

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



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique



Département : Génie Industriel
Entreprise : SLB NAF

Mémoire de Fin d'Études
En vue de l'obtention du diplôme d'Ingénieur d'État en Génie Industriel
Option : Management Industriel

Amélioration de la Supply Chain amont par l'application de
l'apprentissage machine
Application : SLB NAF

ABBOU Imad Eddine Fares
AMEZIANE Sofiane

Sous la direction de M. Iskander ZOUAGHI

Présenté et soutenu publiquement le (11/07/2024)

Composition du jury :

Président	M. Ali BOUKABOUS	MAA	ENP
Examinatrice	Mme. Samia BEKDJOUDI	MCA	ENP
Promoteur	M. Iskander ZOUAGHI	MCA	ENP
Invité	M. Cherif RAHMOUNE	Sourcing Specialist	SLB

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



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique



Département : Génie Industriel
Entreprise : SLB NAF

Mémoire de Fin d'Études
En vue de l'obtention du diplôme d'Ingénieur d'État en Génie Industriel
Option : Management Industriel

Amélioration de la Supply Chain amont par l'application de
l'apprentissage machine
Application : SLB NAF

ABBOU Imad Eddine Fares
AMEZIANE Sofiane

Sous la direction de M. Iskander ZOUAGHI

Présenté et soutenu publiquement le (11/07/2024)

Composition du jury :

Président	M. Ali BOUKABOUS	MAA	ENP
Examinatrice	Mme. Samia BEKDJOUDI	MCA	ENP
Promoteur	M. Iskander ZOUAGHI	MCA	ENP
Invité	M. Cherif RAHMOUNE	Sourcing Specialist	SLB

ملخص

يهدف هذا العمل إلى تعزيز استقرار حصة السوق للموردين. بدأنا بتحليل تطور حصة السوق لكل مورد في SLB NAF، ودراسة الاتجاهات والتغيرات حسب الفئة الفرعية للموردين، وانتهينا بالتشخيص من خلال تقييم SCOR لتحديد أسباب عدم استقرار الموردين. بعد ذلك، قمنا بإعداد وتجهيز الأساس اللازم لاستقرار حصة السوق للموردين من خلال الالتزامات الحجمية. أولاً، تُستخدم تقنيات التعلم غير المراقب لإعادة تصنيف القطع، ودمجها في نظام الطلبات الآلي ومن ثم استقرار تدفق الطلبات. ثانياً، يتم استخدام مجموعة من طرق التنبؤ، بدءاً من النماذج الاقتصادية إلى نماذج التعلم العميق، والتي تم تحسينها باستخدام تقنيات توليد البيانات، لاستقرار التنبؤ بأنماط الطلب المختلفة. تساهم هذه الأساليب مجتمعة في تحسين عمليات سلسلة التوريد وتنتهي باقتراح استقرار حصة السوق للمورد.

الكلمات المفتاحية: سلسلة التوريد، الالتزام بالحجم، إدارة الموردين، التعلم غير المراقب، التنبؤ، التعلم العميق، توليد البيانات.

Abstract

This work aims to enhance supplier market share stability, we started by analyzing respective supplier market share evolution in SLB NAF, studying the tendencies and variations by supplier subcategory, ending the diagnosis with a SCOR evaluation to specify reasons of supplier instability. Subsequently, the study focuses on establishing and preparing the groundwork for stabilizing supplier market share through volume commitments.

Firstly, unsupervised learning techniques are employed to reclassify pieces, integrating them into an automated ordering system and by that stabilizing the flow of the commands. Secondly, a range of forecasting methods, from econometrics to deep learning models which were enhanced using data generation techniques, are utilized to stabilize prediction of various demand patterns. These approaches collectively contribute to optimizing supply chain operations and conclude in a proposal to stabilize the market share of relative supplier.

Key Words: Supply Chain, Supplier management, Volume engagement, Unsupervised Learning, Forecasting, Deep Learning, Data Synthetisation.

Résumé

Ce travail vise à améliorer la stabilité des parts de marché des fournisseurs. Nous avons commencé par analyser l'évolution des parts de marché respectives des fournisseurs chez SLB NAF, en étudiant les tendances et les variations par sous-catégorie de fournisseurs, et en terminant le diagnostic par une évaluation SCOR pour identifier les raisons de l'instabilité des fournisseurs. Ensuite, nous avons établi et préparé le terrain nécessaire pour stabiliser les parts de marché des fournisseurs par le biais d'engagements de volume.

Premièrement, des techniques d'apprentissage non supervisé sont utilisées pour reclasser les pièces, les intégrant dans un système de commande automatisé et par conséquent, stabilisant le flux des commandes. Deuxièmement, une gamme de méthodes de prévision, allant de l'économétrie aux modèles d'apprentissage profond, améliorées grâce à des techniques de génération de données, sont utilisées pour stabiliser la prévision des divers modèles de demande. Ces approches contribuent collectivement à optimiser les opérations de la Supply Chain et aboutissent à une proposition de stabilisation de la part de marché du fournisseur.

Mots-clés : Supply Chain, Gestion des fournisseurs, Engagement de volume, apprentissage non supervisé, prévision, apprentissage profond, synthétisation des données.

Dédicaces

“ Je dédie ce travail à toutes celles et ceux qui me sont chers.

À mon père, dont les conseils et les encouragements ne seront jamais oubliés. Ta mémoire restera à jamais la lumière qui éclairera mon chemin et illuminera les jours les plus sombres de ma route.

À ma mère, mes deux sœurs et mon frère, que Dieu vous protège et que votre avenir soit rempli de bonheur et de réussite. Votre soutien et votre amour m'ont toujours donné la force d'avancer.

À mes chers amis, merci pour votre soutien constant et pour avoir toujours cru en moi. je ferais de mon mieux pour être à la hauteur de votre estime de moi. Tout comme votre aide a toujours été présente lorsque j'en avais besoin, je serai éternellement reconnaissant de vous avoir à mes côtés.

À mon binôme Sofiane, pour ta gentillesse, ton sérieux et ta rigueur tout au long de ce parcours. Aux défis surmontés et aux aventures vécues, May the wind guide but never break.

À l'équipe SLB Hassi, pour les souvenirs que nous chérirons toujours. Merci d'avoir rendu cette expérience mémorable.

À toutes celles et ceux qui ont fait de Polytechnique une expérience aussi agréable. À tout le département, des jeux d'oeufs aux rencontres du club IEC, que ces souvenirs restent à jamais gravés dans notre mémoire.

À tout le monde,

à tout.”

All in all is all we are
Fares

Dédicaces

Louange à Dieu qui m'a permis de terminer ce travail et de voir ce jour.

Par ce travail, je souhaite dédier :

À ma mère, qui m'a tant soutenu et accompagné avec ses mots apaisants, sa tendresse et ses sacrifices durant toutes ces années d'étude. Merci à toi, maman, pour tout ton amour et ton soutien. Que Dieu te préserve et te garde parmi nous.

À toi, papa, même si tu n'es plus parmi nous aujourd'hui, j'imagine que ces moments t'auraient comblé de plaisir. Tu as toujours tenu à voir notre réussite. Que Dieu t'accueille dans son vaste paradis.

À mes sœurs, qui m'ont tant soutenu et aidé avec leur humour. Que Dieu vous préserve.

À ma très chère amie Nour, qui m'a tant aidé durant ces années d'étude en me consolant et en trouvant toujours des solutions. Tes discussions ont grandement contribué à ma réussite.

À la mémoire de mes proches qui nous ont quitté ces dernières années, que Dieu leur accorde son vaste paradis.

À toi, Mehdi, qui a toujours été à mes côtés depuis la primaire. Merci pour ta présence et ta joie de vivre.

Au groupe, Killian, en passant par Aymen, Haithem et Rayhane, avec nos sorties nocturnes et nos débats qui n'ont ni queue ni tête mais qui nous ont fait beaucoup rire.

À notre cher comité d'excellence, grâce à vous j'ai vécu une expérience que je ne revivrai sans doute jamais. Vous êtes devenus plus que des amis. Merci à vous.

À la famille SLB Hassi, sans vous, Yousra, Hiba, Hani et tous ceux déjà mentionnés, cette expérience aurait eu un tout autre goût. Un très grand merci à vous.

Sans oublier notre artiste, mon binôme, Fares, merci à toi pour tes idées, ton sens de l'analyse et pour tous ces moments et ces expériences partagés.

Sofiane

Remerciements

Nous tenons à travers ce message à adresser nos sincères remerciements à notre cher encadrant, Dr. ZOUAGHI Iskander, pour son précieux soutien, ses conseils remarquables et le temps qu'il nous a consacré. Son encadrement et sa présence nous ont permis de réaliser ce modeste projet.

Nous exprimons également notre gratitude à notre honorable jury, présidé par M. BOUKABOUS, et l'examinatrice Mme. BELDJOURI, pour l'honneur qu'ils nous ont fait en consacrant leur temps et en partageant leur précieux savoir lors de l'évaluation de ce projet de fin d'étude.

Nous remercions chaleureusement l'ensemble des équipes de SLB NAF pour leur accueil, en commençant par notre promoteur M. RAHMOUNE Cherif du département Sourcing, pour ses précieuses aides et orientations tout au long de notre stage. Nous remercions également Mme. BOUCHAFAA Yasmine, M. BIDAOUI Kamel, M. Mahmoud Sayed Mohamed Said Mahmoud, M. SIDHOUM Mourad, ainsi que toute autre personne des équipes de SLB pour leur assistance et contribution.

Un profond remerciement à l'ensemble du corps pédagogique du département de génie industriel de l'école, qui nous a transmis leur savoir-faire tout au long de cette formation.

Nous saluons les étudiants du génie industriel, et plus spécialement le club IEC, pour ces trois années remplies de souvenirs et de bons moments.

Enfin, à toute personne ayant contribué de près ou de loin à la réalisation de ce projet, nous vous adressons un grand merci !

Table des matières

Liste des Tableaux

Liste des Figures

Liste des Abréviations

Introduction Générale

1	SLB : Contexte, marché des fournisseurs et formulation de la problématique	18
1.1	Présentation de SLB et de son secteur d'activité.....	18
1.1.1	<i>Le marché pétrolier et parapétrolier.....</i>	<i>18</i>
1.1.2	<i>Présentation de SLB.....</i>	<i>21</i>
1.2	La Supply Chain de SLB.....	27
1.2.1	<i>Organisation internationale de la Supply Chain de SLB.....</i>	<i>28</i>
1.2.2	<i>Les processus clés de la Supply Chain de SLB.....</i>	<i>29</i>
1.3	Analyse du marché concurrentiel des fournisseurs de SLB.....	33
1.3.1	<i>Catégories Fournisseurs ASL.....</i>	<i>33</i>
1.3.2	<i>Analyse des sous catégories de fournisseurs.....</i>	<i>33</i>
1.3.3	<i>Evolution des parts de marché des fournisseurs.....</i>	<i>35</i>
1.3.4	<i>Causes de variation des parts de marché.....</i>	<i>38</i>
1.4	Audit logistique et présentation de la problématique.....	39
1.4.1	<i>Audit logistique.....</i>	<i>39</i>
1.4.2	<i>Présentation de la problématique</i>	<i>42</i>
2	Supply chain, Relations fournisseurs et défis des engagements de volume	45
2.1	Présentation des concepts liés à la gestion de la supply Chain.....	45
2.1.1	<i>Définition de la Supply Chain.....</i>	<i>45</i>
2.1.2	<i>Le Supply chain management.....</i>	<i>46</i>
2.1.3	<i>La Supply Chain dans le secteur du oil & gas.....</i>	<i>47</i>
2.1.4	<i>La fonction approvisionnement et la gestion des relations fournisseurs.....</i>	<i>47</i>
2.2	Relations entre fournisseurs et entreprise.....	48
2.2.1	<i>Gestion de la relation fournisseurs.....</i>	<i>48</i>
2.2.2	<i>Les contrats avec les fournisseurs.....</i>	<i>49</i>
2.2.3	<i>Engagements par volumes.....</i>	<i>50</i>
2.3	Les concepts de stabilisation des commandes et des prévisions	53
2.3.1	<i>Stabilisation des Commandes.....</i>	<i>53</i>
2.3.2	<i>Stabilisation des Prévisions.....</i>	<i>56</i>
3	Exploration des approches de clustering et de prévision temporelle.....	59
3.1	Intelligence artificielle.....	59
3.1.1	<i>Machine Learning.....</i>	<i>59</i>
3.1.2	<i>Types d'apprentissage</i>	<i>60</i>
3.1.3	<i>La méthodologie CRISP-DM</i>	<i>61</i>

3.2	Partitionnement de données (clustering)	61
3.2.1	Choix de l'algorithme	62
3.2.2	Validation des modèles	65
3.3	Méthodes et outils pour les prévisions	68
3.3.1	Les méthodes Auto-Projectives (Série chronologique)	68
3.3.2	Modèles de Machine Learning pour les prévisions	73
4	Approche CRISP-DM pour la stabilisation des commandes	80
4.1	Compréhension des Métiers	80
4.1.1	Répartition des Pièces M&S Prime	80
4.1.2	Objectifs de l'Analyse	81
4.2	Compréhension des données	82
4.3	Préparation de données	84
4.3.1	Normalisation des données	84
4.3.2	Standardisation des données	84
4.4	Modélisation des données	85
4.4.1	Application du K-Means	85
4.4.2	Classification Ascendante Hiérarchique	86
4.4.3	Application du MiniBatch	87
4.4.4	Application du DBSCAN	87
4.4.5	Application de l'algorithme MeanShift	87
4.5	Evaluation des modèles	88
4.5.1	Résultat de chaque modèle par rapport aux métriques	88
4.5.2	Analyse des catégories	89
4.6	Déploiement	91
4.6.1	Réévaluation des quantités de stock de sécurité	91
4.6.2	Tableau de bord	93
5	Engagements contractuels : Approche prévisionnelle	96
5.1	Étude prévisionnelle de la demande	96
5.1.1	Préparation des données pour les prévisions	96
5.1.2	Application des modèles statistiques de prévision	99
5.1.3	Application des modèles de prévision basée sur le Machine Learning	107
5.1.4	Génération de Données Synthétiques	110
5.1.5	Comparaisons entre les modèles prévisionnels	111
5.2	Estimation des gains en coûts à travers l'engagement contractuel en volume	114
	CONCLUSION GENERALE	116
	Références	118
	Annexes :	122

Liste des Tableaux

- Tableau 1.1 – Carte d'Identité de SLB - SLB 2024
- Tableau 1.2 – Description des processus clés de la Supply Chain de SLB - Document Interne
- Tableau 4.1 – Méthode de calcul du stock de sécurité
- Tableau 5.1 – Matrice de la méthode ABC-XYZ
- Tableau 5.2 – Résultats de la classification ABC
- Tableau 5.3 – Résultats de la classification ABC-XYZ
- Tableau 5.4 – Comparaison entre les résultats des différents modèles statistiques
- Tableau 5.5 – Comparaison entre les résultats des modèles pour les pièces sélectionnées
- Tableau 5.6 – Tableau illustratif des différents résultats des pièces à travers FB Prophet
- Tableau 5.7 – Tableau comparatif des résultats des prévisions par l'algorithme XGboost
- Tableau 5.8 – Tableau comparatif des résultats des prévisions par l'algorithme RNN
- Tableau 5.9 – Tableau comparatif des résultats des prévisions par l'algorithme LSTM
- Tableau 5.10 – Tableau comparatif des erreurs MAPE
- Tableau 5.11 – Comparatif des résultats de prévision par modèle étudié
- Tableau C.1 – Synthèse et comparaison des principaux référentiels existants
- Tableau C.2 – Décomposition en niveau 3 du processus de planification
- Tableau C.3 – Décomposition en niveau 3 du processus d'approvisionnement
- Tableau C.4 – Grille d'évaluation du sous processus Planification de la Supply Chain
- Tableau C.5 – Grille d'évaluation du sous processus Planification des approvisionnements
- Tableau C.6 – Grille d'évaluation du sous processus Planification des commandes
- Tableau C.7 – Grille d'évaluation du sous processus Approvisionnement Stratégique
- Tableau C.8 – Grille d'évaluation du sous processus Approvisionnement direct
- Tableau C.9 – Grille d'évaluation du sous processus Approvisionnement indirect
- Tableau G.1 – Résultats de la classification XYZ
- Tableau I.1 – Résultats des prévisions par FB Prophet après génération des données

Liste des Figures

- Figure 1.2 – Evolution des prix du pétrole brut
- Figure 1.3 – Evolution du marché des services pétrolier
- Figure 1.3 – Répartition du Chiffre d’Affaires 2023
- Figure 1.4 – Divisions et Business Lines de SLB - Document Interne
- Figure 1.5 – Carte représentant les différents bassins et GeoUnits de SLB – Document Interne
- Figure 1.6 – Carte de la GeoUnit SLB NAF
- Figure 1.7 – Organisation de SLB NAF - Document Interne
- Figure 1.8 – Metro-Map de la Supply Chain de SLB – Document Interne
- Figure 1.9 – Répartition des DSC de Schlumberger à travers le monde - Document Interne
- Figure 1.10 – Macro Processus du processus Source to Contract - Document Interne
- Figure 1.11 – Décomposition des catégories ASL en sous-catégories
- Figure 1.12 – Distribution des sous-catégories ASL par rapport au nombre de fournisseur et au total des dépenses annuelles
- Figure 1.13 – Formule de calcul de l’indicateur - Editeur DAX
- Figure 1.14 – Liste des sous-catégories après sélection
- Figure 1.15 – Changements des parts de marché des fournisseurs par sous-catégorie
- Figure 1.16 – Illustration des sommes de variance et évolution des deux critères par rapport aux sous-catégories
- Figure 1.17 – Variabilité relative des sous-catégories entre elles selon les deux critères
- Figure 1.18 – Processus de planification : Cartographie de niveau 2
- Figure 1.19 – Processus d’approvisionnement : Cartographie de niveau 2
- Figure 2.1 – Niveaux de complexité de la supply chain
- Figure 2.2 – Composition de la supply chain pétrolière
- Figure 2.3 – Évolution des types de relations fournisseur
- Figure 2.4 – Profile de l’inventaire cyclique
- Figure 2.5 – Profile de l’inventaire avec un stock de sécurité
- Figure 3.1 – Schéma représentant les types d’apprentissage automatique
- Figure 3.2 - Les étapes de la CRISP-DM
- Figure 3.3 – Illustration du déroulement de l’algorithme K-means
- Figure 3.4 – Exemple de dendrogramme
- Figure 3.5 – Exemple de clustering avec Means Shift
- Figure 3.6 – Comparaison des résultats obtenus par DBSCAN et K-Means
- Figure 3.7 – Exemple illustrative de la méthode du coude
- Figure 3.8 – Les distances entre a et b sur un exemple de clustering
- Figure 3.9 – Exemple illustratif de la conception d’un neurone artificiel
- Figure 3.10 – Schématisation du fonctionnement d’une cellule RNN
- Figure 3.11 – Schématisation du fonctionnement d’une cellule LSTM
- Figure 3.12 – Architecture des différents modèles de synthétisation des données
- Figure 4.9 – Les coefficients de silhouette en fonction du nombre de clusters
- Figure 4.10 – Variation de l’inertie en fonction du nombre de clusters
- Figure 4.11 – Le dendrogramme
- Figure 4.12 – Les valeurs de silhouette pour différentes k de CAH
- Figure 4.13 – Capture sur résultat du DBSCAN
- Figure 4.14 – Capture sur les résultats du MeanShift
- Figure 4.15 – Scores des algorithmes par rapport aux métriques d’évaluation

Figure 4.16 – Affectation des notes selon le classement
Figure 4.17 – Visualisation des résultats de l'évaluation
Figure 4.18 – Répartition des clusters par l'algorithme CAH
Figure 4.19 – Evolution de consommation des pièces par rapport à la nouvelle classification
Figure 4.20 – Degré d'automatisation par nouvelles classes
Figure 4.21 – Evolution de par classes et par niveau d'automatisation
Figure 4.22 – Boxplot des lead time pour chaque catégorie
Figure 4.23 – Capture d'écran du calcul du nouveau stock de sécurité
Figure 4.24 – Aperçu du modèle relationnel établi pour la conception du tableau de bord
Figure 4.25 – Interface du tableau de bord
Figure 5.1 – Evolution de la consommation de la pièce de rechange
Figure 5.2 – Corrélogramme de la série
Figure 5.3 – Résultat du test d'Anova
Figure 5.4 – Résultat de l'estimation des paramètres du LES
Figure 5.5 – Résultat de l'estimation des paramètres du LED
Figure 5.6 – Résultat de l'estimation des paramètres de HWNS
Figure 5.7 – Test de la racine unitaire avec tendance
Figure 5.8 – Corrélogramme d'autocorrélation de la série différenciée
Figure 5.9 – Corrélogramme des résidus
Figure 5.10 – Résultats des prévisions du modèle
Figure 5.11 – Graphique des prévisions par Facebook Prophet
Figure 5.12 – Graphique des prévisions par l'algorithme XGboost
Figure 5.13 - Graphique des Prévisions via les Réseaux de Neurones Récurrents (RNN)
Figure 5.15 – Kernel Density plot entre les données réelles et synthétiques
Figure 5.16 – Evolution des moyennes d'amélioration de l'erreur MAPE
Figure 5.17 – Moyenne des résultats par rapport aux pièces étudiées
Figure A.1 – Caractéristiques de qualité des différents pétroles bruts
Figure A.2 – Les plus grands producteurs de pétrole
Figure A.3 – Processus Procure to Pay
Figure A.4 – Processus Material Management (Physical Inventory)
Figure A.5 – Processus Logistics Management (High level process)
Figure B.1 – Nombre des fournisseurs et somme des variances par sous-catégorie de fournisseur
Figure B.2 – Variation des parts de marché des fournisseurs M&S
Figure C.1 – Schématisation des processus SCOR
Figure C.2 – Positionnement des référentiels
Figure C.3 – Domaines couverts par les référentiels logistiques
Figure D.1 – Schéma des méthodes de prévision
Figure D.2 – Stratégie simplifiée de Dicky Fuller
Figure D.3 – Schéma de la méthodologie Box Jenkins
Figure E.1 – Illustration de la fonction Sigmoidale
Figure E.2 – Illustration de la fonction ReLu
Figure E.3 – Illustration de la fonction Tanh
Figure F.1 – Capture du code utilisé pour avoir table du silhouette score
Figure F.2 – Capture de l'algorithme kmeans pour créations de 3 clusters et nombre de pièces par cluster
Figure F.3 – Capture d'écran de l'algorithme CAH
Figure F.4 – Capture de l'algorithme DbSCAN
Figure F.5 – Capture de l'algorithme Mini Batch kmeans
Figure F.6 – Capture de l'algorithme Meanshift pour créations des clusters
Figure H.1 – Résultats des prévisions par le modèle LES

Figure H.2 – Résultats des prévisions par le modèle LED
Figure H.3 - Résultats des prévisions par le modèle HWNS
Figure H.4 – Résultats des prévisions par HWNS
Figure H.5 - Résultat du test DF avec contrainte
Figure H.6 - Résultat du test DF avec tendance et contrainte sur la série différenciée
Figure H.7 - Résultat du test DF avec contrainte sur la série différenciée
Figure H.8 - Résultat du test DF modèle [3] sur la série différenciée
Figure H.9 : Résultat de l'automatic ARIMA
Figure H.10 : Résultat du test de student pour l'automatic ARIMA
Figure H.11 : Résultat du test de student pour ARMA(3,3)
Figure H.12 : Résultat du test de student pour AR(3)
Figure H.13 : Résultat du test de student pour MA(3)
Figure H.14 : Résultats des prévisions par le modèle AR(3)
Figure H.15 : Test de normalité du modèle AR(3)
Figure I.1 – Capture d'écran du code du modèle Facebook Prophet
Figure I.2 – Capture d'écran sur le code utilisé pour les prévisions en utilisant Xgboost
Figure I.3 – Capture d'écran sur le code utilisé pour les prévisions en utilisant les RNN
Figure I.4 – Capture d'écran sur le code utilisé pour les prévisions en utilisant LSTM
Figure I.5 – Capture d'écran sur le code utilisé pour les prévisions en utilisant LSTM
Figure I.6 – Code de la génération de données
Figure I.7 – Performance de la génération de données

Liste des Abréviations

AI	Intelligence Artificielle
AR	Modèle Autorégressif
ARMA	Modèle Autorégressif Moyenne Mobile
ASL	Approved Supplier List
ASLOG	Association française pour la Logistique
BI	Business Intelligence (Intelligence d'Affaires)
BL	Business Line
CAH	Classification Ascendante Hiérarchique
CEO	Chief Executive Officer
CNPC	China National Petroleum Corporation
CPDP	Comité Professionnel du Pétrole
COPS	Compagnie d'Opérations Pétrolières Schlumberger
CRM	Customer Relationship Management (Gestion de la Relation Client)
CRISP-DM	Cross Industry Standard Process for Data Mining
CV	Critère de Variabilité
D&I	Digital & Integration
D2D	Demand to Deliver
DAX	Data Analysis Expressions
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DNN	Réseaux de neurones profonds (Deep Neural Networks)
DS	Differency Stationary (stationnaire en différences)
EIA	Energy Information Administration
EMS	Enterprise Management Systems
ERP	Enterprise Resource Planning (Planification des Ressources de l'Entreprise)
GANs	Generative Adversarial Networks
GOLD	Global Oilfield Logistics and Distribution
GSC	Global Sourcing Centers
HSE	Health, Safety, and Environment
HESP	Halliburton Entreprise de Services aux Puits
HW	Holt-Winters
HWNS	Holt-Winters Non Saisonnier
HWM	Holt-Winters Multiplicatif
HWA	Holt-Winters Additif
IBP	Integrated Business Planning (Planification commerciale intégrée)
ISCM	Internal Supply Chain Management
KPI	Key Performance Indicator (Indicateur de Performance Clé)
LSTM	Réseaux récurrents à mémoire court et long terme (Long Short-Term Memory)
LES	Lissage Exponentiel Simple
LED	Lissage Exponentiel Double
MA	Moyenne Mobile
MAE	Mean Absolute Error (Écart Absolu Moyen)
MAPE	Mean Absolute Percentage Error (Pourcentage d'Erreur Absolu Moyen)

M&S Materials & Supplies (Matériaux et fournitures)
MSE Mean Squared Error (Erreur Quadratique Moyenne)
NAF North Africa
P&SC Planning and Supply Chain
PAR Probabilistic AutoRegressive
PO Purchase Orders (Bon de Commande)
PWC PriceWaterhouseCoopers
RFI Request for Information
RFP Request for Proposal
RFQ Request for Quotation
RNN Réseaux de neurones récurrents (Recurrent Neural Networks)
RMSE Root Mean Squared Error (Erreur Quadratique Moyenne Racine)
SCM Supply Chain Management (Gestion de la Chaîne d'Approvisionnement)
SCOR Supply Chain Operations Reference (Référence des opérations de la Supply Chain)
SLB Schlumberger Limited
SPS Service Pétrolier Schlumberger
SRM Supplier Relationship Management
TLM Technology Lifecycle Management (Gestion du cycle de vie technologique)
TS Trend Stationary (stationnaire en tendance)
TVAE Tabular Variational Auto-Encoder
VAT Visual Assessment of Tendency

Introduction Générale

Introduction Générale

Les produits pétroliers et gaziers représentent des matières premières de grande importance, dont l'industrie occupe une place primordiale dans l'économie mondiale, notamment dans le secteur des énergies. Depuis des décennies, ces produits ont été les principales sources d'énergie consommées, favorisant la croissance et le développement de nombreux pays.

Cette industrie est particulièrement instable et sensible aux événements qui affectent le marché économique mondial, telles que les restrictions réglementaires et environnementales, les crises politiques, les différents événements géopolitiques, la pandémie de Covid-19, qui a considérablement fragilisé sa Supply Chain, ainsi que les divers conflits militaires actuels. Ces éléments ont mis en évidence les différentes failles et vulnérabilités de ce secteur.

Le marché des produits et services pétroliers, crucial pour l'industrie, est marqué par une instabilité considérable. Cette instabilité est principalement due à la sensibilité du secteur d'activité et à la fluctuation des prix du brut. En outre, l'augmentation des frais de transport et des prix des matières premières a gravement perturbé le marché.

Dans cette optique, l'entreprise de services et équipements pétroliers SLB, sur laquelle notre travail s'est concentré, vise à améliorer sa résilience face à ces diverses perturbations. Elle cherche à adapter sa Supply Chain en amont pour garantir la disponibilité des matières premières tout en optimisant le rapport Prix-Qualité-Délais. De ce fait, l'adaptation des approvisionnements est primordiale pour assurer une supply chain résiliente, étant donné leur position en amont et leur poids significatif qui impacte l'ensemble des maillons de la chaîne.

Dans le cadre de notre projet, nous avons intégré l'équipe d'approvisionnement de SLB dans le but d'aider l'entreprise à satisfaire sa demande en termes d'approvisionnements et à en assurer la disponibilité, tout en visant à améliorer les éléments de ce processus, notamment la gestion de la relation fournisseur à travers la stabilisation des parts de marché de ces derniers, qui repose en partie sur deux piliers essentiels : la stabilisation des commandes et la fiabilisation des prévisions des commandes.

Dans une première partie, nous avons commencé par réaliser une analyse des dépenses par rapport aux fournisseurs de SLB, dans le but de cadrer notre sujet et de mettre en contexte les éléments de notre travail. Cette étape nous a permis de nous orienter vers la catégorie des pièces de rechange (M&S) au sein de la division Reservoir Performance (RP). Ensuite, nous avons effectué un audit interne des processus en amont de l'entreprise afin d'identifier les différents dysfonctionnements rencontrés par SLB. Cette analyse nous a conduit à poser la question suivante :

Comment améliorer la gestion et la planification des pièces de rechanges M&S et contribuer à l'amélioration des contrats fournisseurs à travers la proposition d'un engagement en volume ?

Afin de répondre à cette question, nous avons d'abord travaillé sur l'identification de nouvelles catégories pour les pièces de rechange (M&S) et déterminé celles pouvant faire l'objet de commandes automatisées. Cette étape a été réalisée en appliquant des modèles de clustering et en comparant leurs résultats. Ensuite, nous avons recalculé les niveaux de stock de sécurité en nous basant sur l'historique de consommation, tout en proposant un tableau de bord illustrant l'apport de cette nouvelle classification et facilitant la gestion des stocks en combinant de multiples indicateurs.

Dans un second temps, nous avons sélectionné un contrat existant avec un fournisseur pour l'améliorer en proposant un engagement par volume. Cette amélioration a été effectuée en sélectionnant les pièces les plus importantes grâce à une classification combinée ABC-XYZ, puis en appliquant une étude prévisionnelle sur les pièces sélectionnées en se basant sur nos données de base dans un premier temps, puis en proposant une génération de données synthétiques afin de fiabiliser nos prévisions. Ces prévisions nous ont conduit à une estimation probabiliste des réductions du coût de transport, offrant ainsi la possibilité de réaliser des économies.

Le présent travail est illustré et réparti sur cinq chapitres. Le premier chapitre a pour objectif de présenter le marché pétrolier et parapétrolier, l'entreprise, sa structure, et ses processus, ainsi que d'exposer notre diagnostic et la problématique identifiée.

Les deuxième et troisième chapitre explicitent les concepts et définitions théoriques nécessaires à la conception de notre solution, présentent les différents algorithmes utilisés, ainsi que les modèles prévisionnels.

Les quatrième et cinquième chapitre détaillent le développement de notre solution sous ses deux aspects : la révision des classes des produits M&S dans le but d'optimiser les commandes automatiques et l'amélioration des contrats via l'étude prévisionnelle et la proposition d'engagement par volume visant la réalisation d'économies.

Enfin, nous concluons notre travail avec une conclusion.

Chapitre 1 – SLB : Contexte,
marché des fournisseurs et
formulation de la problématique

1 SLB : Contexte, marché des fournisseurs et formulation de la problématique

Dans ce chapitre, nous débuterons par une présentation de l'entreprise SLB et de son secteur d'activité, à savoir, le marché des hydrocarbures. Nous exposerons également ses activités clés ainsi que son organisation. Ensuite, nous nous pencherons plus en détail sur la Supply Chain de l'entreprise, notamment les départements du Procurement & Sourcing (Achats et approvisionnements) ainsi que du Supplier Management (Gestion des fournisseurs). Nous présenterons et analyserons leurs processus clés. Par la suite, nous effectuerons un diagnostic de l'état des lieux afin de mieux appréhender l'environnement de la problématique qui sera abordée dans ce travail. Nous détaillerons cette problématique vers la fin de ce chapitre.

1.1 Présentation de SLB et de son secteur d'activité

1.1.1 Le marché pétrolier et parapétrolier

Nous aborderons dans cette partie le marché pétrolier et ses caractéristiques ainsi que celui des services pétroliers (appelé aussi marché parapétrolier), dans le but d'apporter une vue sur les spécificités du secteur du pétrole et du gaz.

Le pétrole, une source d'énergie fossile, est l'une des ressources les plus convoitées au monde, ayant permis l'émergence des économies industrielles au 20^{ème} siècle grâce à sa capacité à contenir une grande quantité d'énergie dans un petit volume transportable. Le marché du pétrole est structuré de deux manières principales : le marché "spot" ou de gré à gré, où le prix est déterminé par l'offre et la demande. L'offre provient des compagnies d'extraction et de production de pétrole (comme Aramco, ExxonMobil, BP), tandis que la demande vient des raffineurs qui transforment le brut en produits utilisables. L'achat passe souvent par des traders visant des profits à court terme en achetant et revendant le pétrole.

Le second marché est le marché à terme, et signifie que le pétrole est acheté et vendu à terme et à un prix fixé à l'avance (La finance pour tous, 2024).

Le prix du pétrole mondial varie en fonction de la qualité du produit (viscosité, teneur en soufre, etc.). Deux types de pétrole brut servent de référence pour fixer les prix : le West Texas Intermediate (WTI) produit aux États-Unis et le Brent produit en Mer du Nord. D'autres types de pétrole, avec des caractéristiques différentes, influencent également leurs prix (Figure A.1 en Annexe A). Le WTI et le Brent sont privilégiés pour leur liquidité sur les marchés à terme de New York et Londres, permettant des achats et des ventes à différentes échéances.

Le marché pétrolier est influencé par le prix de référence de l'OPEP (Organisation des Pays Exportateurs de Pétrole), créée en 1960 et regroupant actuellement 13 membres, représentant près de 40% de la production mondiale. Cette organisation fonctionne comme un cartel visant à réguler le prix du pétrole en limitant la production de ses membres à des quotas déterminés pour chacun.

La figure 1.1 illustre l'évolution des prix du pétrole brut (Brent) comparé au prix de référence de l'OPEP et du Sahara Blend.

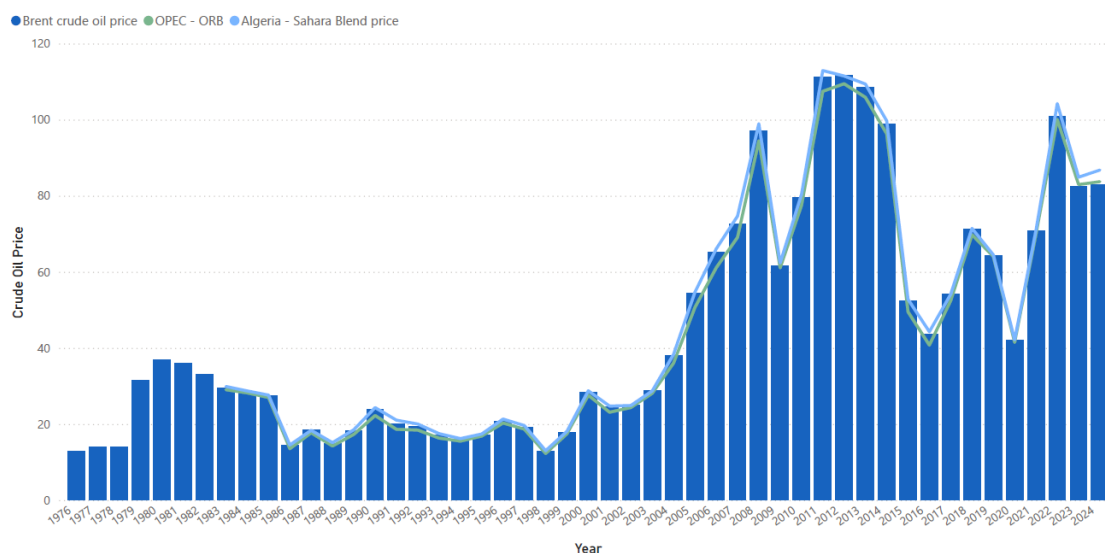


Figure 1.1 – Evolution des prix du pétrole brut (Oil Price, 2024; OPEC, 2024; Statista, 2024)

- Dans le monde :

Suite à la pandémie de Covid-19, le marché mondial du pétrole et du gaz a été durement impacté par une réduction de la consommation de produits pétroliers raffinés et un ralentissement économique global. Cette situation, combinée à une forte volatilité des prix du pétrole brut, a retardé de nombreux projets de raffinerie. Cependant, la reprise économique mondiale et l'augmentation de la consommation dans des pays comme l'Inde et la Chine ont fait grimper la demande mondiale en pétrole brut à 101,7 millions de barils par jour (Mb/j) fin 2023. Selon un récent rapport de l'EIA (Energy Information Administration, 2024), cette tendance devrait se poursuivre avec une consommation prévue de 102,9 Mb/j en 2024 et de 104,3 Mb/j en 2025.

