d'affaires y demeurent (à peu près) stables, ce qui suggère qu'il n'y a pas un effet de cannibalisation, c'est à dire que les magasins voisins, tels que le magasin Ardis ou le magasin Chéraga, n'attirent pas la clientèle de Bab Ezzouar. Cette observation renforce la fiabilité de notre étude.

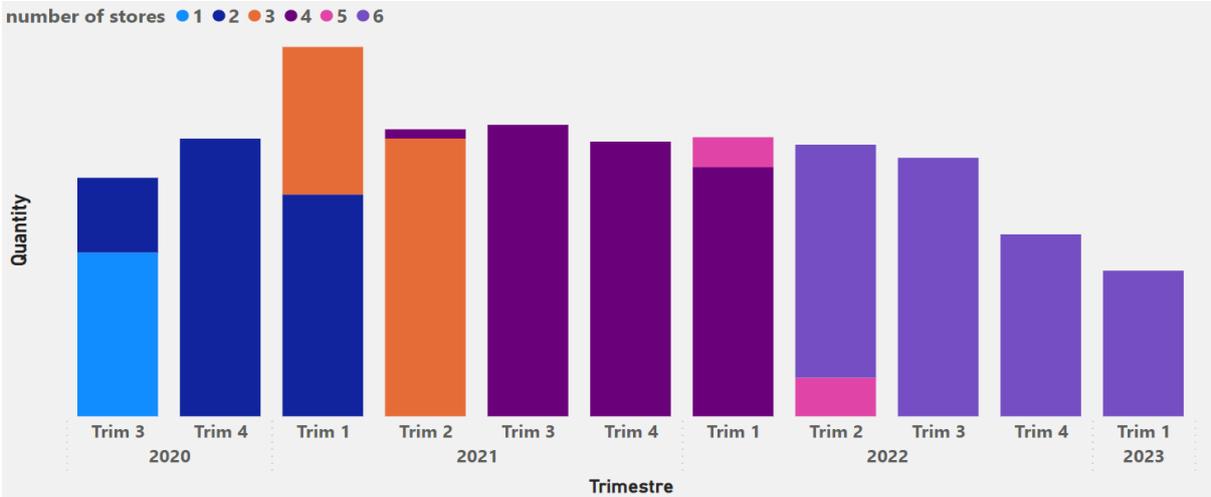


Figure 3.9: Analyse des ventes à Bab Ezzouar en fonction du nombre de magasins Decathlon ouverts durant la même période.

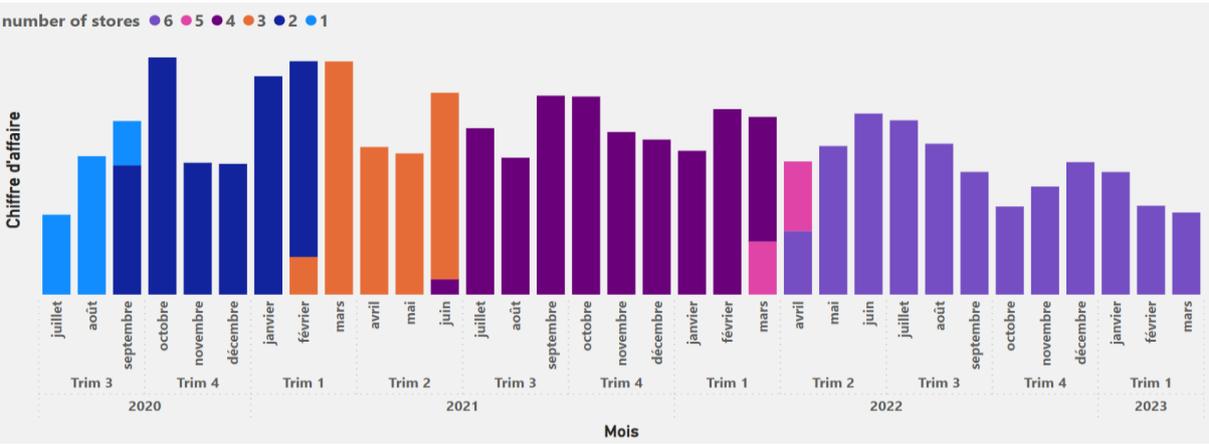


Figure 3.10: Analyse du chiffre d'affaires généré au magasin de Bab Ezzouar en fonction du nombre de magasins de Decathlon durant la même période.

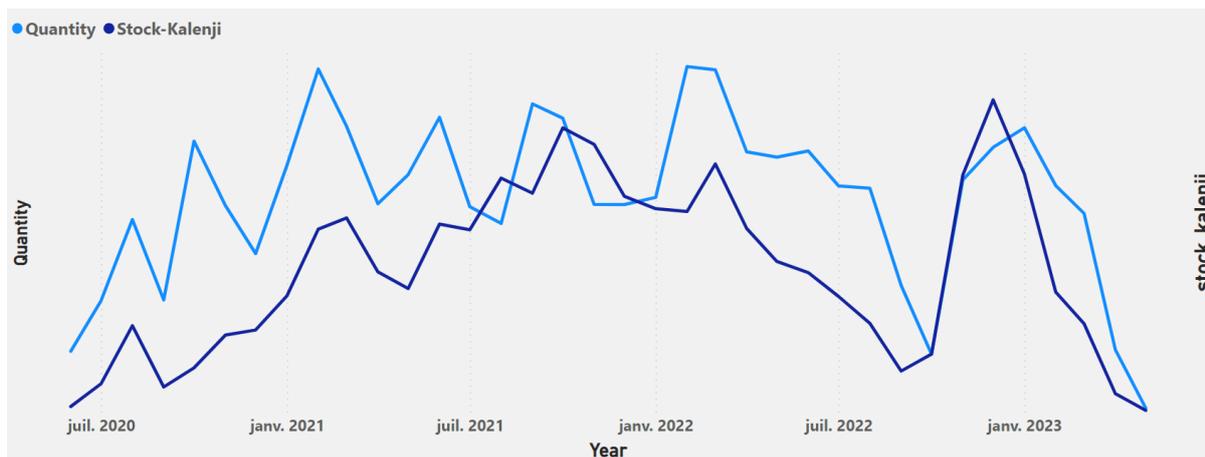


Figure 3.11 : Graphique des ventes mensuelles du stock pour la marque Kalenji au magasin de Bab Ezzouar.

3.5.3. Choix d'un item

Comme décrit dans la section de description des données, chaque produit est identifié par un code modèle et un code article. Il convient de noter que le code modèle varie en fonction de la couleur, tandis que le code article varie en fonction de la pointure pour un même produit.

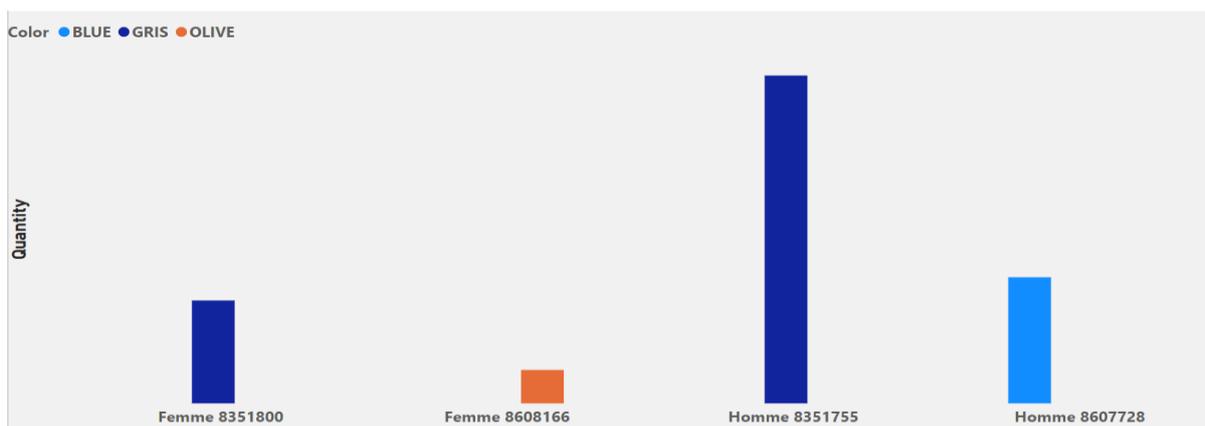


Figure 3.12 : Répartition des quantités par genre et couleur (code modèle).

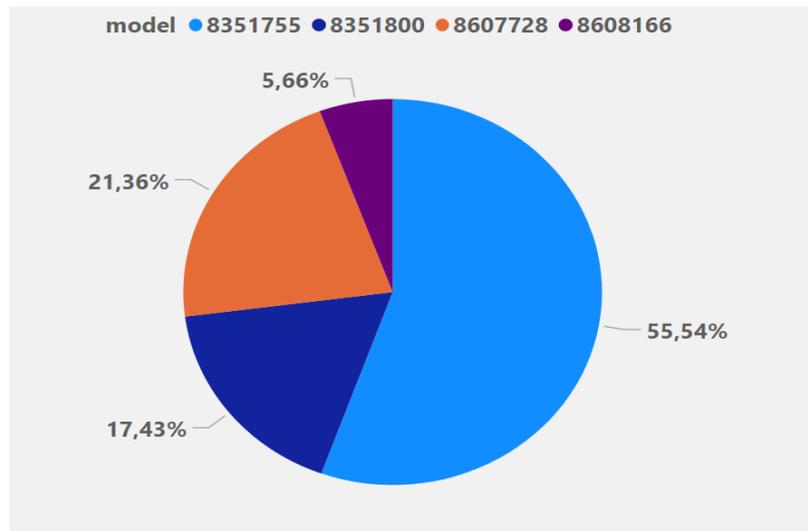


Figure 3.13: Graphique circulaire montrant la répartition des ventes pour les 5 modèles.