Selon le rapport annuel de la SONATRACH (SONATRACH, 2022), les prix du pétrole brut ont fortement augmenté en 2022, atteignant 101,3 \$ par baril, une hausse de 43 % par rapport à 2021, avec un pic à 137,6 \$/b. Cette augmentation est due au conflit russo-ukrainien, aux faibles stocks mondiaux, à la régulation de l'offre par l'OPEP+ et à la hausse des prix du gaz.

Les principaux pays producteurs sont, les Etats-Unis, l'Arabie Saoudite et la Russie qui totalisent une production avoisinant les 44% de la production mondiale, suivi du Canada, de l'Irak, de la Chine et bien d'autres selon le rapport du CPDP (Comité Professionnel du Pétrole, 2023) (Figure A.2 en Annexe A).

- En Algérie :

Le volume de production du marché algérien du pétrole et du gaz devrait atteindre au cours de cette année de 2024 le cap de 1,53 million de barils par jour et selon les prévisions de la période (2024-2029) cette production pourrait atteindre le 1,81 Mb/j avec un TCAC (taux de croissance annuel composé) de 3,41% (Mordor Intelligence, 2024).

Cette augmentation de production est due en majeure partie à l'augmentation des investissements pétroliers et gaziers, ou selon (Ministère de l'Energie et des Mines, 2019) et afin de satisfaire la demande du marché national, le secteur des hydrocarbures prévoyait la réalisation de deux nouvelles raffineries et deux unités d'hydrocraquage. Ces projets ont permis à la capacité totale de traitement de passer de 24,62 à 35,565 Millions Tonnes/an.

En 2023, l'Algérie se classe au 4e rang mondial des exportateurs de gaz et au 16e rang pour ses réserves prouvées de pétrole ou Sonatrach, la société publique énergétique et ses filiales dominent le marché algérien des hydrocarbures avec environ 80% de la production. Le reste est géré par des compagnies pétrolières internationales telles qu'Eni, CNPC, Equinor, BP, entre autres, qui dépendent largement des entreprises parapétrolières pour leurs opérations.

Le marché parapétrolier

Le marché parapétrolier englobe les services qui soutiennent l'extraction et la production de pétrole et de gaz sur terre et en mer. Cela inclut la construction, la complétion et la production des puits. Les principaux services incluent les opérations sous-marines, sismiques, de forage, de réparation et de complétion, ainsi que la fourniture d'équipements de production et de services de traitement et de séparation (The Business Research Company, 2024). Le secteur est évalué à 119,85 milliards de dollars en 2024 et devrait croître à un taux annuel composé (TACT) de 5,13%, pour atteindre 153,49 milliards de dollars d'ici 2029 (Mordor Intelligence., 2023).

La demande mondiale croissante en énergie devrait dynamiser le marché des services pétroliers. Cette demande accrue stimulera les activités d'exploration et de production, créant ainsi des opportunités pour les services pétroliers tels que le forage et la gestion des réservoirs. Par exemple, aux États-Unis, la consommation d'énergie dans le secteur industriel devrait augmenter de 5 % à 32 % entre 2022 et 2050, alimentant ainsi la croissance du marché des services pétroliers pour répondre aux besoins énergétiques mondiaux en expansion.

Oilfield Services (OFS) Market: Growth Rate by Region, 2023-2028

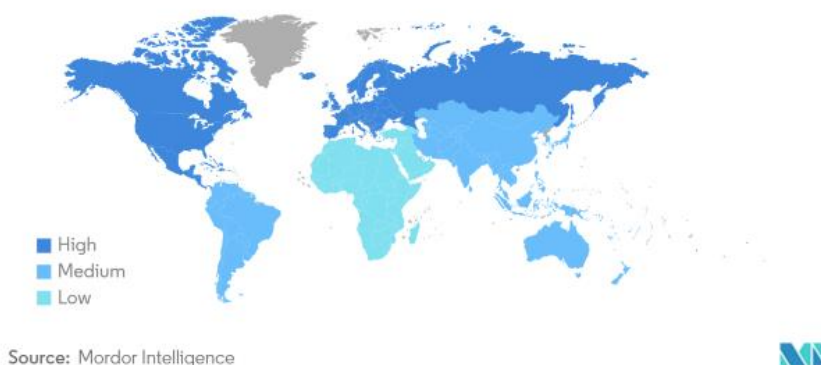


Figure 1.2 – Evolution du marché des services pétrolier (Mordor Intelligence., 2023)

Le marché des services pétroliers est fragmenté, ce qui signifie qu'il est hautement concurrentiel sans acteur dominant. Parmi les principaux acteurs du marché, on trouve Schlumberger Limited avec la plus grande part de marché de 13.82% en 2023 (CSI Market, 2024), Baker Hughes Company, Halliburton Company, Weatherford International Plc, et China Oilfield Services Limited, entre autres.

Dans le marché algérien, on peut citer :

- Slb ex Schlumberger
- HESP, Halliburton Entreprise de Services aux Puits
- Baker Hughes
- OilServ
- Nesr
- Weatherford

1.1.2 Présentation de SLB

A travers cette partie, nous allons présenter l'établissement d'accueil ou nous avons effectué notre stage de fin d'étude, à savoir, Slb ex Schlumberger. Nous nous étalerons sur sa présentation en générale, puis nous aborderons sa structure organisationnelle en GeoUnit et ses différentes divisions, nous aborderons ensuite SLB NAF ou North Africa et slb Algérie. Nous présenterons par la suite la supply chain de l'entreprise.

SLB Limited

SLB est une grande entreprise multinationale parapétrolière qui s'est spécialisée dans les services pétroliers et gaziers. Elle a été fondée en 1926 par les frères Schlumberger, à savoir, Conrad et Marcel, sous le nom de "Société de Prospection Electrique" ou "Pros" et cela après plus d'une quinzaine d'années d'essais en laboratoire sur la résistivité électrique de la roche à huile pour pouvoir déterminer les différents types de sous-sol (SLB, 2024).

Avec un effectif de 111 000 salariés représentant plus de 160 nationalités et un chiffre d'affaires de 33,14 milliards de dollars en 2023, en hausse de 18% par rapport à l'année précédente, cette entreprise pétrolière est présente dans plus de 120 pays à travers le monde, avec des bureaux exécutifs à Paris, Houston, Londres et La Haye (Zone Bourse, 2024). Schlumberger, leader mondial des services pétroliers, offre une gamme complète de technologies, de solutions d'information et de services de gestion de projets qui améliorent les performances des réservoirs dans l'industrie pétrolière et gazière. Ses services couvrent un large éventail, allant de la sismique de surface au forage, en passant par l'évaluation des formations, la complétion de puits et les services de stimulation, l'optimisation de la production, les études de réservoir, la construction de puits et la gestion de projets (SLB, 2024).

En fin 2022, Schlumberger a changé d'identité en devenant SLB et a annoncé se réorienter en une entreprise technologique mondiale axée sur l'innovation énergétique. SLB se concentrera désormais sur quatre domaines qui sont, Les nouveaux systèmes énergétiques, La décarbonisation industrielle, Le numérique à grande échelle et L'innovation pétrolière et gazière.

Le tableau suivant résume les principales informations sur l'entreprise :

Nom de l'entreprise	SLB Ltd anciennement Schlumberger
Date de création	1926
Membres fondateur	Marcel et Conrad Schlumberger
Forme juridique	Société anonyme avec appel public à l'épargne
Siège social	Bureaux principaux à Houston, Paris, et la Haye.
CEO	Olivier le Peuch
Secteur d'activité	Prestation de Services pétroliers
Capital Social	74 327 millions \$
Chiffres d'affaires	33 135 millions \$
Résultats Net	4 203 millions \$

Tableau 1.1 – Carte d'Identité de SLB - SLB 2024

Le CA par activité se répartit comme suit :

- développement et construction de puits (39,9%)
- vente d'équipements et de systèmes de production pétroliers (27,6%)
- vente de solutions et de services d'optimisation de la performance et du rendement des réservoirs (19,5%)
- vente de solutions et de services technologiques et géophysiques (13%) : vente de logiciels, acquisition et traitement de données sismiques, prestations conseil en caractérisation des réservoirs, etc.

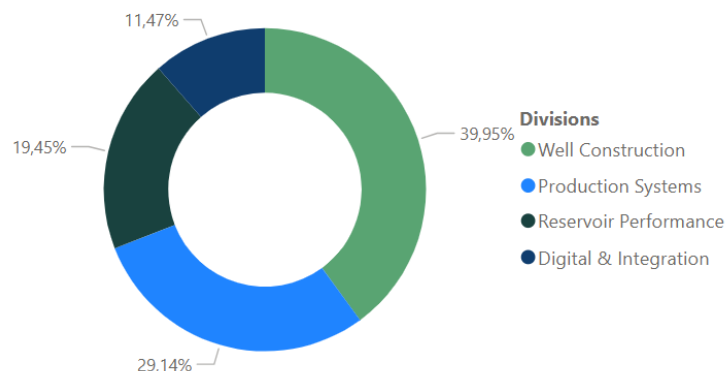


Figure 1.3 – Répartition du Chiffre d’Affaires 2023 (Zone Bourse, 2024)

Le chiffre d'affaires est réparti géographiquement comme suit : Amérique du Nord (21,3%), Moyen-Orient et Asie (32,2%), Europe-CEI-Afrique (25,6%), Amérique latine (20,2%) et autres régions (0,7%) (Zone Bourse, 2024).

Les divisions de l’entreprise

Au fil du temps, SLB a pu se démarquer sur le marché parapétrolier en touchant tous les marchés significatifs du secteur. Elle gère ses activités à travers des divisions composées de business lines. Slb est structurée en quatre divisions qui sont à leur tour composées de plusieurs Business Lines (BL) qui incluent différents produits, services et technologies qui répondent à l’ensemble du cycle de vie du réservoir.

Ces divisions en question sont : Well Construction, Reservoir Performance, Production Systems et Digital & Integration.

1. Digital & Integration :

La division Digital & Integration (D&I) réunit les activités logicielles, sismiques et numériques. Son objectif est d’intégrer les données, les technologies numériques et les processus pour améliorer les actifs de l’entreprise. Cela inclut l’exploitation des données sismiques, l’automatisation du forage et de la production, la gestion des projets de production et de construction de puits, ainsi que le suivi des performances des actifs. Avec la croissance numérique en cours, cette division offre un fort potentiel de développement pour l’entreprise.

Elle est composée de cinq Business Lines

- Digital subsurface solutions : Géosciences et ingénierie des réservoirs.
- Exploration Data : Sismique multi client et traitement associé en plus des services d’exploration de données.
- Digital Operations Solutions : Permet d’offrir une gamme d’opérations digitales comme l’Automatisation du forage et de la production.
- Integrated Well Construction : Gestion et pilotage intégré de projet de construction de puits.
- Integrated Reservoir Performance : Gestion de la production, de la récupération et des performances des actifs.

2. La division Production Systems :

La division Production Systems vise à offrir et à développer des technologies avancées, ainsi qu'à fournir une expertise pour maintenir les puits en production et optimiser la récupération des ressources des réservoirs souterrains à la surface. Elle couvre divers domaines, notamment les complétions, les remontées artificielles, les équipements de surface, le traitement et les systèmes sous-marins.

Cette division se constitue des Business Lines suivantes :

- Well Production Systems : Services de complétion et systèmes de Portance Artificielle en fond de puits.
- Surface Production Systems : Offre les technologies et services comme les têtes de puits (wellhead), services de fracturation et pompes de production en surface.
- Subsea Production Systems : Systèmes de production et de traitement sous-marins.
- Midstream Production Systems : Vannes, systèmes de traitement, chimie de production et installations.

3. Reservoir Performance :

La division Reservoir Performance vise à évaluer la performance des réservoirs en comprenant le comportement des fluides, leurs propriétés, leur composition, ainsi que certaines variables du puits telles que la température et la pression. Cette évaluation permet d'identifier les techniques et méthodes qui maximisent la récupération et la production des hydrocarbures. Elle est composée de trois BLs :

- Reservoir Performance Evaluation : Services de prélèvement et d'analyse des fluides et des échantillons de roches, des services d'évaluation par câble et des tests Down Hole et cela sous un réseau de laboratoires de roches et de fluides.
- Reservoir Performance Intervention : Combine des technologies dans la production et l'intervention à travers les services de Coiled tubing (désigne un long tuyau métallique utilisé pour les interventions dans les puits), de tests en surface, d'interventions slickline, perforation et interventions par câble.
- Reservoir Performance Simulation : Qui comprend la gestion et simulation du sable, la fracturation hydraulique, l'acidification de la matrice.

4. Well Construction

Cette division se spécialise dans la construction de puits pour l'extraction des hydrocarbures. Elle propose une gamme de produits et de services visant à réaliser des forages précis et efficaces tout en minimisant les risques. La WC met en œuvre des méthodes et des technologies innovantes pour la conception et la construction des puits.

Ses Business Lines sont de l'ordre de quatre :

- Well Construction Measurement : Mesure et acquisition de données de forage.
- Well Construction Drilling : Forage directionnel et trépan.
- Well Construction Fluids : Fluides de forage et cimentation des puits.
- Well Construction Equipment : Foreuses et équipements, équipements de contrôle de pression.

La figure ci-dessous résume les différentes divisions et leurs Business Lines :

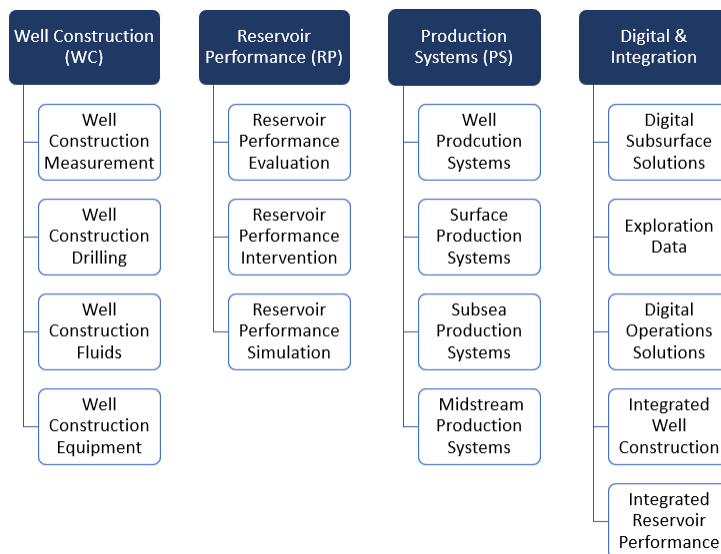


Figure 1.4 – Divisions et Business Lines de SLB - Document Interne

Organisation de l'entreprise

En 2020, SLB a opéré une importante réorganisation pour s'adapter aux défis de l'industrie pétrolière et faire face au ralentissement qu'elle a connu. Cette transformation a conduit à la création d'une organisation plus agile et réactive, structurée en quatre divisions et cinq bassins géographiques, avec un total de 30 GeoUnits, qui correspondent à un pays ou à un groupe de pays gérés au sein d'un des cinq bassins. En consolidant ses Products Lines et en remplaçant les équipes GeoMarkets par des GeoUnits en groupes de bassins, SLB a simplifié son organisation et amélioré son efficacité opérationnelle. Cette nouvelle structure permet à l'entreprise de mieux répondre aux besoins des clients locaux et de fournir des solutions personnalisées.

La figure suivante illustre la distribution des cinq bassins dans le monde.



Figure 1.5 – Carte représentant les différents bassins et GeoUnits de SLB - Document Interne

SLB NAF

SLB est présente en Afrique du Nord à travers sa division géographique ou GeoUnit SLB NAF, une GeoUnit du bassin MENA (Ou Middle East and North Africa), et qui regroupe cinq pays au total : l'Algérie, la Tunisie, la Libye, le Maroc et le Tchad, comme l'illustre la figure ci-dessous.

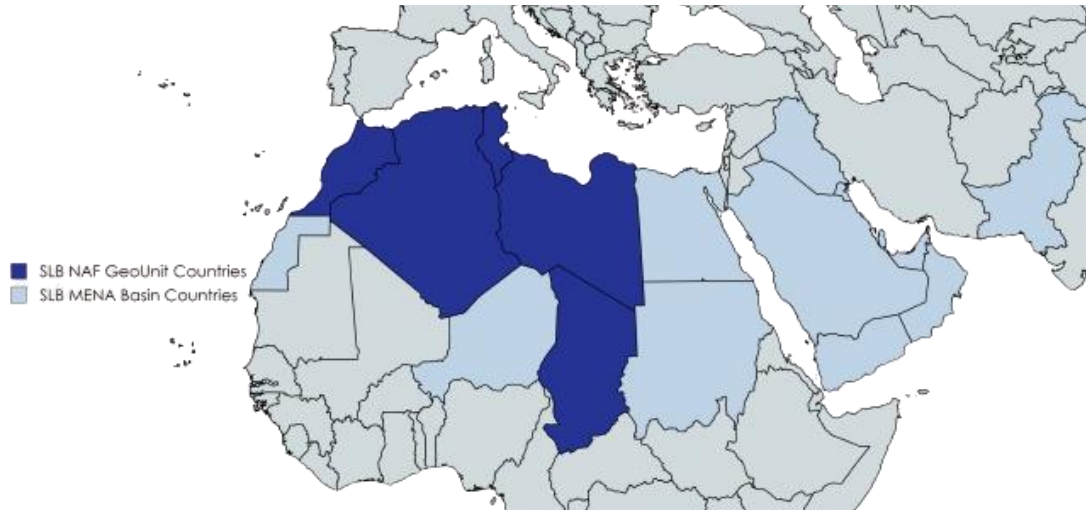


Figure 1.6 – Carte de la GeoUnit SLB NAF

Étant une région stratégique dotée de réserves importantes d'hydrocarbures, notamment dans le désert Libyque et le Sahara Algérien, elle constitue un marché à fort potentiel pour l'entreprise. Afin d'exercer ses activités et de fournir sa gamme de technologies et de services à ses clients, SLB NAF dispose d'une structure organisationnelle comme suit :

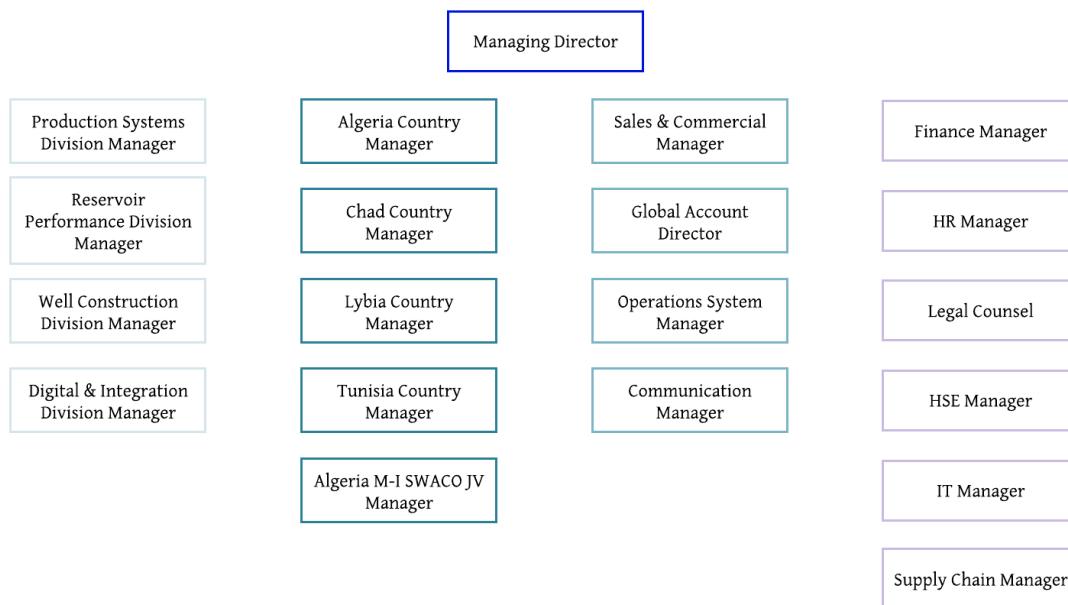


Figure 1.7 – Organisation de SLB NAF - Document Interne

SLB Algérie :

SLB Limited est présente en Algérie à travers la division géographique (GeoUnit) SLB NAF. Installée dans le pays depuis 1955, elle a son siège à Hydra, Alger. Elle opère sous quatre entités :

- “Service Pétrolier Schlumberger (SPS)”
- La “Compagnie d’Opérations Pétrolières Schlumberger (COPS)”
- Cameron Algeria SPA
- Sahara Well Construction Services SPA.

Pour exercer ses activités, l’entreprise dispose de plusieurs bases opérationnelles situées dans les régions pétrolières du sud du pays, afin de faciliter l’accès aux puits. Ces régions comprennent Hassi Messaoud, Hassi Berkine et In Amenas. Auparavant, SLB était également présente à In Salah, mais cette base a été fermée en raison de la baisse d’activité dans le marché parapétrolier.

Les bases sont réparties comme suit :

- Quatre bases à Hassi Messaoud (MD1, MD2, MD3 (Base 2 et Base 3) et la base WesternGeco)
- Une base à Hassi Berkine
- Une base à In Amenas

Grâce à sa présence sur le territoire national, SLB fournit divers services pétroliers à une multitude d’entreprises, notamment Total Energies, British Petroleum, Oxy (Ex Anadarko), et surtout SONATRACH et ses groupements, l’entreprise pétrolière et gazière nationale. Les activités de SLB en Algérie sont présentées à travers ses quatre divisions et incluent principalement :

- Études sismiques et géologiques
- Construction et test de puits
- Vente de technologies et d’équipements pétroliers

1.2 La Supply Chain de SLB

Dans le but de bien cerner la problématique abordée et apporter plus de clarté vis-à-vis des processus de SLB, nous présenterons en détail dans cette partie la Supply Chain de l'entreprise, sa structure et son organisation en mettant en lumière ceux du service d'Achats et d'Approvisionnement.

La supply chain constitue un élément crucial pour SLB. Elle vise à organiser et à connecter les différentes divisions et départements pour garantir un transfert fluide des flux physiques et informationnels, assurant ainsi la disponibilité des équipements et des matières premières pour les diverses Business Lines de l'entreprise. Elle est considérée comme une fonction support parmi les fonctions de finances, de ressources humaines et d'information et technologie (IT). La performance de sa supply chain est considérée comme capitale pour la performance globale de l'entreprise et donc sa bonne gestion constitue un axe stratégique.

La gestion de la supply chain est effectuée par l'organisation Planning and Supply Chain (P&SC), et est représentée avec les autres organisations de l'entreprise dans la Metro-Map suivante, illustrant leurs différentes activités et interconnexions.

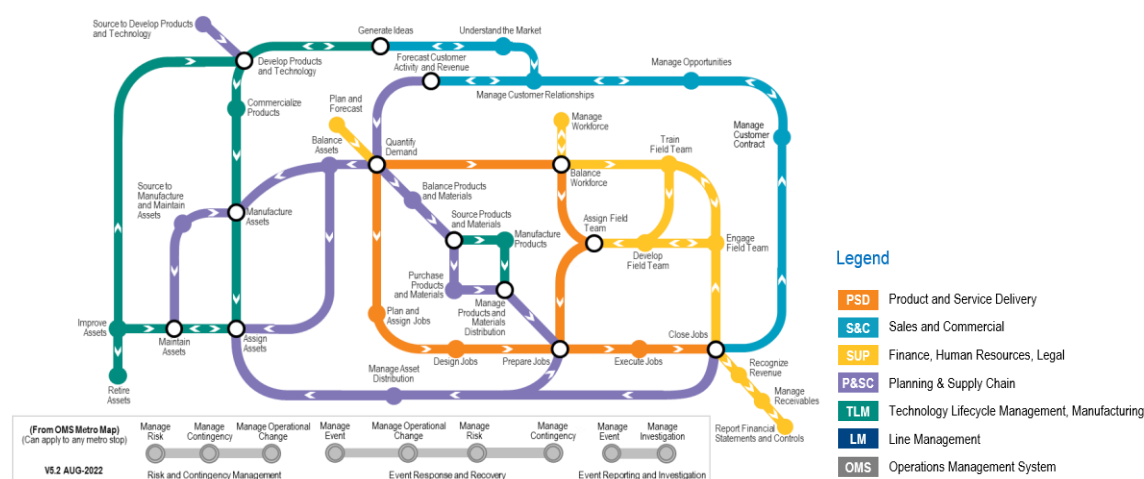


Figure 1.8 – Metro-Map de la Supply Chain de SLB – Document Interne

La Supply Chain de SLB est structurée principalement en cinq départements travaillant en osmose avec le reste des départements de l'entreprise, et sont :

- 1) Approvisionnement (Sourcing) : Ce département vise à répondre aux besoins en matières premières et équipements des divisions et des business lines. L'équipe a comme fonction la sélection des fournisseurs en fonction des critères de délai, de qualité et de coût ainsi que la contractualisation de ces derniers, tout en établissant des relations d'achats avec eux.
- 2) Gestion des fournisseurs (Supplier Management) : Ce département intervient après la contractualisation avec les fournisseurs pour gérer les relations, surveiller régulièrement leur performance et maintenir la base de données fournisseurs à jour.
- 3) Import-Export : Elle assure la gestion et le suivi des opérations d'importation et d'exportation des besoins nécessaires aux activités des Business Lines ainsi que les procédures de dédouanement.
- 4) Logistique domestique (Domestic Logistics) : Cette fonction assure la planification et l'optimisation du transport local des équipements et des matières entre les différentes bases et chantiers de l'entreprise.

- 5) Materials Management : Elle s'occupe de la gestion des flux physiques de matériel et des stocks des entrepôts de l'entreprise. Ces matériels comprennent les "materials and supplies" (M&S), les actifs et l'inventaire financier.

1.2.1 Organisation internationale de la Supply Chain de SLB

Sa présence dans plus de 120 pays lui a permis de développer une supply chain globale lui permettant de s'approvisionner de nombreux fournisseurs de différentes régions dans le monde en maintenant des stocks dans des emplacements stratégiques pour rester réactive face à son activité.

L'évolution de sa Supply Chain a contribué à la création d'une structure qui est la "Global Oilfield Logistics and Distribution" (GOLD) en 2002. C'est une structure mondiale créée par SLB pour gérer sa logistique internationale, c'est aussi un processus métier et un système qui organise, coordonne et contrôle la Supply Chain (logistique et distribution) relative à tous les équipements, matériels et fournitures nécessaires aux sites de l'entreprise (sites opérationnels, centres EMS) à travers une organisation logistique mondiale. Cette organisation logistique consiste en des centres de services distribution (Distribution Service Center DSC) qui sont des hubs de stockage localisés stratégiquement à travers le globe dans le but de fournir un emplacement commun à un groupement de GeoUnits pour consolider les commandes et stocker les équipements et matériels ayant une demande et une criticité élevée.

Les principales activités de la structure sont :

- L'achat auprès des fournisseurs internationaux et la vente aux entités de SLB.
- La consolidation des commandes d'expédition pour les différents centres.
- Le transfert des commandes vers les centres en utilisant le mode de transport le plus efficace.

Il existe deux types d'expédition à l'international :

- Les expéditions GOLD se réfèrent à toutes les commandes internationales passées auprès de fournisseurs qui ne se trouvent pas dans le même pays que le site demandeur.
- Les expéditions non-GOLD se réfèrent à toutes les commandes (produits) passées auprès de fournisseurs situés dans le même pays que le site demandeur.

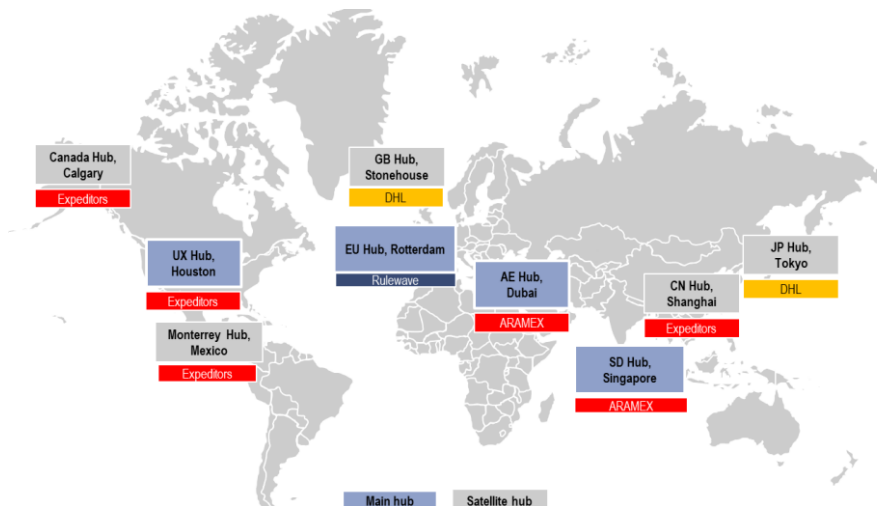


Figure 1.9 – Répartition des DSC de Schlumberger à travers le monde - Document Interne

1.2.2 Les processus clés de la Supply Chain de SLB

La Supply Chain de l'entreprise dépend de quatre processus clés qui sont listés dans le tableau suivant. Ces processus interviennent dans la recherche des fournisseurs de matières premières et équipements à leur livraison et suivi des stocks tout en passant par le management des fournisseurs, le suivi et l'acheminement des commandes, les stratégies de paiement ainsi que le transport.

Intitulé	Description	Processus
Source to contract	Processus pour la stratégie d'approvisionnement à travers des plans de sourcing, des activités de sourcing, la gestion des fournisseurs et l'exécution et la mise en œuvre des contrats.	Source to Contract
Source to Pay	Processus pour exécuter les exigences d'achat grâce à une gestion efficace des bons de commande, à la collaboration avec les fournisseurs et aux processus de comptes fournisseurs.	Purchase To Pay
Warehouse Management	Processus pour gérer les opérations d'entrepôt, de la réception des marchandises jusqu'à la préparation et l'emballage des pièces pour une mission ou un transfert.	Warehouse Management
Logistics Management	Processus de planification et d'exécution du transport des marchandises entre deux sites.	Logistics Management

Tableau 1.2 – Description des processus clés de la Supply Chain de SLB - Document Interne

En ce qui concerne notre projet, nous nous intéresserons plus particulièrement au processus "Source to Contract" qui est sous la responsabilité du département Procurement and Sourcing et qui sera détaillé dans la partie qui suit. Les autres processus seront détaillés en annexe A.

Processus Source to Contract :

Ce processus est crucial pour créer de la valeur dans la Supply Chain de l'entreprise. Il repose sur trois éléments : le "Category Management" pour l'élaboration des plans d'approvisionnement pour des catégories particulières de produits et services gérées globalement, le "Sourcing" pour l'exécution des activités d'approvisionnement, et le "Supplier Management" pour la gestion des relations avec les fournisseurs, l'exécution et la mise en œuvre des contrats. Le processus Source to Contract (S2C) se compose de quatre sous-processus, qui sont les suivants.

1. Gestion de la Stratégie de Catégorie :

La gestion de catégories (Category Management) définit la stratégie d'approvisionnement à long terme pour une catégorie de dépenses, en tenant compte des objectifs de l'entreprise, des exigences des parties prenantes, de la demande prévue et du coût total de possession.

- Pour proposer une stratégie de sourcing efficace pour chaque catégorie de dépenses, le processus intègre les inputs suivants :
 - Consolidation de la demande nette globale pour les dépenses directes et indirectes (par le Division Demand to Deliver Manager).
 - Function Management fournit des détails sur la demande pour les catégories de dépenses indirectes (IT, RH, etc.).
 - Le division Technology Development Manager fournit des informations sur les initiatives technologiques et leurs impacts sur la demande.
 - Le GU SC Manager apporte des aperçus du marché et des projets de sourcing.
 - Le Division GU P&SC Manager consolide la demande de la GeoUnit qu'il gère.
 - Le Supplier Manager fournit des informations sur la performance des fournisseurs.
- Analyse et Élaboration du Plan de groupe : Le Category Manager analyse les données (inputs) des parties prenantes pour développer des stratégies de catégorie et des plans globaux adaptés au mieux aux commodités des catégories de dépenses.

Le plan de groupe inclut :

- Un plan de sourcing (incluant la rationalisation, le resourcing et les exigences des nouveaux fournisseurs).
 - Un plan de gestion des fournisseurs avec des objectifs et des exigences de développement pour le fournisseur.
 - Des fournisseurs préférés pour chaque catégorie de dépenses.
- Approbation du Plan de Groupe : Le Category Manager obtient l'approbation des Gestionnaires de la Planification (P&SC Manager) et de la Supply Chain de la Division (GU SC Manager), représentant au moins 10% des dépenses du groupe afin d'assurer la cohérence entre le plan de dépense et les objectifs de la division.
 - Communication et Suivi : Le Category Manager communique le plan (sourcing plan et le plan de gestion des fournisseurs) aux parties prenantes internes. L'équipe de Category Management suit la mise en œuvre et ajuste les actions en fonction de l'impact et du consensus des parties prenantes.

2. Définition du « Sourcing & Supply Base Plan » :

Ce processus établit la conception du plan de base d'approvisionnement et la création du Plan de Sourcing spécifique à une Division dans une Unité Géographique (GeoUnit) ou un Centre. Cela permet d'assurer l'alignement entre la stratégie d'achat et les exigences des divisions.

Le plan de base d'approvisionnement (Supply base plan) doit assurer une concurrence adéquate et une concentration optimale des fournisseurs, tout en minimisant les risques liés à la Supply Chain. Il doit également répondre aux exigences locales et réglementaires, réduire les coûts et garantir la livraison des biens et services avec les performances requises.

Les étapes suivantes doivent être assurées :

- Le Category Manager propose un Guide d'Achat et un Plan de Sourcing spécifique à la Division.
- Le « Demand to Deliver » (D2D) de la division recueille les exigences relatives au Supply Base auprès des parties prenantes, notamment la performance et engagement des fournisseurs ; besoins en nouveaux fournisseurs ; contraintes de localisation ; adéquation de la base de fournisseurs existante ; et autres facteurs.
- Le D2D révisé et ajuste les Guides d'Achat et le Plan de Sourcing avec le Category Manager, puis les approuve.

- Le Category Manager met à jour trimestriellement les Guides d'Achat et communique le Plan de Sourcing aux équipes de Sourcing globales et des Unités Géographiques pour exécution.

Remarque :

- L'équipe centrale de gestion de catégories (Central Category Management Team) définit et maintient les Guides d'Achat pour les catégories globales.
- Catégories Locales : Chaque équipe sourcing locale des GeoUnits conçoit et maintient la base de fournisseurs pour leur région pour les catégories locales.
- Les P&SC et Supply Chain managers des GeoUnits et de la Division sont responsables de garantir la conformité aux Guides d'Achat.

3. Gestion du projet de sourcing :

Ce processus consiste à gérer les projets de sourcing initiés par le Category Manager en tant que plan d'achat. Selon les dépenses et l'impact des projets, ceux-ci peuvent être traités par les Global Sourcing Centers (GSC) ou par l'équipe de sourcing de la GeoUnit.

- Par la GSC : concerne les commodités considérées comme globales ou à fort impact. Les projets de sourcing sont centralisés et gérés par le GSC, garantissant une optimisation des ressources et une gestion globale des risques et des coûts.
- Par GeoUnits : Pour les commodités locales, les GeoUnits gèrent directement les projets de sourcing. Cela permet une adaptation aux spécificités locales et une meilleure intégration des fournisseurs régionaux.

Le processus comprend les étapes suivantes :

- Identification des opportunités et lancement du projet de sourcing : En démarrant par la planification des besoins d'approvisionnement par les Category Managers, qui élaborent des plans d'achat détaillés pour chaque famille de dépenses. Chaque projet de sourcing inclut des objectifs spécifiques, des critères de qualification des fournisseurs, des estimations de ressources et des dates clés pour assurer une exécution efficace.
- Identification et préqualification des fournisseurs : Cette étape implique l'évaluation des fournisseurs existants à travers les guides d'achat ou les systèmes d'information de SLB. Si nécessaire, de nouveaux fournisseurs sont préqualifiés via des demandes d'information (RFI), en se basant sur des critères tels que l'historique de l'entreprise, sa santé financière, sa capacité opérationnelle et ses certifications.
- Appel d'offres et sélection du fournisseur : Pour sélectionner le meilleur fournisseur, des RFX (RFQ et RFP) sont utilisés pour solliciter des offres détaillées. Une grille d'évaluation compare les propositions selon le prix, la qualité, la conformité, l'expérience du fournisseur et les délais de livraison, permettant ainsi de choisir le fournisseur optimal tout en maximisant la valeur pour chaque dépense. La décision finale est prise en accord avec le plan stratégique de SLB, en utilisant la grille d'évaluation pour évaluer objectivement les performances des fournisseurs.
- Qualification des produits et services : Avant d'entamer toute collaboration opérationnelle, les fournisseurs sélectionnés sont soumis à un audit rigoureux pour évaluer leur capacité technique et leur conformité aux normes HSE, de Management et légales. Ce processus assure que seuls les fournisseurs qualifiés sont intégrés à la Supply Chain de SLB, avec des ré-audits réguliers pour maintenir la conformité continue.