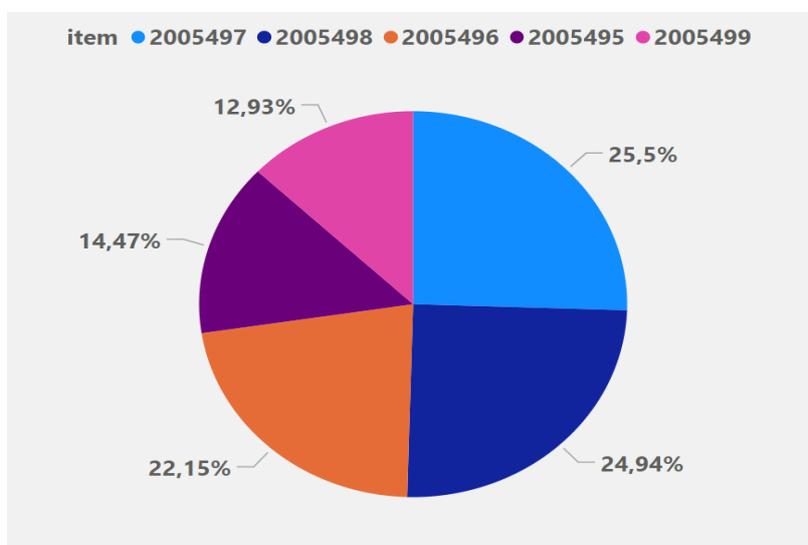


Figure 3.14: Graphique circulaire montrant la répartition des ventes pour les 5 "items"

D'après les graphiques présentés ci-dessus, il est clair que la majorité des acheteurs de Kalenji sont des hommes, et que leur couleur préférée est le "Gris", qui correspond au code de modèle 8351755. Après avoir analysé les ventes de ce modèle, nous avons identifié l'article le plus vendu comme étant le 2005497.



Figure 3.15: Graphique en barres montrant les quantités vendues pour chaque prix de vente pour “item” 2005497.

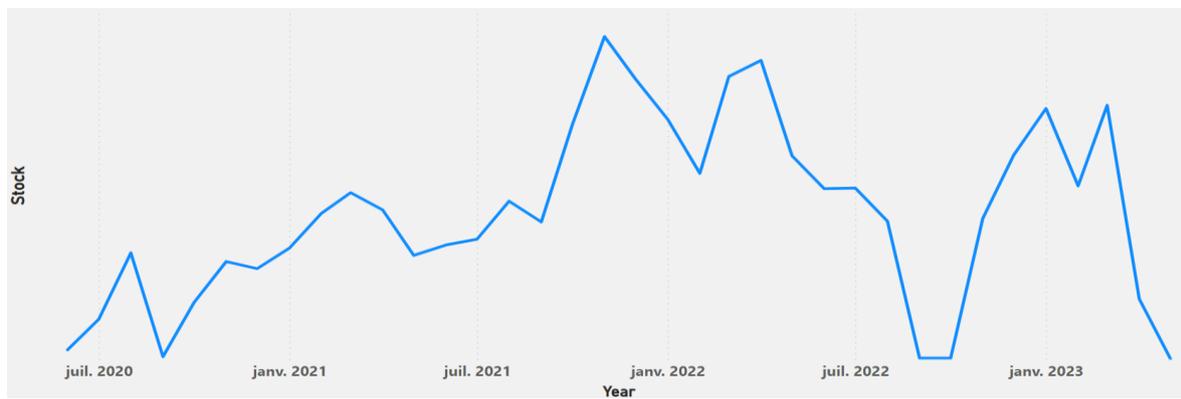


Figure 3.16 : Graphique en ligne montrant l'évolution de l'état du stock pour “item” 2005497 au magasin de Bab Ezzouar.

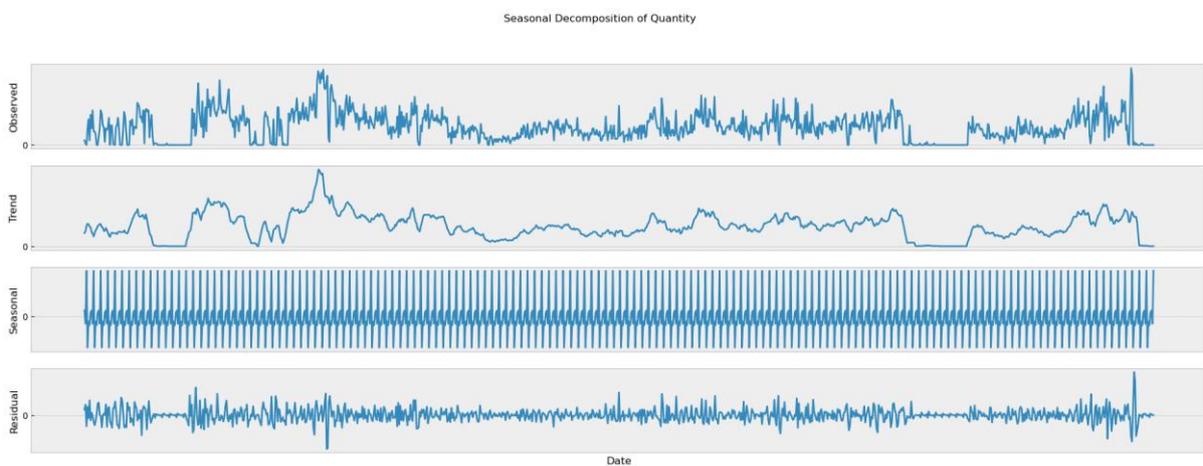


Figure 3.17 : Décomposition de la série temporelle en : tendance, saisonnalité et composante irrégulière pour une période de 7 jours.

Il est intéressant de noter que cet article a été soumis à quatre changements de prix et que son stock est principalement disponible à Bab Ezzouar. De plus, Ce produit est caractérisé par une tendance de vente non linéaire et une forte saisonnalité de période égale à sept jours, c'est-à-dire que le comportement des consommateurs suit un schéma hebdomadaire.

Afin de résumer les résultats de l'analyse exploratoire des données, nous avons sélectionné l'article 2005497 de la marque Kalenji pour notre étude. Ce choix s'est basé sur plusieurs critères importants : tout d'abord, cet article contribue de manière significative au chiffre d'affaires de Decathlon. De plus, il dispose d'un stock disponible la plupart du temps, ce qui nous permettra d'éliminer les effets d'une rupture de stock sur les ventes. En outre, étant donné sa fréquence de vente élevée, nous disposons d'un ensemble de données plus important pour entraîner notre modèle par rapport aux autres produits. Enfin, nous avons remarqué que le prix de cet article avait changé à quatre reprises, ce qui est bénéfique pour notre modèle d'apprentissage automatique ou d'apprentissage profond.

3.6.Conclusion

Dans ce chapitre, nous avons décrit les étapes de collecte et de traitement des données. Nous avons également présenté les résultats du calcul de l'élasticité-prix et expliqué pourquoi nous avons opté pour une solution basée sur l'intelligence artificielle. Enfin, nous avons discuté des critères qui ont guidé notre choix de produits qui est l'objet de cette étude.

Chapitre 4 :
Méthodologie de la résolution de la
problématique et présentation des
résultats

4.1. Introduction

Cette section décrit toutes les étapes précédant la modélisation, de l'évaluation de la stationnarité de la série temporelle au test des modèles de prédiction. Les résultats de l'évaluation sont présentés à travers les métriques de R carré et de R carré ajusté. Nous avons également effectué une simulation de prédiction pour de nouveaux prix de vente, mettant en évidence la valeur ajoutée de notre solution.

4.2. Méthodologie de la résolution de la problématique

Voici les étapes que nous avons suivies pour la création de notre modèle de prédiction :

4.2.1. Évaluation de la stationnarité de la série temporelle

Pour évaluer la stationnarité de notre série temporelle, nous avons utilisé le test de Dickey-Fuller augmenté (ADF), qui est couramment utilisé pour déterminer si une série temporelle possède des racines unitaires, ce qui indiquerait une non-stationnarité.

Nous avons d'abord évalué la stationnarité de notre série temporelle sans faire de data augmentation, et avons obtenu une p-value de 5.55×10^{-5} , ce qui est inférieur à la valeur de seuil couramment utilisée de 0.05. Par conséquent, nous avons rejeté l'hypothèse nulle (H_0) selon laquelle la série temporelle est non-stationnaire et conclu que la série temporelle est stationnaire.

Ensuite, après avoir appliqué une data augmentation, nous avons obtenu une p-value de 3.38×10^{-7} , ce qui est également inférieur à 0.05. Cela implique que l'hypothèse nulle H_0 est rejetée et que la série est donc stationnaire.

Bien que la série temporelle soit déjà stationnaire, mais l'application d'une transformation supplémentaire pourrait améliorer la précision du modèle de prédiction. Alors on a appliqué une transformation logarithmique à la série temporelle et avons obtenu une p-value de

9.92×10^{-6} .

Cependant, il est important de noter que la p-value seule ne suffit pas pour déterminer la stationnarité d'une série temporelle. Nous avons donc combiné le test statistique avec une évaluation visuelle de la série temporelle pour confirmer la stationnarité. Nous avons observé visuellement que la série ne présente pas de tendance, ce qui implique que les propriétés statistiques telles que la moyenne et la variance restent stables en fonction du temps.

Cette analyse visuelle renforce les résultats obtenus par le test ADF et nous permet de conclure que les trois séries temporelles, dont les graphiques figurent ci-dessous, sont stationnaires.

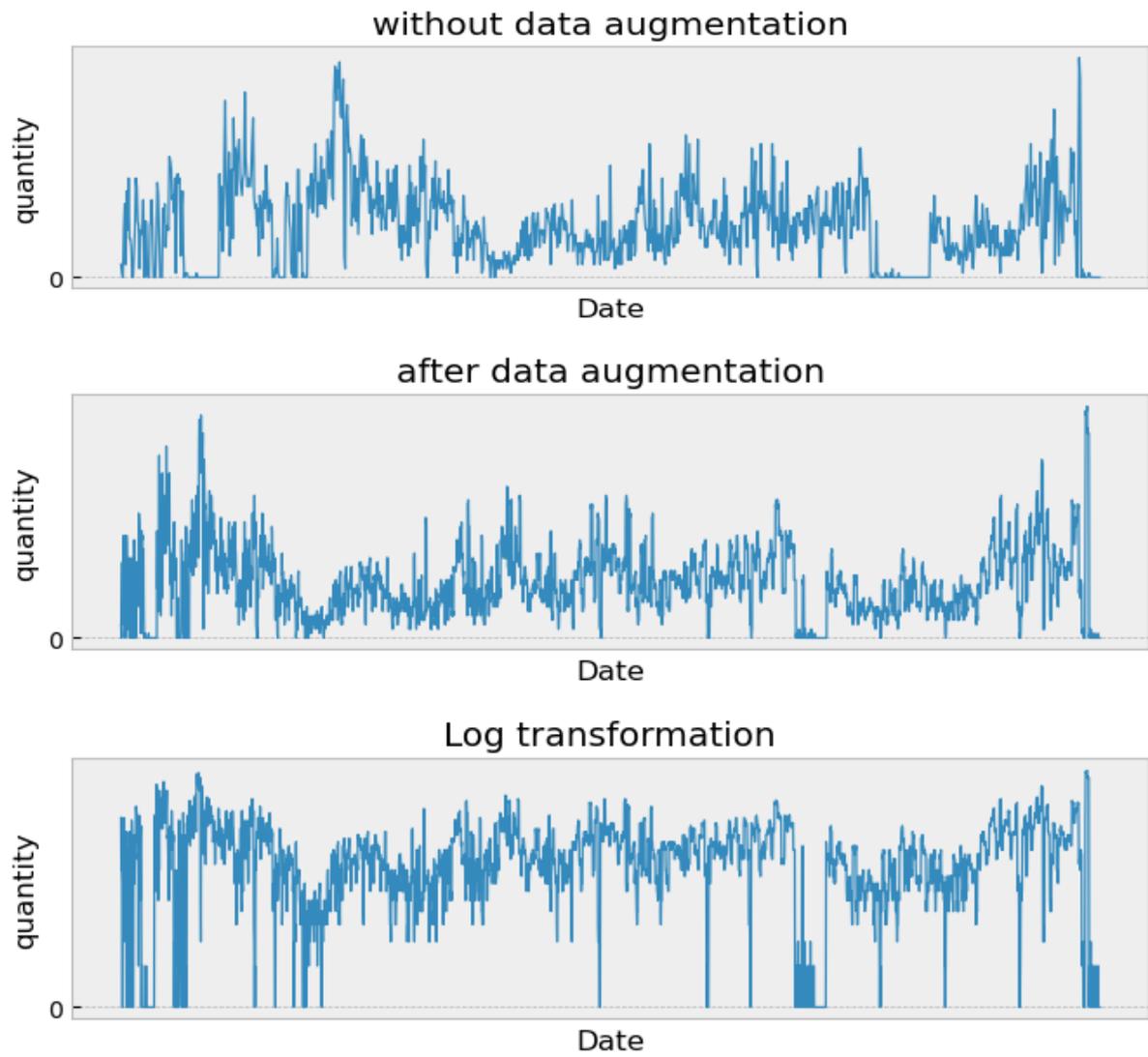


Figure 4.1: L'évolution des ventes après application des transformations.

4.2.2. Sélection des paramètres pour le modèle de prédiction

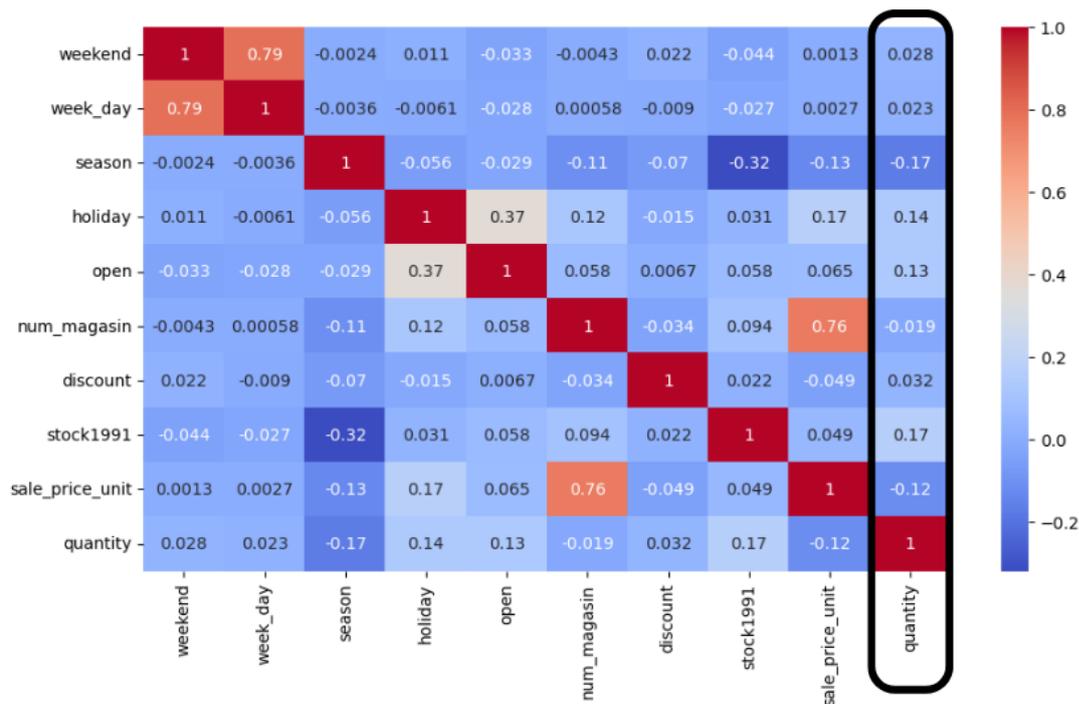


Figure 4.2: Matrice des corrélations.

Dans cette heatmap, on peut remarquer qu'il n'y a pas de corrélations significatives entre les variables explicatives, à savoir : weekend, week_day, season, holiday, open, num_magasin, discount, stock1991 et sale_price_unit, et la variable expliquée, à savoir quantity. À partir de cette observation, nous pouvons conclure qu'il n'y a pas de relation linéaire évidente entre la variable expliquée et les variables explicatives.

Pour explorer des relations potentielles non linéaires, nous avons utilisé le modèle de machine learning Random Forest qui est capable de détecter des relations non linéaires entre les variables, ce qui en fait un outil puissant pour la modélisation de données complexes. En utilisant la fonction intégrée "random_forest.feature_importances_" de ce modèle, nous avons sélectionné les variables suivantes comme entrées pour notre modèle de prédiction : season, stock1991 et sale_price_unit. Ces variables représentent les meilleurs prédicteurs de la quantité de notre produit vendu.

4.2.3. Création des variables retardées

Les variables retardées (lagged variables) sont une technique courante pour capturer l'effet historique dans les séries temporelles, améliorant ainsi la précision des modèles. Ces variables représentent les valeurs passées d'une variable à différents moments dans le temps, et sont utilisées pour prendre en compte les variations passées, notamment la saisonnalité, afin de mieux comprendre les schémas récurrents. En incluant ces variables dans le modèle, on peut capturer

les tendances et les dépendances temporelles, ce qui améliore la capacité du modèle à faire des prédictions précises.

Pour améliorer la précision de notre modèle de prédiction, nous avons créé des variables retardées pour les variables sélectionnées précédemment (stock1991, sale_price_unit) ainsi que pour notre variable cible (quantity).

Date	quantity	Lag1	Lag2
T1	10	Nan	Nan
T2	6	10	Nan
T3	8	6	10
T4	13	8	6
T5	17	13	8

Figure 4.3: Illustration explicative de la création des variables retardées.

4.2.4. Division des données

En apprentissage automatique, la division de données est une pratique courante pour entraîner et évaluer les performances d'un modèle. Elle consiste à séparer l'ensemble de données en trois parties distinctes : l'ensemble d'entraînement, l'ensemble de validation et l'ensemble de test.

L'ensemble d'entraînement est utilisé pour entraîner le modèle, l'ensemble de validation est utilisé pour ajuster les hyperparamètres du modèle et éviter le surapprentissage, et l'ensemble de test est utilisé pour évaluer les performances du modèle sur des données qu'il n'a jamais vues auparavant, car si nous évaluions le modèle sur les mêmes données sur lesquelles il a été entraîné, cela ne nous donnerait pas une idée précise de ses performances sur de nouvelles données.

Cependant, dans notre cas, la taille de l'ensemble de données est limitée, donc il est difficile de se permettre de diviser les données en trois ensembles distincts. Alors, on a divisé les données en ensembles d'entraînement (train) et de test uniquement. Voici la plage de date pour chaque ensemble de données :

train : 18-06-2020 jusqu'au 31-12-2022

test : 01-01-2023 jusqu'au 06-05-2023

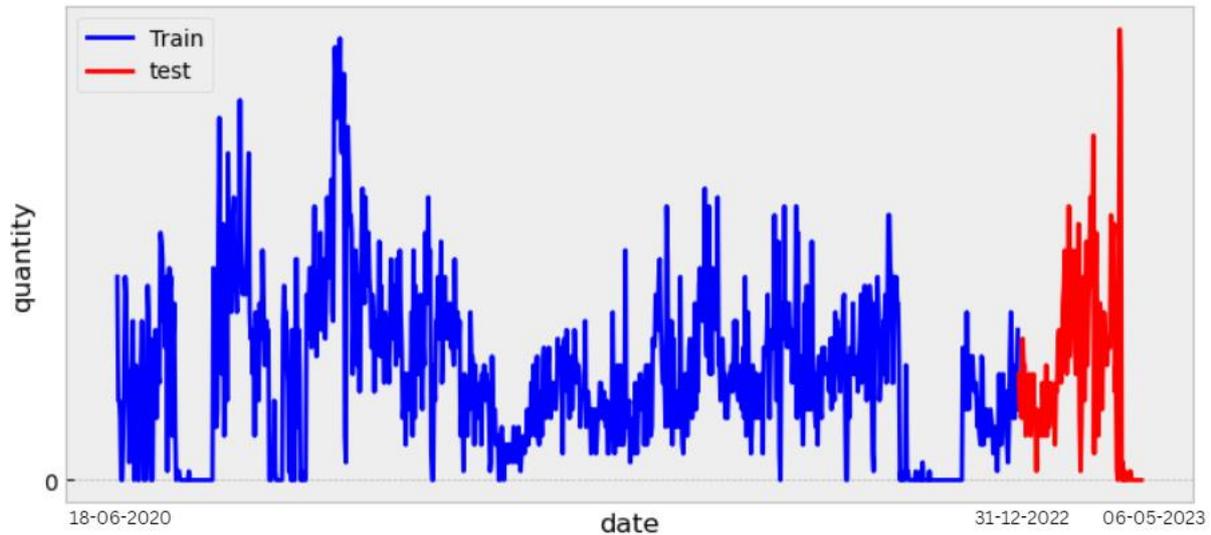


Figure 4.4: Graphiques linéaire de l'évolution des ventes séparées entre les ensembles d'entraînement et de test.