- Intégration du fournisseur et suivi : Une fois qualifiés, les fournisseurs sont intégrés au système d'information de SLB (SAP Ariba) avec une documentation appropriée. Les guides et catalogues d'achat sont mis à jour pour inclure les nouvelles offres des fournisseurs, facilitant ainsi une collaboration efficace et transparente qui répond aux exigences évolutives du marché. Le suivi continu des performances des fournisseurs et des résultats des projets de sourcing est assuré, avec des révisions périodiques des stratégies de sourcing pour optimiser les coûts, réduire les risques et améliorer la qualité des services et des produits.
- Mise à jour des données et communication : Les bases de données fournisseurs et les catalogues de produits sont régulièrement mis à jour. Une communication continue avec les parties prenantes assure une compréhension mutuelle et une coopération efficace tout au long du processus.

4. Gestion des fournisseurs :

La gestion des fournisseurs chez SLB vise à établir des relations efficaces et mutuellement bénéfiques avec les fournisseurs, tout en optimisant les coûts et en assurant la qualité des produits et services livrés. Elle définit un cadre pour engager, gérer la performance, développer et collecter des informations sur les fournisseurs.

SLB cherche à devenir un client de choix en recevant un traitement préférentiel sur la technologie, la capacité et les prix grâce à des partenariats solides avec ses fournisseurs.

Actions clés :

- Établissement de relations solides avec les fournisseurs, en mettant l'accent sur l'innovation, la collaboration et la gestion des performances.
- Suivi continu pour assurer une livraison à temps de produits et services sans défaut.
- Planification et mise en œuvre de plans pour améliorer les capacités et les processus des fournisseurs.
- Maintenance et mise à jour des données de base et des informations d'approvisionnement pour une gestion efficace.

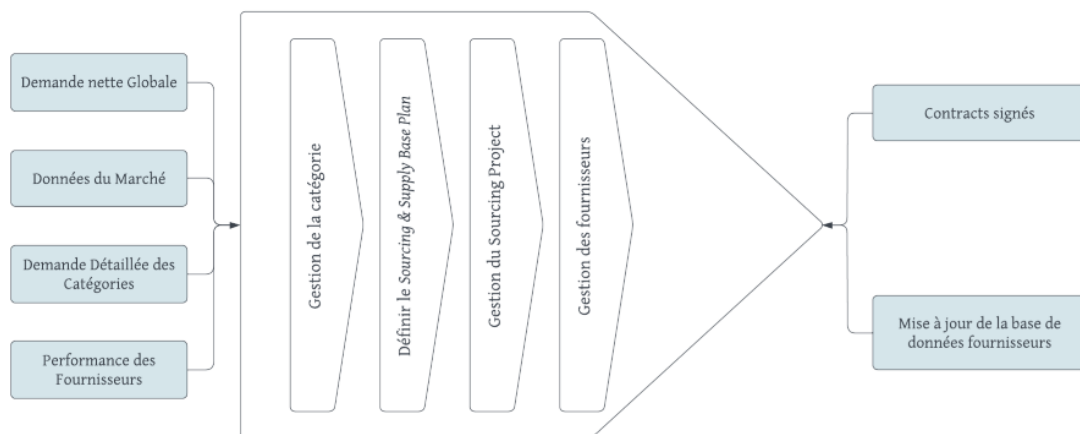


Figure 1.10 – Macro Processus du processus Source to Contract - Document Interne

1.3 Analyse du marché concurrentiel des fournisseurs de SLB

1.3.1 Catégories Fournisseurs ASL

L'objectif initial de l'étude est de comprendre la distribution des fournisseurs approuvés par SLB et d'analyser leur comportement en termes de parts de marché. Il est important de noter qu'il existe principalement cinq catégories de fournisseurs : Field Equipment & Services, Indirect, Chemicals, Manufacturing, et Logistics. Chaque catégorie est subdivisée en plusieurs sous-catégories comme l'illustre la figure suivante :

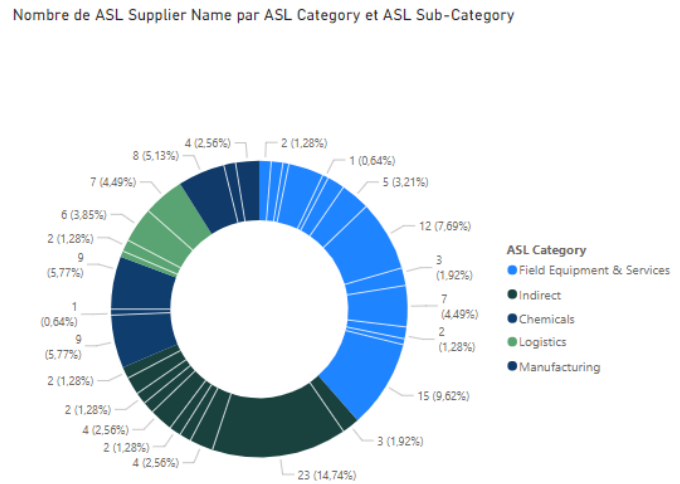


Figure 1.11 – Décomposition des catégories ASL en sous-catégories

Pour les fins de notre étude, nous allons analyser les fournisseurs selon leur sous-catégorie permettant une analyse plus fine et détaillée des dynamiques spécifiques à chaque catégorie.

1.3.2 Analyse des sous catégories de fournisseurs

Il existe 35 sous-catégories ASL, et chaque sous-catégorie comporte un certain nombre de fournisseurs. La figure suivante illustre le nombre de fournisseurs par sous-catégorie, ainsi que leur part respective dans le total des dépenses (spend).

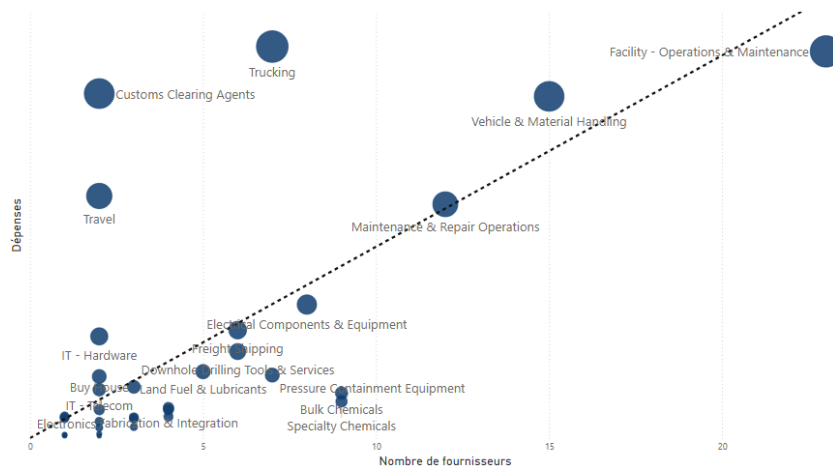


Figure 1.12 – Distribution des sous-catégories ASL par rapport au nombre de fournisseur et au total des dépenses annuelles

Ce graphique révèle des disparités entre les catégories de fournisseurs : certaines présentent un nombre élevé de fournisseurs mais des dépenses totales relativement faibles, tandis que d'autres affichent une situation inverse.

À partir de cette observation, nous constatons que SLB applique différentes stratégies en fonction de la nature de la catégorie et de ses caractéristiques, notamment sur le marché algérien.

Identification des stratégies d'approvisionnement appliquées chez SLB : Les approvisionnements revêtent une importance cruciale dans les économies réalisées sur toute la Supply Chain de SLB, leur gestion efficace étant essentielle pour assurer la rentabilité des projets. Notre diagnostic met en lumière la nécessité de comprendre les différentes approches et stratégies en matière d'approvisionnement et de sélection des fournisseurs. Après analyse des pratiques d'approvisionnement, il est apparu qu'il n'y avait pas de stratégies clairement définies, mais que certaines approches étaient utilisées dans le processus décisionnel, à savoir :

- Stratégie du mieux disant : Simple à mettre en œuvre avec une multitude de fournisseurs sur le marché. Elle consiste à sélectionner le fournisseur offrant le prix le plus bas, tout en tenant compte des frais de transport.
- Stratégie de contrat moyenne durée (3 ans maximum) : Adaptée aux commodités critiques sujettes à des fluctuations de disponibilité ou de prix. Elle implique la signature de contrats à moyen ou long terme pour garantir un prix fixe ou assurer une disponibilité constante des marchandises.
- Stratégie de changement de fournisseur : Utilisée en cas de performance insatisfaisante des fournisseurs actuels, lorsque d'autres options sont facilement disponibles sur le marché pour répondre aux besoins de l'entreprise.
- Stratégie de développement des fournisseurs actuels : Implique d'encourager et d'assister les fournisseurs existants à améliorer leurs performances, notamment lorsque le choix de fournisseurs sur le marché est limité. Cette approche favorise le développement et l'innovation tout en renforçant la stabilité de la Supply Chain.

Il en découle plusieurs types de sélection de fournisseurs :

1. **Sole Source** (Fournisseur unique) :

Un fournisseur unique est sélectionné pour fournir un produit ou un service spécifique, soit en raison de sa présence exclusive sur le marché, soit en raison des exigences de conformité aux standards appliqués.

Avantages : Simplification des processus d'achat, relations renforcées avec le fournisseur, prix potentiellement réduits grâce à des volumes d'achat plus importants.

Inconvénients : Risque accru en cas de défaillance du fournisseur, dépendance élevée.

2. **Main and Backup Supplier** (Fournisseur Principal et de Secours) :

Un fournisseur principal est désigné ayant la plus grande part de marché, avec un fournisseur de secours détenant une petite part de marché en cas de besoin.

Avantages : Continuité de l'approvisionnement en cas de problème avec le fournisseur principal, réduction du risque de rupture de stock dans les cas des fournisseurs de produits.

Inconvénients : Complexité accrue de la gestion des fournisseurs, risque d'augmentation des coûts.

3. Multiple Source Suppliers (Fournisseur à Sources Multiples) :

Plusieurs fournisseurs sont sélectionnés pour un même produit ou service.

Avantages : Diversification des risques, meilleure négociation des prix et flexibilité accrue.

Inconvénients : Complexité de gestion accrue et risque de qualité inégale entre les fournisseurs.

Après la sélection des fournisseurs et la signature éventuelle de contrats avec ces derniers, un "Contract Handover" ou manuel des fournisseurs est remis aux opérations de SLB. Ce document détaille les parts de marché prévues pour chaque fournisseur pour une commodité donnée. Ces "Contracts Handover" concrétisent les stratégies de sélection des fournisseurs, permettant de réaliser des optimisations des coûts et d'améliorer la performance de la Supply Chain. Le non-respect de ces **parts de marché** peut entraîner des manques à gagner et des pertes.

1.3.3 Evolution des parts de marché des fournisseurs

Dans l'objectif d'analyser les fournisseurs au sein de leurs sous-catégories respectives, nous avons décidé de les évaluer en calculant la part de marché de chacun, ce qui nous permettra de réaliser une comparaison inter-catégorie.

- Comment le mesurer : Pour évaluer la part de marché des fournisseurs dans leurs catégories respectives, nous avons créé un indicateur appelé "Market Share". Cet indicateur suit la part de marché des fournisseurs dans leurs sous-catégories respectives par mois, calculé comme le rapport entre les dépenses réalisées avec un fournisseur et les dépenses totales de la sous-catégorie, ce qui permet d'éliminer l'effet de la demande sur la sous-catégorie et de donner purement les parts de marchés des fournisseurs entre eux.

```
1 MarketShare =
2 DIVIDE(
3   CALCULATE(
4     SUM('Export'[Spend (USD)]),
5     ALLEXCEPT('Export', 'Export'[ASL Sub-Category], 'Export'[Hrchy_Year_Date - YearMonth].[Mois])
6   ),
7   CALCULATE(
8     SUM('Export'[Spend (USD)]),
9     ALLEXCEPT('Export', 'Export'[ASL Supplier Number], 'Export'[Hrchy_Year_Date - YearMonth].[Mois])
10  )
11 )
```

Figure 1.13 – Formule de calcul de l'indicateur - Editeur DAX

$$\text{Market Share} = \frac{\sum \text{dépenses fournisseur}}{\sum \text{dépenses dans la sous catégorie}}$$

La figure précédente illustre la formule appliquée pour le calcul de l'indicateur "Market Share".

Analyse des parts de marché et identification des fournisseurs par sous-catégorie : Après avoir calculé la part de marché relative à chaque fournisseur et discuté avec le Suppliers Manager (responsable de la gestion des fournisseurs), nous avons décidé de sélectionner un ensemble de sous-catégories pour poursuivre notre analyse. L'objectif est d'observer les changements de préférence des fournisseurs et d'affiner notre analyse vis-à-vis des sous-catégories les plus importantes.

Nous nous sommes basés pour cette sélection sur les deux critères suivants :

- Somme des Dépenses > 200k Dollars
- Les sous-catégories ayant plus de 2 fournisseurs

ASL Sub-Category	Nombre de ASL Supplier Name	%TG Somme de Spend (USD)
Bulk Chemicals	9	13,14%
Downhole Drilling Tools & Services	6	4,28%
Freight Shipping	6	2,16%
Institutional Services	4	3,78%
Land Fuel & Lubricants	5	8,88%
Maintenance & Repair Operations	12	2,66%
Oilfield Service Companies & Alliances	3	5,84%
Pressure Containment Equipment	7	6,48%
Rig Management Services and OCTG	2	11,91%
Specialty Chemicals	9	8,14%
Trucking	7	18,91%
Vehicle & Material Handling	15	13,83%

Figure 1.14 – Liste des sous-catégories après sélection

A l'issue de cette première phase, nous avons abouti à un ensemble de douze catégories totalisant un nombre de 85 fournisseurs comme l'illustre la figure précédente.

Et pour mieux voir le changement des parts de marché des fournisseurs, nous avons élaboré un indicateur de changement de part de marché des fournisseurs, ou nous avons calculé la différence entre ceux du début d'année 2023, et ceux de la fin d'année pour voir l'évolution ou la détérioration de ces parts de marché. Nous pouvons apercevoir dans la figure suivante l'évolution des fournisseurs par sous-catégorie.



Figure 1.15 – Changements des parts de marché des fournisseurs par sous-catégorie

A travers la figure précédente, nous pouvons constater la présence d'une variabilité importante et une non stabilité des parts de marché des fournisseurs entre le début et la fin d'année étudiée.

Nous poursuivons donc notre analyse en comparant en détail les sous-catégories afin de comprendre leurs différences, lesquelles peuvent être expliquées par divers facteurs. Quantitativement, en tenant compte des variations des fournisseurs dans chaque sous-catégorie, nous avons évalué en utilisant les deux critères suivants :

- **La variance de la part de marché des fournisseurs**, dans la sous-catégorie afin de comprendre la non stabilité des parts de marché des fournisseurs présents.
- **La somme en valeur absolue des différences de parts de marché entre le début et la fin de l'année** de chaque fournisseur, pour évaluer les taux de changement dans la sous-catégorie au fil du temps.

Ces critères nous permettront d'affiner notre analyse des variations des parts de marché des fournisseurs et ainsi de mieux comprendre leur dynamique. Les figures suivantes permettent d'illustrer ces variations par rapport aux sous-catégories.

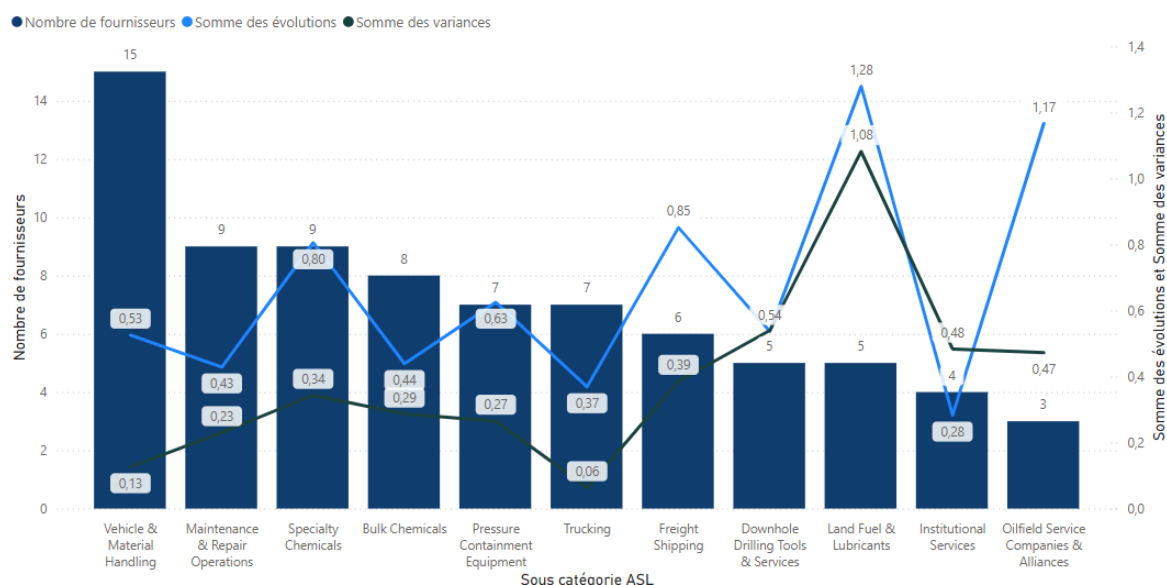


Figure 1.16 – Illustration des sommes de variance et évolution des deux critères par rapport aux sous-catégories

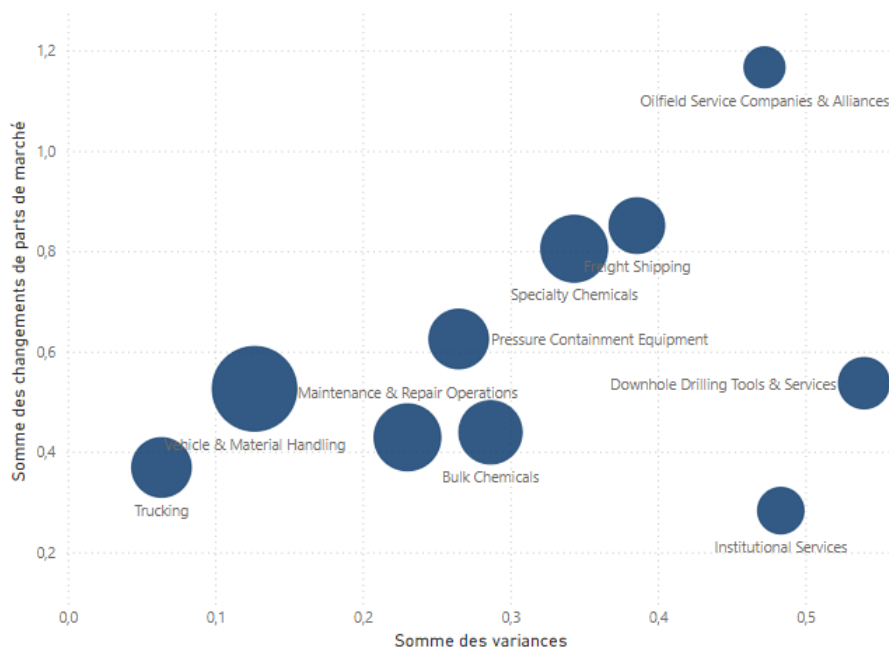


Figure 1.17 – Variabilité relative des sous-catégories entre elles selon les deux critères

L'analyse de la variabilité des parts de marché des fournisseurs des catégories présélectionnées nous permet de constater qu'un haut niveau de variabilité est présent dans la majorité des sous-catégories analysées, indépendamment du nombre de fournisseurs impliqués.

Ces analyses nous ont permis, dans un premier temps, de mettre en évidence la part de marché des fournisseurs de SLB à travers l'indicateur Market Share proposé. Dans un second temps, elles ont illustré la présence d'une variabilité et d'une instabilité de cet indicateur au fil du temps.

En discussion avec les parties prenantes du projet et les responsables d'approvisionnement, il nous a été suggéré d'investiguer les fluctuations des parts de marché spécifiquement pour la catégorie de produits M&S de la fonction TLM (ou Technology Lifecycle Management, chargé des activités de maintenance). Cette suggestion découle de la constatation de variabilités et d'instabilités dans certains contrats avec les fournisseurs de cette sous-catégorie. En particulier, le contrat avec le représentant des pièces de rechange Renault en Algérie, qui représente un pourcentage significatif des dépenses de TLM, a attiré notre attention. L'objectif est d'analyser en profondeur les causes de cette instabilité, d'identifier les facteurs qui influencent ces variations, et de proposer une solution qui vise à améliorer ce contrat.

1.3.4 Causes de variation des parts de marché

Il existe plusieurs raisons pour la variation des parts de marché de ces fournisseurs, parmi ces raisons que nous avons constaté lors de nos séries de discussions avec les équipes de SLB, nous pouvons citer :

- Lead time (délai de livraison) : Représente les délais de livraison des produits des fournisseurs vers les sites de SLB, une forte variabilité de ce dernier impacte fortement la sélection du fournisseur lors de l'achat et donc induit une perturbation des parts de marché.
- Cash-flow (Flux de trésorerie) : La santé financière d'un fournisseur, notamment sa capacité à maintenir un flux de trésorerie stable, peut avoir un impact sur sa compétitivité. Les fournisseurs ayant des difficultés financières peuvent être moins fiables et perdre des parts de marché.
- Disponibilité : La disponibilité constante des produits ou des services est importante pour maintenir la satisfaction des clients dans ce cas SLB et lui permettre d'assurer une continuité de son business. Les fournisseurs capables de maintenir des niveaux de stock adéquats peuvent attirer et conserver plus de clients.
- Demande sur le marché : Les variations de la demande sur le marché peuvent influencer la demande des clients pour les produits ou services d'un fournisseur spécifique. Les fournisseurs capables de s'adapter rapidement aux fluctuations de la demande peuvent mieux préserver leur part de marché.
- Préférences subjectives des opérations : Les préférences, basées sur l'expérience et l'expertise des responsables des opérations, peuvent aussi influencer les décisions. Par exemple, des relations de longue date avec un fournisseur ou des expériences passées positives peuvent influencer la décision de rester fidèle à un fournisseur spécifique.

1.4 Audit logistique et présentation de la problématique

1.4.1 Audit logistique

Afin d'analyser et de mieux comprendre l'état de fonctionnement actuel de la fonction TLM de la division Reservoir Performance en ce qui concerne la catégorie de produits M&S, il était nécessaire d'effectuer un audit logistique afin de mesurer et d'identifier les raisons de l'instabilité des parts de marchés des fournisseurs qui est directement liée à l'activité de la TLM, des approvisionnements et des achats. Cette étape aura pour objectif de mettre en lumière les différents dysfonctionnements relatifs à cette catégorie de produits.

Pour être efficace et efficient, un audit doit se baser sur un référentiel, indispensable pour formuler un jugement précis.

Choix du référentiel d'audit

Selon la définition donnée par la norme ISO 9000 - 2015, le référentiel d'audit ou bien aussi les critères d'audit représentent l'ensemble des politiques, procédures et exigences.

Un référentiel constitue également une base de comparaison essentielle à la formulation du jugement, et donc ce système d'évaluation est indispensable pour la réalisation d'un bon audit logistique (Voir Annexe C).

Pour notre étude, nous avons choisi le modèle SCOR comme référentiel. Ce choix est justifié par la capacité de SCOR à offrir une approche par processus, qui fournit une vision générale et intégrée du système logistique, permettant ainsi une analyse approfondie et une amélioration significative de la supply chain.

Pour approfondir notre sélection, nous avons comparé les différents référentiels d'audit selon des critères objectifs en corrélation avec le milieu d'étude. Les résultats de cette comparaison sont en annexe C.

Déroulement de l'audit

Cette section explicite la démarche suivie pour l'audit au sein de SLB et en précise le périmètre. Suite à la consultation du corps managérial et une première évaluation des potentiels d'amélioration des processus, nous avons décidé de restreindre l'audit aux processus amonts de la Supply Chain pour la catégorie de produits M&S de la division RP, en raison de l'importance majeure de la phase d'approvisionnement pour cette catégorie. Cette phase englobe la gestion des stocks, la sélection des fournisseurs, la planification des besoins et l'affectation des commandes.

Les processus amonts constituent l'une des principales sources de dépenses pour SLB, tout en apportant une valeur ajoutée considérable à ses activités principales, une entreprise de services dont la qualité repose en partie sur les produits et équipements utilisés.

Par conséquent, les processus de planification et d'approvisionnement seront les principaux objets de notre audit. La schématisation des processus SCOR (cartographie de niveau 1) est illustrée dans la figure C.1 en annexe.

Pour commencer, nous allons décomposer chaque processus selon les trois niveaux définis par le référentiel SCOR V14.0. Ensuite, nous évaluerons les processus en utilisant la grille SCOR, en prenant en compte deux paramètres : les indicateurs de performance et les pratiques utilisées. Les grilles complétées, ainsi qu'une description des indicateurs, des bonnes pratiques et des compétences SCOR, sont présentées en annexe. Pour réaliser cette évaluation, nous avons utilisé deux sources d'information :

- La documentation interne de l'entreprise : tous les processus y sont documentés et expliqués.
- Les interviews des acteurs de la Supply Chain : ces entretiens ont permis de recueillir des informations non documentées.

Les parties prenantes interviewées comprennent des Sourcing Specialists, des Supplier Managers, des M&S Coordinators, des TLM Planners et Managers, des Maintenance Engineers, et des Domestic Logistics Planners.

1) Processus de planification :

Le processus de planification est crucial et son optimisation peut apporter un avantage concurrentiel significatif à l'entreprise. Dans le cadre de la décomposition au niveau 2, nous avons identifié trois sous-processus pertinents pour notre étude, qui serviront de référentiel d'audit pour les processus de SLB : la planification de la Supply Chain, la planification de l'approvisionnement, et la planification des commandes.

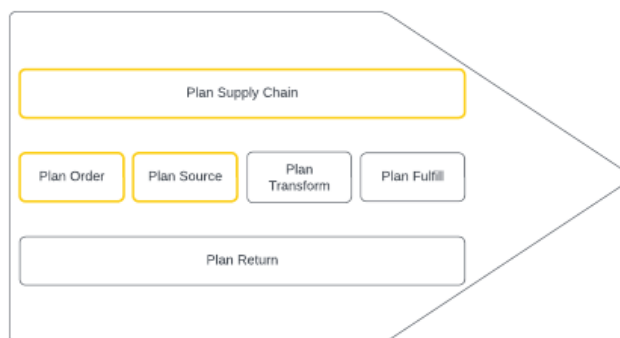


Figure 1.18 – Processus de planification : Cartographie de niveau 2

Nous abordons maintenant le niveau 3 du référentiel, qui traite des sous-processus des processus sélectionnés précédemment. Le tableau C.1 en annexe C présente la décomposition en niveau 3 du processus de planification.

Dysfonctionnements constatés :

1. Manque de prévision et de planification :

- Absence de planification ou de prévisions pour les commandes et les besoins de maintenance.
- Faible fiabilité des prévisions réalisées à travers l'outil dédié IBP entraînant des réajustements majeurs.
- Exécution des PO uniquement selon les demandes opérationnelles immédiates, sans considération des prévisions à long terme.

2. Dépendance aux fournisseurs internationaux :

- Les longs délais de livraison (lead time) pour les produits M&S provenant des hubs internationaux complique la planification des besoins et des réapprovisionnements.

3. Coordination et communication insuffisantes :

- Les besoins globaux des opérations sont discutés pour établir un budget, mais cette approche réactive ne permet pas une planification proactive et structurée.
- Manque de communication entre les départements achats et opérations, entraînant une absence de coordination entre les besoins en termes de missions et de maintenance.

4. Défaut d'engagement contractuel avec les fournisseurs :

- En raison de la non fiabilité des prévisions réalisées, il est difficile de suivre les engagements en volume préétablis avec les fournisseurs.

- La variabilité des parts de marché des fournisseurs, causée par les choix des opérations, empêche de suivre certains engagements pris avec eux.

2) Processus d'approvisionnement :

L'approvisionnement dépasse la simple nécessité pour devenir un véritable atout lorsqu'il est optimisé pour réduire les coûts, éviter les retards et générer des économies. Après des discussions approfondies et des recherches sur ce processus au sein de la division, il a été conclu que les sous-processus les plus pertinents pour notre étude, en lien avec l'environnement de l'entreprise, dans la décomposition au niveau 2 du référentiel SCOR, sont les sous-processus d'approvisionnement stratégique, direct et indirect.

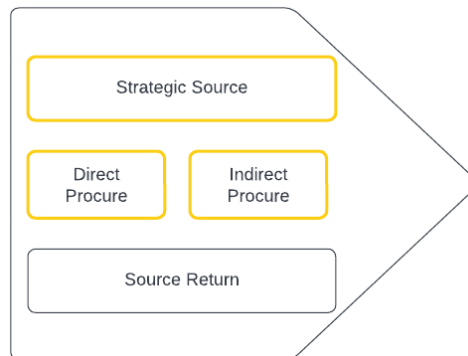


Figure 1.19 – Processus d'approvisionnement : Cartographie de niveau 2

De même, nous passons ensuite au niveau 3 qui décompose les sous-processus sélectionnés précédemment. Le tableau C.2 en annexe C décrit les sous-processus audités par rapport au processus d'approvisionnement.

Dysfonctionnements constatés :

1. Problèmes de réapprovisionnement et gestion des stocks :

- Les produits inclus dans l'outil M&S Prime (un outil permettant l'automatisation des commandes chez SLB) bénéficient d'un réapprovisionnement automatique basé sur l'historique de consommation, alors que les autres produits nécessitent des PO (Purchase Orders) créés selon les besoins après des réunions d'équipe, ce qui peut introduire des délais d'approvisionnement.
- La stratification des pièces n'est pas mise à jour selon l'historique de consommation induisant des approvisionnements en pièces non nécessaires.
- Les stocks de sécurité sont maintenus uniquement pour les produits inclus dans M&S Prime augmentant le risque de rupture pour les autres produits.
- La majorité des M&S proviennent de fournisseurs internationaux, entraînant des lead times de 3 à 6 mois, ce qui complique la gestion des stocks et la satisfaction des besoins immédiats.
- Contraintes liées aux budgets par rapport aux besoins réels.
- Certaines pièces se trouvent exclues du M&S Prime après 6 mois de non-validation des PO (Purchase Order) ou de non consommation, empêchant l'auto-réapprovisionnement.
- Des PO lancés mais non réceptionnés à temps, obligeant les opérateurs à trouver des solutions alternatives et entraînant des stocks inutilisés et ainsi une possibilité d'exclusion de l'outil M&S Prime.

2. Problèmes de gestion des fournisseurs :

- Absence de prévisions lors de l'édification des contrats avec les fournisseurs.
- Suivi de performance des fournisseurs très insuffisant, avec près d'une année d'absence de monitoring.

3. Délais de réception des commandes :

- Les retards de livraison dus aux procédures douanières et au traitement interne des commandes.
- Les pièces critiques pour la maintenance corrective ayant un lead time de 70 jours, compliquant la planification et l'exécution des maintenances.

4. Budget et disponibilité des pièces :

- La majorité du budget est consommée par des maintenances correctives imprévisibles, limitant la possibilité de planifier des maintenances préventives adéquates.
- Beaucoup de véhicules lourds nécessitant des réparations globales ne reçoivent que des maintenances superficielles, en raison de la disponibilité limitée des pièces et la mauvaise allocation des ressources.

Synthèse des résultats du diagnostic

A l'issue de notre diagnostic réalisé et l'audit effectué sur la gestion de la catégorie de produit M&S, nous avons pu constater quelques dysfonctionnements résumés comme suit :

- Les prévisions et la planification sont globalement imprécises et insuffisantes, entraînant un déséquilibre entre les engagements signés avec les fournisseurs et la réalité des besoins.
- Les stratégies de réapprovisionnement et de gestion des stocks sont inefficaces, avec une mauvaise stratification des produits induisant souvent des commandes non nécessaires.
- Le suivi de la performance et l'évaluation des fournisseurs sont insuffisants, avec une absence de retour d'information des opérations, entraînant une forte variabilité des parts de marché pour une commodité donnée.
- La coordination insuffisante entre les différents départements impacte l'allocation efficace des ressources financières aux besoins urgents des missions et aux activités de la TLM.

1.4.2 Présentation de la problématique

L'approvisionnement est d'une importance capitale pour toute organisation et joue un rôle vital dans sa Supply Chain. Une gestion efficace de cette fonction permet à l'entreprise d'assurer un flux continu de ses activités commerciales.

SLB vise à optimiser cette fonction pour réduire les coûts, répondre efficacement aux besoins en matières premières, et améliorer la rentabilité globale de l'entreprise tout en réduisant la fluctuation des parts de marché de ses fournisseurs.

L'analyse approfondie des dysfonctionnements révélés par le diagnostic et les différentes analyses, combiné aux consultations avec le corps managérial de l'entreprise, nous a permis d'identifier les besoins de la division Reservoir Performance, particulièrement en ce qui concerne l'activité de Technology Lifecycle Management (TLM).

Cette démarche nous a conduits à mettre en lumière une problématique majeure liée à l'instabilité des parts de marché des fournisseurs des pièces de rechange (M&S) et à la gestion et à la planification de ces pièces de manière générale, caractérisée par une satisfaction insuffisante de la demande. Cela soulève la question suivante :

Comment améliorer la gestion et la planification des pièces de rechanges M&S et contribuer à l'amélioration des contrats fournisseurs à travers la proposition d'un engagement en volume ?

Pour pouvoir mieux répondre à cette problématique, nous la décomposons en multiple sous question :

- Quelles seront les nouvelles stratifications des pièces M&S ? Quel modèle attribue le mieux les stratifications des pièces ?
- Quel sera le nouveau stock de sécurité appliqué ?
- Comment pouvons-nous proposer un engagement en volume ? Quel est le moyen idéal pour répondre à cela ?

C'est dans ce cadre que s'inscrit notre travail. L'objectif est d'améliorer la Supply Chain amont de SLB en se concentrant sur l'approvisionnement et la gestion des stocks des pièces de rechange M&S, en visant à réduire les commandes inutiles d'une part et à améliorer l'établissement des contrats avec les fournisseurs d'autre part.

Pour répondre à cette problématique, la première partie consistera à catégoriser les pièces de rechange M&S en fonction de leur fréquence d'utilisation, en mettant un accent particulier sur les six derniers mois visant l'automatisation des commandes. La deuxième partie portera sur une nouvelle planification visant à renforcer les contrats existants et à obtenir des engagements de volume avec les fournisseurs afin de réaliser d'éventuelles économies.

Pour conclure, dans ce chapitre, nous nous sommes familiarisés avec la structure de l'entreprise d'accueil, ainsi qu'avec son domaine d'activité et le marché dans lequel elle opère. Nous avons également examiné en détail sa structure organisationnelle et celle de sa Supply Chain.

Nous avons également effectué une analyse détaillée des dépenses par fournisseurs et leurs catégories, permettant ainsi de mieux cibler notre sujet d'étude. En parallèle, nous avons utilisé le référentiel SCOR pour diagnostiquer plusieurs dysfonctionnements. À partir de ces analyses, nous avons défini la problématique portant sur l'amélioration de la gestion des pièces de rechange M&S et des contrats fournisseurs.

Les deux prochains chapitres exploreront l'état de l'art et la littérature existante afin d'introduire les concepts et théories pertinents pour résoudre ces défis.

Chapitre 2 – Supply chain,
Relations fournisseurs et défis des
engagements de volume

2 Supply chain, Relations fournisseurs et défis des engagements de volume

Ce chapitre vise à offrir une vue d'ensemble approfondie des concepts clés liés à notre étude. Nous commencerons par explorer les concepts de la supply chain et de sa gestion, de l'approvisionnement et de la gestion des relations avec les fournisseurs, ainsi que nos objectifs de stabilisation des commandes et des prévisions, qui constitueront le cœur de notre travail.

2.1 Présentation des concepts liés à la gestion de la supply Chain

Dans cette première partie, nous aborderons les concepts essentiels pour comprendre notre problématique. Nous explorerons d'abord ceux de la supply chain, en particulier dans le secteur du pétrole et du gaz. Ensuite, nous nous concentrerons sur la fonction d'approvisionnement, qui constitue le pilier central de notre étude. Nous examinerons également les notions de gestion des relations entreprise-fournisseur et de gestion des contrats, afin de détailler les aspects clés de notre travail visant à stabiliser ces relations fournisseurs.

2.1.1 Définition de la Supply Chain

La supply chain est un terme anglo-saxon apparu dans les années 1980 et devenu d'usage courant dans les années 1990 (Hugos, 2018). La Supply chain (chaîne d'approvisionnement) est le processus qui englobe les entreprises et les activités commerciales nécessaires pour concevoir, fabriquer, livrer et utiliser un produit ou un service. Les entreprises dépendent de leurs Supply Chain pour obtenir ce dont elles ont besoin. Chacune s'intègre dans une ou plusieurs chaînes et joue un rôle dans chacune d'elles.

D'après (Lambert et al., 1998), une supply chain est l'alignement des entreprises qui amènent des produits ou des services sur le marché. D'un autre côté, une supply chain comprend toutes les étapes impliquées, directement ou indirectement, dans la satisfaction d'une demande client. Elle inclut non seulement le fabricant et les fournisseurs, mais aussi les transporteurs, les entrepôts, les détaillants et les clients eux-mêmes (Chopra & Meindl, 2016). Ganeshan et Harrison la définissent comme étant un réseau d'installations et d'options de distribution qui effectue les fonctions d'approvisionnement, de transformation en produits finis, et de distribution de ces produits finis aux clients (Ganeshan & Harrison, 1995).

Donc, même s'il y a une légère divergence parmi les différentes définitions de la supply chain, elles convergent toutes vers l'idée qu'elle est un système de fournisseurs, producteurs, distributeurs et détaillants ainsi que de clients, étant tous en relation entre eux par différents flux : physiques, informationnels et financiers.

Les Supply Chains peuvent être classées en trois niveaux de complexité : directe, étendue et ultime comme l'illustre la figure suivante.

Chaque niveau de complexité montre comment les relations et les interactions s'étendent au-delà des liens directs pour inclure des partenaires supplémentaires, illustrant la nature interconnectée et en réseau de la supply chain.

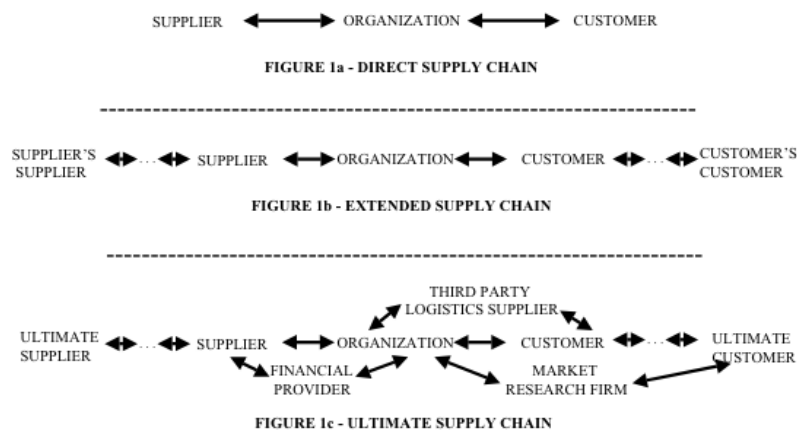


Figure 2.1 – Niveaux de complexité de la supply chain (Mentzer et al., 2001, p.5)

2.1.2 Le Supply chain management

La gestion de la chaîne d'approvisionnement (Supply Chain Management) se concentre principalement sur l'intégration efficace des fournisseurs, usines, centres de distribution, entrepôts et magasins afin que les marchandises soient produites et distribuées en quantités adéquates, aux bons endroits et au bon moment, tout en minimisant le coût total du système et en répondant à certaines exigences de service

Bien que ses définitions diffèrent selon les auteurs, elles peuvent être classées en trois catégories : une philosophie de gestion, la mise en œuvre d'une philosophie de gestion et un ensemble de processus de gestion (Mentzer et al., 2001). Ces catégories montrent que la SCM peut prêter à confusion. Selon (Cooper et al., 1997), c'est une philosophie intégrative pour gérer le flux total d'un canal de distribution, du fournisseur à l'utilisateur final. Monczka, Trent, et Handfield soulignent que la SCM nécessite que les fonctions de matériaux traditionnellement séparées relèvent d'un cadre responsable de la coordination de l'ensemble du processus des matériaux, et nécessite également des relations conjointes avec les fournisseurs sur plusieurs niveaux (R. M. Monczka et al., 1998).

Tous les processus de la SCM peuvent être classés en trois macro-processus principaux, qui se répartissent dans les trois parties de la supply chain : amont, interne et aval (Chopra & Meindl, 2016).

- Dans la supply chain aval, le Customer Relationship Management (CRM) inclut tous les processus entre l'entreprise et ses clients. Son objectif principal est de stimuler la demande client et de faciliter le placement ainsi que le suivi des commandes, y compris le marketing, la tarification, les ventes, la gestion des commandes et les centres d'appels.
- L'Internal Supply Chain Management (ISCM) gère les processus internes de l'entreprise pour répondre efficacement à la demande générée par le CRM. Cela comprend la planification de la production, la gestion des capacités de stockage, la préparation des plans de demande et d'approvisionnement, ainsi que l'exécution des commandes.
- En amont, le Supplier Relationship Management (SRM) permet de gérer les processus entre l'entreprise et ses fournisseurs afin d'organiser et d'optimiser les sources d'approvisionnement. Cela inclut l'évaluation des fournisseurs et la négociation des termes d'approvisionnement.

Ces trois processus garantissent la gestion fluide des informations, des produits et des fonds nécessaires pour répondre à la demande client et assurer le succès de la Supply Chain, en mettant l'accent sur leur intégration pour une efficacité maximale.

2.1.3 La Supply Chain dans le secteur du oil & gas

L'industrie pétrolière est une partie critique du secteur énergétique mondial, car ses processus ont des effets significatifs sur tous les secteurs productifs de l'économie et sur la population.

La supply chain de l'industrie pétrolière englobe les activités liées à l'exploration, l'extraction, le raffinage, le transport et la distribution des produits pétroliers raffinés à travers le monde, depuis l'amont jusqu'à l'aval.

Le Supply Chain Management dans l'industrie pétrolière est la configuration, la coordination et l'amélioration continue des opérations organisées de manière séquentielle impliquées dans l'amont, le milieu et l'aval (Chima, 2007). Ainsi, la supply chain pétrolière se compose de trois segments fonctionnels, à savoir l'amont, le milieu et l'aval (figure suivante).

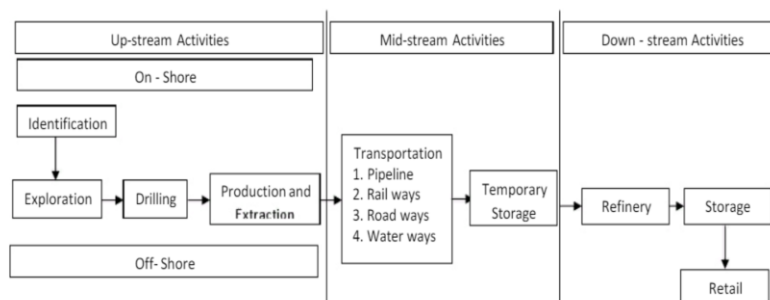


Figure 2.2 – Composition de la supply chain pétrolière (Lisitsa et al., 2019, p.6)

La gestion de la Supply Chain (SCM) est essentielle dans l'industrie pétrolière en raison de la demande croissante pour le pétrole et ses dérivés à l'échelle mondiale. Cette industrie, caractérisée par sa complexité et son inflexibilité logistique, doit améliorer la gestion pour réduire les coûts et accroître l'efficacité. Coordonner efficacement l'approvisionnement, le stockage et la distribution des produits pétroliers à travers le globe tout en maintenant des niveaux élevés de service représente un défi majeur dont la SCM joue un rôle crucial à le surélever. Les entreprises de services pétroliers, ou parapétrolières, se positionnent dans la partie "amont" de la Supply Chain, qui inclut l'extraction et la production de matières premières, en contraste avec les activités "aval" qui sont plus proches de l'utilisateur final.

2.1.4 La fonction approvisionnement et la gestion des relations fournisseurs

Dans le secteur du pétrole et du gaz, une gestion efficace des approvisionnements est essentielle pour assurer la continuité des opérations et maximiser la rentabilité. La complexité de l'industrie pétrolière nécessite une supply chain robuste et bien gérée, en particulier en ce qui concerne les approvisionnements en raison de leur criticité pour les opérations. Une bonne gestion de ces derniers permet de réduire les coûts, notamment à travers des négociations efficaces, d'assurer la continuité des opérations en garantissant la disponibilité des équipements, et de renforcer les relations avec les fournisseurs. L'optimisation de la gestion des approvisionnements est donc un levier pour assurer la résilience et l'agilité de l'ensemble de la Supply Chain.

Présentation de la fonction approvisionnement

- L'approvisionnement stratégique, également connu sous le nom de sourcing, est le processus de recherche, d'évaluation et de sélection de fournisseurs pour obtenir des matériaux, des produits ou des services. Le sourcing implique une analyse approfondie des marchés, des tendances et des fournisseurs potentiels. L'objectif principal est de trouver les meilleures sources offrant une combinaison optimale de qualité, de coût, de capacité et de fiabilité. Le sourcing est l'ensemble des processus commerciaux nécessaires pour acheter des biens et des services. Les managers doivent

d'abord décider si chaque tâche sera exécutée par une source réactive ou efficiente, puis si la source sera interne à l'entreprise ou tierce (Chopra & Meindl, 2016).

- Les achats, également connus sous le terme de procurement, englobent l'ensemble des activités nécessaires pour acquérir les biens et services dont une organisation a besoin pour ses opérations. Les achats couvrent l'acquisition, la réception et le paiement des biens et services. Ce sont l'achat de matériaux et de services auprès d'organisations externes pour soutenir les opérations de l'entreprise, de la production à la commercialisation, aux ventes et à la logistique (Lambert et al., 1998).

Différenciation entre achats et approvisionnement :

L'approvisionnement (sourcing) et les achats, bien que souvent confondus, ont des rôles distincts dans la supply chain. Le sourcing se concentre sur la sélection stratégique des fournisseurs et l'établissement de relations à long terme, cherchant à optimiser les coûts et la qualité sur le long terme. En revanche, les achats se concentrent sur le processus transactionnel d'acquisition de biens et services, avec une perspective à court terme. Ainsi, le sourcing se positionne en amont des achats et joue un rôle clé dans la définition des stratégies d'approvisionnement.

Traditionnellement, les principales activités d'un responsable des achats consistaient à faire pression sur les fournisseurs potentiels pour obtenir des prix plus bas (Hugos, 2018). Cette activité reste importante, mais d'autres activités deviennent également tout aussi importantes. En raison de cela, l'activité d'achat est désormais considérée comme faisant partie d'une fonction plus large appelée approvisionnement (sourcing), qui se fractionne en cinq principales catégories d'activités : achats, gestion de la consommation, sélection des fournisseurs, négociation de contrats et gestion des contrats.

Selon (R. Monczka et al., 2008), le Supplier Relationship Management (SRM) devient essentiel dans un environnement commercial de plus en plus complexe et mondialisé. L'intégration, qui consiste à unir différents groupes, fonctions ou organisations pour collaborer efficacement vers un objectif commun, joue un rôle crucial dans l'optimisation des opérations. La gestion des relations est donc essentielle pour établir des liens solides avec les fonctions internes et les fournisseurs externes. Ainsi, le SRM est indispensable pour harmoniser les relations avec les fournisseurs et répondre aux exigences de l'entreprise.

2.2 Relations entre fournisseurs et entreprise

La gestion des fournisseurs est essentielle dans toute entreprise. Cependant, la relation fournisseur représente également un atout majeur pour une entreprise, nécessitant une relation de qualité constante (Bothin, 2024). Le fournisseur est l'acteur qui assure la disponibilité des matières premières et des intrants pour les différentes activités de l'entreprise. Ainsi, il est impératif de mettre en place un système performant pour assurer une gestion optimale de cette relation.

2.2.1 Gestion de la relation fournisseurs

La relation fournisseurs fait référence au lien de collaboration entre un fournisseur et son client, qu'il soit une personne physique ou morale. Ce lien repose sur la confiance mutuelle entre les parties. Le client doit avoir confiance dans la capacité du fournisseur à livrer les commandes selon les exigences. De son côté, le fournisseur doit pouvoir compter sur le fait que son client honorera les paiements selon les termes du contrat.

Le SRM (Supplier Relationship Management) est le processus d'engagement dans des activités visant à établir, développer, stabiliser et dissoudre les relations avec les fournisseurs internes, ainsi que l'observation des fournisseurs externes pour créer et améliorer la valeur au sein de ces relations (Benoit (née Moeller) et al., 2006).

Il existe deux types de relations fournisseurs. La relation verticale implique que le client travaille avec des fournisseurs spécifiques pour satisfaire ses besoins, ce qui permet de réduire les coûts, sécuriser l'approvisionnement, diminuer les charges de l'entreprise et améliorer ses performances. La relation horizontale, où plusieurs fournisseurs collaborent pour satisfaire les besoins du client sous un seul contrat, permet à l'entreprise de gagner du temps.

Selon un récent rapport de PriceWaterhouseCoopers, les principaux objectifs du SRM (Supplier Relationship Management) sont la création de valeur conjointe basée sur la confiance et la collaboration, la réduction des coûts qui demeure une priorité, et l'assurance de la sécurité des approvisionnements malgré les fluctuations du marché. D'autres objectifs incluent l'augmentation de l'innovation, la réduction du fonds de roulement, la prévention des dommages à la réputation, la gestion de la Supply Chain en boucle fermée, la réduction du délai de mise sur le marché, ainsi que l'amélioration du service des fournisseurs et de la qualité des produits (PWC, 2013).

Pour formaliser et sécuriser les relations avec les fournisseurs ainsi que stabiliser leurs parts de marché, les entreprises utilisent fréquemment des contrats pour plusieurs raisons essentielles :

- Établir un cadre formel de collaboration définissant clairement les obligations, les modalités et les clauses spécifiques.
- Réduire les risques en clarifiant dès le départ les responsabilités et les attentes mutuelles.
- Optimiser les performances en identifiant des indicateurs clés à respecter tout au long de la collaboration.
- Intégrer une flexibilité stratégique, notamment pour la révision des prix, afin de s'adapter aux conditions changeantes du marché.

2.2.2 Les contrats avec les fournisseurs

Étant donné l'évolution rapide et dynamique des contrats au cours des dernières décennies, il n'est pas surprenant que différents experts et organisations de recherche aient des points de vue et des définitions variés des contrats modernes. Selon la recherche de Gartner, un contrat est la somme de toutes les transactions et interactions qui ont eu lieu entre les parties, tant avant qu'après l'attribution du contrat. Une définition assez globale et proposée qui stipule que c'est un ensemble de documents, régis et limités par la loi, qui établissent clairement les limites, l'étendue et l'intention de la relation entre les parties contractantes, ainsi que les droits et responsabilités des entités impliquées (Saxena, 2008).

Les contrats peuvent être classés en différentes catégories en fonction de leurs caractéristiques et de leurs objectifs. Cette classification est influencée par plusieurs facteurs, notamment le type de projet, son horizon, le pricing, ainsi que la planification du projet (Saxena, 2008). D'autres critères peuvent également être pris en compte, tels que la flexibilité et le niveau de risque toléré.

Presque tous les contrats d'approvisionnement sont basés sur une forme quelconque de mécanisme de tarification et peuvent être catégorisés comme une variation de deux types de base : les contrats à prix fixe et les contrats basés sur les coûts (R. Monczka et al., 2008).

- 1) Contrats à prix fixe
 - Prix ferme : Le prix indiqué dans l'accord ne change pas, quelle que soit toute modification environnementale.
 - Prix fixe avec escalade/déescalade : Les prix de base peuvent augmenter ou diminuer en fonction de changements spécifiques et identifiables dans les prix des matériaux.

- Prix fixe avec redétermination : Prix cible initial basé sur des estimations approximatives des coûts de main-d'œuvre et des matériaux, puis renégocié une fois qu'un niveau spécifique ou un volume de production est atteint.
 - Prix fixe avec incitations : Prix cible initial basé sur des estimations approximatives des coûts de main-d'œuvre et des matériaux, puis les économies de coûts résultant des initiatives du fournisseur sont partagées à un taux prédéterminé pour une période de temps désignée.
- 2) Contrats basés sur les coûts
- Coût plus rémunération incitative : Le prix de base est basé sur les coûts admissibles du fournisseur, et toutes les économies de coûts sont partagées entre l'acheteur et le fournisseur selon un taux prédéterminé pour une période de temps désignée.
 - Partage des coûts : Les coûts admissibles réels sont partagés entre les parties sur une base de pourcentage prédéterminée et peuvent inclure des objectifs d'amélioration de la productivité des coûts.
 - Contrat temps et matériaux : Le fournisseur est payé pour toute la main-d'œuvre et les matériaux selon un taux spécifié pour la main-d'œuvre, les frais généraux, le profit et les matériaux.
 - Coût plus frais fixes : Le fournisseur reçoit un remboursement pour tous les coûts admissibles jusqu'à un montant prédéterminé, plus des frais fixes, qui représentent un pourcentage du coût ciblé du bien ou du service.

Les contrats avec les fournisseurs peuvent être considérés comme des engagements mutuels, jouant un rôle significatif dans la stabilisation de la part de marché du fournisseur concerné au sein de l'entreprise. Les parts de marché fournisseur représentent la part relative d'un fournisseur spécifique dans les approvisionnements totaux d'une entreprise. Pour renforcer l'engagement avec un fournisseur, la mise en place d'engagements de volume est une option courante. Ces engagements peuvent offrir des avantages en termes de prévisibilité et de stabilité des approvisionnements, tout en introduisant des défis tels que la gestion des risques liés aux fluctuations de la demande et aux changements de marché.

2.2.3 Engagements par volumes

Un engagement de volume est un accord formel entre un client et un fournisseur dans lequel le client s'engage à acheter une certaine quantité de produits ou de services sur une période déterminée. En contrepartie de cet engagement à un volume d'achat élevé, le fournisseur propose des tarifs plus compétitifs ou des remises sur le prix unitaire.

Dwyer, Schurr et Oh définissent l'engagement comme un engagement implicite ou explicite de continuité relationnelle entre les partenaires d'échange (Dwyer et al., 1987). L'engagement est un ingrédient essentiel pour les relations à long terme réussies qui sont une composante de la mise en œuvre du SCM (Gundlach et al., 1995). Lambert, Stock et Ellram soulignent également que l'engagement nécessaire en termes de ressources et de pouvoir pour atteindre les objectifs déclarés est important pour une bonne gestion de la supply chain (Lambert et al., 1998).

Depuis quelque temps, de nombreuses firmes indiquent leur intention de se diriger vers des relations plus profondes avec les fournisseurs selon un dernier rapport de McKinsey & Company (2023). Elles considéraient la « collaboration et le développement des fournisseurs » parmi leurs trois principales priorités. Par exemple, ce rapport a révélé que des relations plus profondes (y compris des engagements de volume à long terme, des plans stratégiques partagés sur trois à cinq ans et des partenariats de collaboration) représentent 43 % de la base totale des fournisseurs

L'étude reflète également que le nombre de relations avec les fournisseurs dans ce secteur, incluant des engagements en volume, devrait augmenter (Figure 2.3), notamment aux dépens des relations transactionnelles habituelles, comme l'illustre la figure suivante.

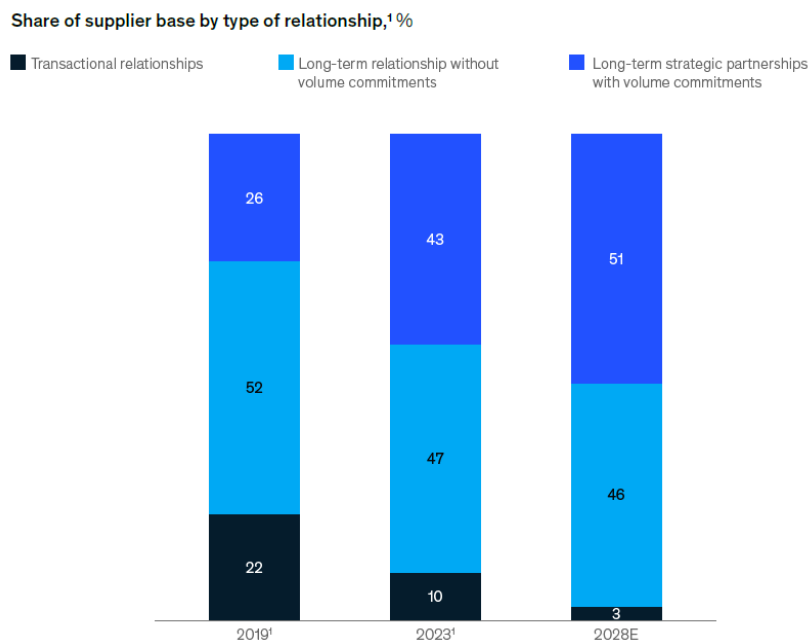


Figure 2.3 – Évolution des types de relations fournisseur (McKinsey, 2024)

La clé de ces accords est que le fournisseur s'attend à ce que le client respecte les engagements de volume afin de bénéficier des tarifs réduits, favorisant ainsi une relation commerciale continue et mutuellement bénéfique. Plusieurs avantages peuvent être identifiés (Festa, 2021), notamment :

- **Relations à long terme** : Favorisent des partenariats durables et une coopération renforcée entre les départements opérationnels, établissant des schémas de communication et une confiance mutuelle.
- **Réduction des coûts de transaction** : Simplifient le processus de négociation en établissant les termes dès le départ, minimisant ainsi les ressources humaines nécessaires pour les discussions ultérieures.
- **Expertise** : Permet aux fournisseurs de mieux comprendre les besoins spécifiques de l'entreprise, conduisant à des économies d'échelle et à une efficacité améliorée grâce à une meilleure utilisation des ressources.
- **Stabilité des prix** : Assurent la disponibilité de produits et services à des prix prévisibles, même en cas de fluctuations des coûts, ce qui facilite la planification budgétaire.
- **Priorité accrue** : Accroissent la priorité accordée à vos besoins, renforçant l'engagement des fournisseurs à répondre efficacement à vos demandes.

Par ailleurs, ces avantages n'empêchent pas l'existence de quelques inconvénients :

- **Prévisions incorrectes** : Des écarts importants par rapport aux prévisions de dépenses peuvent entraîner des violations des engagements de volume, affectant ainsi les relations et entraînant des conséquences financières.
- **Volatilité des prix** : L'engagement à des tarifs fixes peut conduire à un sur-paiement si les prix du marché fluctuent à la baisse, compromettant la rentabilité attendue.

Pour obtenir des engagements fiables sur les volumes, il est essentiel de résoudre les problématiques liées à la stabilisation des prévisions et à la consolidation des commandes. Ces deux aspects jouent un rôle important dans la gestion efficace des Supply Chain et dans la garantie d'une production fluide.

- Stabilisation des commandes : Automatiser les processus de commande est essentiel pour réduire les erreurs et garantir une régularité dans les approvisionnements. Cela permet d'éviter les variations et les interruptions dans la Supply Chain.
- Stabilisation de la demande : En stabilisant les prévisions de demande, il devient plus facile de planifier la production et de s'engager sur des volumes précis. Cela implique l'utilisation d'outils avancés de prévision et d'analyse des tendances du marché, qui permettent de mieux anticiper les fluctuations de la demande.

Ces deux aspects sont essentiels pour établir un engagement solide et fiable sur les volumes de commandes, assurant ainsi une gestion optimisée et prévisible de la Supply Chain.

2.3 Les concepts de stabilisation des commandes et des prévisions

2.3.1 Stabilisation des Commandes

Aujourd'hui, de nouvelles technologies telles que l'automatisation créent de nouvelles opportunités pour les industries afin de construire des avantages compétitifs dans divers domaines de la gestion, y compris le SCM.

N. Chaudhari met en avant l'évolution du processus d'approvisionnement dans les organisations. Autrefois axé sur le choix du fournisseur offrant les prix les plus bas, ce processus s'est transformé avec l'introduction de l'e-procurement et l'automatisation des commandes (Chaudhari, 2019). Cette approche moderne permet de mener les transactions en ligne et de développer des relations stratégiques à long terme avec des fournisseurs de qualité à travers des contrats. L'e-procurement facilite également la stabilisation des commandes en automatisant les processus, ce qui optimise les approvisionnements selon les besoins spécifiques des entreprises grâce aux ERP et aux technologies avancées.

Automatisation des commandes

L'automatisation des processus de commande vise à réduire les erreurs humaines et à accélérer le traitement des commandes. L'utilisation de systèmes ERP (Enterprise Resource Planning) ou de solutions de gestion des commandes permet d'automatiser et de standardiser ces processus. Ces systèmes intègrent les différentes fonctions de l'entreprise, facilitant ainsi la communication entre les départements et assurant une meilleure gestion des flux de travail.

- **Réapprovisionnement Automatique** : Le réapprovisionnement automatique est un système qui déclenche automatiquement des commandes lorsque les stocks atteignent un certain niveau. Ce processus utilise des seuils prédéfinis pour chaque produit afin de maintenir des niveaux de stock optimaux, évitant ainsi les ruptures de stock et les surplus.

Cette automatisation apporte quelques avantages à la gestion de la supply chain en général selon N. Chaudhari, notamment la réduction des coûts d'exploitation, l'amélioration de la précision en fournissant des informations en temps réel, et permet des gains de temps significatifs en rationalisant les processus. Elle facilite également l'intégration avec les systèmes des fournisseurs, améliorant ainsi la collaboration entre partenaires.

L'engagement en volume et l'automatisation des commandes apportent un nombre considérable d'avantages pour la relation entreprise-fournisseurs, qui peuvent être cités comme suit :

- **Renforcement du partenariat** : Les engagements de volume établissent une base pour un partenariat stable et à long terme. L'automatisation, via l'échange de données informatisé et le traitement automatisé des commandes, peut renforcer cette dépendance. En déclenchant des commandes lorsque les stocks atteignent un certain niveau, cela garantit des livraisons ponctuelles et constantes. De plus, cela facilite l'engagement à l'amélioration en fournissant des données et des analyses en temps réel que les deux parties peuvent utiliser pour optimiser les processus et résoudre les inefficacités.
- **Économies d'échelle** : L'automatisation permet de réaliser des économies d'échelle en optimisant les calendriers de production ou de livraison. Ces économies peuvent être répercutées sur

l'entreprise sous forme de prix plus bas. De plus, les engagements de volume combinés aux systèmes d'achat automatisés peuvent améliorer le pouvoir de négociation d'une entreprise en comparant les prix et les conditions de plusieurs fournisseurs, garantissant ainsi le meilleur accord possible.

- **Meilleure gestion des stocks** : L'automatisation des commandes permet une meilleure gestion des stocks, minimisant les risques de rupture de stock, notamment grâce à l'inventaire cyclique automatisé lorsque les niveaux tombent en dessous d'un seuil. Cela garantit des niveaux de stock optimaux et réduit les coûts de stockage.
- **Renforcement de la communication et de la confiance** : Les systèmes automatisés assurent la fiabilité et la cohérence dans le traitement des commandes et la livraison, en garantissant que les engagements signés seront respectés. Cette fiabilité peut conduire à un traitement préférentiel en période de forte demande ou de pénurie d'approvisionnement.

L'automatisation, si elle n'est pas bien gérée, peut engendrer certains risques pour l'entreprise et le fournisseur :

Pour l'entreprise : L'automatisation peut entraîner des risques de sur-engagement. Des prévisions peu fiables peuvent conduire à des stocks excédentaires. Elle réduit également la flexibilité en rendant difficile l'adaptation rapide aux conditions changeantes et le changement de fournisseurs.

Pour le fournisseur : L'automatisation peut fixer les fournisseurs dans des prix désavantageux en période de volatilité des prix. De plus, elle peut accroître la dépendance au marché si une grande partie des revenus du fournisseur provient d'une seule entreprise.

Pour des engagements solides et une automatisation fiable, il est essentiel de bien gérer cette automatisation et ses paramètres. Une gestion rigoureuse permet de minimiser les risques et d'assurer une collaboration efficace et une bonne gestion des relations entre l'entreprise et le fournisseur.

Gestion des stocks

En raison de son étroite relation avec la gestion des stocks, il est essentiel de présenter ses concepts et d'expliquer les notions d'inventaire cyclique et de stock de sécurité.

La gestion des stocks est le processus de gestion du flux des biens au sein d'une entreprise dans un cycle continu de commandes, de stockage, de production, de vente et de réapprovisionnement. Pour déterminer leur stratégie de gestion des stocks, les entreprises font des compromis entre la réduction du montant des liquidités liées aux stocks et la détention de stocks supplémentaires afin d'optimiser l'efficacité du service à la clientèle ou de la production.

- L'inventaire cyclique existe parce que produire ou acheter en grandes quantités permet à une étape de la supply chain de tirer parti des économies d'échelle et ainsi de réduire les coûts (Chopra & Meindl, 2016). La présence de coûts fixes associés aux commandes et au transport, les remises quantitatives dans la tarification des produits et les réductions ou promotions à court terme encouragent les différentes étapes de la supply chain à exploiter les économies d'échelle et à commander en grandes quantités.

Un lot ou une taille de lot est la quantité qu'une étape de la Supply Chain produit ou achète à un moment donné. L'inventaire cyclique est l'inventaire moyen dans une Supply Chain en raison de la production ou des achats en lots plus grands que ceux demandés par le client.

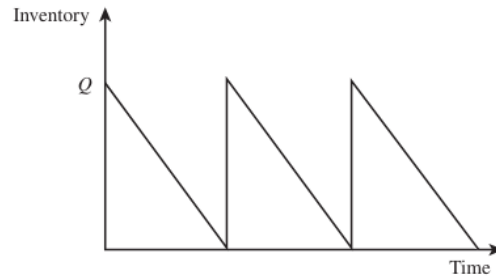


Figure 2.4 – Profile de l'inventaire cyclique (Chopra & Meindl, 2016).

- L'inventaire de sécurité, ou stock de sécurité, est l'inventaire détenu pour répondre à une demande qui dépasse la quantité prévue. Il est nécessaire en raison de l'incertitude de la demande, où une pénurie de produits peut survenir si la demande réelle dépasse la demande prévue.

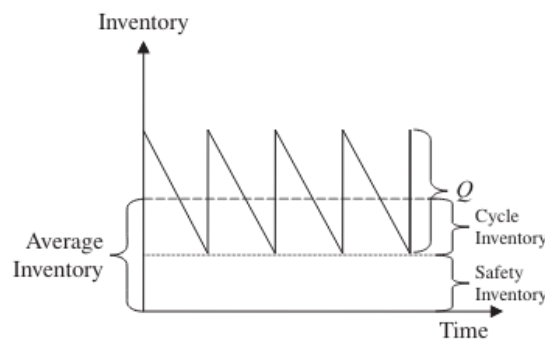


Figure 2.5 – Profile de l'inventaire avec un stock de sécurité (Chopra & Meindl, 2016).

Le niveau approprié de stock de sécurité est déterminé par deux facteurs principaux :

- L'incertitude tant de la demande que de l'approvisionnement.
- Le niveau souhaité de disponibilité des produits.

À mesure que l'incertitude de l'approvisionnement ou de la demande augmente, le niveau requis de stocks de sécurité augmente également. De même, si le niveau souhaité de disponibilité des produits augmente, le niveau requis de stocks de sécurité augmente aussi.

Conclusion : Pour conclure sur la stabilisation des commandes, il est important de souligner l'importance de bien choisir les produits adaptés à l'automatisation des commandes. Cette sélection nécessite une compréhension approfondie des paramètres spécifiques impliqués dans l'automatisation, tels que la prévisibilité de la demande et la régularité des commandes. Il est essentiel de mentionner que tous les produits ne peuvent pas être automatisés de manière efficace, nécessitant un processus de sélection.

L'utilisation d'outils et de moyens appropriés pour cette sélection est nécessaire pour maximiser les avantages de l'automatisation tout en minimisant les risques et les coûts associés. Ces outils peuvent inclure des systèmes ERP avancés, ainsi que des algorithmes de machine learning non supervisés, qui seront détaillés dans le chapitre suivant. En intégrant judicieusement l'automatisation dans la gestion des stocks, les entreprises peuvent améliorer leur efficacité opérationnelle, réduire les erreurs humaines, et renforcer leurs relations avec les fournisseurs en optimisant les niveaux de commande, renforçant ainsi la résilience de leur supply chain.

2.3.2 Stabilisation des Prévisions

Le concept de prévision de la demande

Les prévisions de la demande sont essentielles pour la planification de la Supply Chain. Dans le modèle push/pull, les processus "push" anticipent la demande client, tandis que les processus "pull" y répondent directement. Ainsi, prévoir la demande est la première étape qui doit être franchie dans les deux approches. (Chopra & Meindl, 2016).

Les prévisions possèdent plusieurs caractéristiques importantes :

- Elles sont toujours inexactes et doivent inclure la valeur attendue et une mesure de l'erreur.
- Les prévisions à long terme sont généralement moins précises que celles à court terme.
- Les prévisions agrégées sont plus précises.
- Plus une entreprise est éloignée du consommateur final, plus l'information qu'elle reçoit est distordue. L'effet coup de fouet amplifie cette variation. Les prévisions collaboratives basées sur les ventes finales réduisent l'erreur de prévision.

Prédire l'avenir est une étape primordiale pour permettre aux cadres de la supply chain d'adopter une approche proactive plutôt que réactive. Divers types de prévisions sont nécessaires, notamment la demande, l'approvisionnement et les prix et cela afin de pouvoir optimiser la gestion de son organisation mais aussi ses relations. Ces prévisions sont essentielles à différents horizons : à long terme pour la planification stratégique des approvisionnements et des ventes, à moyen terme pour la budgétisation et les plans, et à court terme pour la planification opérationnelle détaillée (Lambert et al., 1998).

Importance de la stabilisation des prévisions pour la relation fournisseur

La gestion efficace de la supply chain repose de plus en plus sur l'établissement de relations à long terme avec les fournisseurs. Pyke et Johnson (Pyke et al., 2004) analysent divers types de relations d'achat, allant des transactions au comptant aux approvisionnements à long terme, en identifiant les facteurs clés influençant chaque type de partenariat. Parallèlement, (Plambeck & Taylor, 2006) explorent également les relations client-fournisseur à travers des contrats de relation optimisés pour maximiser les profits anticipés des clients. Leurs recherches ont mis en avant l'importance de ces contrats adaptatifs dans la performance au sein des supply chains modernes, en intégrant des mécanismes de prévision et de gestion stratégique des relations (Ren et al., 2006)

Les contrats à engagement sur prévision permettent à l'entreprise d'inciter le fournisseur à répondre à ses besoins, limitant ainsi les risques de rupture d'approvisionnement. Ces contrats obligent les clients à fournir des prévisions de demande et à s'engager sur une fraction de ces prévisions, assurant ainsi une meilleure planification de la production pour les fournisseurs. En conséquence, cela réduit l'incertitude et améliore la fiabilité des livraisons (Durango-Cohen, 2003).

J. Durango-Cohen souligne que la structure de pénalité d'un contrat d'engagement de prévision capture les pertes liées aux engagements et aux livraisons. Les pénalités servent à aligner la valeur relative des clients et les considérations de profit à court terme, fournissant ainsi au fournisseur un mécanisme explicite pour favoriser les "meilleurs" clients. Cela permet non seulement de sécuriser les approvisionnements, mais aussi de renforcer les relations stratégiques entre les parties, créant ainsi une supply chain plus résiliente et plus performante (Durango-Cohen & Yano, 2006).

La stabilisation des prévisions est un élément important dans la gestion des relations fournisseurs et l'établissement d'engagements contractuels. Elle permet de réduire les incertitudes, d'améliorer la planification des approvisionnements et de renforcer les partenariats stratégiques à long terme, essentiels pour une supply chain efficace.

Enfin, la combinaison de diverses méthodes de prévision s'avère être une approche plus fiable que l'utilisation isolée d'une méthode unique, permettant de pallier les limitations individuelles et d'améliorer la précision globale. L'intégration d'outils avancés tels que l'intelligence artificielle et l'apprentissage machine facilite l'analyse de grandes quantités de données, optimisant ainsi les processus de prévision et renforçant la capacité à anticiper les fluctuations du marché avec précision.

L'analyse réalisée par J. Ren (Ren et al., 2006) a mis en lumière l'impact bénéfique du partage honnête des prévisions sur la coordination de la supply chain. À court terme, des prévisions imprécises peuvent conduire à une sous-optimalité des contrats, où le fournisseur ajuste rationnellement sa capacité en fonction de prévisions surestimées par le client. En revanche, à long terme, une communication sincère des prévisions favorise un équilibre plus efficace, permettant une allocation optimale de la capacité et des résultats améliorés pour toutes les parties impliquées. Ainsi, cela souligne l'importance critique d'une coordination fondée sur des informations de prévision fiables pour des Supply Chains durables et efficaces.