4.2.5. Modélisation

Dans ce projet, nous avons utilisé quatre modèles populaires pour la prévision des séries temporelles : ARIMA, Prophet, Random Forest et LSTM.

ARIMA est un modèle classique qui est bien adapté pour capturer les tendances linéaires et les patterns saisonniers dans les données. Cependant, il peut également être étendu pour traiter les tendances non linéaires en transformant les données avant d'ajuster le modèle. Par ailleurs, si les données présentent des schémas non linéaires complexes, Prophet ou Random Forest peuvent s'avérer de meilleurs choix. Prophet est conçu pour traiter les tendances linéaires et non linéaires, tandis que Random Forest peut capturer des relations complexes entre les variables d'entrée et la variable cible. Il peut traiter un grand nombre de caractéristiques d'entrée et n'est pas sensible à la distribution des variables d'entrée.

Le dernier modèle c'est LSTM; c'est un modèle d'apprentissage profond spécialement conçu pour traiter les données de séries temporelles avec des dépendances à long terme et des modèles non linéaires. Il utilise une cellule de mémoire pour stocker des informations sur les entrées passées, ce qui lui permet de capturer des modèles complexes et des irrégularités dans les données.

Génération des séquences pour LSTM

Pour utiliser les LSTM efficacement avec des séries chronologiques, il est crucial de comprendre leur dépendance temporelle. Chaque observation dans une série chronologique dépend des valeurs précédentes dans la séquence, ce qui rend les LSTM particulièrement adaptés pour capturer ces dépendances complexes. En générant des séquences, on permet aux LSTM d'exploiter cette caractéristique essentielle des séries chronologiques.

En outre, les LSTM disposent d'une mémoire à long terme qui leur permet de stocker des informations importantes sur de longues séquences. En générant des séquences, on leur donne la possibilité de tirer parti de cette mémoire pour modéliser les motifs temporels et effectuer des prédictions précises.

Dans notre cas, la quantité (Q_t) représente notre target, supposons qu'on veut générer des séquences de quatre jours. Voici un schéma illustratif :

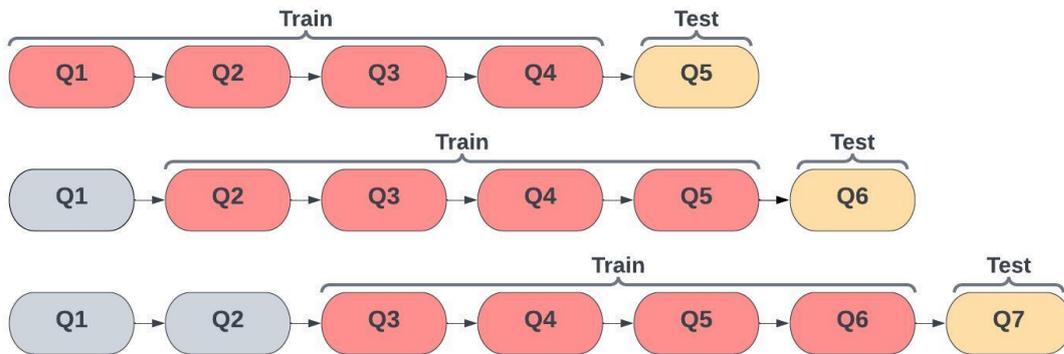


Figure 4.5: Schéma illustrant la génération des séquences.

Architecture LSTM

Pour notre projet, nous avons utilisé une architecture simple composée de 10 nœuds LSTM avec une fonction d'activation ReLu, suivi par 4 nœuds Dense(fully connected)avec une fonction d'activation toujours ReLu et à la fin on trouve un noeud Dense car c'est un problème de régression . Nous avons choisi cette architecture en raison de la taille limitée des données dont nous disposons. Les techniques d'apprentissage profond nécessitent souvent de grands ensembles de données pour atteindre des performances optimales, mais nous avons dû travailler avec ce que nous avons. Néanmoins, même avec cette architecture simple, nous avons pu obtenir des résultats prometteurs que nous présentons par la suite.

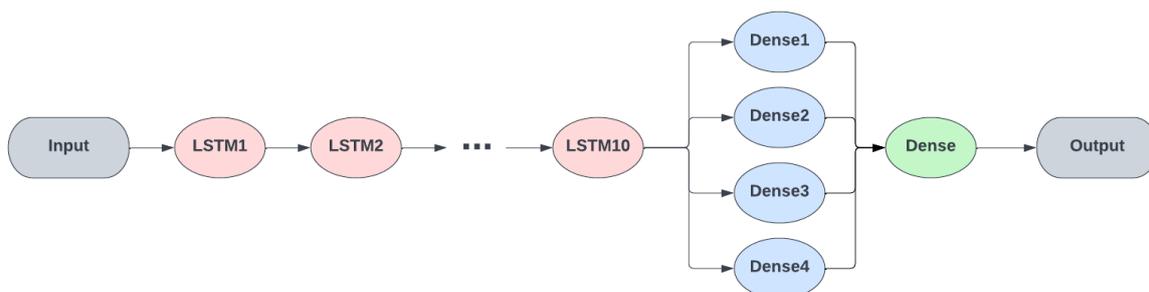


Figure 4.6: Architecture du modèle LSTM utilisé dans l'étude.

Pour s'assurer que notre modèle n'est pas surajusté (overfitted) aux données d'apprentissage, nous avons tracé la perte d'apprentissage et la perte de validation et comparé leurs tendances. Il y a surajustement lorsqu'un modèle apprend à s'adapter trop étroitement aux données d'apprentissage, ce qui se traduit par des performances médiocres sur des données nouvelles et inédites. Pour détecter un surajustement(overfitting), nous avons comparé le schéma de la perte d'apprentissage et de la perte de validation au cours de l'apprentissage. Les deux courbes diminuent progressivement sans qu'il n'y ait de différence significative entre les deux pertes. Ces résultats suggèrent que le modèle LSTM ne souffre pas de surajustement aux données d'entraînement. Pour être sûr, on testera la capacité du modèle à généraliser sur un nouvel ensemble de données.

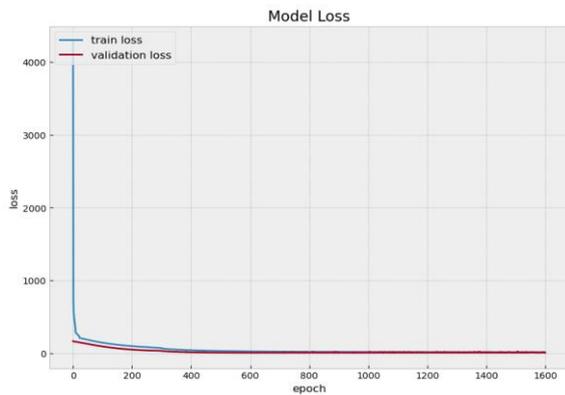


Figure 4.7: La perte avant l’augmentation des données .

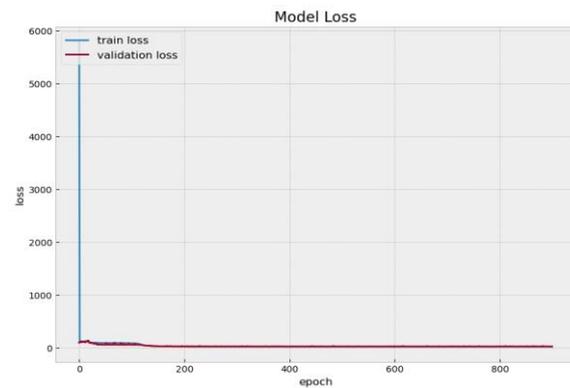


Figure 4.8: La perte après l’augmentation des données.

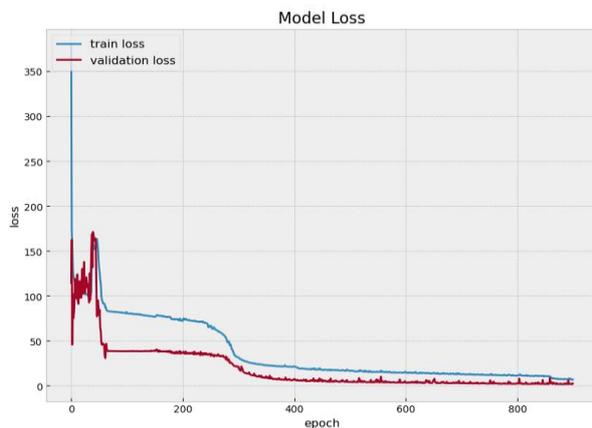


Figure 4.9: La perte après la transformation Log.

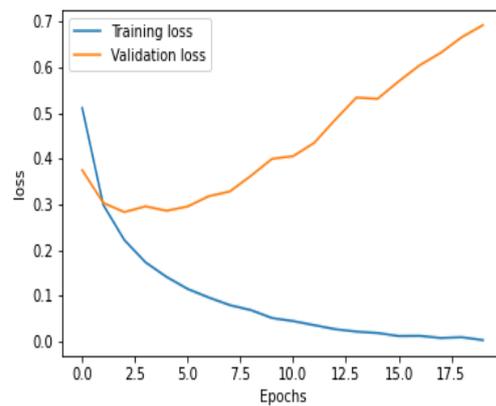


Figure 4.10: Exemple d’un surajustement.

4.3. Présentation et discussion des résultats :

Les résultats de notre projet de fin d'études mettent en évidence l'impact de l'augmentation des données et de l'application d'une transformation logarithmique sur les performances de quatre modèles de prédiction : Prophet, Random Forest, ARIMA et LSTM. Nous avons conservé les mêmes variables et la même taille des ensembles d'entraînement et de test pour toutes les

expériences sauf pour les lstm car ils nécessitent une structure de données spéciale. Ensuite, nous avons augmenté la taille des données et appliqué une transformation logarithmique.

Avant l'augmentation des données, les modèles ont affiché des valeurs modérées du coefficient de détermination (R-carré ou R2), indiquant leur capacité à expliquer une partie de la variance de la variable dépendante (la variable cible ou expliquée). Cependant, les valeurs ajustées du R-carré étaient légèrement inférieures. En effet, les méthodes d'apprentissage automatique et d'apprentissage profond nécessitent généralement des ensembles de données de taille élevée pour obtenir des performances optimales.

Après l'augmentation des données, nous avons observé une amélioration significative des performances des quatre modèles. Les valeurs du R-carré ont augmenté dans une fourchette de 0.78 à 0.81, ce qui indique un meilleur ajustement aux données et une plus grande proportion de variance expliquée. Les valeurs ajustées du R-carré ont également augmenté, ce qui suggère que les variables conservées dans les modèles ont contribué de manière plus significative à leurs capacités prédictives.

De plus, l'application d'une transformation logarithmique a encore amélioré les performances des modèles. Les valeurs du R-carré ajusté ont atteint une fourchette de 0.80 à 0.90, ce qui indique une plus grande puissance explicative. Nous pensons que la transformation logarithmique a permis de stabiliser la variance des données et d'établir une relation plus linéaire entre les prédicteurs et la variable cible, ce qui a renforcé les performances des modèles.

Table 4.1: Tableau comparatif des résultats du R carré et R carré ajusté.

Modèle de prédiction	Avant l'augmentation des données		Après l'augmentation des données		Après la transformation log	
	R carré	R carré ajusté	R carré	R carré ajusté	R carré	R carré ajusté
PROPHET	0.36	0.21	0.81	0.76	0.84	0.82
RANDOM FOREST	0.44	0.38	0.79	0.77	0.91	0.90
ARIMA	0.41	0.32	0.81	0.78	0.84	0.83
LSTM	0.59	0.52	0.78	0.74	0.92	0.90

- Avant l'augmentation des données:

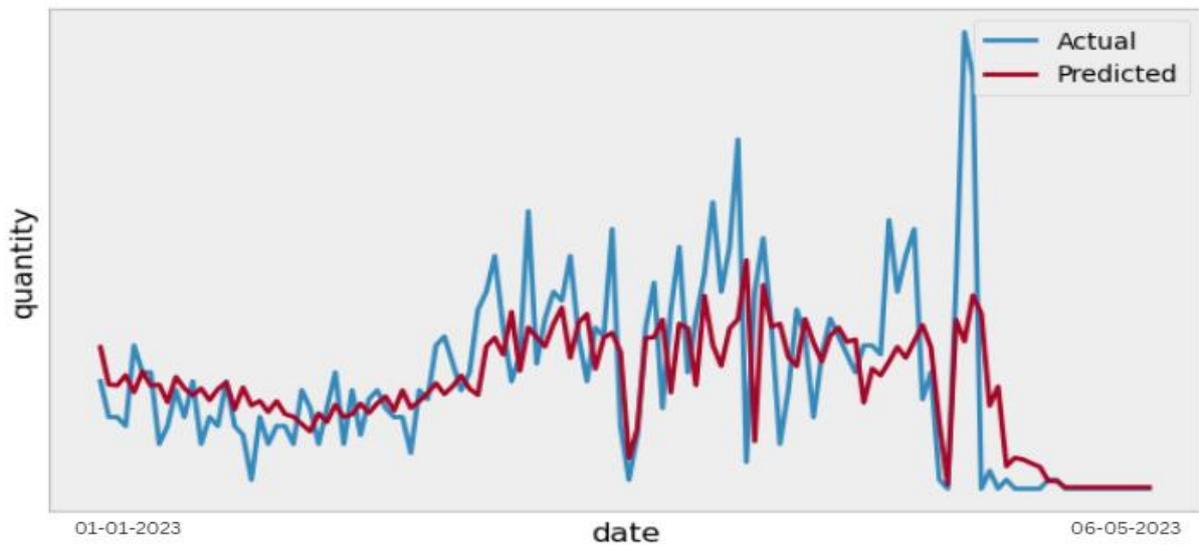


Figure 4.11: Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle Random forest avant l'augmentation des données.

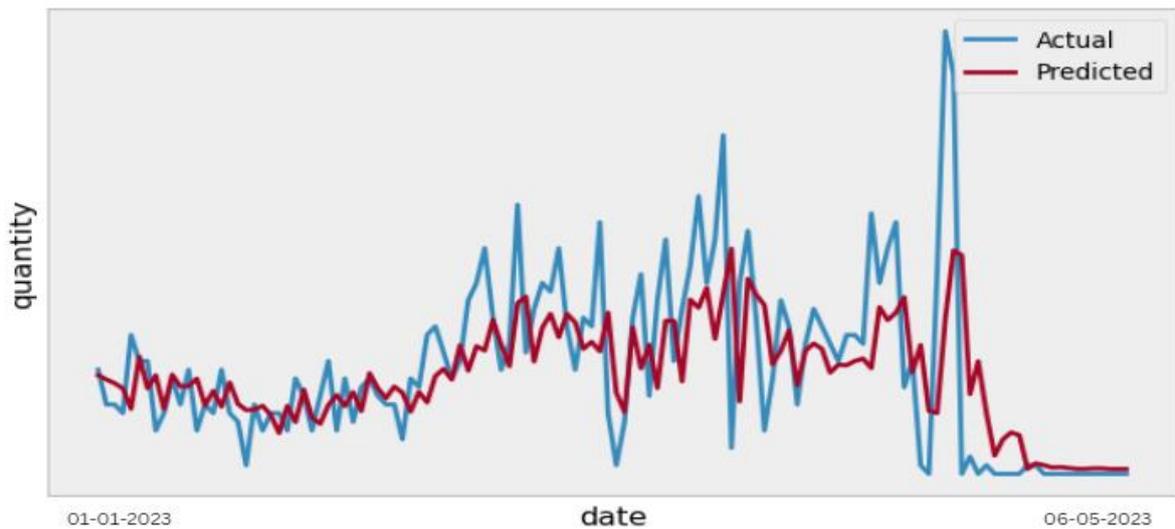


Figure 4.12: Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle ARIMA avant l'augmentation des données.

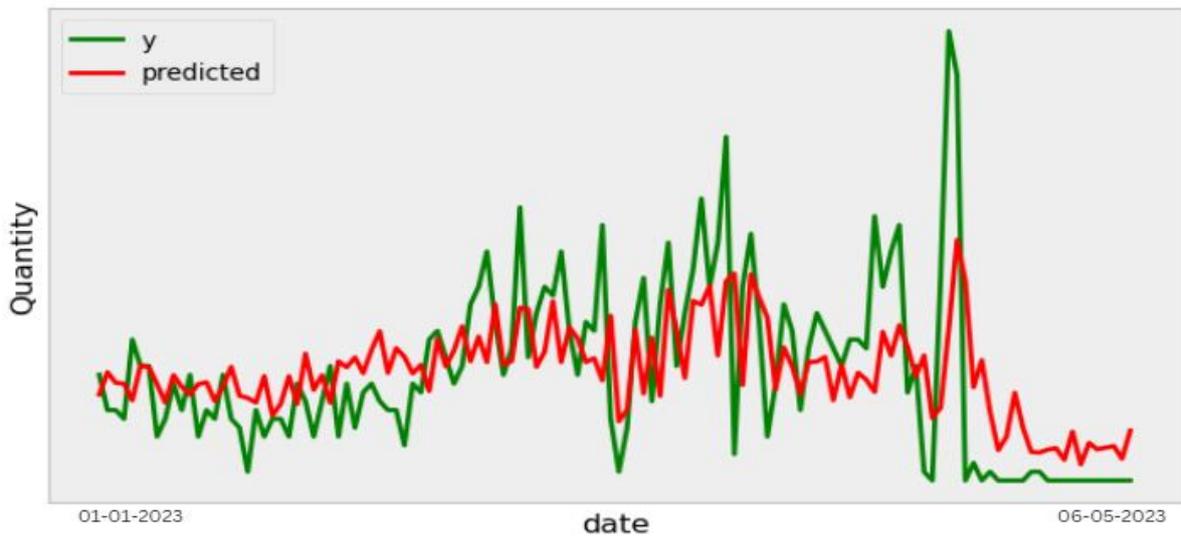


Figure 4.13: Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle Fb Prophet avant l'augmentation des données.