Pour conclure et comme mentionné précédemment dans ce chapitre, dans les supply chains modernes, la stabilisation des commandes par l'automatisation et les prévisions joue un rôle critique, où les engagements en volume avec les fournisseurs assurent une prévisibilité des approvisionnements tout en réduisant les risques liés aux fluctuations de la demande. Pour maximiser ces avantages, il est essentiel d'adopter des modèles prévisionnels performants et de choisir soigneusement les produits adaptés à l'automatisation des commandes. Grâce aux outils d'automatisation et à l'adoption de contrats basés sur des engagements, les relations fournisseur-client sont renforcées, garantissant à la fois une bonne gestion des besoins et une transparence durable des relations bilatérales.

Dans le chapitre suivant, nous explorerons en détail les techniques et méthodes visant à atteindre ces objectifs, à savoir la stabilisation des commandes et des prévisions.

Chapitre 3 – Exploration des approches de clustering et de prévision temporelle

3 Exploration des approches de clustering et de prévision temporelle

Dans ce chapitre, nous allons présenter les concepts et méthodes que nous utiliserons par la suite pour résoudre la problématique identifiée dans le premier chapitre. Tout d'abord, nous introduirons l'intelligence artificielle, le machine learning, et plus spécifiquement le partitionnement de données. Ensuite, nous aborderons les modèles et outils de prévision, ainsi que le concept de génération de données synthétiques.

3.1 Intelligence artificielle

L'intelligence artificielle est définie par le dictionnaire Le Robert comme un ensemble de théories et de techniques développant des programmes informatiques complexes capables de simuler certains traits de l'intelligence humaine (raisonnement, apprentissage...).

”Pour comprendre la renaissance actuelle de ce que nous appelons « l'intelligence artificielle », qui remonte clairement aux débuts de l'informatique, nous devons nous intéresser à la convergence de trois tendances : i) le Big Data, ii) le machine learning (apprentissage machine) et iii) le cloud supercomputing (calcul de haute performance sur le cloud). En ce sens, l'essor de l'IA est bel et bien une manifestation de la révolution numérique. L'une de ses lois principales, prédites en 1965 par Gordon Moore, cofondateur du fabricant de puces Intel, stipule que la puissance informatique double tous les deux ans en moyenne, à coût constant. Cette croissance exponentielle résulte des prouesses technoscientifiques continues de la miniaturisation. Celle-ci a permis l'avènement de la micro-informatique, puis celui de la nano-informatique, qui ont débouché, avec leur puissance toujours accrue, sur l'apparition des smartphones et de l'Internet des Objets” (Miaillhe & Hodes, s. d.)

3.1.1 Machine Learning

Le machine learning désigne toute méthode permettant de développer un modèle de la réalité à partir des données, que ce soit en améliorant un modèle existant ou en créant un modèle entièrement nouveau (Cornuéjols & Miclet, 2010).

L'apprentissage automatique est un sous-domaine de l'intelligence artificielle qui fait référence au développement de méthodes qui permettent à une machine d'évoluer grâce à un processus d'apprentissage, et ainsi de résoudre les problèmes pour lesquels les algorithmes classiques n'offrent pas de solution. En d'autres termes, lorsque nous ne connaissons pas de solution exacte, nous ne pouvons pas écrire de programme informatique. L'apprentissage automatique consiste alors à programmer des algorithmes permettant d'apprendre automatiquement de données et d'expériences passées pour proposer des solutions à ces problèmes (Gacôgne, 2015).

Bien qu'il soit actuellement promu par les nouvelles technologies et les nouveaux usages, l'apprentissage automatique n'est pas un domaine de recherche récent. Nous avons trouvé la première définition en 1959, grâce à l'un des pionniers de l'intelligence artificielle, Arthur Samuel, qui a défini l'apprentissage automatique comme un domaine de recherche visant à renforcer l'apprentissage de la machine sans programmation explicite. En 1997, Tom Mitchell de l'Université Carnegie Mellon a fourni une définition plus précise : *”On dit qu'un programme d'ordinateur apprend de l'expérience E par rapport à une certaine classe de tâches T et à la mesure de performance P , si sa performance aux tâches en T , telle que mesurée par P , s'améliore avec l'expérience E ”*.

La tâche T représente l'objectif d'un algorithme d'apprentissage, souvent intégré par l'utilisateur sous forme d'exemples historiques. L'utilisateur effectue une mesure de performance à travers plusieurs métriques sélectionnées par lui-même en fonction de la situation de recherche et de l'algorithme d'apprentissage utilisé, il s'agit de la mesure de performance P.

3.1.2 Types d'apprentissage

Le machine Learning nous permet de répondre à des problématiques diverses et variées, voici les 3 types d'apprentissages les plus utilisés en pratique :

Supervisé : On dispose d'un ensemble d'objets, chacun associé à une valeur cible. L'objectif est d'apprendre un modèle capable de prédire correctement la valeur cible pour un nouvel objet. Cette méthode peut être utilisée pour la classification d'images, la détection de spams, la reconnaissance vocale, la prédiction de pannes de machines, ainsi que pour les prévisions de séries temporelles.

Non supervisé : On dispose d'un ensemble d'objets sans aucune valeur cible associée. Il faut apprendre un modèle capable d'extraire les régularités présentes au sein des objets pour mieux visualiser ou appréhender la structure de l'ensemble des données.

Par renforcement : On dispose d'un ensemble de séquences de décisions (politiques ou stratégiques) dans un environnement dynamique, et, pour chaque action de chaque séquence, d'une valeur de récompense. La valeur de récompense de la séquence est alors la somme des valeurs des récompenses des actions qu'elle met en œuvre. Il faut apprendre un modèle capable de prédire la meilleure décision à prendre étant donné un état de l'environnement.

La figure suivante représente les différents types d'algorithmes ML existants ainsi que leurs applications courantes :

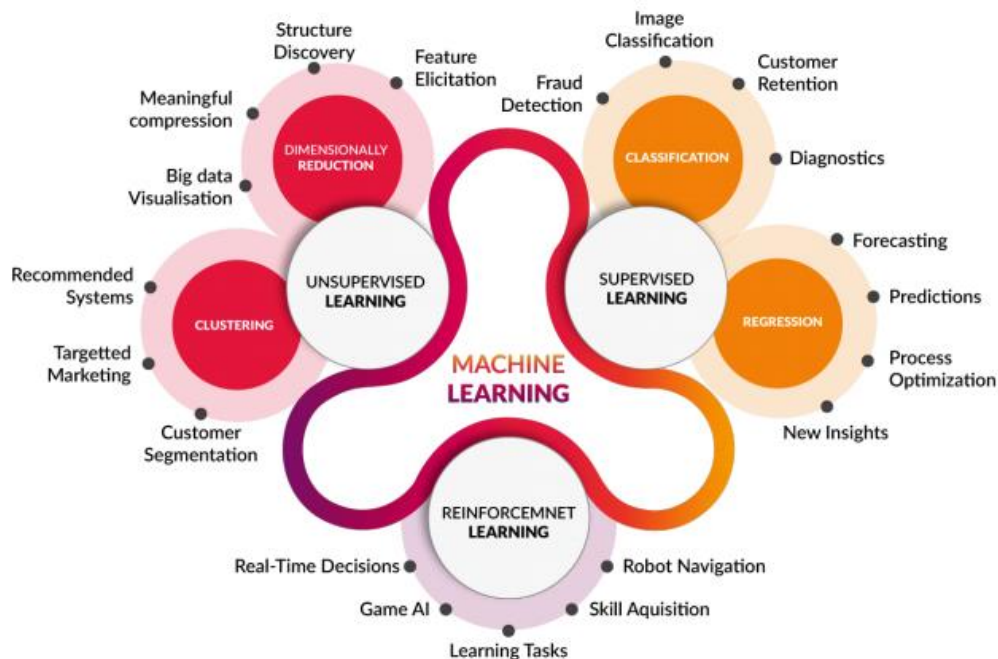


Figure 3.1 – Schéma représentant les types d'apprentissage automatique (Krzyk, 2018)

Après avoir exploré les différents types d'apprentissage automatique, nous présenterons une méthodologie structurée pour aborder le machine Learning, connue sous le nom de CRISP-DM.

3.1.3 La méthodologie CRISP-DM

CRISP-DM, acronyme de (Cross Industry Standard Process for Data Mining), est une méthodologie développée par Chapman et al. (2000). Vingt ans plus tard, cette méthodologie est devenue la norme dans les projets de data mining et de data science (Martínez-Plumed et al., 2019). Elle englobe six grandes phases :

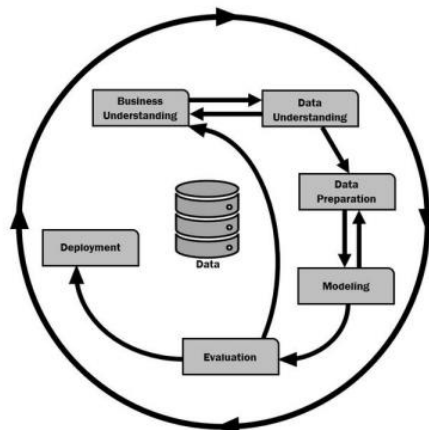


Figure 3.2 - Les étapes de la CRISP-DM (Nima, 2019)

1. **Compréhension des métiers** : la première phase a pour but de définir l'objectif du projet d'un point de vue business, ensuite formuler une problématique d'exploration de données ainsi qu'un plan préliminaire pour atteindre ces objectifs.

2. **Compréhension des données** : parmi les tâches essentielles de cette phase, la collecte des données à partir des sources de données, leur exploration et leur description ainsi que la vérification de la qualité des données.

3. **Préparation des données** : cette phase consiste à préparer les données pour le modèle final. La préparation des données englobe plusieurs étapes comme le nettoyage des données, la construction de nouveaux attributs ainsi que la transformation des données.

4. **Modélisation** : cette phase a pour but de tester plusieurs modèles ainsi que la recherche des paramètres optimaux de chaque modèle, cette phase est fortement liée avec la phase de préparation des données.

5. **Évaluation** : après avoir essayé plusieurs modèles, il est nécessaire de les évaluer d'un point de vue business et non seulement d'une perspective d'analyse de données.

6. **Déploiement** : la création du modèle ne représente pas la fin du projet, car il faut le déployer et faciliter son exploitation par les utilisateurs.

3.2 Partitionnement de données (clustering)

Le partitionnement de données (clustering, en anglais) est une technique d'apprentissage non supervisé qui englobe plusieurs méthodes mathématiques de base permettant d'identifier des groupes d'objets similaires, appelés clusters, en fonction des variables caractérisant chaque objet. On note $X = (x_1, x_2, \dots, x_n)$ l'ensemble des n variables caractérisant m objets. Les données sont généralement représentées par une matrice de taille $m * n$. Le clustering a pour objectif de regrouper les objets jugés similaires dans un même cluster et de séparer les objets dissimilaires dans des clusters distincts.

3.2.1 Choix de l'algorithme

Il existe une multitude d'algorithmes qui peuvent être utilisés pour un projet de partitionnement de données, l'utilisation de ces algorithmes dépend principalement des données d'entrées. Afin de choisir les algorithmes adéquats, les scientifiques se basent sur des indicateurs qui évaluent la performance des modèles. Il existe une multitude d'algorithmes qui permettent d'effectuer un partitionnement de données, parmi eux :

Algorithme K-Means

K-Means est un algorithme d'apprentissage non supervisé développé par le mathématicien Hugo Steinhaus en 1957, il regroupe l'ensemble de données non étiquetées en différents clusters. K définit le nombre de clusters qui doivent être créés dans le processus, si K=2, il y aura deux clusters, et pour K=3, il y aura trois clusters, et ainsi de suite. Cet algorithme nécessite que le nombre de clusters soit spécifié dès le départ par l'utilisateur. Chaque cluster est associé à des centroïdes qui représentent les moyennes de chaque cluster. L'Algorithme K-means vise à choisir des centroïdes qui minimisent l'inertie, ou le critère de la somme des carrés à l'intérieur du groupe :

$$\sum_{i=0}^n \min_{u_j \in C} (\|x_i - u_j\|^2)$$

K-means est souvent appelé l'algorithme de Lloyd. En termes simples, l'algorithme comporte trois étapes. La première étape consiste à choisir les centroïdes initiaux, la méthode la plus simple consistant à choisir des échantillons dans l'ensemble de données. Après l'initialisation, K-means effectue une boucle entre les deux autres étapes. La première étape affecte chaque échantillon à son centroïde le plus proche. La deuxième étape crée de nouveaux centroïdes en prenant la valeur moyenne de tous les échantillons affectés à chaque centroïde précédent. La différence entre les anciens et les nouveaux centroïdes est calculée et l'algorithme répète ces deux dernières étapes jusqu'à ce que cette valeur soit inférieure à un seuil. En d'autres termes, il répète jusqu'à ce que les centroïdes ne se déplacent plus de manière significative (Figure 3.3).

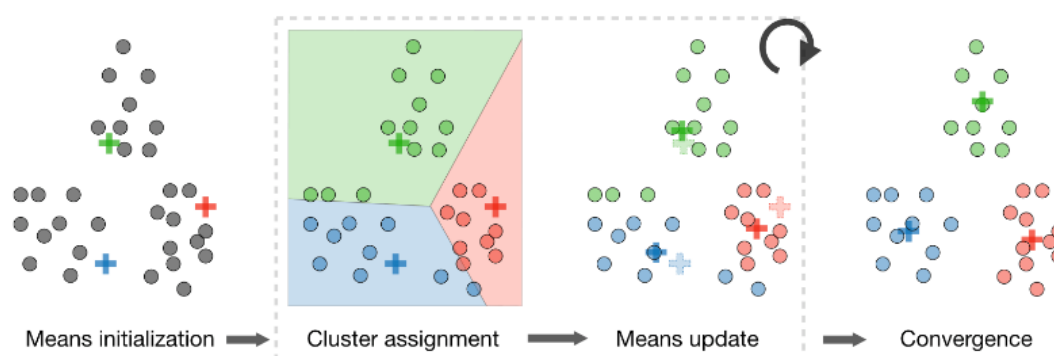


Figure 3.3 – Illustration du déroulement de l'algorithme K-means (Stanford, 2024)

Algorithme Minibatch K-Means

Le MiniBatch KMeans est une variante de l'algorithme KMeans qui utilise des petits lots pour réduire le temps de calcul, tout en essayant d'optimiser la même fonction objective. Les mini-batch sont des sous-ensembles de données d'entrée, échantillonnés de manière aléatoire à chaque itération d'apprentissage. Ces

Contrairement à l'algorithme k-means et à la classification ascendante hiérarchique, le nombre de cluster est estimé par l'algorithme lui-même par contre l'utilisateur doit introduire la valeur d'un paramètre appelé bandwidth, ce dernier dicte la taille de la région à parcourir par means shift et il peut être estimé à son tour en utilisant un algorithme spécifique. L'algorithme Means shift n'est pas très évolutif, car il effectue plusieurs recherches de voisins les plus proches pendant son exécution. La convergence de l'algorithme est garantie mais il arrête d'itérer lorsque la modification des centroïdes devient très faible. Voici un exemple de clustering en utilisant means shift :

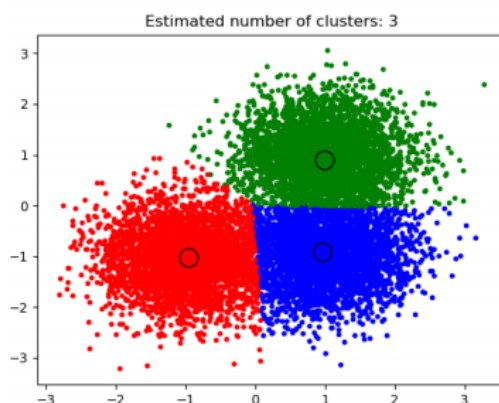


Figure 3.5 – Exemple de clustering avec Means Shift (Scikit-Learn, 2024)

Algorithme DBSCAN

DBSCAN C'est un algorithme de clustering basé sur la densité. Il peut découvrir des clusters de différentes formes et tailles à partir d'une grande quantité de données contenant du bruit et des valeurs aberrantes. L'algorithme DBSCAN utilise deux paramètres :

- **min Pts** : Le nombre minimum de points (un seuil) regroupés pour qu'une région soit considérée comme dense.
- **Eps** : Une mesure de distance qui sera utilisée pour localiser les points dans le voisinage de tout point.

DBSCAN est capable d'estimer par lui-même le nombre de clusters nécessaires lors du clustering contrairement à k-means, l'algorithme produit également des résultats plus raisonnables que k-means sur une variété de distributions différentes. La figure ci-dessous illustre ce fait : Fig. 22 : Comparaison des résultats obtenus par DBSCAN et K-Means (*Scikit-learn*, 2024)) Au niveau de la figure précédente, nous pouvons remarquer que DBSCAN peut traiter des jeux de données ayant une structure circulaire contrairement au K-Means.



Figure 3.6 – Comparaison des résultats obtenus par DBSCAN et K-Means (*Scikit-learn*, 2024)

Au niveau de la figure précédente, nous pouvons remarquer que DBSCAN peut traiter des jeux de données ayant une structure circulaire contrairement au K-Means.

3.2.2 Validation des modèles

Mesure de la tendance au partitionnement

Avant d'évaluer la performance du clustering, il est très important de s'assurer que la base de données a une tendance au clustering et ne contient pas de points uniformément distribués. Si les données ne contiennent pas de tendance au clustering, les clusters identifiés par les algorithmes les plus avancés peuvent ne pas être pertinents. La distribution non-uniforme des points dans l'ensemble des données devient importante dans le clustering. Pour effectuer ce genre d'évaluation, il existe deux principales méthodes :

- **Test statistique de Hopkins :**

Dans le but de tester l'aptitude d'une base de données à être partitionnée en plusieurs groupes, nous avons opté pour le test statistique de Hopkins qui permet de voir si les données suivent une distribution uniforme. En suivant la formulation du test d'hypothèse suivante :

$$\text{Hopkins statistics} \begin{cases} H > 0,5 \\ H < 0,5 \end{cases}$$

- Si $H > 0.5$ cela veut dire que les données sont uniformément distribuées et donc la tendance au clustering est faible.
- Si $H < 0.5$ nos données ont un caractère aléatoire spatial et ils ont tendance à être partitionnées en groupes.

- **L'algorithme VAT :** qui permet de visualiser le nombre de clusters possible pour une base de données précise à travers un graphique.

Détermination du nombre de clusters optimal

Certains algorithmes de clustering, comme les K-means, requièrent le nombre de clusters, k , comme paramètre d'entrée. Obtenir le nombre optimal de clusters est très important dans l'analyse. Si k est trop élevé ou trop faible, alors les points de données seront incorrectement regroupés. Trouver le bon nombre de clusters est un problème très subjectif, cependant il existe quelques méthodes qui permettent d'obtenir le K optimal, il s'agit de :

- Exigences liées au cas étudié ou avis d'experts.
- Approche mathématique en utilisant la méthode du coude.
- Approche statistique par la méthode du gap.

Méthode du coude

La variance intra-groupe est une mesure de la compacité du groupe. Plus la valeur de la variance intra-groupe est faible, plus la capacité du groupe formé est élevée. La somme de la variance intra-classe, W , est calculée pour les analyses de clustering effectuées avec différentes valeurs de k . W est une mesure

cumulative de la qualité du clustering des points dans l'analyse. Le tracé des valeurs de k et de la somme correspondante de la variance intra-classe aide à trouver le nombre de clusters.

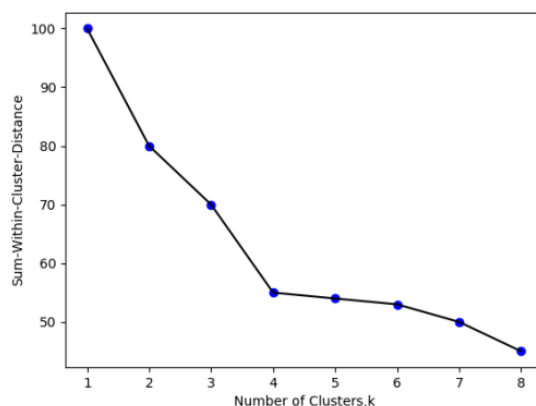


Figure 3.7 – Exemple illustrative de la méthode du coude

Nous remarquons sur cet exemple que le coude se casse sur $k = 4$, le nombre de cluster optimal dans ce cas est de 4.

La méthode Gap statistique

La méthode a été publiée par Tibshirani et al. (2001). L'approche peut être appliquée à toute méthode de clustering. La méthode consiste en le calcul des variances intra-clusters pour différentes valeurs de K (nombre de cluster s), sa valeur optimale sera la valeur qui maximise les écarts.

Mesure de la qualité du clustering

Après avoir effectué le clustering il est nécessaire d'évaluer la qualité de ce dernier en utilisant des métriques. Un clustering idéal est caractérisé par une distance intra cluster minimale et une distance inter cluster maximale. Pour l'évaluation du clustering effectué sur un ensemble de données non étiqueté. Il existe trois principaux indicateurs qui permettent de mesurer la performance de ce type de clustering :

- **Coefficient de silhouette :**

Il a été publié par Rousseeux en 1987, c'est une métrique utilisée pour mesurer la qualité d'un partitionnement de données, il est compris entre -1 et 1. Les résultats du coefficient de silhouette sont interprétés de la manière suivante :

- 0.71 - 1 : Une structure solide a été trouvée, la qualité du clustering est bonne.
- 0.52 - 0.7 : Structure raisonnable, la qualité du clustering est moyenne.
- 0.26 - 0.51 : Faible structure.
- < 0.25 : Très faible structure.

Le coefficient de silhouette se constitue de deux autres métriques il est calculé par la manière suivante :

$$s = \frac{b - a}{\max(a, b)}$$

Tel que :

a : Représente la distance moyenne entre un échantillon et tous les autres points de la même classe.

b : Il s'agit de la distance moyenne entre un échantillon et tous les autres points du cluster le plus proche.

La figure suivante illustre les deux distances a et b sur un exemple de clustering :

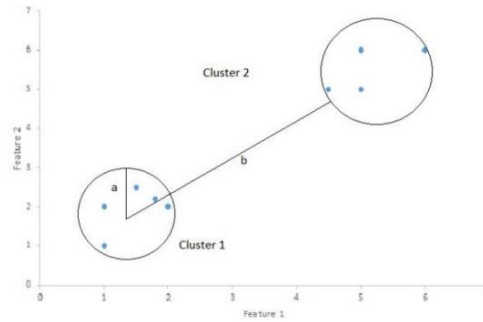


Figure 3.8 - Les distances entre a et b sur un exemple de clustering

- **Indice de Calinski-harabasz :**

Introduit par Calinski et Harabasz en 1974, il s'agit du rapport entre la variance inter-groupes et la variance intra-groupe. La valeur de l'indice est plus élevée lorsque les clusters sont denses et bien séparés, ce qui correspond au concept de cluster. La formule mathématique de l'indice est la suivante :

$$s = \frac{\text{tr}(B_k)}{\text{tr}(W_k)} \times \frac{n_E - 1}{k - 1}$$

Tel que :

s : Indice de Calinski-Harabasz.

k : Nombre de clusters.

B_k : Matrice des variances intra-groupes avec : $B_k = \sum_q (c_q - c)(c_q - c)^T$

W_k : Matrice des variances inter-groupes avec : $W_k = \sum_{q=1}^k \sum_{x \in C_q} (x - c_q)(x - c_q)^T$

C_q : L'ensemble des points dans un cluster q .

c_q : Le centre d'un cluster.

c : Le centre de l'ensemble de données E .

Indice de Davies-Bouldin :

C'est une métrique utilisée pour la mesure de la qualité d'un groupe de données, il s'agit de la moyenne du rapport maximal entre la distance d'un point au centre du cluster et la distance entre deux centres de cluster. L'indice est calculé à partir d'un coefficient de similarité :

$$R_{ij} = \frac{s_i + s_j}{d_{ij}}$$

Avec : R_{ij} : Coefficient de similarité. s_i : La distance moyenne entre chaque point et son centroïde. d_{ij} : Distance entre les deux centroïdes. L'indice est donc défini par la relation suivante :

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{ij}$$

Avec k : Le nombre de clusters.

3.3 Méthodes et outils pour les prévisions

En parlant des outils et méthodes utilisés pour les prévisions, avant d'aborder les méthodes et modèles de machine learning, il est important de considérer d'abord les modèles classiques de prévisions. On distingue principalement deux groupes : les modèles de lissage et les modèles de contrôle de Box-Jenkins. Cependant, avant cela, nous allons définir les séries chronologiques. En annexe, vous trouverez également les familles de modèles de prévisions, qualitatives et quantitatives. Les modèles quantitatifs se divisent en modèles causaux et auto-projectifs (Hubert & Thibault, 2013). Dans le cadre de cette étude, nous nous concentrerons sur les modèles auto-projectifs.

3.3.1 Les méthodes Auto-Projectives (Série chronologique)

La modélisation auto-projective, quant à elle, est fondée exclusivement sur les données historiques de la variable à prédire. Son concept de base consiste à filtrer la série Y_t avec des outils mathématiques et statistiques pour trouver les valeurs futures y_{t+h} . Il s'agit, en effet, de prédire la valeur de l'endogène Y à partir de sa série chronologique uniquement, d'où la qualification de "modélisation auto-projective".

$$Y_t = S(t) + B(t) + U(t)$$

- $S(t)$: Fonction de saisonnalité.
- $B(t)$: Fonction de tendance.
- $U(t)$: Fonction aléatoire.

La tendance correspond à une évolution au cours du temps. Cette tendance est représentée par une fonction qui marque, indépendamment de la saisonnalité, le comportement des données de long terme. Le paramètre saisonnalité suit les transformations survenues au cours du temps, autrement dit, à des périodes de temps récurrentes. Les méthodes auto-projectives se séparent en deux parties : les techniques de lissage et les techniques de contrôle (Box-Jenkins).

Les techniques du lissage

Filtrant les valeurs historiques pour éliminer ces variations. Un filtre linéaire est la combinaison linéaire d'un nombre précis d'anciennes données de la série chronologique initiale. Ce nombre est défini comme étant le nombre de périodes avant t à considérer pour évaluer le filtre (Hubert & Thibault, 2013).

Les deux grandes classes de filtres utilisées pour le traitement des séries chronologiques sont : les moyennes mobiles et les lissages exponentiels.

Moyenne Mobile : Cette méthode utilise l'historique des N dernières demandes en calculant la moyenne de ces dernières demandes. Cet outil est très efficace lorsqu'il s'agit d'une faible variation de la demande, en d'autres termes, une demande quasi constante (ni saisonnalité, ni changement de structure).

$$P_t = \frac{\sum_{k=1}^t D_{t-k}}{n}$$

Avec

P_t : Prédiction pour la période t ; D_t : la demande réelle pour la période t et N : le nombre de périodes prises en compte, généralement calculé comme suite : $N = \frac{1}{5}n$; Avec n le nombre de périodes.

Lissage Exponentiel : Le lissage permet une pondération particulière des données du passé récent. Il est adéquat pour des séries sans saisonnalité ni tendance apparentes (séries stationnaires) pour éliminer leur contenu aléatoire et estimer une valeur de prévision. La méthode donne le plus grand poids aux données récentes et des poids décroissants aux valeurs les plus anciennes (Hubert, 2013).

Lissage Exponentiel Simple (LES) : Ce type de modèle est généralement utilisé lorsque la série ne présente ni saisonnalité ni tendance. Cette méthode permet de donner un poids décroissant aux données passées en fonction de leur antériorité.

Lissage Exponentiel Double (LED) : Cette technique consiste à effectuer un lissage de la série déjà lissée en utilisant le même principe que celui du LES, tout en prenant en compte le caractère tendanciel de la série chronologique, mais sans la saisonnalité.

Lissage Exponentiel Triple (Holt-Winters) : Cette méthode peut prendre en compte la présence de tendance et de saisonnalité dans la série chronologique. Il est néanmoins nécessaire de définir la périodicité des données à prendre en compte. Traitant des séries homogènes en tendance et saisonnalités, elle est considérée comme une amélioration des techniques de lissage exponentiel et de la moyenne mobile. Elle comporte 3 paramètres à estimer : paramètre de tendance, paramètre de variation et paramètre de saisonnalité. Sa forme additive est la suivante :

$$P_{t+h} = (a_{0t} + a_{1t} \cdot h) \cdot S_{t-p+h} \quad \text{si } 1 \leq h \leq p$$

$$P_{t+h} = (a_{0t} + a_{1t} \cdot h) \cdot S_{t-p+2h} \quad \text{si } p+1 \leq h \leq 2p$$

Avec :

$$\begin{cases} a_{0t} = \alpha \left(\frac{D_t}{S_{t-p}} \right) + (1 - \alpha)(a_{0(t-1)} + a_{1(t-1)}) \\ a_{1t} = \beta(a_{0t} - a_{0(t-1)}) + (1 - \beta)a_{1(t-1)} \\ S_t = \gamma \left(\frac{D_t}{a_{0t}} \right) + (1 + \gamma)S_{t-p} \end{cases}$$

Où P_{t+h} la prévision à $t + h$; p la période ; D_t la demande réelle pour la période t ; α , β et γ des paramètres à estimer.

Les techniques de contrôle (Box Jenkins)

La méthode de prévision de Box-Jenkins (BJ) permet de traiter des séries chronologiques perturbées et beaucoup plus complexes. Elle utilise également des tests statistiques et des indicateurs de performance sévères pour déterminer le meilleur modèle de type ARMA qui restitue au mieux le comportement d'une série temporelle stochastique selon une procédure en trois étapes : identification, estimation, et enfin validation et choix du modèle (Gourieroux & Monfort, 1995).

Avant de rentrer plus profondément dans les détails de la méthode, quelques fondements de base doivent être définis.

1. Concept de stationnarité

Une série chronologique Y_t avec $t = 1, \dots, T$; est dite stationnaire si ses caractéristiques stochastiques sont invariantes par rapport au changement d'origine :

- $E(Y_t) = \mu \forall t$ (constante indépendante du temps)

- $Var(Y_t) = \sigma^2 < \infty$ Constante indépendante du temps
- $Cov(y_t, y_{t-k}) = E(y_t - \mu)(y_{t-k} - \mu) = \gamma_k$

Ceci implique qu'une série stationnaire ne présente ni tendance, ni saisonnalité. Dans le cas contraire, la série est non stationnaire et peut être de type DS (déterministe stochastique) ou de type TS (tendance stochastique).

Remarque : Une série ε_t avec $E(\varepsilon_t) = 0$ et $Var(\varepsilon_t) = \sigma^2$ est une série stationnaire à bruit blanc.

2. Fonction d'autocorrélation simple et partielle

Définition 1 : La fonction d'autocorrélation, notée P_k , mesure la corrélation de la série avec elle-même décalée de k périodes (Gourieroux & Monfort, 1983).

$$P_k = \frac{Cov(y_t, y_{t-k})}{\sigma_{y_t} \sigma_{y_{t-k}}} = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sqrt{\sum_{t=k+1}^n (y_t - \bar{y})^2} \sqrt{\sum_{t=k+1}^n (y_{t-k} - \bar{y})^2}}$$

Où \bar{y} représente la moyenne empirique et de la série calculée sur n-k périodes avec n le nombre d'observations. Nous pouvons en déduire que

$$p_0 = 1, p_k = p_{t-k}$$

Définition 2 : La fonction d'autocorrélation partielle de retard k mesure la corrélation entre y_t et y_{t-k} , en retirant l'influence des autres variables ($y_{t1}, y_{t2}, \dots, y_{tk+1}$). A été retirée.

Le graphe de ces fonctions est appelé respectivement corrélogramme simple et corrélogramme partiel. Le premier permet de déterminer la typologie de la série par rapport à la décomposition MA(q), où q est la valeur à partir de laquelle les termes ne sont plus significativement différents de 0.

En effet, si aucun terme n'est significativement différent de 0, le processus étudié est sans mémoire et donc il n'est affecté ni de tendance ni de saisonnalité. De la même façon, le second corrélogramme permet de détecter la typologie de la série par rapport à la décomposition AR(p).

3. Test de Dickey-Fuller

Les tests de Dickey-Fuller permettent non seulement de détecter l'existence d'une tendance (tests de racine unitaire) mais aussi de déterminer la bonne manière de stationnariser une chronique. Il existe deux types de processus sont distingués (Hamisultane, 2016) :

Le processus TS (Trend Stationary) : Un processus pouvant s'écrire de la façon suivante :

$$y_t = \alpha + \beta t + \varepsilon t$$

Où : εt représente l'erreur du modèle à la période t.

Le processus est de nature déterministe et non stationnaire, non stationnaire car $E(y_t) = \alpha + \beta t$ dépend donc du temps. Cependant un processus TS peut être stationnaire et cela en soustrayant une valeur estimée $\hat{\alpha} + \hat{\beta} t$ en utilisant la méthode des moindres carrés.

Le processus DS (Differency Stationary) : aussi appelé marche au hasard, ce processus présente une non stationnarité de type stochastique et peut s'écrire de la sorte :

$$y_t = y_{t-1} + \beta t + \varepsilon t$$

Deux cas alors sont considérés :

- $\beta t \neq 0$ (Nous parlons de DS avec dérivé) : ici la non stationnarité est alors causée par $E(yt) = yt + \beta t$
 Implique que $E(yt)$ dépend du temps.

- $\beta t = 0$ (DS sans dérive) : dans ce cas la non stationnarité vient de la variance de yt , en effet ; implique que si $t \rightarrow \infty$ alors $V(yt) \rightarrow \infty$.

Enfin, pour stationnaires une DS (avec ou sans dérive), il suffit de le passer en différence première.

A noter qu'une série est dite intégrée d'ordre d (notée $yt \sim I(d)$) s'il convient de la différencier d fois afin qu'elle soit stationnaire.

- **Test de Dickey-Fuller simple et augmenté**

Le test de racine unitaire de Dickey-Fuller est un test statistique qui vise à savoir si une série temporelle est stationnaire. Au cours de ce test, nous supposons que l'erreur de la série (ϵ_t) suit une voie normale. Les modèles servant de base à la construction de ce test sont au nombre de trois :

[1] $X_t = \phi_1 \cdot X_{t-1}$: Modèle autorégressif d'ordre 1

[2] $X_t = \phi \cdot X_{t-1} + c + \epsilon_t$: Modèle autorégressif d'ordre 1 avec constante

[3] $X_t = \phi \cdot X_{t-1} + b_t + c + \epsilon_t$: Modèle autorégressif d'ordre 1 avec tendance et constante

Les hypothèses du test sont les suivantes :

- $H_0 : \phi_1 = 1$, La série est non stationnaire, existence d'une racine unitaire
- $H_1 : \phi_1 \neq 1$, La série est stationnaire, Absence d'une racine unitaire

Avec : ϕ_1 : la racine unitaire ; c : la constante ; b : coefficient de la tendance ; ϵ_t : l'erreur prévisionnelle à bruit blanc.

Il existe une version du Test de Dickey-Fuller Augmenté, qui est similaire au précédent mais ne suppose pas que l'erreur est à Bruit Blanc. Le schéma qui récapitule l'ensemble des étapes à suivre lors du test de racine unitaire de Dickey Fuller (Annexe D).

4. Typologie des modèles AR, MA et ARMA

Autorégressif (AR) :

Cette composante implique que la variable d'intérêt est modélisée en fonction de ses propres valeurs passées. Ou une valeur à un moment donné peut être exprimée comme une combinaison linéaire de ses valeurs passées (Hamisultane, 2016).

$$AR(p): y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \epsilon_t$$

Moyenne mobile (MA) :

Cette composante suggère que la variable d'intérêt dépend des termes d'erreur passés. Ou chaque observation y_t est générée par une moyenne pondérée d'aléas jusqu'à la q -ième période (Hamisultane, 2016).

$$MA(q) = y_t = \epsilon_t - \alpha_1 \epsilon_{t-1} - \alpha_2 \epsilon_{t-2} - \dots - \alpha_q \epsilon_{t-q}$$

Modèle ARMA (mélange de processus AR et MA) :

Les modèles ARMA sont donc représentatifs d'un processus généré par une combinaison des valeurs passées et des erreurs passées. Ils sont définis par l'équation

$$ARMA(p,q): y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p+1} + \epsilon_t - \alpha_1 \epsilon_{t-1} - \alpha_2 \epsilon_{t-2} - \dots - \alpha_q \epsilon_{t-q+1}$$

Intégré (I) : La partie intégrée du modèle concerne le nombre de différenciations nécessaires (d) pour rendre la série temporelle stationnaire (les valeurs des données sont remplacées par la différence entre les valeurs des données et les valeurs précédentes).

L'évaluation de la stationnarité des données repose sur le paramètre d'auto régression (p). Si les données sont stationnaires, la prévision est simplifiée. Sinon, elles doivent être différenciées (d). On vérifie également si les données conviennent à un modèle de moyenne mobile (q).