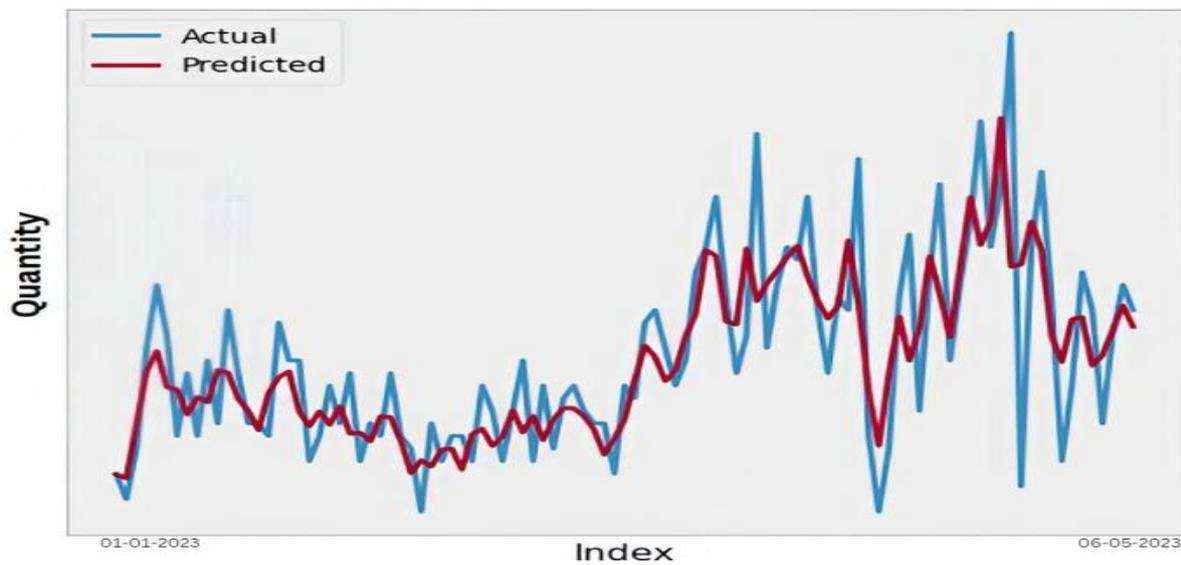


Figure 4.14: Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle LSTM avant l'augmentation des données.

- Après l'augmentation des données:

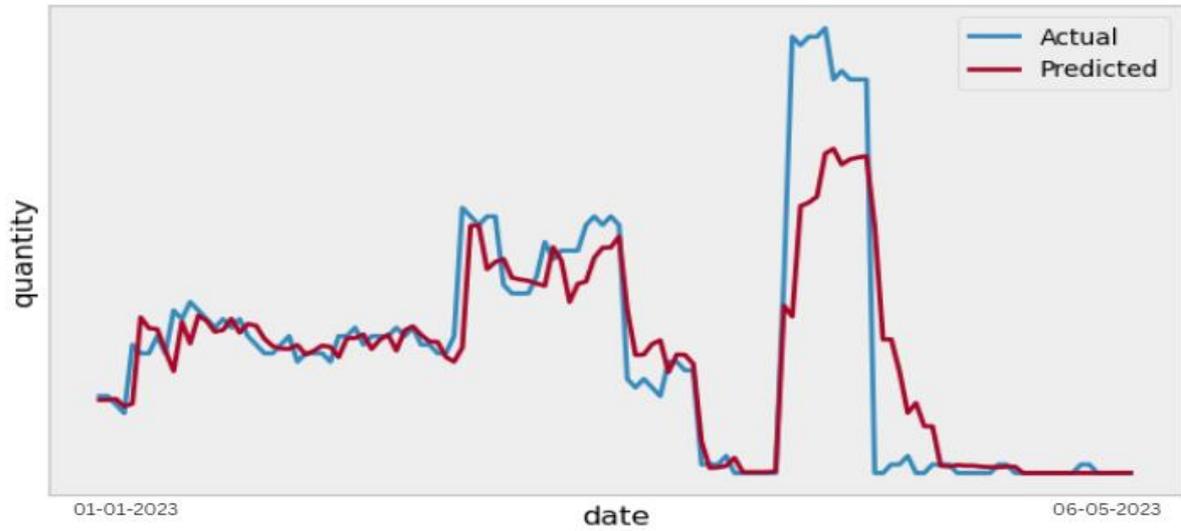


Figure 4.15: Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle Random forest après l'augmentation des données.

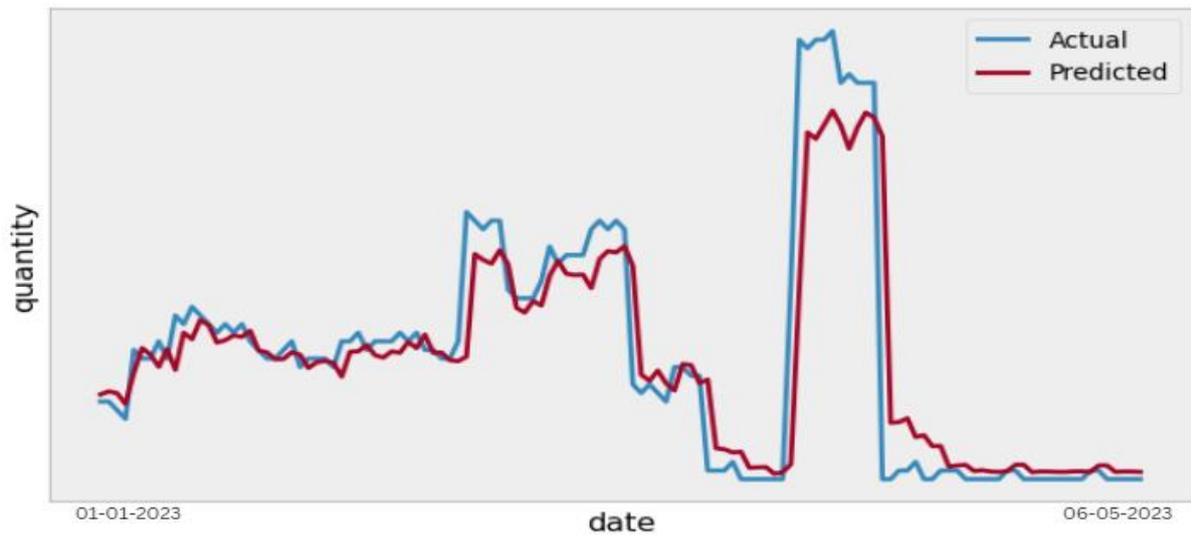


Figure 4.16: Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle ARIMA après l'augmentation des données.

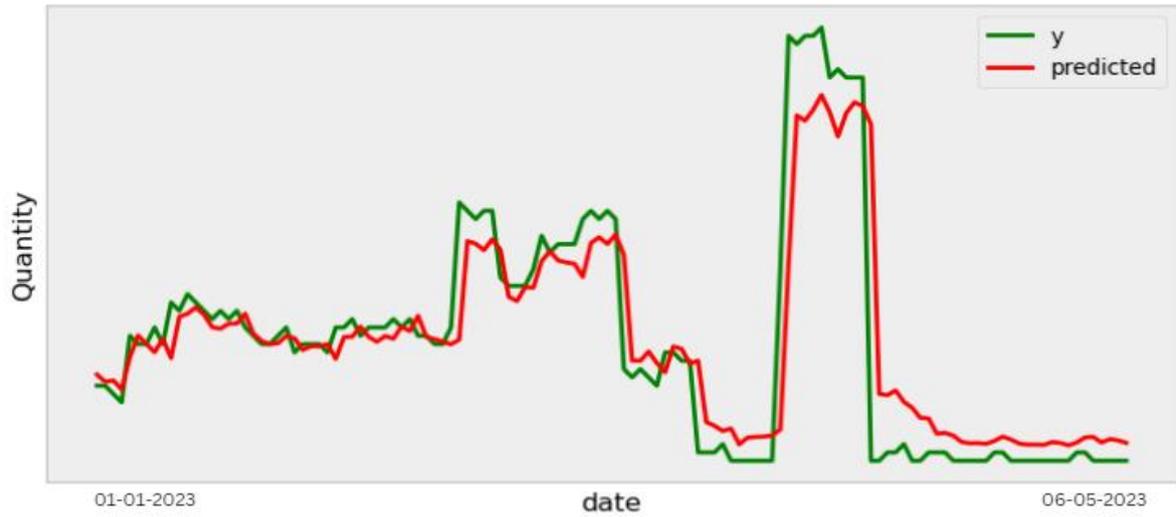


Figure 4.17: Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle Fb Prophet après l'augmentation des données.

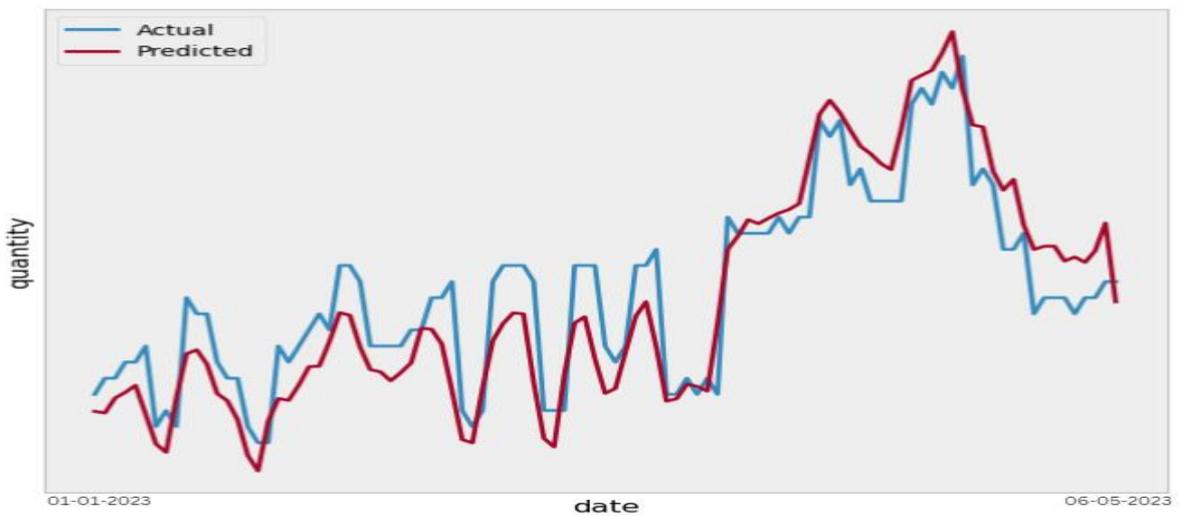


Figure 4.18: Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle LSTM après l'augmentation des données.

- Après la transformation Log:

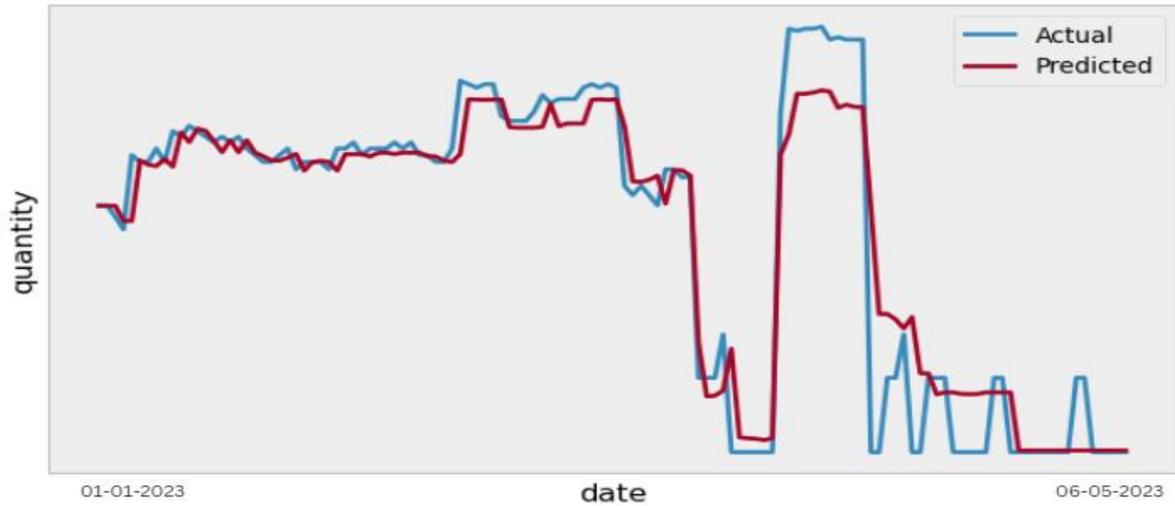


Figure 4.19: Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle Random Forest après la transformation log.

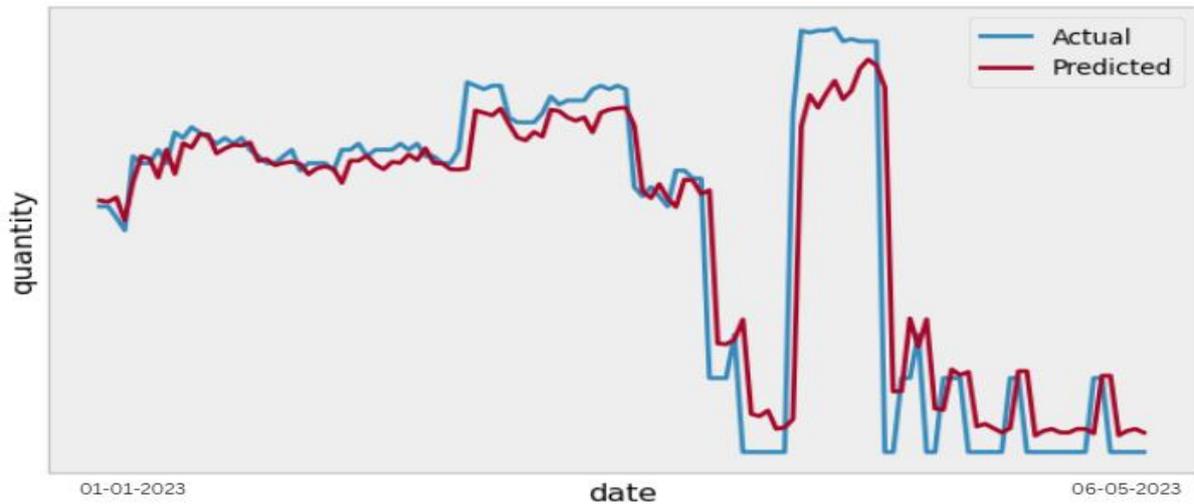


Figure 4.20: Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle ARIMA après la transformation log.

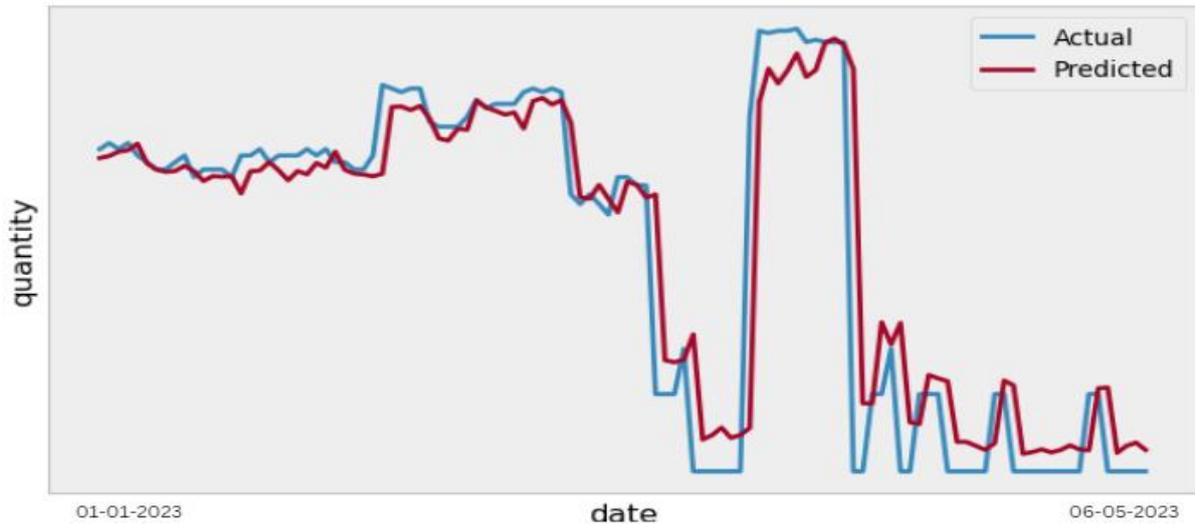


Figure 4.21: Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle FB Prophet après la transformation log.

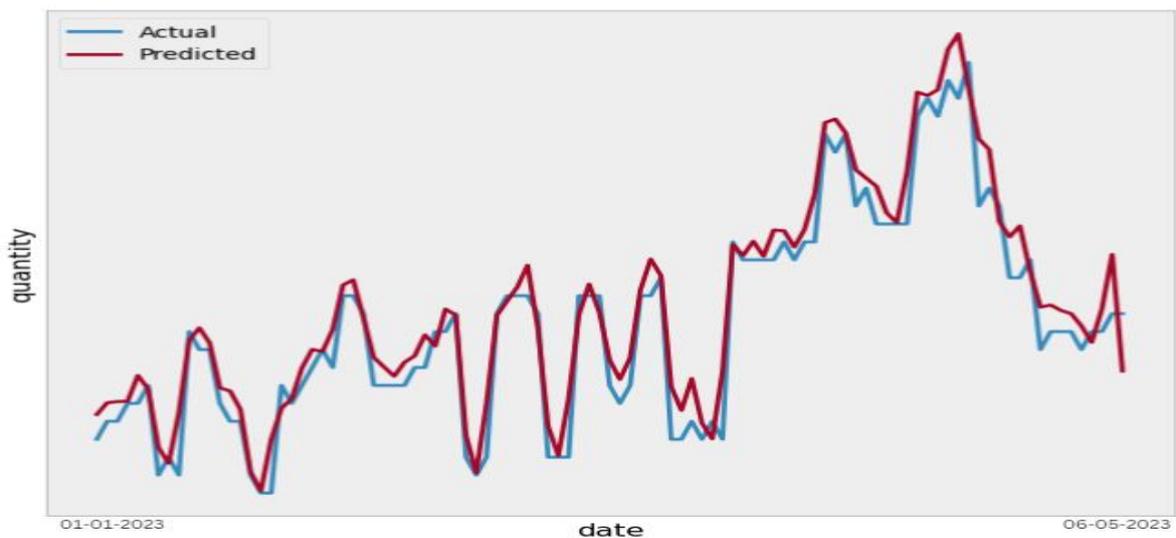


Figure 4.22: Graphique linéaire montrant l'évolution des ventes réelles et celles prédites par le modèle LSTM après la transformation log

En testant les modèles avec de nouveaux prix (1000, 3500 et 4000 dinars algériens), nous avons constaté que Prophet a donné les meilleurs résultats. Nous pouvons expliquer ces résultats en observant que Prophet a bien compris la relation entre les ventes et les prix. En effet, nous avons remarqué que si le prix est élevé, les ventes diminuent et vice versa. Nous pouvons expliquer cette relation par la notion d'élasticité-prix de la demande, qui indique que si le produit est élastique, une augmentation du prix entraînera une baisse de la demande,

et inversement. Nous pensons que la capacité de Prophet à modéliser des tendances saisonnières et à prendre en compte les événements récurrents dans les données de vente a également contribué à sa performance supérieure.

D'un point de vue managérial, nos modèles de prédiction peuvent être utilisés pour prédire les ventes avec une précision élevée, ce qui permettra à l'entreprise de gérer efficacement son stock, de planifier les commandes à l'étranger et d'estimer le chiffre d'affaires généré. La simulation des prix permet également aux dirigeants d'estimer les performances de ventes pour chaque prix choisi. Par exemple, si le leader décide de fixer les prix à 3500 DA, nous avons observé une diminution des ventes de 37%, car pour un prix de 2900 DA, le total des ventes durant une période d'environ 4 mois était de 1625 unités vendues. Selon notre estimation, pour un prix de 3500 DA, les ventes chuteront jusqu'à 1020 unités vendues, ce qui représente une perte de chiffre d'affaires égale à 1142500 DA.