La préparation initiale des données pour la prévision implique de déterminer ces paramètres (p, d, q) à travers une analyse approfondie. Une fois déterminés, ces paramètres sont utilisés dans la méthode de Box-Jenkins pour élaborer des prévisions précises. La procédure de prévision avec la méthode de Box-Jenkins implique plusieurs étapes schématisées, comprenant notamment l'identification du modèle, l'estimation des paramètres, et la vérification des résidus pour assurer la qualité de la prévision. Un récapitulatif de la méthode Box-Jenkins se trouve en annexe.

Mesure de l'erreur

Faire des prévisions implique de s'exposer au risque d'erreur, pour cela il est impératif de continuellement remettre en question la fiabilité des modèles utilisés et des prévisions qui en découlent. Cela peut être fait en mesurant le niveau d'erreur, et plusieurs méthodes existent à cet effet, on cite

Erreur	Formule	Description
Erreur brute de prévision : E_t	$E_t = F_t - D_t$	Où F_t est la prévision pour la période t , et D_t est la valeur réelle observée pour la période t .
Erreur quadratique moyenne (Mean squared error) : MSE	$MSE = \frac{1}{n} \sum_{t=1}^n E_t^2$	Mesure la moyenne des carrés des erreurs de prévision, pénalisant davantage les grandes erreurs et mettant en évidence les prévisions très inexactes.
Ecart absolu moyen (Mean absolute Error) : MAE	$MAE = \frac{1}{n} \sum_{t=1}^n E_t $	Mesure l'erreur moyenne absolue en pourcentage entre les prévisions et les valeurs réelles, fournissant une indication relative de la précision des prévisions. Utile pour estimer la variance de l'erreur en supposant que l'erreur suit une loi normale.
Pourcentage d'erreur moyenne absolue (Mean absolute percentage error) : MAPE	$MAPE = \frac{1}{n} \sum_{t=1}^n \left \frac{E_t}{D_t} \right \cdot 100$	Calcule la moyenne des écarts absolus entre les valeurs prévues et les valeurs réelles, offrant une mesure directe de la précision des prévisions sans considération pour la taille des valeurs réelles. Pertinente pour la mesure de l'erreur dans le cas où il existe une forte saisonnalité avec des grandes variations de la demande.
Erreur quadratique moyenne racine (Root Mean squared error) : RMSE	$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n E_t^2}$	Combine les propriétés de la MSE en fournissant une mesure de l'erreur de prévision en unités des valeurs observées. Utile pour avoir une mesure de l'erreur qui a la même unité que les données d'origine.

3.3.2 Modèles de Machine Learning pour les prévisions

Après avoir exploré les méthodes statistiques pour la prévision, déjà considérées comme classiques, qui peuvent présenter certaines limitations lorsqu'il s'agit de capturer les nuances, la robustesse et les motifs complexes des séries temporelles, nous allons donc passer à un autre niveau d'analyse en explorant de nouveaux algorithmes qui ont émergé ces dernières années. Ces algorithmes réussissent à construire des modèles prédictifs de séries temporelles avec variabilité saisonnière et capturent les nuances les plus complexes des séries temporelles. Parmi ces méthodes, on trouve Facebook Prophet (Taylor & Letham, 2018), XGBoost (Chen & Guestrin, 2016), et un ensemble de méthodes de réseaux neuronaux artificiels (ANN), en particulier la méthode Long Short-Term Memory (LSTM) développée par Hochreiter et Schmidhuber en 1997 (Kramar & Alchakov, 2023).

Nous allons donc découvrir les principaux outils de machine learning les plus fréquemment utilisés pour la prévision des séries temporelles : FbProphet, XGBoost et les réseaux neuronaux récurrents (RNNs), ainsi que les LSTMs.

L'Algorithme Facebook Prophet

Facebook Prophète (FB Prophet) 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 en commerce électronique ou des modèles météorologiques. À la base, FB Prophet 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 (Taylor & Letham, 2018)

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, FB Prophet utilise une approche de régression linéaire par morceaux qui s'adapte aux données. La composante saisonnière, quant à elle, permet de capturer les schémas récurrents dans les données, tels que les tendances hebdomadaires ou mensuelles. FB Prophet utilise la série de Fourier pour modéliser de manière flexible différents types de modèles saisonniers, en s'adaptant à diverses fluctuations temporelles. Enfin, la composante "bruit" tient compte des fluctuations aléatoires des données qui ne peuvent être attribuées ni à la tendance ni à la saisonnalité.

Le modèle XgBoost

Le modèle XGBoost a été proposé pour la première fois par Chen Tianqi et Carlos Guestrin en 2011 et a été continuellement optimisé et perfectionné dans des recherches ultérieures par de nombreux scientifiques. L'algorithme de boosting de gradient extrême (XGBoost) est un système d'apprentissage automatique évolutif pour le boosting des arbres, qui est une bibliothèque de boosting de gradient optimisée et distribuée et qui peut examiner efficacement l'importance de toutes les caractéristiques d'entrée. Au fil des années, il s'est avéré être une solution fiable et efficace pour les problèmes d'apprentissage automatique. Comparé à d'autres algorithmes de boosting de gradient, XGBoost peut rassembler un classificateur fort à partir d'un ensemble de classificateurs faibles et présente des avantages tels que la gestion efficace des valeurs manquantes, la capacité de prévenir le surapprentissage ainsi que le calcul parallèle et distribué qui réduit le temps d'exécution. L'objectif de XGBoost est d'employer une méthodologie d'optimisation par descente de gradient et des fonctions de perte différentiables arbitraires pour minimiser la fonction de perte en ajoutant des apprenants faibles, c'est-à-dire de définir et d'optimiser la fonction objective. XGBoost tente de minimiser l'objectif régularisé comme suit :

$$Obj_m = \sum_{i=1}^n l(\hat{y}_i, y_i) + f_m(x_i) + \Omega(f_m), f \in$$

Où l est la fonction de perte d'entraînement qui mesure la déviation entre la valeur prédite \hat{y}_i par notre modèle et la valeur réelle y_i . Ω est la fonction de régularisation qui mesure la complexité du modèle, ce qui tend à prévenir le surapprentissage. Afin de minimiser la structure, un algorithme de recherche glouton est utilisé.

Apprentissage approfondi et les Réseaux de neurones artificiels

Un réseau de neurones est une technique d'apprentissage automatique inspirée des réseaux de neurones biologiques, c'est donc un sous-domaine de l'apprentissage automatique. Les réseaux de neurones artificiels sont composés de neurones artificiels simples qui sont des petites fonctions mathématiques montés en réseau et qui nous permettent de former des fonctions complexes.

Un réseau de neurones profond est un réseau de neurones multicouches (en général plus de 4 couches) qui peut comporter des millions de neurones, répartis en plusieurs dizaines de couches. Les réseaux de neurones profonds peuvent apprendre des caractéristiques à différents niveaux d'abstraction, des contours (aux couches inférieures) aux caractéristiques très complexes (aux couches plus profondes).

Le réseau neuronal

Neurone formel (Perceptron) La conception des réseaux de neurones se base sur les neurones artificiels qui sont inspirés du vrai neurone existant dans notre cerveau. La figure 3.9 montre une représentation d'un neurone artificiel.

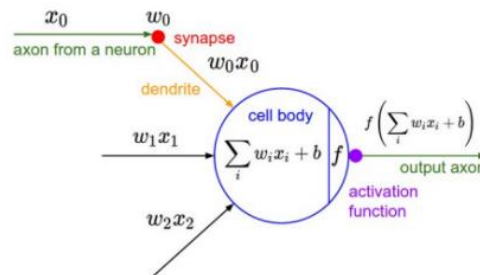


Figure 3.9 - Exemple illustratif de la conception d'un neurone artificiel

Les x_i représentent soit les données d'entrée, soit les valeurs de sorties d'autres neurones.

- Les poids w_i représentent les valeurs de puissance des entrées.
- Le neurone artificiel fera un produit entre les poids w_i et les valeurs d'entrée x_i , puis ajoutera un biais b .
- Le résultat est transmis à une fonction d'activation f qui ajoutera une certaine non-linéarité.

Les Fonctions d'activation

Après la sommation pondérée de ses entrées, le neurone applique également une non linéarité sur ce résultat. Cette fonction non-linéaire s'appelle la fonction d'activation. La fonction d'activation est essentielle pour le fonctionnement des réseaux de neurones, c'est elle qui fait que si le neurone est activé ou non. La sortie de cette fonction est envoyée à la couche suivante. Le détail des fonctions d'activations se trouve en annexe E.

Les réseaux de neurones récurrents (RNNs)

Les réseaux récurrents (RNN) sont un type des réseaux de neurones qui sont adaptés à plusieurs tâches de la prédiction des informations séquentielles. Pour les réseaux RNN, l'information est donnée par une connexion en boucle ce qui permet de prendre en compte à l'étape courante les informations prédites dans les étapes précédentes. Ils sont appelés récurrents car ils effectuent la même tâche pour chaque bloc RNN (AMIDI, s. d.). Les réseaux récurrents à mémoire court et long terme (LSTM) introduits par Hochreiter et Schmidhuber en 1997 sont un type des réseaux récurrents (RNN) capable de prendre en compte des dépendances à long terme et se souvenir des informations pendant une longue période. C'est l'architecture de réseau de neurones récurrents la plus utilisée en pratique et qui permet de répondre au problème de disparition de gradient (gradient vanishing).

Fonctionnement d'un réseau de neurones récurrents RNN

Un réseau neuronal récurrent est une répétition d'une seule cellule. Cette cellule représente le cœur du réseau, toutes les opérations qui assurent le bon fonctionnement du réseau se font dedans. Pour comprendre le fonctionnement de chaque cellule RNN, on donne la figure suivante :

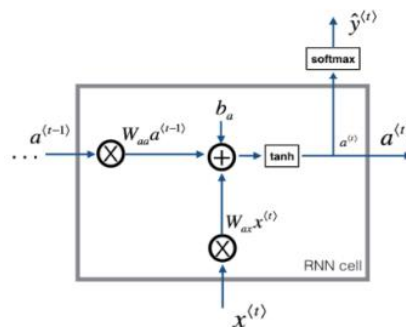


Figure 3.10 - Schématisation du fonctionnement d'une cellule RNN

Où les x représentent les entrées du réseau, Les a représentent les sorties de la fonction d'activation, ils seront des entrées pour les blocs qui viennent après et Les y représentent les sorties du réseau (le nombre d'entrées et sorties peuvent être différents).

Les sorties de la cellule RNN sont données par :

$$a^t = \tanh(W_{ax}x^t + W_{aa}a^{t-1} + b_a)$$

$$y^t = \text{softmax}(W_{ax}a^t + b_a)$$

- a^{t-1} représente la sortie de la fonction d'activation de la cellule précédente.
- Les W et b représentent respectivement les poids et les biais du réseau. Les cellules RNN prennent en compte les informations courantes contenues dans l'entrée x , ainsi que les informations prédites dans les étapes précédentes contenu dans la fonction d'activation a à $t-1$.

Fonctionnement LSTM

L'architecture des réseaux LSTM est un modèle séquentiel basé sur deux composants essentiels : les états et les portes. Les états incluent : L'état caché qui représente la valeur de la couche cachée précédente et

est également appelé vecteur de sortie, noté s . et l'état d'entrée qui est une combinaison linéaire des données d'entrée actuelles et de l'état caché.

La cellule LSTM se compose de trois portes. Une Porte d'entrée (notée i), une Porte d'oubli (notée f) et une porte de sortie (notée o). Chaque unité du réseau LSTM comprend ces trois portes et utilise une fonction d'optimisation pour mettre à jour les poids associés aux unités du réseau.

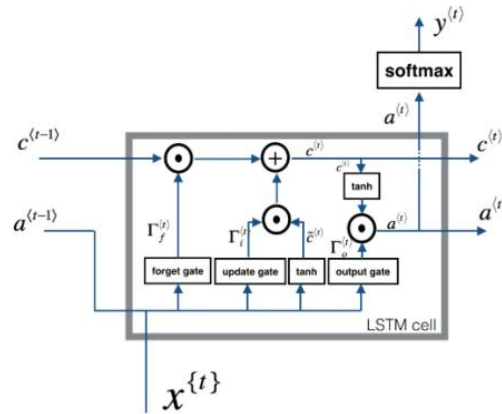


Figure 3.11 - Schématisation du fonctionnement d'une cellule LSTM

Porte d'oubli :

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Où σ désigne la fonction d'activation sigmoïde. Cette équation permet de déterminer quelle information de l'état précédent doit être conservée pour les calculs futurs

Porte d'entrée :

Cette porte est utilisée pour trouver un paramètre intermédiaire et décider si la valeur de l'état interne servira de cellule mémoire.

$$i = \sigma(W_i x_t + W_{i_s} s_{t-1} + b_i)$$

$$c_t = \tanh(W_c x_t + W_{c_s} s_{t-1} + b_c)$$

Mise à jour de l'état cellulaire : l'état cellulaire est mis à jour en combinant les sorties des portes d'entrée et d'oubli pour décider de l'information à conserver.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot c_t$$

Porte de sortie :

$$o = \sigma(W_o x_t + W_{o_s} s_{t-1} + b_o)$$

$$\tilde{s}_t = o_t \cdot \tanh(C_t)$$

Les équations capturent la nouvelle information à stocker dans l'état cellulaire et la sortie finale pour la prédiction. Où $W_{f_x}, W_{f_s}, W_{i_x}, W_{i_s}, W_{c_x}, W_{c_s}, W_{o_x}, W_{o_s}, b_f, b_i, b_c$ sont les poids et biais correspondants utilisés dans les différentes couches, et \tilde{s}_t désigne la sortie du réseau LSTM au signal temporel t .

Fonction Objective :

$$\min_{(W,b)} \sum_{t=1}^{t_{train}} \|s_t o_t \cdot \tanh(f_t \cdot C_{t-1} + i_t \cdot c_t)\|_F^2$$

L'objectif d'optimisation pour la plupart des problèmes de régression, tels que la prévision de séries temporelles, est une fonction de perte quadratique. Étant donné que la fonction de perte quadratique n'est pas robuste aux valeurs aberrantes dans le jeu de données, la fonction de perte absolue est également intégrée comme fonction objective de notre modèle. Les métriques telles que l'erreur quadratique moyenne (RMSE) sont utilisées pour minimiser la fonction de perte.

Limites des modèles d'apprentissage profond

Le principal inconvénient des modèles d'apprentissage profond est leur incapacité à bien performer lorsque les bases de données de séries temporelles sont limitées par la quantité de données disponibles, c'est-à-dire pour des ensembles de données de petite à moyenne taille. Dans de tels cas, les modèles d'apprentissage profond peuvent ne pas atteindre leur plein potentiel en termes de précision (Bandara et al., 2020).

Pour pallier cette limitation, l'application de stratégies d'augmentation de données (DA) a prouvé son efficacité dans divers domaines de l'apprentissage automatique. Par exemple, l'augmentation de données a considérablement amélioré les performances en classification d'images (Krizhevsky et al., 2012), reconnaissance vocale (Wang et al., 2020), classification de textes (Zhang et al., 2015), classification semi-supervisée générale (Donyavi & Asadi, 2020), et recherche liée aux séries temporelles (Kang et al., 2020; Talagala et al., 2022).

Les Données synthétiques

Les données synthétiques sont définies comme des données générées à l'aide d'un modèle mathématique ou d'un algorithme spécialement conçu, dans le but de résoudre une ou plusieurs tâches de science des données (Jordon et al., 2022). La synthétisation de données par la suite dans ce contexte permet de générer une quantité importante de données qui peuvent améliorer les performances des modèles de machine learning sans les biaiser, car ces données sont basées sur les propriétés des données réelles.

Génération des données synthétique à travers des réseaux de neurones artificiels (deep learning) :

La synthétisation de données à travers le deep learning est un domaine de recherche d'actualité, les réseaux de neurones, combinés à des techniques statistiques ou de l'apprentissage non supervisé permettent de générer des données réalistes et très proches des données originales, plusieurs modèles ont été mis en place ces dernières année (entre 2014 et 2024), nous allons présenter les 3 grandes familles :

— **Tabular Variational Auto-Encoder (VAE)** : un algorithme non supervisé qui peut apprendre la distribution d'un ensemble de données original et générer des données synthétiques via une double transformation, une première transformation à travers un réseau qui encode les données dans un espace avec des dimensions plus petites que les dimensions initiales (un espace latent) et un décodeur qui essaye de reconstruire la donnée originale depuis l'espace latent. Le modèle ajoute une erreur de reconstruction avant le décodage, qui peut être minimisée par un apprentissage itératif. Ce processus permet d'entraîner un réseau à deux blocs (un encodeur et un décodeur) qui peut générer des données.

- **Conditional Tabular Generative Adversarial Networks (CTGAN)** : Un réseau de neurones génératif développé par Xu et al. (2019) et inspiré de l'architecture des GANs (Generative Adversarial Networks) (Goodfellow et al., 2014). Le principe est d'entraîner un générateur G et un classificateur (appelé discriminateur D), le générateur essaie de générer des données aussi proches de la réalité afin de tromper le classificateur D. Ce processus permet de générer des données très réalistes et très semblables à la réalité. Les GANs ont révolutionné la génération de données en permettant de produire des échantillons de haute qualité dans des domaines tels que les images et les données.

— **Modèles de diffusion** : Un type de réseau de neurones inspiré de la thermodynamique et qui se base sur l'apprentissage non supervisé (Sohl-Dickstein et al., 2015). Ce type de modèle prend en entrée une donnée et ajoute une erreur gaussienne à cette dernière, à chaque itération le réseau rajoute une erreur plus importante jusqu'à ce que la donnée soit détruite et devient identique à un bruit aléatoire. Par la suite, le réseau essaiera de reconstruire la donnée initiale. Cet apprentissage lui permet de générer des données réalistes dans plusieurs formats.

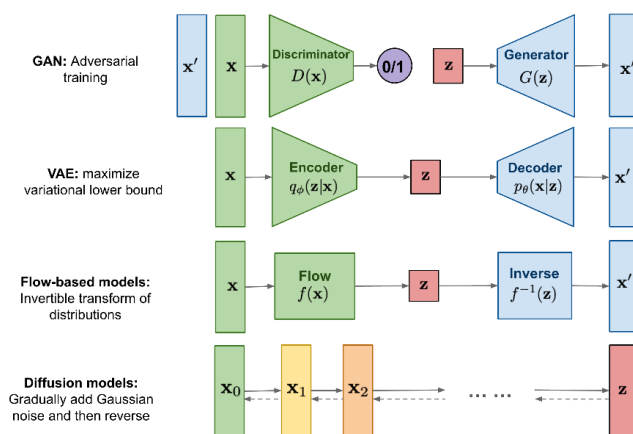


Figure 3.12 - Architecture des différents modèles de synthèse des données (Weng, 2021)

Chapitre 4 – Approche CRISP-DM pour la stabilisation des commandes

4 Approche CRISP-DM pour la stabilisation des commandes

Nous visons dans cette partie la stabilisation des commandes des pièces de rechange M&S afin d'avoir un contrôle optimal sur la gestion de stock et d'éviter des achats inutiles, et les ruptures de stock. Pour procéder à cette stabilisation, nous avons opté pour une approche de sélection des pièces nécessitant des commandes automatiques. Cette sélection nécessite l'application de modèles de clustering, et à cet effet, nous avons choisi d'utiliser des modèles non supervisés de Machine Learning.

Dans un premier temps, pour résoudre cette problématique, nous avons utilisé la démarche CRISP-DM, dont les étapes seront détaillées une à une dans cette partie.

La seconde partie de résolution consiste à l'évaluation des nouveaux stocks de sécurité des pièces.

4.1 Compréhension des Métiers

Pour procéder à notre analyse, les équipes de la division RP nous ont fourni l'historique de consommation en M&S provenant du système SAP de l'entreprise, via le HUB de SLB, concernant la division RP.

4.1.1 Répartition des Pièces M&S Prime

Nous avons d'abord analysé la répartition des pièces qui font partie du programme M&S Prime et qui sont soumises à des commandes automatiques par rapport à celles qui n'en font pas partie (Figure 4.1, où le groupe 1 désigne les pièces ayant des commandes automatiques). Cette analyse a mis en évidence une fluctuation significative du pourcentage de pièces couvertes par le M&S Prime et soumises à des commandes automatisées. Cela souligne l'importance de comprendre comment ces différentes catégories de pièces impactent la gestion des stocks et la planification de la demande.

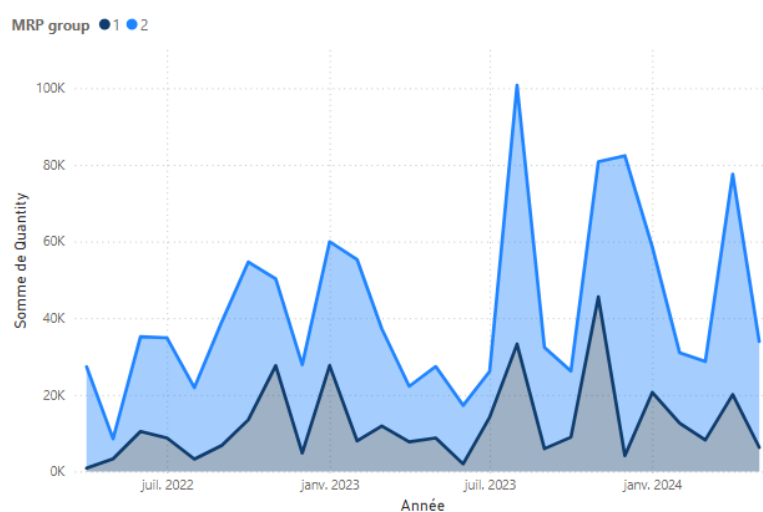


Figure 4.1 – Répartition des pourcentages des pièces inclus dans M&S Prime

Nous avons par la suite visualisé les pourcentages de pièces qui font partie du M&S Prime en fonction de la catégorie de leur comportement de consommation (Figure 4.2), les regroupant en catégories telles que

"runner" (pièces à forte rotation), "repeater" (pièces répétitives), "stranger" (pièces peu fréquentes) et "slow mover" (pièces à faible rotation). En comparant la distribution et en observant le manque de corrélation entre ces catégories et celles du M&S Prime, nous cherchons à développer une nouvelle classification spécifiquement adaptée aux pièces du M&S Prime.

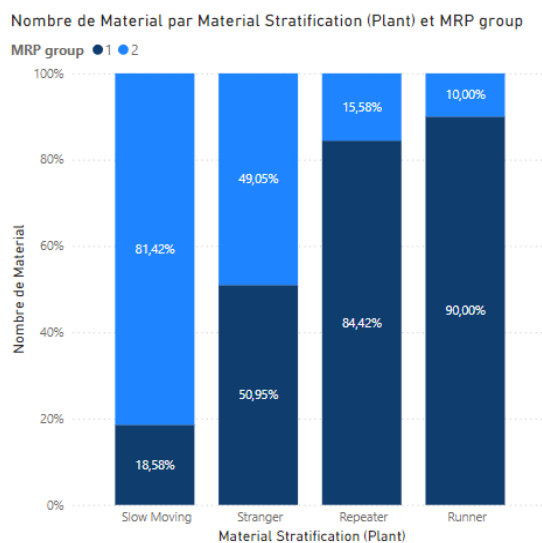


Figure 4.2 – Part des pièces inclus dans M&S Prime par catégorie de comportement

4.1.2 Objectifs de l'Analyse

Identification d'une Nouvelle Catégorisation des Pièces

- **Approche Holistique** : L'un des principaux objectifs est de développer une nouvelle catégorisation des pièces qui ne repose pas uniquement sur les données des six derniers mois. Cette approche holistique prend en compte l'historique complet de 32 mois, permettant ainsi une vision plus complète et précise des comportements de consommation.
- **Critères multiples** : Cette nouvelle catégorisation sera basée sur divers critères tels que la fréquence de consommation, les quantités utilisées, et les périodes d'inactivité, afin de mieux capturer les nuances dans l'utilisation des pièces.

Optimisation de la Gestion des Stock

- **Classification Avancée** : En élaborant une nouvelle classification des pièces du M&S Prime, nous visons à :
 - **Optimiser la Gestion des Stocks** : La nouvelle classification permettra d'améliorer l'allocation des ressources, de minimiser les stocks excédentaires et d'éviter les ruptures de stock.
 - **Réduire les coûts opérationnels** : Une gestion optimisée des stocks contribuera à diminuer les coûts liés aux stocks inutiles, y compris les coûts de stockage et de gestion, tout en évitant les achats en urgence à des prix plus élevés.
 - **Faciliter l'automatisation des commandes** : Grâce à une compréhension plus précise et adaptée des besoins réels, nous pourrions automatiser les commandes de manière plus efficace. Cela impliquera de mettre en place des systèmes d'approvisionnement automatisés qui tiennent compte des tendances de consommation à long terme et des variations saisonnières.

4.2 Compréhension des données

Après avoir importé les données couvrant une période de 32 mois, nous avons procédé à l'étape de feature engineering ou ingénierie des caractéristiques en extrayant deux principales caractéristiques pour chaque pièce :

- Somme des consommations mensuelles (Row Sum) : Cette variable représente la quantité totale consommée par la pièce sur les 32 mois.
- Nombre de mois d'utilisation : Indique combien de mois au cours des 32 mois il y a eu une consommation pour la pièce en question.

Nous avons ensuite calculé une troisième variable :

- Quantité moyenne utilisée : Obtenu en divisant la somme des consommations par le nombre de mois d'utilisation. Cette mesure permet de normaliser la quantité consommée en fonction de la durée pendant laquelle la pièce était active.

$$RS = \sum_{i=1}^{32} \text{Quantités sorties} \qquad NM = \sum_{i=1}^{32} [qté \neq null] \qquad AQ = \frac{RS}{NM}$$

Nous avons ensuite examiné les boxplots des deux caractéristiques afin de tenter d'extraire, à partir de ces deux critères, la distribution des variables décrivant les pièces. Nous constatons que la majorité des pièces ont un nombre de mois d'utilisation inférieur à 15 mois. Concernant la distribution de la quantité moyenne utilisée, nous observons que leur valeur est proche du zéro.

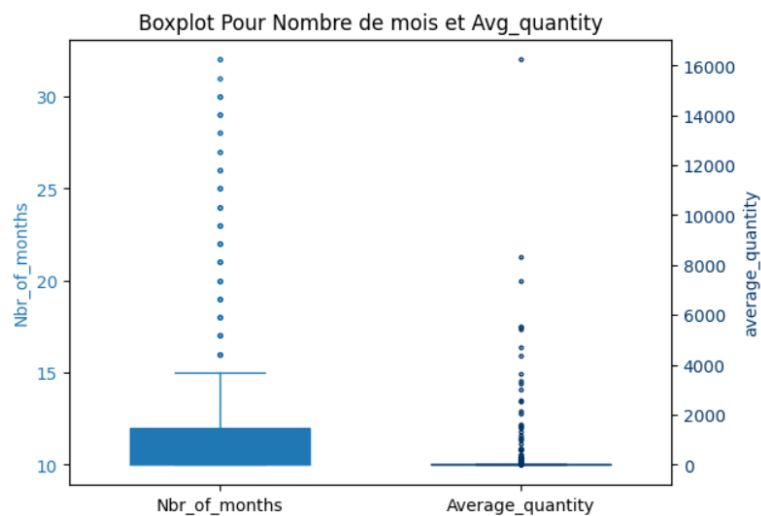


Figure 4.3 – Boxplot des caractéristiques

Par la suite, nous avons accompagné les boxplots d'un kernel density plot (ou tracé de la densité du noyau). On peut remarquer que pour les deux critères, il y a des pics où entre 40 % et 50 % des pièces tendent à avoir approximativement le même nombre de mois d'utilisation et la même quantité moyenne utilisée (Figure 4.5).

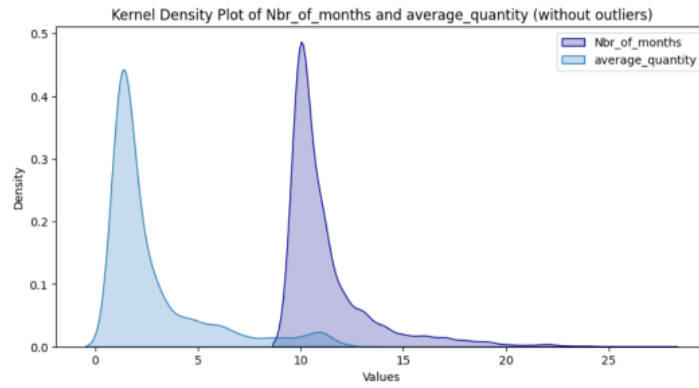


Figure 4.4 – Kernel Density plot des caractéristiques

Nous avons aussi évalué la corrélation mensuelle pour identifier d'éventuelles relations ou répétitions claires entre les différentes variables. Cela permet de déterminer si certains mois présentent des schémas de sortie répétitifs (Figure 4.5).

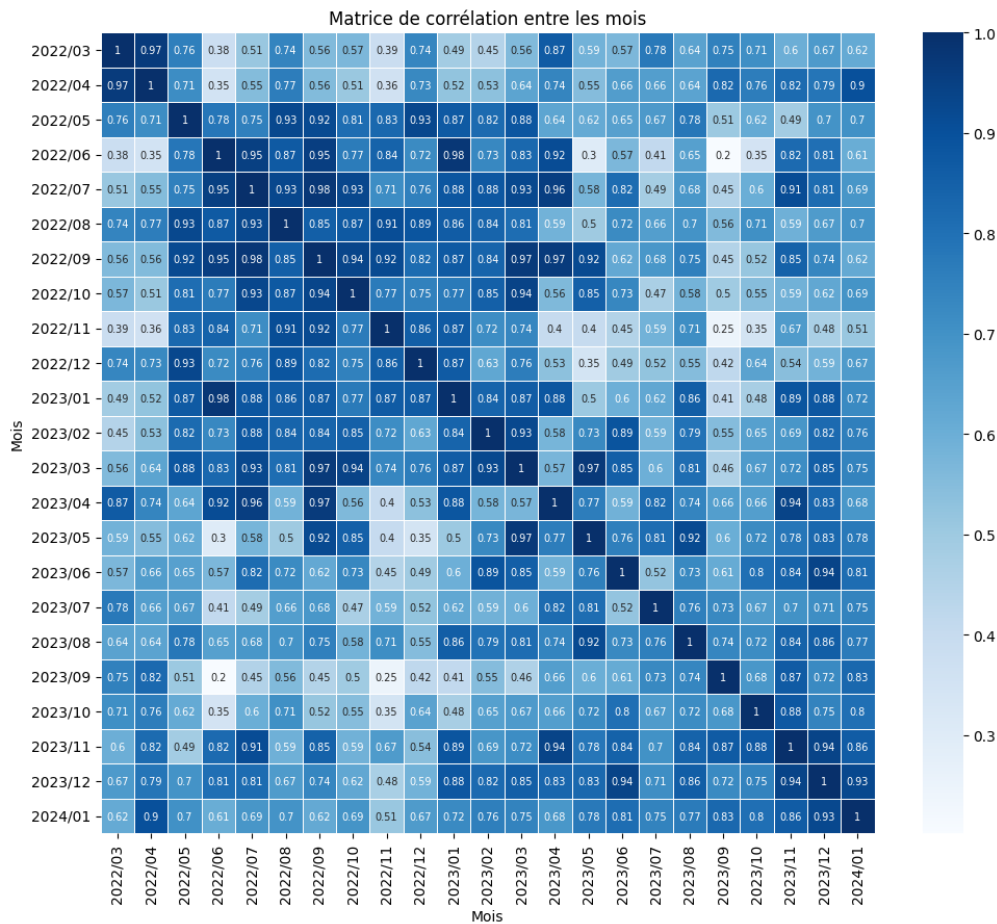


Figure 4.5 – Matrice de corrélation mensuelle des sorties des pièces

Test statistique de Hopkins

Dans le but de tester l'aptitude d'une base de données à être partitionnée en plusieurs groupes, nous avons opté pour test statistique de Hopkins qui permet de voir si les données suivent une distribution uniforme. Nous avons utilisé la fonction `hopkins()` prédéfinie dans la bibliothèque `pyclustertend` pour appliquer le test de Hopkins au jeu de données dont nous disposons. Nous avons obtenu les résultats suivants (Figure 4.6) :

Hopkins Test Result

```
resultat = hopkins(h_df)
print(f"La statistique de Hopkins pour la base de données est : {resultat:.4f}")
✓ 1.3s
La statistique de Hopkins pour la base de données est : 0.9981
```

Figure 4.6 – Capture d'écran du résultat du test de Hopkins

H est largement supérieur à 0.5, l'hypothèse H0 est donc rejetée, les données ne sont pas uniformément distribuées, elles ont donc tendance à être partitionnées en groupes.

4.3 Préparation de données

Pour garantir des modèles d'apprentissage performants, il est crucial de préparer soigneusement les données. Cette étape essentielle permet de filtrer et de conserver les informations les plus pertinentes qui seront utilisées comme variables d'entrée dans nos modèles de clustering ultérieurs.

4.3.1 Normalisation des données

Afin de prévenir toute distorsion causée par les quantités utilisées et pour que le clustering reflète principalement la fréquence d'utilisation, nous avons adopté une approche de normalisation. Nous avons divisé chaque colonne de données par sa moyenne mensuelle respective (Figure 4.7). En procédant ainsi, nous nous assurons que la fréquence d'utilisation influence le clustering de manière proportionnelle. Cela garantit une analyse équilibrée où chaque pièce est évaluée selon sa fréquence relative plutôt que selon la quantité absolue, assurant ainsi une représentation précise des données dans l'étude de clustering.

```
def divide_all_columns_by_column(df, divisor_column):
    # Avoid division by zero
    if (df[divisor_column] == 0).any():
        raise ValueError(f"'{divisor_column}' ")

    for col in df.columns:
        if col != divisor_column: # Excluding divisor column
            result_df[col] = df[col] / df[divisor_column]

    return result_df

df = divide_all_columns_by_column(df, 'average_quantity')
```

Figure 4.7 – Capture d'écran sur la normalisation des données

4.3.2 Standardisation des données

Nous avons choisi d'appliquer le MinMax Scaler pour standardiser les valeurs. Ce procédé a pour effet de mettre toutes les données à l'échelle de 0 à 1, préservant ainsi les distributions relatives tout en facilitant une comparaison équitable entre les différentes caractéristiques.

Pour accorder une importance accrue à la consommation des six derniers mois dans notre analyse, nous avons introduit un facteur de pondération ajusté $p = 1.5$. Cette approche assure que la période récente influence davantage le processus de clustering, tout en maintenant une échelle uniforme pour toutes les variables.

```

last_6_months_cols = df.columns[-7:]
remaining_cols = df.columns.difference(last_6_months_cols)

scaler = MinMaxScaler()
scaled_values = scaler.fit_transform(df)

df_scaled = pd.DataFrame(scaled_values, columns=df.columns)

# weights
last_6_months_weight = 1.5
remaining_cols_weight = 1

df_scaled[last_6_months_cols] = df_scaled[last_6_months_cols] * last_6_months_weight
df_scaled[remaining_cols] = df_scaled[remaining_cols] * remaining_cols_weight

```

Figure 4.8 – Capture d’écran sur le scaling avec pondération

Le code correspondant à ces standardisations est présenté dans la Figure 4.8. Il inclut la sélection des colonnes pertinentes, la normalisation des données à l'aide du MinMax Scaler, et l'application de poids différents aux périodes récentes et antérieures pour mieux aligner l'analyse avec les données récentes.

4.4 Modélisation des données

Après la fin des étapes de préparation des données, nous passons maintenant à la modélisation, nous allons présenter l'ensemble des algorithmes de clustering utilisés afin de pouvoir sélectionner celui qui donne le meilleur résultat.

4.4.1 Application du K-Means

Pour déterminer précisément le nombre de clusters requis par l'algorithme KMeans, nous avons utilisé deux méthodes complémentaires :

Tout d'abord, nous avons exploré divers nombres de clusters et évalué le coefficient de silhouette pour mesurer la cohérence des clusters, comme détaillé dans la figure 4.9.

Ensuite, nous avons confirmé cette analyse en appliquant la méthode du coude, qui consiste à examiner la variation de l'inertie en fonction du nombre de clusters, illustrée dans la figure 4 ;10. Cette approche a mis en évidence que le nombre optimal de clusters est de 3, confirmé par une cassure nette dans le coude et un coefficient de silhouette de 0.81 pour cette configuration.

Une fois le nombre de clusters déterminé, nous avons initialisé le modèle KMeans de la bibliothèque Sklearn avec 3 clusters, entraîné le modèle sur nos données et utilisé la fonction predict() pour obtenir les labels (étiquettes) des clusters sous forme d'un tableau numpy.

k	silhouette_score
2	0.53686
3	0.57328
4	0.64806
5	0.68975
6	0.72054
7	0.77585
8	0.78582
9	0.79177
10	0.81715
11	0.84002

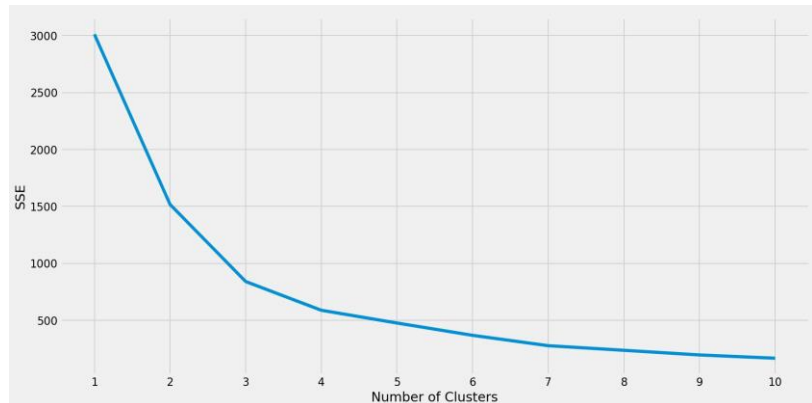


Figure 4.9 – Les coefficients de silhouette en fonction du nombre de clusters

Figure 4.10 – Variation de l’inertie en fonction du nombre de clusters

4.4.2 Classification Ascendante Hiérarchique

Nous avons appliqué l'algorithme Agglomerative Clustering de la bibliothèque Scikit-Learn pour effectuer une Classification Ascendante Hiérarchique (CAH). Nous avons spécifié le nombre de clusters préalablement déterminé et avons choisi d'utiliser la distance euclidienne comme mesure de similarité entre les données. Après l'entraînement du modèle, nous avons procédé à la visualisation du dendrogramme à l'aide de la fonction dendrogram du module Scipy.

Le dendrogramme obtenu (Figure 4.11) illustre les étapes de fusion des clusters tout au long de l'analyse. Chaque branche du dendrogramme représente une fusion de clusters à une certaine distance, permettant ainsi de visualiser la hiérarchie des regroupements de nos données. Cette représentation graphique nous a permis de mieux comprendre la structure intrinsèque des données et d'identifier des seuils significatifs pour la découpe en clusters, en fonction des distances de fusion observées.

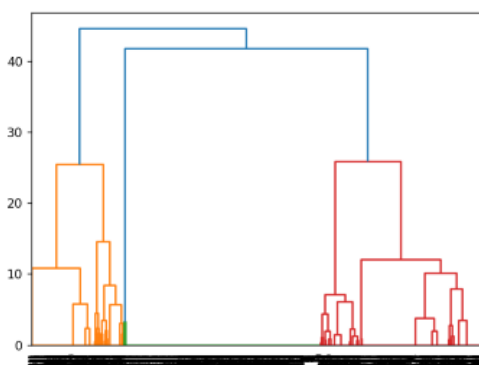


Figure 4.11 – Le dendrogramme

k	silhouette score
2	0.51293
3	0.58494
4	0.66397
5	0.68978
6	0.69075
7	0.72464
8	0.73726
9	0.75487
10	0.82433
11	0.83342

Figure 4.12 – Les valeurs de silhouette pour différentes k de CAH.

La décision de former 5 clusters par Classification Ascendante Hiérarchique repose sur deux critères clés : le Silhouette Score (Figure 4.12) et l'analyse du dendrogramme (Figure 4.11). Le Silhouette Score évalue la qualité de la séparation des clusters, tandis que le dendrogramme illustre visuellement les fusions de

clusters à différentes distances. Ensemble, ces critères ont confirmé que le nombre de **5 clusters** était optimal pour notre analyse, assurant une segmentation claire et pertinente des données.

4.4.3 Application du MiniBatch

Pour entraîner l'algorithme Minibatch K-Means, une variante efficace de K-Means adaptée au traitement de grands ensembles de données, nous avons simplement précisé le nombre de clusters requis. Pour déterminer ce nombre de clusters optimal, nous avons utilisé la méthode du coefficient de silhouette. Cette méthode a recommandé l'utilisation de 4 clusters, avec un coefficient de silhouette mesuré à 0.82.

4.4.4 Application du DBSCAN

Pour appliquer l'algorithme de DBSCAN, il faudrait introduire deux paramètres d'entrée parmi eux "eps" qui représente la distance entre deux points du jeu de données, celle-ci sera prise en considération par l'algorithme dans la construction des groupes de données. Afin de choisir le bon epsilon nous avons utilisé la distance euclidienne :

Soit : $p = (p_1, \dots, p_n)$ et $q = (q_1, \dots, q_n)$: $d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_i - q_i)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$.

Afin d'optimiser eps, nous avons calculé pour chaque point la distance avec son voisin le plus proche, par la suite on a fixé un eps tel que :

- Une grande partie des points à une distance inférieure à eps.
- Une grande partie des points à au moins un voisin dans son eps-voisinage.

La bibliothèque Scikit-Learn met à disposition les méthodes de la classe Nearest Neighbors qui permettent de déterminer les voisins les plus proches ainsi que les distances que l'eps choisi est de 0.2.

Le modèle estimé 8 clusters et 316 points de bruit (Figure 4.13).

```
Estimated number of clusters: 8
Estimated number of noise points: 316
Silhouette Coefficient: 0.720
Calinski-harabasz : 515.820
Indice de Davies-Bouldin : 0.892
```

Figure 4.13 – Capture sur résultat du DBSCAN

4.4.5 Application de l'algorithme MeanShift

Avant d'appliquer MeanShift, il est crucial de déterminer un paramètre appelé "bande passante" (bandwidth). Cette bande passante définit la portée spatiale autour de chaque point, influençant la taille des clusters obtenus. La fonction `estimate_bandwidth()` est souvent utilisée pour estimer cette valeur à partir des données. MeanShift commence par choisir des centres de clusters initiaux dans l'espace des données. Ensuite, pour chaque point de données, il calcule la moyenne des points situés à une distance inférieure ou égale à la bande passante. Ce processus déplace itérativement chaque centre de cluster vers une région où la densité de points est maximale.

Ce processus a permis de générer automatiquement 11 clusters en fonction des densités observées dans les données, offrant ainsi une segmentation adaptée des pièces de rechange selon leurs caractéristiques communes (Figure 4.14).

```
number of estimated clusters : 11
Silhouette Coefficient: 0.597
Calinski-harabasz : 1193.781
Indice de Davies-Bouldin : 0.728
```

Figure 4.14 – Capture sur les résultats du MeanShift

4.5 Evaluation des modèles

Afin d'évaluer la performance des modèles que nous avons obtenu en utilisant les 5 algorithmes (K-Means, Minibatch, CAH, DBSCAN et Meanshift), nous avons utilisé les trois métriques expliquées déjà au niveau du Chapitre X, ces derniers sont :

- Coefficient de silhouette.
- Indice de Davies-Bouldin.
- Indice de Calinski Harabasz.

4.5.1 Résultat de chaque modèle par rapport aux métriques

Ces indices sont déjà implémentés dans le module Python sklearn.metrics, nous les avons donc utilisés en instanciant les objets suivants : silhouette_score, calinski_harabasz_score et davies_bouldin_score.

Le tableau suivant regroupe les scores des cinq algorithmes par rapport aux trois critères d'évaluation.

	Silhouette	Calinski-harabasz	Davies-Bouldin
kmeans	0.641	2073.865	0.728
Minibatch	0.629	1635.762	0.795
CAH	0.69	2057.963	0.676
DBSCAN	0.882	989.126	0.772
Meanshift	0.597	1193.781	0.795

Figure 4.15 – Scores des algorithmes par rapport aux métriques d'évaluation

- **Utilisation d'une méthode d'aide multicritère à la décision (AMD)**

Concernant les évaluations et les comparaisons entre les résultats pour les trois critères, il faut noter que le Score Silhouette et le Score Calinski-Harabasz sont des critères à maximiser, tandis que le score de Davies-Bouldin est un critère à minimiser.

Ensuite, afin de choisir le modèle le plus performant vis-à-vis des trois critères, nous avons utilisé la méthode d'aide multicritère à la décision de Borda, où nous avons attribué un poids à chacun des classements. Le poids du premier classement est de 5 et celui du cinquième est de 1. Le modèle obtenant le plus de points est considéré comme le meilleur. Les résultats de cette analyse sont résumés dans le tableau suivant :

Indicateur/ Algorithme	Silhouette		Calinski-harabasz		Davies-Bouldin		Note
	Score		Score		Score		
kmeans	0.641	2	2073.865	5	0.728	4	11
Minibatch	0.629	3	1635.762	3	0.795	1	7
CAH	0.69	4	2057.963	4	0.676	5	13
DBSCAN	0.882	5	989.126	1	0.772	3	9
Meanshift	0.597	1	1193.781	2	0.795	1	4

Figure 4.16 – Affectation des notes selon le classement

L'algorithme CAH possède le score le plus élevé (13 points) suivi du kmeans avec 11 points, nous pouvons visualiser les scores à travers le diagramme suivant :

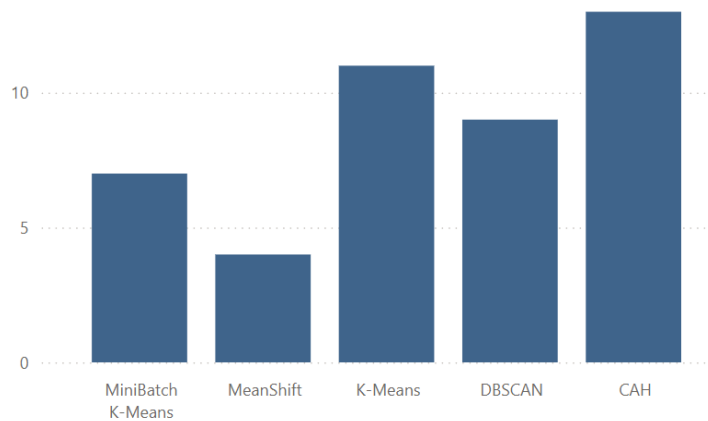


Figure 4.17 – Visualisation des résultats de l'évaluation

L'algorithme CAH donne des résultats très convaincants sur les trois indices, nous l'avons donc maintenu afin de l'utiliser dans le déploiement de l'outil d'aide à la décision. La visualisation des différents clusters est comme suit :

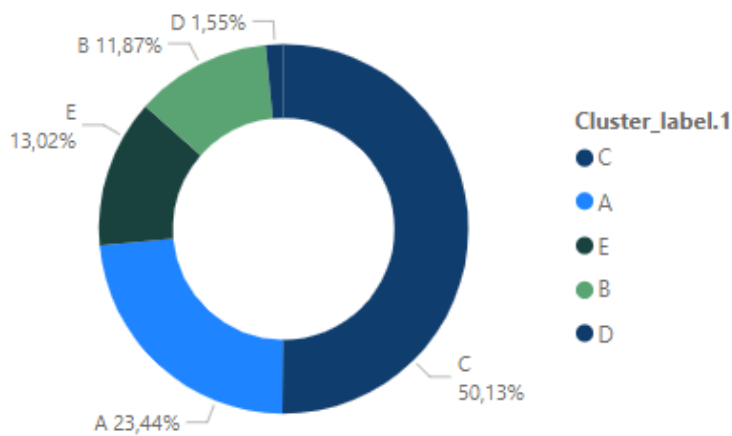


Figure 4.18 – Répartition des clusters par l'algorithme CAH

4.5.2 Analyse des catégories :

En examinant les résultats de la nouvelle classification par rapport à l'historique de consommation de notre étude, nous avons d'abord analysé l'évolution des sorties en termes de quantité par catégorie (Figure 4.19). Il est notable que la catégorie B, bien que représentant seulement 12 % des pièces, correspond à 86 % des quantités consommées au cours des deux derniers trimestres.

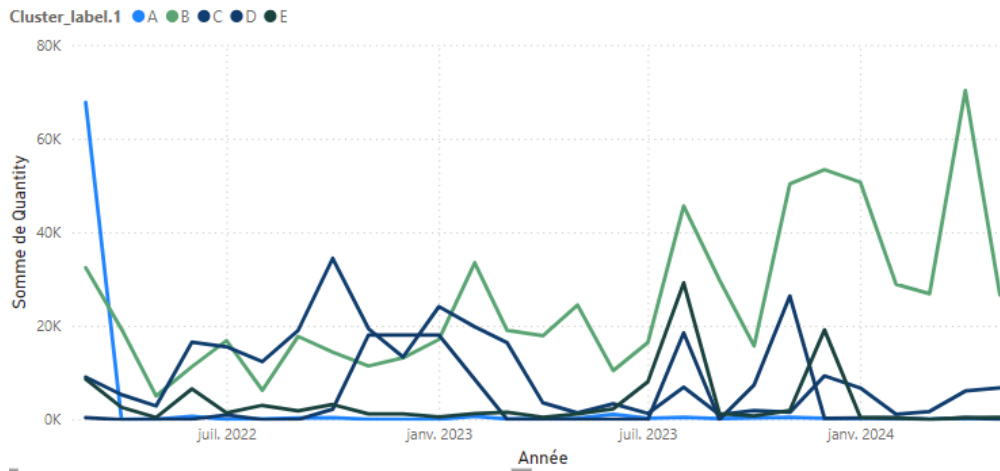


Figure 4.19 – Evolution de consommation des pièces par rapport à la nouvelle classification

Nous avons également observé, à travers la figure 4.20, que 37 % des pièces classées dans la catégorie B ne font pas partie du M&S Prime, qui est responsable de l'automatisation des commandes.

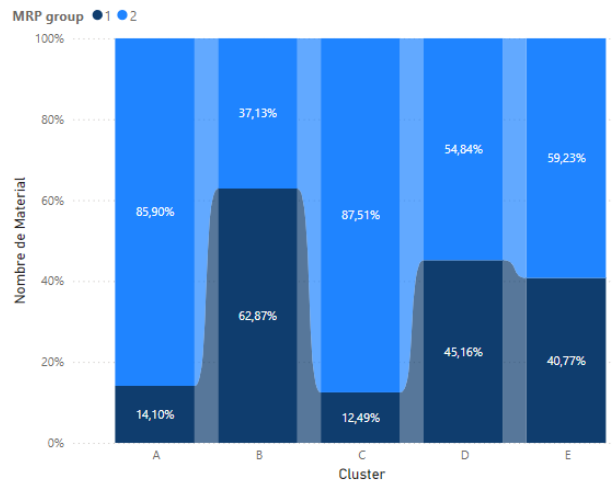


Figure 4.20 – Degré d'automatisation par nouvelles classes

À travers la figure 4.21, qui illustre l'historique de consommation en fonction de leur inclusion dans l'outil d'automatisation, nous pouvons observer que la catégorie B présente une augmentation considérable de consommation pour les pièces non incluses dans l'automatisation. Cela souligne la nécessité d'intégrer ces pièces dans le système M&S Prime.

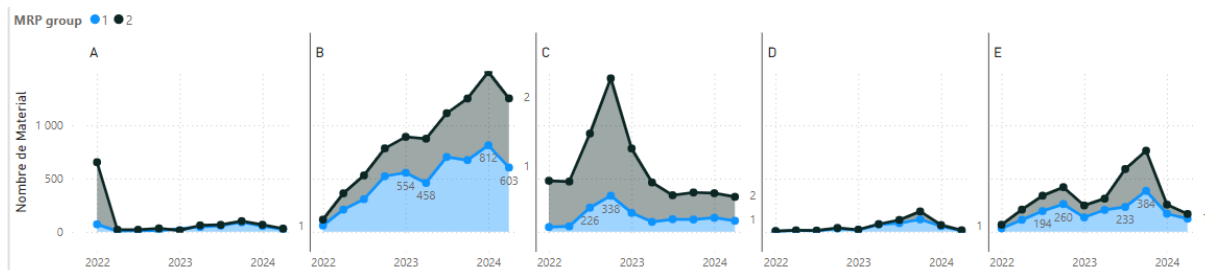


Figure 4.21 – Evolution de par classes et par niveau d'automatisation

4.6 Déploiement

4.6.1 Réévaluation des quantités de stock de sécurité :

La réévaluation des quantités de stock de sécurité (Safety Stock) pour les pièces de rechange est une étape importante dans la gestion efficace des stocks, visant à aligner les niveaux de stock avec les besoins réels. Cette démarche a été entreprise à la suite d'une observation indiquant que les seuils actuels étaient souvent bien supérieurs aux besoins réels.

Méthodologie de réévaluation

La méthode de réévaluation s'est basée sur l'analyse de plusieurs données clés, à savoir :

Paramètres de quantité de sortie mensuelles : Nous avons déjà examiné la distribution de densité de la quantité de sorties mensuelles par pièce.

- Minimum/Maximum de sortie mensuelles ($\text{Min}(N.\text{Sortie}/\text{Mois})$, $\text{Max}(N.\text{Sortie}/\text{Mois})$) : Correspond aux quantités minimales et maximales sorties pour une pièce donnée sur une période de 32 mois, représentant respectivement le mois où la consommation a été la plus faible (min) et le mois où elle a été la plus élevée (max).
- Moyenne de sortie mensuelle ($\text{Moyenne}(N.\text{Sortie}/\text{Mois})$) : La moyenne des sorties mensuelles pour la même pièce.

Paramètres de Délai de livraisons : Dans le cas des délais de livraison, les valeurs minimales et maximales peuvent être très influencées par des événements exceptionnels ou des anomalies dans les données de délais de livraison. En revanche, les quantiles q25 et q75 représentent respectivement les délais au 25e et 75e percentiles de la distribution des délais. Ils sont moins sensibles aux valeurs extrêmes et offrent une meilleure représentation de la tendance centrale des délais observés. La figure représente un boxplot segmenté par catégorie, mettant en évidence la répartition des délais de livraison où l'on peut observer la variabilité et la distribution des délais, incluant les médianes, les percentiles et les valeurs extrêmes pour chaque catégorie.

- Percentile 25 du Délai de Livraison (q25) : Le délai de livraison associé au 25e centile des commandes de pièces.
- Percentile 75 du Délai de Livraison (q75) : Le délai de livraison associé au 75e centile des commandes de pièces.

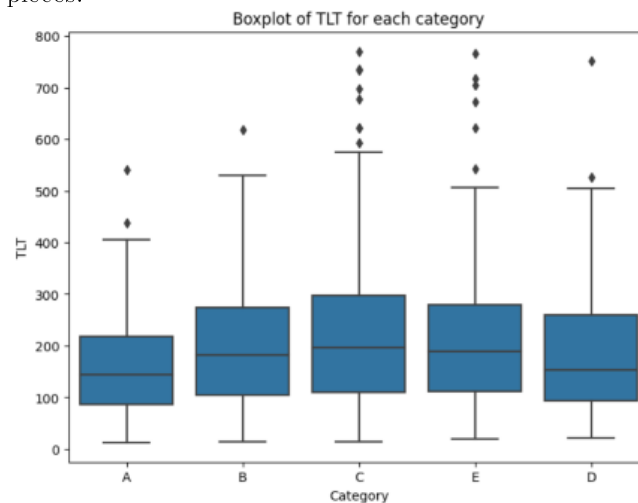


Figure 4.22 – Boxplot des lead time pour chaque catégorie

Les Équations utilisé pour la réévaluation

La réévaluation des quantités de stock de sécurité (Safety Stock) repose sur des équations spécifiques conçues pour ajuster les niveaux de stock en fonction des données historiques de sortie mensuelle et des délais de livraison observés.

Quantité Minimale	Quantité Maximale
$Q_{min} = \left(\frac{\text{Min}(N.\text{Sortie}/\text{Mois}) + \text{Moyenne}(N.\text{Sortie}/\text{Mois})}{2} \right) * \left(\frac{d_{q25} + d_{Moy}}{2} \right)$	$Q_{max} = \left(\frac{\text{Max}(N.\text{Sortie}/\text{Mois}) + \text{Moyenne}(N.\text{Sortie}/\text{Mois})}{2} \right) * \left(\frac{d_{q75} + d_{Moy}}{2} \right)$
La formule combine le minimum de sorties mensuelles avec la moyenne des sorties. Le résultat est ensuite multiplié par une moyenne des délais de livraison au 25 ^{ème} percentile et du délai moyen, ce qui permet de calculer un niveau de stock minimum plus représentatif des besoins courants.	
Quantité de Stock de sécurité	
$Q_{ss} = \frac{Q_{min} + Q_{max}}{2} = \left(\frac{\text{Max}(N.\text{Sortie}/\text{Mois}) + \text{Min}(N.\text{Sortie}/\text{Mois}) + \text{Moyenne}(N.\text{Sortie}/\text{Mois}) * 2}{4} \right) * \left(\frac{d_{q25} + d_{q75} + 2 * d_{Moy}}{4} \right)$	
La formule finalisée du stock de sécurité combine les deux formules afin de déterminer des niveaux plus représentatifs des besoins réels. Elle intègre les variations et les tendances des délais de livraison ainsi que des sorties de stock mensuelles.	

Tableau 4.1 – Méthode de calcul du stock de sécurité

Résultats des nouveaux stocks de sécurité proposés

Après avoir implémenté les équations proposées, nous avons pu recalculer les nouveaux niveaux de stock pour les produits concernés. En comparant avec les anciens niveaux de stock proposés, la nouvelle méthode a permis de réduire l'utilisation de 2869 unités de stock soit une réduction du coût global de l'inventaire de 30,11 %.

Material	Safety Stock	Current Grouping	New Grouping	Min Qte/Month	Max Qte/Month	Avg Qte/Month	Dq25	Dq75	Dm	Qmin	Qmax	Qss
100213000	5 000,000	Repeater	B	0	24548	2826	3	4	4	1416	13691	7554
E027910	1 296	Stranger	B	0	2592	648	4	5	5	328	1625	977
H543851	5 000,000	Repeater	B	0	14500	886	5	15	11	451	7706	4079
E016031	400,000	Stranger	B	0	600	100	3	7	5	54	356	205
H400299	500,000	#N/A	B	0	1050	158	3	5	4	82	608	345
H350729	600,000	#N/A	B	0	1200	135	4	6	5	72	673	372
H125979	200,000	#N/A	B	0,5	464	60	3	4	3	33	266	150
B013113	100,000	Repeater	B	0	280	24	3	6	5	16	157	86
B012416	100,000	Stranger	B	0	375	36	3	5	4	21	210	116
B017168	100,000	Slow Moving	B	0	100	18	6	6	6	14	64	39
H014201	100,000	Slow Moving	B	0	220	47	4	4	4	28	138	83
H125975	200,000	#N/A	B	0	400	73	3	5	4	40	241	140
B012764	100,000	Repeater	B	0	370	34	2	4	3	19	205	112
0078410	100,000	Slow Moving	B	0	160	25	3	7	5	17	98	58
B036200	100,000	Slow Moving	B	0	160	18	3	6	4	13	94	54
101760981	100,000	#N/A	B	7,25	176	31	3	4	4	22	107	65
0078519	200,000	#N/A	B	0	500	80	3	5	5	44	295	169
H400311	200,000	#N/A	B	0	295	54	2	4	4	30	178	104
B015547	80,000	#N/A	B	0	110	27	3	5	4	17	73	45
44000-009-00	80,000	#N/A	B	0	222	21	2	3	2	12	124	68
H712973	50,000	#N/A	B	0	100	15	3	4	3	11	61	36
100746089	80,000	#N/A	B	0	140	26	3	9	6	18	90	54
B034904	100,000	#N/A	B	0	100	22	2	3	2	13	63	38
B036206	75,000	#N/A	B	0	140	30	3	4	4	19	89	54
B024202	100,000	Repeater	B	0	293	31	3	4	4	18	166	92
H400313	100,000	#N/A	B	0	298	25	3	4	4	16	165	90
H432128	50,000	Stranger	B	0	175	39	2	4	3	22	110	66

Figure 4.23 – Capture d'écran du calcul du nouveau stock de sécurité

4.6.2 Tableau de bord

Afin de mieux illustrer la nouvelle classification des pièces de rechange M&S comparé à l'ancienne utilisée par SLB. Nous avons opté vers l'utilisation d'un tableau de bord en utilisant l'outil PowerBI.

L'objectif principal de la réalisation de ce tableau de bord BI est de mettre en lumière les niveaux de stock entre l'ancienne classification et la nouvelle, les coûts associés, les délais de livraison, et les ordres de commande. Ce tableau de bord permettra de visualiser ces informations sur une période donnée, facilitant ainsi la prise de décisions informées concernant la gestion des stocks et les processus d'approvisionnement. De plus, ce tableau de bord vise à résoudre le problème initial des commandes automatiques de pièces non nécessaires, en utilisant une nouvelle classification pour déterminer quelles pièces doivent être automatisées ou non.

Pour commencer, nous allons examiner la conception d'une base de données relationnelle. Pour cela, nous avons les tables suivantes :

- Table SOH : Contient les données sur le stock disponible sur site, extraites automatiquement de SAP, associées à des dates spécifiques.
- Table PO : L'historique des Purchase Orders (PO) ou bons de commandes passés, y compris les détails sur les fournisseurs et les produits commandés, est extrait de SAP Ariba.
- Table CR : Enregistre les résultats du clustering effectué sur les pièces M&S, incluant les nouvelles classifications des produits générées par l'algorithme de la CAH, ainsi que les résultats du calcul du stock de sécurité pour les nouvelles pièces.
- Table M&S : Liste des produits faisant partie de M&S Prime avant la nouvelle catégorisation, pour comparaison des résultats.
- Table Consommation : Enregistre l'historique de consommation quotidienne des pièces.
- Table Lead Time : Qui représente l'historique des délais de livraison (lead time) pour les pièces de la division RP et qui permet d'obtenir le lead time moyen des pièces.

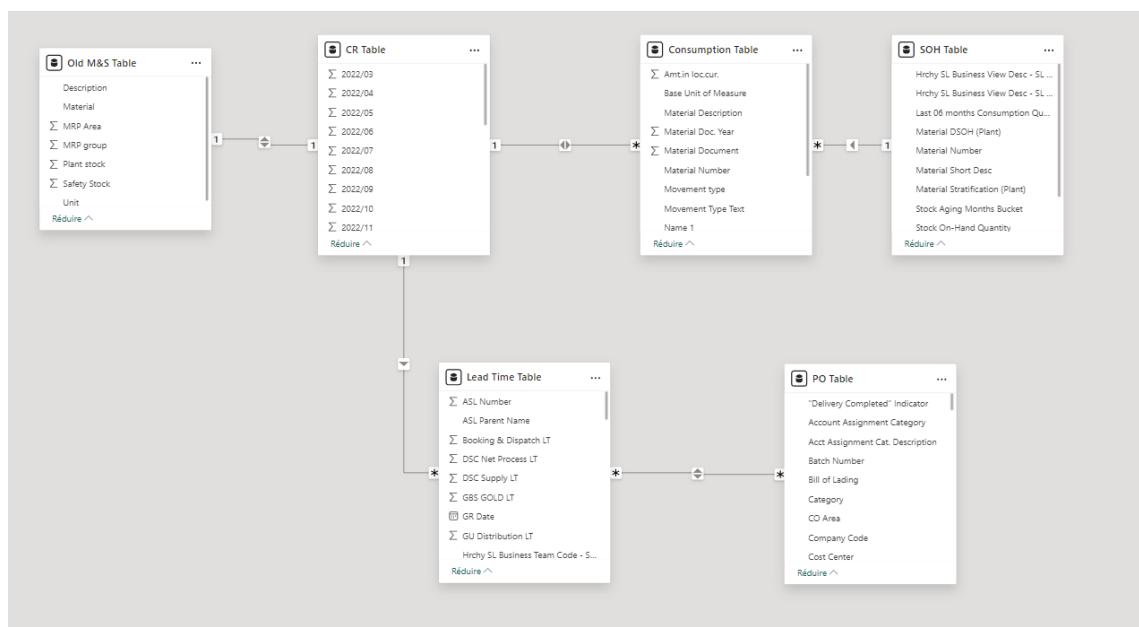
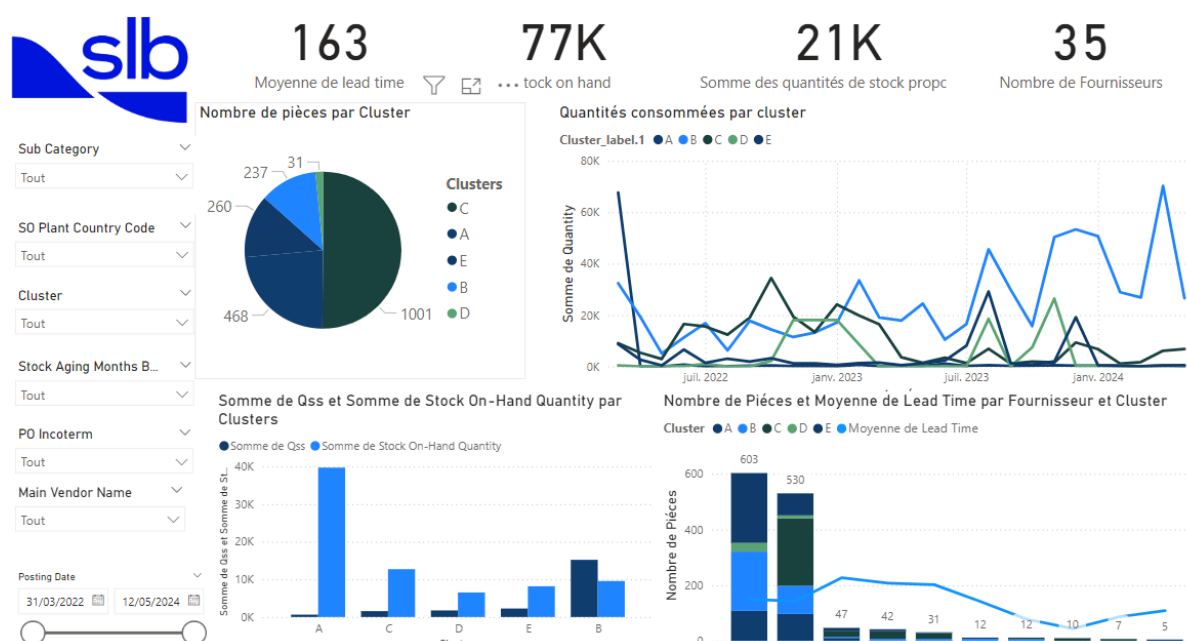


Figure 4.24 – Aperçu du modèle relationnel établi pour la conception du tableau de bord

Dans la conception de notre tableau de bord, nous avons choisi d'utiliser un modèle tabulaire de données en raison de la structure des données extraites des systèmes SAP chez Schlumberger. Notre objectif est

d'illustrer de manière efficace les nouvelles classes de pièces, en mettant l'accent sur leurs caractéristiques de disponibilité et de délais de livraison. Ce modèle nous permet d'organiser les informations de manière optimale, d'établir des relations pertinentes entre les données déjà existantes, assurant ainsi une analyse détaillée et précise des pièces nouvellement classifiées.

Remarque : Dans la mise en œuvre de notre modèle tabulaire, nous avons rencontré des défis liés à l'intégration d'une table de dates, particulièrement complexe en raison de sa connexion avec plusieurs autres tables simultanément. Une table date joue un rôle essentiel dans la modélisation des données et l'analyse temporelle en apportant une segmentation temporelle, un support pour les agrégations et une intégration des informations des différentes dates. Malgré ces défis, l'approche tabulaire nous offre une structure flexible et performante pour analyser efficacement les nouvelles classes.



La figure 4.25 illustre le tableau de bord Power BI réalisé sur l'état de stock des pièces de rechange M&S. Ce tableau de bord permet d'illustrer le résultat de la nouvelle segmentation des pièces ainsi que différents KPI relatifs à ces pièces, tels que le lead time, le nombre de fournisseurs, les quantités en stock et celles proposées à travers le nouveau calcul des stocks de sécurité, tout en montrant les variations de ces différents indicateurs. Des filtres temporels ainsi que d'autres filtres permettent d'affiner l'analyse pour obtenir des informations plus précises et pertinentes.

À travers l'analyse effectuée sur les pièces de rechange M&S, nous avons proposé une nouvelle segmentation visant d'une part à corriger la classification actuelle, qui est loin de la réalité du terrain, et d'autre part à orienter la proposition des pièces à inclure dans le système d'automatisation. De plus, nous avons introduit une nouvelle mesure des stocks de sécurité pour les pièces qui seront incluses dans ce système. Enfin, nous avons développé un tableau de bord permettant le suivi de ces pièces selon la nouvelle segmentation.

Chapitre 5 – Engagements
contractuels : Approche
prévisionnelle

5 Engagements contractuels : Approche prévisionnelle

Dans le cadre de l'amélioration du contrat réalisé avec un fournisseur de pièces de rechange automobile local, nous visons à améliorer les prévisions de consommation de ces pièces et cela dans le but de stabiliser le contrat réalisé en termes de quantité dans la perspective de réaliser des économies futures avec le fournisseur.

L'analyse des prévisions chez SLB montre qu'ils n'atteignent pas encore le niveau de performance souhaité. C'est pourquoi nous avons jugé nécessaire de les rendre plus fiables en utilisant des méthodes quantitatives capables de gérer la complexité de la demande, en particulier pour les produits M&S.

5.1 Étude prévisionnelle de la demande

Afin de répondre au besoin de stabilisation des contrats soulevé dans notre premier chapitre, nous nous sommes intéressés à l'historique de consommation des pièces de rechange du fournisseur choisi. Pour analyser et comprendre la consommation de ses pièces les plus critiques, une étude prévisionnelle est nécessaire.

Au cours de cette partie, nous débuterons par la définition du champ d'application de l'étude prévisionnelle en appliquant une classification des pièces ayant le plus fort besoin. Nous poursuivrons avec une prévision de la demande pour ces pièces, en utilisant d'abord des modèles statistiques, puis des outils de Machine Learning et de réseaux de neurones artificiels. Enfin, nous validerons les modèles à l'aide d'indicateurs de performance adaptés.

5.1.1 Préparation des données pour les prévisions

Pour procéder à notre analyse, nous avons extrait l'historique de consommation des pièces de rechange M&S du fournisseur concerné de la division Reservoir Performance (RP), issue du système d'information utilisé au sein de SLB qui est SAP. Les données récoltées représentaient la consommation de 57 pièces du fournisseur sur une période s'étalant sur 32 mois, contenant la quantité consommée par jour tout au long de cette période. Pour des fins d'adaptabilité des données aux outils prévisionnels, nous avons regroupé les données par mois.

De plus, étant donné que les données disponibles se limitent à l'historique de consommation des produits M&S, avec une seule variable, à savoir la quantité consommée au fil du temps, nous nous orientons vers l'utilisation des modèles auto-projectifs pour notre étude prévisionnelle, plutôt que d'autres types de modèles.

Classification des pièces

Afin d'agir sur les pièces de rechange les plus influentes ayant le plus d'importance sur la Supply Chain, nous avons procédé à une analyse ABC XYZ. Cette analyse nous permettra d'un côté, d'illustrer l'application de notre étude prévisionnelle et de l'autre côté d'assurer un engagement avec le fournisseur

sur les pièces les plus urgentes et ainsi accroître l'efficacité des contrats tout en réduisant l'horizon de contrôle.

Dans un premier temps, la classification ABC, permettra de diviser les pièces en catégories basées sur différents critères donnant la classification suivante :

- Classe A : 20% des produits ayant 80% d'impact
- Classe B : 30% des produits ayant 15% d'impact
- Classe C : 50% des produits ayant 5% d'impact

Dans un second temps, la classification XYZ, permet d'illustrer et de diviser les pièces en catégorie selon la variation de consommation de ces dernières, donnant les classes suivantes :

- Classe X : de 0% à 10% des pièces dont la demande peut être estimée avec précision
- Classe Y : 10 % à 25 % pour les pièces dont la demande peut être relativement bien prédite
- Classe Z : plus de 25 % des pièces dont la demande peut être prédite avec très peu de précision

Pour la méthode ABC, il est important de définir le ou les critères de classification. Dans ce volet, nous avons opté pour deux critères de classification, qui sont la valeur moyenne des quantités consommées (en multipliant le prix de la pièce avec sa consommation moyenne) d'une part et de la fréquence de consommation d'une autre durant la période d'étude qui est de 32 mois. Nous avons choisi ces critères-là dans le but de sélectionner les pièces ayant le plus d'impact s'ils subissent des ruptures de stock.

Afin de simplifier l'analyse et d'améliorer l'efficacité, nous avons combiné les deux critères, à savoir, la moyenne des quantités et la fréquence de consommation, en un seul nouveau critère. Pour ce faire, nous avons multiplié pour chaque pièce les deux critères rapportés au maximum enregistré pour toutes les pièces.

$$\text{Critère de classification} = \frac{\text{Valeur Moyenne}}{\text{Max des Valeurs}} * \frac{\text{Fréquence de consommation}}{\text{Max des Fréquences}}$$

Dans l'analyse XYZ, le classement est effectué selon le critère de variabilité de la consommation par rapport à la consommation moyenne. Il est nécessaire de déterminer le coefficient de variation, qui est calculé comme le rapport entre l'écart type et les moyennes. Le coefficient de variation est une mesure relative de la dispersion de la distribution de probabilité comme suit :

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} ; CV = \frac{\sigma}{\bar{x}}$$

L'approche intégrée ABC-XYZ est utilisée pour déterminer les activités pour chacun des groupes d'articles définis dans une matrice de comparaison par paires. Le tableau suivant résume les caractéristiques des neuf différentes classes de matériaux après avoir combiné l'analyse ABC avec l'analyse XYZ.