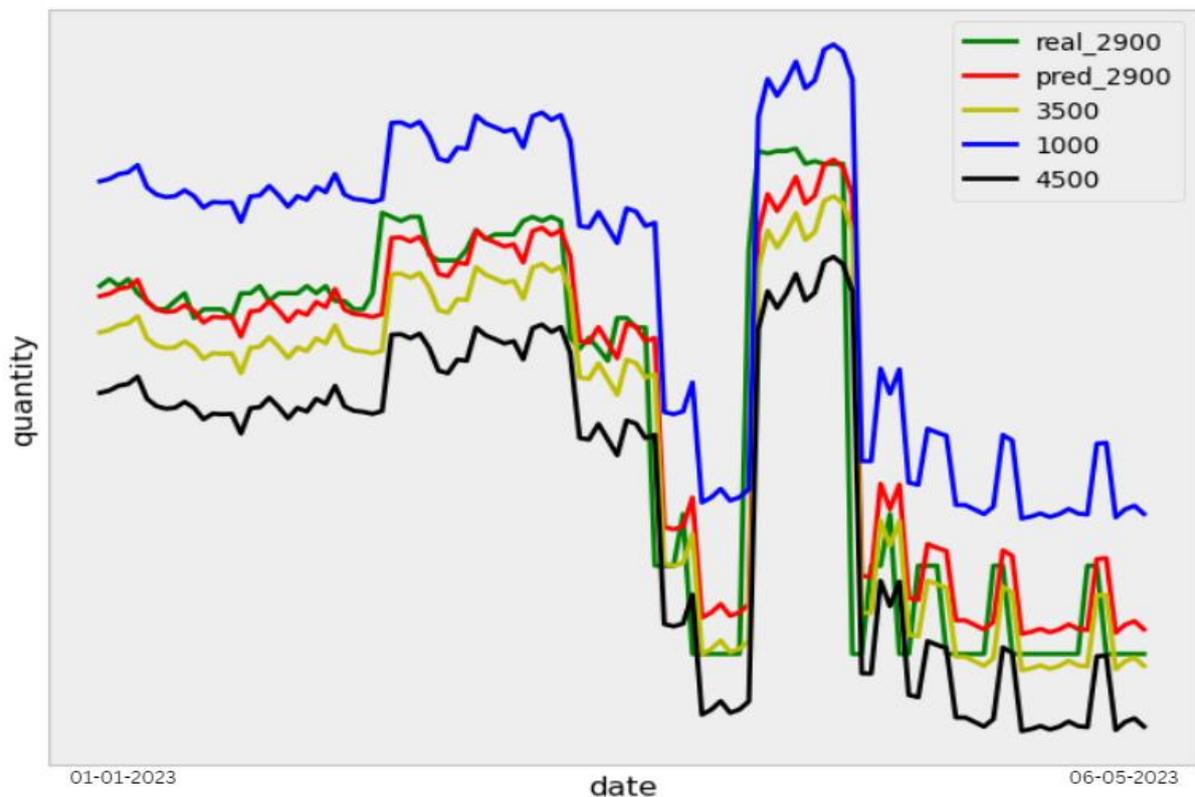


Figure 4.23 : L'évolution des ventes prédites en fonctions de plusieurs prix donnés aux modèles

4.4.Conclusion:

Ce chapitre présente les étapes suivies pour appliquer les modèles de prédiction des ventes. Nous avons ensuite présenté les résultats obtenus à travers les métriques de R carré et R carré ajusté. Enfin, nous avons conclu ce chapitre par une simulation qui met en évidence l'importance de notre solution pour soutenir les décisions de tarification.

Conclusion générale et travaux futurs

Dans ce projet de fin d'études, l'objectif était d'étudier l'impact du changement de prix sur les ventes. Cette problématique a été proposée par Decathlon El Djazair, car l'entreprise avait besoin d'une solution pour déterminer le juste prix de ses produits. Nous avons appliqué quatre modèles de prédiction des séries temporelles sur trois versions de données (avant l'augmentation, après l'augmentation et après la transformation log). Les modèles ont fourni des performances élevées après l'augmentation des données, et ces performances ont été encore améliorées après la transformation log. Cette précision de prédiction des ventes facilite plusieurs tâches de gestion, telles que la gestion du stock.

Ensuite, pour mettre en évidence l'impact du changement de prix sur les ventes, nous avons effectué une prédiction des ventes pour de nouveaux prix, afin d'estimer les ventes pour chaque prix. Cette fonctionnalité permet aux managers de choisir le meilleur prix pour augmenter le bénéfice de l'entreprise.

Nous avons identifié plusieurs pistes pour améliorer la précision de nos prévisions et mieux comprendre les facteurs qui influencent le comportement des clients. Tout d'abord, prévoyons d'appliquer deux transformations, la différenciation et la transformation logarithmique, en même temps pour évaluer les performances de notre modèle. Nous sommes convaincus que cette approche peut nous aider à améliorer encore plus les résultats que nous avons obtenus jusqu'à présent.

Ensuite, nous avons prévu de lancer une campagne d'étude de marché pour recueillir des données sur les préférences des clients, le prix psychologique que les clients sont prêts à payer pour un produit et des données sur les concurrents. Nous pensons que ces informations peuvent nous aider à mieux comprendre les facteurs qui influencent les décisions d'achat des clients et à améliorer la précision de nos prévisions.

Nous considérons également que les données sur l'extension d'utilisation sont extrêmement importantes, surtout dans notre pays. Nous prévoyons donc de collecter des données sur la capacité d'un produit à être utilisé à des fins autres que celles pour lesquelles il a été conçu, par exemple, une chaussure de course peut être utilisée pour des activités autres que la course, comme la randonnée ou le travail.

Enfin, nous voulons rendre notre modèle capable de faire des prévisions pour tous les produits. Pour cela, nous prévoyons d'ajouter les caractéristiques du produit en entrée. Nous pensons que la génération d'embeddings pour les caractéristiques des produits qui sont généralement des textes et ensuite réduire leur dimension à l'aide des auto-encodeurs peuvent aider le modèle à apprendre le comportement des clients pour chaque produit et à différencier correctement les produits.

En somme, ces travaux futurs visent à améliorer la précision de nos prévisions et à mieux comprendre les facteurs qui influencent le comportement des clients. Nous sommes convaincus que l'intégration de ces idées dans notre travail peut contribuer significativement à l'amélioration des performances du modèle de prédiction.

Références bibliographiques

Dudu, O., Agwu, E. *A Review of The Effect of Pricing Strategies on The Purchase of Consumer Goods*, International Journal of Research in Management, Science & Technology. 2014, volume: 2, Issue: 2, pp: 88-101. Consulté le: 10-03-2023. Disponible sur: https://www.researchgate.net/publication/269632507_A_Review_of_The_Effect_of_Pricing_Strategies_on_The_Purchase_of_Consumer_Goods.

Ariton, L. *A Thorough Introduction To ARIMA Models*, Analytics Vidhya. 2021. Consulté le: 07-04-2023. Disponible sur <https://medium.com/analytics-vidhya/a-thorough-introduction-to-arima-models-987a24e9ff71>

Chung, J., Gulcehre, C., Cho, K., Bengio, Y. *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*, NIPS Workshop on Deep Learning. 2014. Consulté le 30-03-2023. Disponible sur <https://arxiv.org/abs/1412.3555>

Toppr, *Components of Time Series Analysis*, Time Series Analysis. Consulté le 14-03-2023. Disponible sur <https://www.toppr.com/guides/business-mathematics-and-statistics/time-series-analysis/components-of-time-series/>

Amherd, F., Rodriguez, E. *Heatmap-based Object Detection and Tracking with a Fully Convolutional Neural Network*. Computer Science. 2021. Consulté le 29-03-2023. Disponible sur : https://www.researchgate.net/figure/Dense-Neural-Network_fig5_348402885

Dilmegani, C. *Top Data Augmentation Techniques: Ultimate Guide for 2023*, AIMultiple. 2021. Consulté le : 07-06-2023. Disponible sur <https://research.aimultiple.com/data-augmentation-techniques/>

Dolgui, A., Proth, J. *Pricing strategies and models*, Annual reviews in control. 2010, volume: 34, Issue: 1, pp: 101-110. Consulté le: 01-03-2023. Disponible sur <https://www.sciencedirect.com/science/article/abs/pii/S1367578810000118>

Facebook. *Prophet | Forecasting at scale*, Meta Open Source. 2023. Consulté le : 20-03-2023. Disponible sur <https://facebook.github.io/prophet/>

Fernandez, J. *Creating an ARIMA Model for Time Series Forecasting*, Towards Data Science. 2022. Consulté le: 23-04-2023. Disponible sur <https://towardsdatascience.com/creating-an-arima-model-for-time-series-forecasting-ff3b619b848d>

Hayes, A., Stapeleton, C., Munichello, K. *Autoregressive Integrated Moving Average (ARIMA) Prediction Model*, Advanced technical analysis concepts. Consulté le 03-04-2023. Disponible sur <https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp>

Liu, C., Sustik, M. A. *Elasticity Based Demand Forecasting and Price Optimization for Online Retail*. The University of Texas at Austin. 2021. Consulté le : 02-05-2023. Disponible sur : <https://arxiv.org/abs/2106.08274>

Pascanu, R., Mikolov, T., Bengio, Y. . On the difficulty of training recurrent neural networks. *Proceedings of the 30th International Conference on International Conference on Machine Learning*; 2013, Vol. 28, Issue :3, pp : 1310-1318. Date de consultation : 30-04-2023. Disponible sur le site : <https://proceedings.mlr.press/v28/pascanu13.html>

Help Center. *Quel outil d'analyse de séries temporelles choisir ?*. Consulté le 01-05-2023. Disponible sur : <https://help.xlstat.com/fr/6504-quel-outil-danalyse-de-series-temporelles-choisir>

Sepp, H., Jürgen, S. Long Short-Term Memory, *Neural Computation*. 1997, Volume: 9, Issue: 8, pp : 1735 - 1780. Consulté le 20-04-2023. Disponible sur : <https://ieeexplore.ieee.org/abstract/document/6795963/authors>

Verma, Y. *Complete Guide To Dickey-Fuller Test In Time-Series Analysis*. Consulté le : 25- 04-2023. Disponible sur : <https://analyticsindiamag.com/complete-guide-to-dickey-fuller-test-in-time-series-analysis/>

Thirtieth International Joint Conference on Artificial Intelligence. *Time Series Data Augmentation for Deep Learning: A Survey*, Montreal, 2021. . Consulté le : 10- 04-2023. DOI: 10.24963/ijcai.2021/63.

Zivot, E., & Wang, J. *Modeling Financial Time Series with S-PLUS®*. Springer, 2006.

Goodfellow, I., Bengio, Y., Courville, A. *Deep Learning*, MIT. 2016