	A	B	C
X	Haute valeur, demande continue à haute prévisibilité	Valeur moyenne, demande continue à haute prévisibilité	Faible valeur, demande continue à haute prévisibilité
Y	Haute valeur, demande fluctuante à prévisibilité moyenne	Valeur moyenne, demande fluctuante à prévisibilité moyenne	Faible valeur, demande fluctuante à prévisibilité moyenne
Z	Haute valeur, demande irrégulière à faible prévisibilité	Valeur moyenne, demande irrégulière à faible prévisibilité	Faible valeur, demande irrégulière à faible prévisibilité

Tableau 5.1 – Matrice de la méthode ABC-XYZ

Classification ABC :

Après le calcul de la moyenne de consommation pour chaque pièce ainsi que la fréquence de consommation, nous calculons le nouveau critère de classification défini précédemment. On les classe par ordre décroissant et calculons la somme cumulée ainsi que son pourcentage. On finit par attribuer les classes selon les pourcentages définis précédemment.

N°	Numéro de la pièce	Moyenne de consommation	Valeur	Fréquence	Critère de classification	Pourcentage du critère	Pourcentage du critère cumulé	Classe ABC
1	A100917567	5572,80	96537,85	30	1	0,39870	0,39870	A
...
5	H432650	1546,74	29706,09	23	0,21279	0,08484	0,91367	A
6	E023908	1099,17	3784,76	12	0,07890	0,03146	0,94513	B
...
19	B011975	39,80	1626,4	15	0,00357	0,00142	0,98162	C
20	B077178	30,79	644,44	19	0,00350	0,00140	0,98302	C
...
56	A101723526	1,54	40,23	13	0,00012	0,00005	0,99997	C
57	P799414	1,08	177,06	13	8,4E-05	3,3E-05	1	C

Tableau 5.2 – Résultats de la classification ABC

Classification XYZ : Pour la classification XYZ, on débute par le calcul du critère de variabilité CV puis de même manière que la classification ABC, nous obtenons le tableau G.1 en annexe G.

Classification multicritère ABC-XYZ : A travers la combinaison des deux critères, nous obtenons le tableau suivant :

N°	Numéro de la pièce	Classe XYZ	Classe ABC
1	A100917567	X	A
2	A100019788	X	A

3	H432650	X	A
4	A100213000	X	A
5	T6016548	Y	A
6	E023908	Y	B
...
20	H125975	Y	C
21	P778944	Y	C
...
56	0050832	Z	C
57	H712973	Z	C

Tableau 5.3 – Résultats de la classification ABC-XYZ

Dans ce qui suit, nous optons pour l'utilisation des classes combinées AX, BX, AY et BY pour notre étude, en raison de leur importance par rapport aux autres classes. Les pièces sélectionnées représentent plus de 79% de la valeur des pièces et près de 70% des quantités consommées.

5.1.2 Application des modèles statistiques de prévision

Pour effectuer ces prévisions, nous optons dans un premier temps l'utilisation du logiciel Eviews ou nous débuterons par une analyse informelle de la consommation de la pièce de rechange sélectionnée auparavant, suivie par la modélisation de cette consommation à l'aide de deux méthodes distinctes, adaptées à la nature de la série chronologique.

L'application du modèle prévisionnel sera effectuée sur la série temporelle Part A représentant l'évolution de la consommation de la pièce de rechange sélectionnée, et cela sur une période s'étalant d'octobre 2021 jusqu'à mai 2024.

Analyse Informelle

Nous commençons dans un premier temps notre étude à travers la visualisation graphique de l'évolution de la consommation de la pièce en question dans le but de définir sa nature et pouvoir identifier les modèles les mieux adaptés à notre série. Les figures suivantes illustrent l'évolution de consommation ainsi que le corrélogramme de la série chronologique.

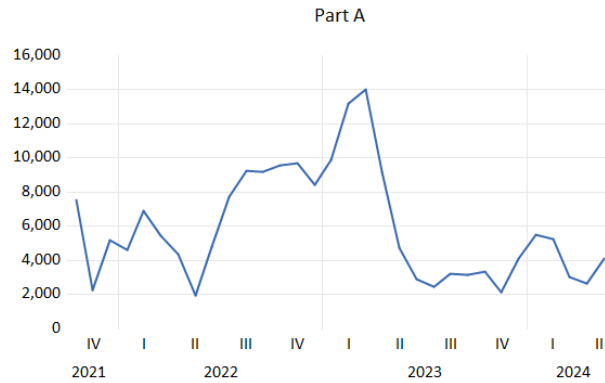


Figure 5.1 – Evolution de la consommation de la pièce de rechange

Ce graphique d'évolution permet dans un premier temps de détecter une éventuelle tendance mais qui reste à vérifier analytiquement.

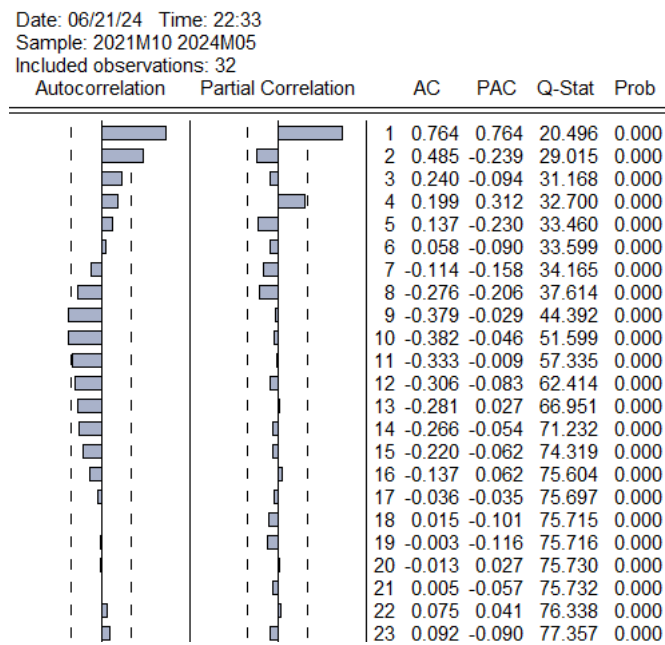


Figure 5.2 – Corrélogramme de la série

Interprétation Informelle

Le corrélogramme indique que la série n'est probablement pas stationnaire, malgré la présence de plusieurs pics significatifs sur le corrélogramme simple et partiel.

De plus, le critère de probabilité "prob" (Q-stat) est inférieur dans sa majorité à la valeur 0,05, ce qui peut indiquer une potentielle stationnarité de la série.

Test de stationnarité de Fisher à travers l'ANOVA

Dans le but de confirmer ou non l'existence d'une saisonnalité, un test d'Anova réalisé sous excel le permet. La figure ci-présente illustre les résultats de l'analyse :

ANALYSE DE VARIANCE						
Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	29083015,69	3	9694338,562	0,912002083	0,447753007	2,946685266
A l'intérieur des groupes	297632521,8	28	10629732,92			
Total	326715537,5	31				

Figure 5.3 – Résultat du test d'Anova

La probabilité est supérieure à 0.05 et la valeur du F-test $F = 0.912$ est inférieure à la valeur critique du test qui est de 2.947. Ce qui nous permet de rejeter l'hypothèse de présence de saisonnalité, et donc notre série n'est pas saisonnière.

Application des modèles : A travers l'analyse informelle effectuée sur notre série ainsi que le test statistique de saisonnalité, nous avons pu déterminer que notre série temporelle est de tendance quelconque, qu'elle est non-saisonnière, et qu'elle présente des changements de structure. Et donc, nous avons opté pour les méthodes de lissage exponentiel, à savoir, le modèle de lissage exponentiel simple, double et le modèle de Holt-Winters non-saisonnière pour les critères adéquats, ainsi que la méthodologie de contrôle de Box-Jenkins, en raison de sa capacité à gérer des modèles complexes et variés.

Prévision à travers les modèles de lissage exponentiel

Dans cette partie nous allons appliquer les modèles de prévisions de lissage exponentiel. L'estimation des paramètres des modèles choisis en raison de leur adéquation à notre série a été effectuée à travers l'outil Eviews, les résultats de l'estimation sont comme suit :

a) Le modèle de LES :

```
Date: 06/22/24 Time: 01:18
Sample: 2021M10 2024M05
Included observations: 32
Method: Single Exponential
Original Series: PART_A
Forecast Series: PART_ASM_SINGLE
```

Parameters: Alpha	0.9990
Sum of Squared Residuals	1.49E+08
Root Mean Squared Error	2157.103

End of Period Levels: Mean	4081.545
----------------------------	----------

Figure 5.4 – Résultat de l'estimation des paramètres du LES

b) Le modèle de LED :

L'estimation des paramètres du modèle nous a donné les résultats illustrés dans la figure :

```
Date: 06/22/24 Time: 01:22
Sample: 2021M10 2024M05
Included observations: 32
Method: Double Exponential
Original Series: PART_A
Forecast Series: PART_ASM_DOUBLE
```

Parameters: Alpha	0.4660
Sum of Squared Residuals	2.14E+08
Root Mean Squared Error	2587.691

End of Period Levels: Mean	3641.476
Trend	-37.56429

Figure 5.5 – Résultat de l'estimation des paramètres du LED

c) Le modèle de HWNS :

De même manière, les résultats de l'estimation des paramètres du modèle sont présentés dans la figure qui suit :

Date: 06/22/24 Time: 01:05		
Sample: 2021M10 2024M05		
Included observations: 32		
Method: Holt-Winters No Seasonal		
Original Series: PART_A		
Forecast Series: PART_ASM		
<hr/>		
Parameters:	Alpha	1.0000
	Beta	0.0100
	Sum of Squared Residuals	1.54E+08
	Root Mean Squared Error	2195.287
<hr/>		
End of Period Levels:	Mean	4083.000
	Trend	228.4445
<hr/>		

Figure 5.6 – Résultat de l'estimation des paramètres de HWNS

Afin d'estimer la précision des modèles prévisionnels, nous avons appliqué les modèles à notre série de données et évalué les résultats à l'aide de plusieurs métriques (Annexe G). Le taux d'erreur MAPE obtenu est de 35,4 %, 61,07 % et 32,24 % pour les modèles LES, LED et HWNS, respectivement.

Prévision à travers la méthode Box Jenkins (BJ)

Après avoir analysé la série chronologique informellement au début de l'étude ce qui nous a permis d'identifier que la série est de tendance quelconque, nous allons désormais appliquer le modèle de Box Jenkins en analysant formellement dans un premier temps notre série à travers le test de Dicker Fuller.

Test de stationnarité de Dickey Fuller : Dans le but de choisir les modèles les mieux adaptés à notre série, il est nécessaire de vérifier l'existence d'une tendance et d'effectuer une stationnarisation si la série n'est pas stationnaire. Et afin de la vérifier, nous utilisons le test de Dickey Fuller ou bien le test de racine unitaire.

Nous commençons donc par le modèle [3] sur le coefficient de la tendance (ou test sur b).

En posant le test d'hypothèse suivant :

- Si $\text{Prob}(b) > 5\%$; $H_0 : b = 0$; b n'est pas significativement différent de 0.
- Si $\text{Prob}(b) < 5\%$; $H_1 : b \neq 0$; b est significativement différent de 0.

Les hypothèses peuvent être acceptées en fonction de la valeur de la P-value. Si cette valeur est inférieure à 5% nous rejetons l'hypothèse nulle H_0 et nous acceptons l'hypothèse alternative H_1 qui signifie que le modèle possède une racine unitaire.

Après avoir appliqué le test de DF sur la série chronologique, nous obtenons les résultats suivants :

Null Hypothesis: PART_A has a unit root
 Exogenous: Constant, Linear Trend
 Lag Length: 1 (Automatic - based on SIC, maxlag=7)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.678876	0.2516
Test critical values:		
1% level	-4.296729	
5% level	-3.568379	
10% level	-3.218382	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(PART_A)
 Method: Least Squares
 Date: 06/21/24 Time: 22:36
 Sample (adjusted): 2021M12 2024M05
 Included observations: 30 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PART_A(-1)	-0.303035	0.113120	-2.678876	0.0126
D(PART_A(-1))	0.290901	0.164935	1.763731	0.0895
C	2840.298	1089.321	2.607403	0.0149
@TREND("2021M10")	-56.63224	39.95560	-1.417379	0.1682

Figure 5.7 – Test de la racine unitaire avec tendance

Interprétation des résultats obtenus par Eviews :

On remarque que la valeur de probabilité du coefficient de la tendance @TREND est égale à 0.1682, ce qui est supérieur à 5%. Donc nous acceptons l'hypothèse H0 que la série possède une tendance non significative.

A partir de ces résultats et en suivant le schéma du test de DF, nous passons au modèle [2], qui est le test de la constante c.

En postant le test d'hypothèse suivant :

- Si $\text{Prob}(c) > 5\%$; $H_0 : c = 0$; c n'est pas significativement différent de 0.
- Si $\text{Prob}(c) < 5\%$; $H_1 : c \neq 0$; c est significativement différent de 0.

En suivant la même approche du test d'hypothèse du coefficient de la tendance en fonction de la valeur de P-value de la constante "c", on obtient à travers Eviews les résultats représentés par la Figure en annexe.

Nous pouvons constater que la probabilité de la constante "c" qui est égale à 0.0325 est inférieure à 5%. Et donc nous rejetons l'hypothèse nulle H0 et nous acceptons l'hypothèse alternative H1 qui implique la présence d'une constante.

A partir de cette conclusion nous passons au test de la racine unitaire comme suit en postant l'hypothèse suivante :

- Si $\text{Prob}(\phi) > 5\%$; $H_0 : \phi = 0$; présence de racine unitaire ; Processus DS.
- Si $\text{Prob}(\phi) < 5\%$; $H_1 : \phi \neq 0$; pas de racine unitaire ; Série stationnaire.

Comme affiché sur la figure, la valeur de (Phi) qui est de 0.149 est supérieure à 5% et donc nous rejetons l'hypothèse nulle H0 et acceptons l'hypothèse alternative qui confirme la présence d'une racine unitaire. Et donc nous devons passer par le processus DS pour stationnariser notre série.

Nous appliquons sur notre série d'étude une différenciation d'ordre 1 et nous repassons de nouveau par le test de DF, cette fois-ci sur la nouvelle série D(Part_A). Et nous commençons par le modèle [3] comme suit :

On remarque d'après la figure x en annexe que la probabilité de la tendance @TREND est supérieure à 5%, on accepte donc l'hypothèse H0 que la série différenciée possède une tendance non significative. Et nous passons au modèle [2] (Annexe H).

On peut constater que la probabilité de la constante C qui est de 0,8549 est supérieure à 5%, et donc on accepte l'hypothèse H0 qu'induit l'absence de constante significative. Nous passons au modèle [1] pour la série différenciée (Annexe H).

La probabilité de la racine unitaire est inférieure au seuil critique 5%, et donc on rejette l'existence de racine unitaire. **Notre série différenciée est bien stationnaire.**

Estimation des ordres p et q du modèle ARMA(p,q)

Après avoir vérifié que la série est stationnaire, l'étape qui suit est l'estimation du modèle ARMA le plus adapté à notre série qui se fait par la détermination des paramètres p et q du modèle. Pour cela, nous utilisons le corrélogramme de la série d'analyse différenciée afin d'analyser les fonctions d'autocorrélation simple et partielle.

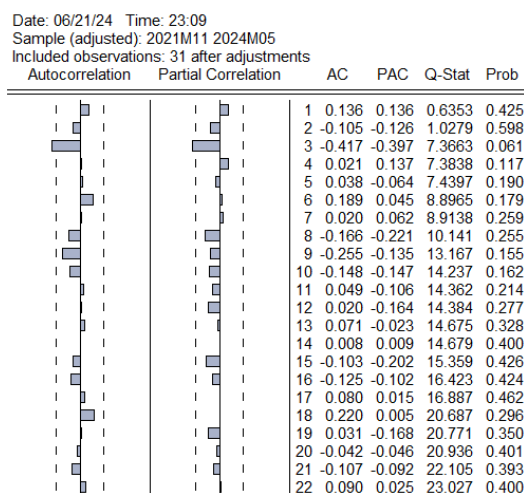


Figure 5.8 – Corrélogramme d'autocorrélation de la série différenciée

L'analyse des fonctions d'autocorrélation permet d'identifier les modèles candidats à travers l'identification l'ordre des pics significatifs et qui sont : MA(3), AR(3) et ARMA(3,3).

Pour valider ces modèles candidats, il est nécessaire de de retenir que ceux ayant les termes significativement différents de 0, et cela peut être réalisé à l'aide du test de student en vérifiant que la probabilité P est inférieure à 5%.

Les modèles retenus après vérification de la significativité des termes sont : le modèle AR(3).

Nous décidons également de faire appel au modèle ARMA automatique, qui permet de proposer un modèle expert. Nous obtenons le modèle suivants : ARMA(2,3) (Annexe H).

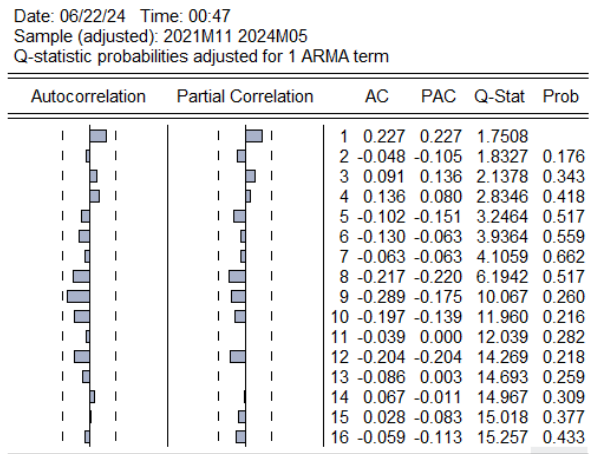


Figure 5.9 – Corrélogramme des résidus

D'après les résultats de la figure précédente, on remarque que les probabilités des Q-stat sont toutes supérieures à 5%. Et donc les résidus ne sont pas significativement différents de zéro (sont à bruit blanc).

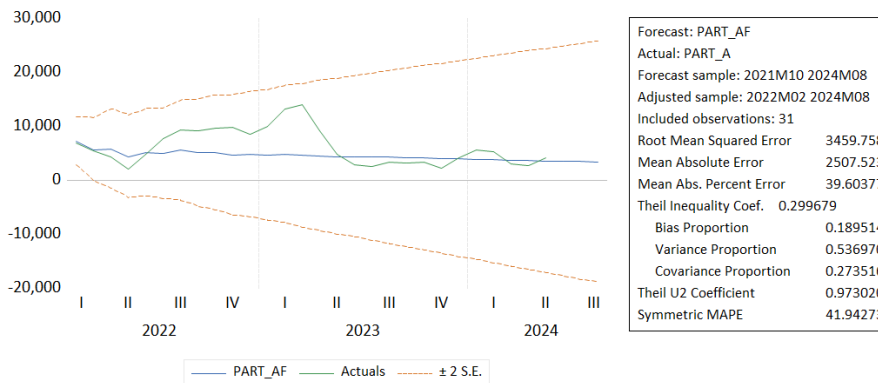


Figure 5.10 – Résultats des prévisions du modèle

Validation du modèle : Pour valider ou pas le modèle expert proposé par Eviews, nous effectuons le test de Student (Annexe H). On remarque que pour :

- AR(2) : $|t\text{-stat}(\text{AR})| < t(\text{tab})$ (la probabilité p est supérieure à 5%)
- MA(3) : $|t\text{-stat}(\text{MA})| < t(\text{tab})$, donc le modèle est non significatif.

Afin de choisir le meilleur modèle à partir des modèles sélectionnés, nous comparons les résultats des critères Akaike, Schwarz et Hannan-Quinn et le R^2 , sachant que l'objectif est de minimiser ces trois premiers critères et de maximiser le dernier. Cependant, notre modèle ARMA non-validé nous passons directement à la comparaison des modèles traités précédemment.

Mesure de la fiabilité des modèles

Afin de déterminer lequel des modèles développés précédemment, c'est à dire les modèles prévisionnels de lissage exponentiel, ainsi que la méthodologie de Box Jenkins, offre de meilleurs résultats vis à vis de la consommation du produit traité, nous évaluerons leur fiabilité en utilisant les indicateurs de performance RMSE, MAE et MAPE, présentés dans la partie état de l'art. Les résultats de ces indicateurs sont illustrés dans le tableau suivant :

Modèle	LES	LED	HWNS	AR(3)
RMSE	2.16e+3	2.58e+3	2.19e+3	3.46e+3
MAE	1.67e+3	2.11e+3	1.66e+3	2.51e+3
MAPE	35.4%	61.07%	32.24%	54.65%

Tableau 5.4 – Comparaison entre les résultats des différents modèles statistiques

Nous constatons que parmi les quatre modèles utilisés, le modèle de HWNS s'avère être le plus performant avec la série chronologique étudiée, affichant un taux d'erreur de 32,24 %, soit un taux de précision de 67,76 %.

Le tableau suivant illustre le résultat obtenu pour le restant des pièces sélectionné précédemment

Modèle	Modèle de lissage		ARMA	
	Modèle choisi	MAPE	Modèle ARMA	MAPE
Part A	HWNS	32.24%	Modèle non significatif	
Part B	HWM	20.07%	Modèle non significatif	
Part C	HWNS	24.45%	ARMA(2,4)	28.04%
Part D	HWM	26.41%	AR(10)	23.6%
Part E	HWM	19.28%	ARMA(3,1)	19.48%
Part F	HWM	26.9%	MA(3)	27.9%

Tableau 5.5 – Comparaison entre les résultats des modèles pour les pièces sélectionnées

De même, nous constatons que les modèles de lissage s'avèrent être les plus performants pour l'instant, avec une moyenne des erreurs de 25%.

Bien que les résultats de notre analyse prévisionnelle soient acceptables, ils ne répondent pas pleinement à nos exigences. Pour affiner notre étude, nous prévoyons maintenant d'appliquer des modèles de Machine Learning avancés. Ces modèles nous permettront d'exploiter mieux les données en notre possession, améliorant ainsi la précision de nos prévisions et notre capacité à identifier les tendances importantes pour les pièces sélectionnées. Cela nous aidera à proposer des engagements de volume avec les fournisseurs en question.

5.1.3 Application des modèles de prévision basée sur le Machine Learning

Dans cette section, compte tenu de la faible précision des modèles de prévision statistiques appliqués à nos séries chronologiques, nous avons décidé de recourir à des modèles de prévision basés sur le Machine Learning. Les algorithmes que nous utiliserons varieront en complexité, allant de l'algorithme FB Prophet aux réseaux de neurones artificiels. En adoptant cette approche, nous visons à augmenter la précision des prévisions et à améliorer la fiabilité des résultats.

Prévision à travers l'algorithme FB Prophet

Nous décidons d'appliquer dans un premier temps l'algorithme Facebook Prophet, qui se présente comme un outil hybride en se basant sur des modèles statistiques ainsi que sur le Machine Learning et qui permet de gérer automatiquement les tendances saisonnières et les changements dans les données, ce qui peut être complexe à modéliser avec d'autres méthodes comme le lissage exponentiel ou ARIMA.

Pour appliquer le modèle prévisionnel à notre série temporelle, nous utilisons le langage de programmation Python avec la bibliothèque Prophet. Nous entraînons l'algorithme sur notre série chronologique en spécifiant des données d'entraînement et de test. Une fois le modèle entraîné et les prévisions effectuées, nous évaluons sa performance en utilisant les métriques suivantes : R^2 , MAE, RMSE et MAPE. Le déroulement complet de cette procédure est détaillé dans le script en annexe I.

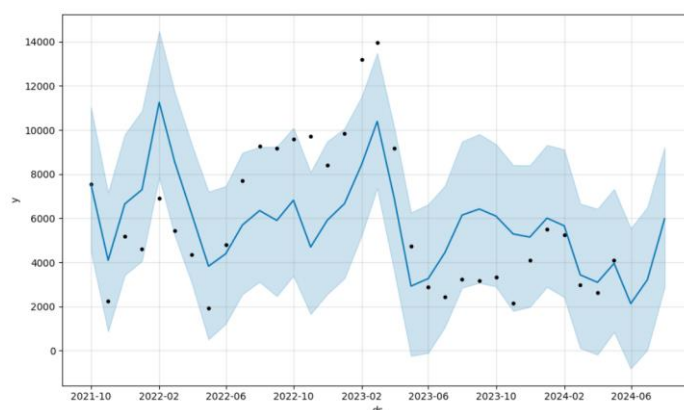


Figure 5.11 – Graphique des prévisions par Facebook Prophet

Validation du modèle : Nous évaluons la performance du modèle en calculant les métriques RMSE, MAE, MAPE pour la série traitée avec les résultats obtenus pour les autres séries de travail. Les résultats des indicateurs de performance sont récapitulés dans le tableau suivant :

Modèle	Part A	Part B	Part C	Part D	Part E	Part F
RMSE	2.55e+3	1.18e+3	554.64	494.64	8.54	8.022
MAE	2.16e+3	1.02e+3	457.38	421.93	6.81	6.11
MAPE	43.55%	32.51%	27.55%	27.93%	16.49%	27.62%

Tableau 5.6 – Tableau illustratif des différents résultats des pièces à travers FB Prophet

Prévision à travers l’algorithme XGBoost

Nous avons créé une fonction "dt_decomp" pour extraire des informations trimestrielles, mensuelles et annuelles à partir de la série chronologique. Ces informations ont été ajoutées comme nouvelles colonnes dans le dataframe. Ensuite, nous avons divisé la série en deux parties comme nous l'avons mentionné auparavant. La fonction "dt_decomp" a été appliquée à la fois sur l'ensemble d'entraînement et l'ensemble de test pour extraire ces caractéristiques supplémentaires.

Ces caractéristiques ont été utilisées comme entrées pour entraîner le modèle XgBoost, dans le but d'améliorer la précision des prédictions. Une fois le modèle entraîné avec la fonction XGBRegressor() et ajusté avec les paramètres appropriés pour éviter le surapprentissage, nous avons utilisé ce dernier pour effectuer des prédictions sur l'ensemble de test en utilisant les caractéristiques correspondantes.

Nous avons initialisé le modèle XGBoost avec des hyperparamètres choisis. Pour éviter le surapprentissage, nous avons effectué une validation croisée et une recherche par grille pour trouver les meilleurs hyperparamètres. Le modèle entraîné a été utilisé pour faire des prédictions sur l'ensemble de test.

Les prédictions ont ensuite été comparées aux valeurs réelles de l'ensemble de test avec le graphique suivant :

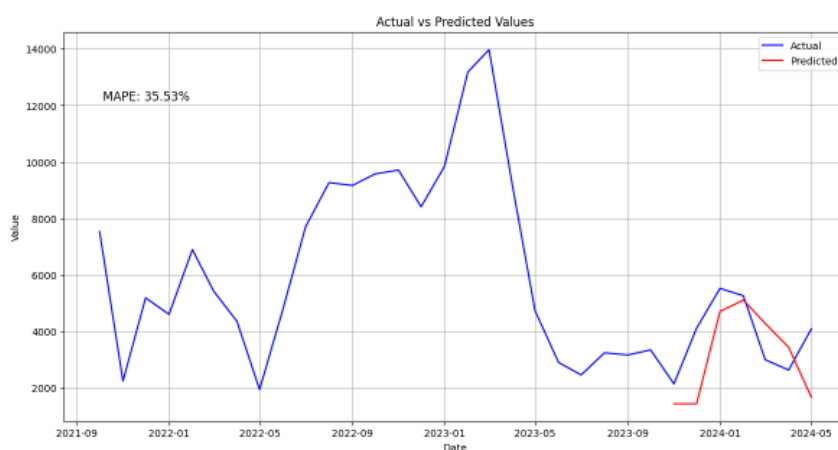


Figure 5.12 – Graphique des prévisions par l’algorithme XGboost

En examinant les résultats, nous avons constaté que les variations dans les données d'entraînement ont été également répliquées dans les données de test, bien que les valeurs diffèrent légèrement. Enfin, nous avons tracé deux courbes : la courbe de la série temporelle d'origine et la courbe de la série temporelle prédite.

Les performances du modèle ont été évaluées en utilisant plusieurs métriques telles que l'erreur quadratique moyenne (MSE), l'erreur absolue moyenne (MAE) et l'erreur absolue moyenne en pourcentage (MAPE). Ces métriques ont fourni une évaluation complète de la précision du modèle.

	Part A	Part B	Part C	Part D	Part E	Part F
MAPE	48,47%	48,11%	32,97%	34,15%	26,09%	27,07%
MAE	2945,40677	2928,0659	479,0151	1864,7979	15,4707	12,3566
RMSE	3250,924144	3297,81981	533,6629367	2268,995493	19,03125062	13,01047655

Tableau 5.7 – Tableau comparatif des résultats des prévisions par l’algorithme XGboost

Prévision en utilisant les RNNs

Initialement, les données sont importées et prétraitées pour inclure des caractéristiques temporelles telles que le mois et l'année. Cette étape permet d'enrichir les données en capturant les variations saisonnières et annuelles qui peuvent influencer les prévisions. Ensuite, les données sont normalisées à l'aide d'une mise à l'échelle MinMax Scaler, ce qui garantit que toutes les variables sont ramenées à une plage commune, facilitant ainsi l'apprentissage du modèle.

Une fois les données prétraitées, elles sont structurées en séquences temporelles de longueur fixe. Une modélisation est ensuite réalisée à l'aide de réseaux neuronaux récurrents (RNN), tels que le modèle SimpleRNN. Ces modèles sont configurés avec plusieurs couches récurrentes et des couches de dropout pour optimiser l'apprentissage et prévenir le surapprentissage. Cette architecture permet au modèle de capturer les dépendances temporelles complexes et d'améliorer la précision des prévisions.

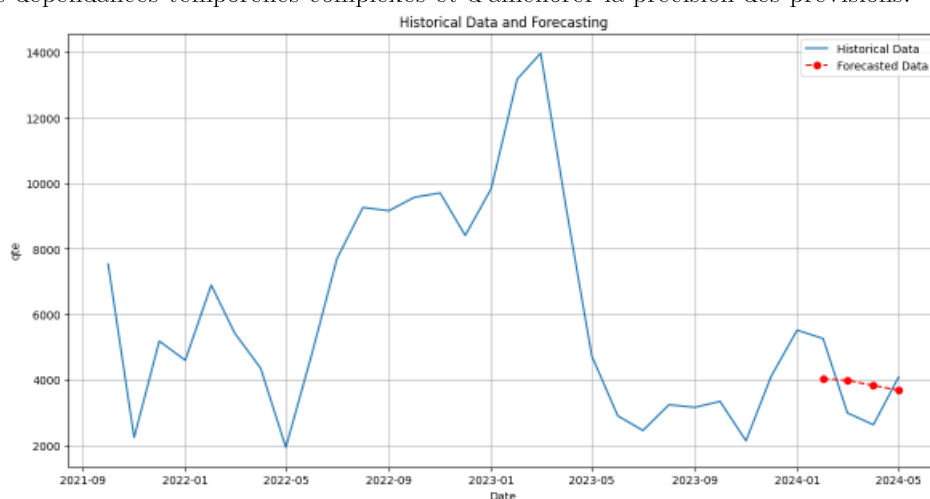


Figure 5.13 - Graphique des Prévisions via les Réseaux de Neurones Récurrents (RNN)

Une fois le modèle construit, il est entraîné sur les séquences temporelles préparées. L'entraînement utilise des techniques d'optimisation comme Adam et une fonction de perte comme l'erreur quadratique moyenne (MSE) pour ajuster les poids du modèle. Ce processus itératif permet d'améliorer les performances du modèle en minimisant l'erreur de prédiction. Après l'entraînement, le modèle est évalué en utilisant des données de validation ou de test distinctes. L'évaluation inclut la comparaison des prédictions du modèle avec les valeurs réelles, mesurée par des métriques telles que l'erreur absolue moyenne (MAE) et l'erreur quadratique moyenne (MSE) et MAPE.

Métrique	Part A	Part B	Part C	Part D	Part E	Part F
MAPE	44.23%	27.32%	9.37%	38.72%	30.34%	16.85%
MAE	1251.33	1747.45	105.20	2207.77	18.63	7.59
RMSE	1516.50	1921.47	109.67	2253.54	21.56	8.76

Tableau 5.8 – Tableau comparatif des résultats des prévisions par l'algorithme RNN

Prévision à travers le LSTM

Dans cette partie, nous décrivons notre approche pour prédire les séries temporelles en utilisant un modèle LSTM avec TensorFlow et Python. Nous avons commencé par charger et normaliser les données en utilisant un scaler MinMax pour l'adaptation au modèle LSTM. Les données ont été divisées en ensembles d'entraînement et de test. Le modèle LSTM a été construit en utilisant TensorFlow, avec une architecture comprenant plusieurs couches pour capturer les dépendances temporelles à long terme. Des couches Dropout ont été ajoutées pour régulariser le modèle et éviter le surapprentissage. Le modèle a été compilé avec l'optimiseur Adam et la fonction de perte MSE (Mean Squared Error), tout en calculant également les métriques MAE (Mean Absolute Error) et MAPE (Mean Absolute Percentage Error) pendant l'entraînement.

Après l'entraînement, nous avons évalué le modèle sur les ensembles d'entraînement et de test en inversant les prédictions pour obtenir les valeurs réelles

Les résultats ont révélé que notre modèle LSTM a montré une capacité limitée à capturer les motifs dans les données de séries temporelles, avec des performances mitigées sur les ensembles d'entraînement et de test. Les métriques calculées ont indiqué une précision relativement insuffisante dans la prédiction des valeurs de quantité. En conclusion, l'approche utilisée n'a pas démontré l'efficacité attendue des réseaux LSTM pour la modélisation et la prédiction de séries temporelles complexes.

Nous pouvons dire que la performance médiocre des modèles de machine Learning et des réseaux de neurones est relativement due au manque de données. Pour remédier à cela, nous nous sommes tournés vers des modèles de génération de données synthétiques afin de fiabiliser les modèles. Ensuite, nous procéderons à une comparaison des performances des différents modèles.

Métrique	Part A	Part B	Part C	Part D	Part E	Part F
MAPE	46.59%	27.14%	18.23%	18.05%	17.91%	17.48%
MAE	1431.08	1478.84	305.37	1000.90	11.86	6.91
RMSE	1493.40	1851.49	449.03	1130.04	16.11	7.94

Tableau 5.9 – Tableau comparatif des résultats des prévisions par l’algorithme LSTM

Nous avons remarqué dans la littérature que les performances médiocres des modèles de machine Learning et des réseaux de neurones étaient souvent attribuables à un manque de données adéquates. Pour pallier cette limitation, nous avons exploré l'utilisation de modèles de génération de données synthétiques afin de renforcer la robustesse de nos modèles. Par la suite, nous avons procédé à une comparaison des performances entre les différents jeux de données.

Après avoir utilisé plusieurs algorithmes, le minimum de MAPE extrait est de 17%, ce qui souligne la nécessité de créer de nouvelles données synthétiques afin d'augmenter la précision des modèles de prévision.

5.1.4 Génération de Données Synthétiques

Création et Détection des Métadonnées

La création des métadonnées est une étape nécessaire pour la génération de données synthétiques. Les métadonnées décrivent la structure et les types des colonnes du DataFrame. Nous avons utilisé la bibliothèque SDV pour détecter les métadonnées à partir du DataFrame.

Nous avons ensuite mis à jour le type de colonne pour certaines colonnes spécifiques, comme la colonne entity, que nous avons définie comme un identifiant (id). De plus, nous avons spécifié la colonne « ds » comme index de séquence et la colonne entity comme clé de séquence. Ces étapes garantissent que la

génération de données synthétiques respecte la structure temporelle et relationnelle des données réelles (Annexe I).

Génération de Données Synthétiques

Pour générer les données synthétiques, nous avons utilisé le synthétiseur PAR (Probabilistic AutoRegressive) de la bibliothèque SDV. Ce modèle est particulièrement adapté pour les données séquentielles, où l'ordre des enregistrements est important.

Le processus de génération se déroule en trois étapes principales. La Création du synthétiseur : Initialisation du modèle de synthèse avec les métadonnées. Puis l'entraînement du synthétiseur : Le modèle est entraîné sur les données réelles pour apprendre les distributions et les relations entre les colonnes. Et finalement la Génération de données synthétiques : Une fois entraîné, le modèle génère de nouvelles données qui imitent les propriétés des données réelles.

Évaluation de la Qualité des Données Synthétiques

L'évaluation de la qualité des données synthétiques générées est une étape également importante pour s'assurer que les données synthétiques sont statistiquement similaires aux données réelles. Pour cela, nous avons utilisé la bibliothèque SDMetrics, qui permet de comparer les données synthétiques aux données réelles sur divers aspects statistiques. Le rapport de qualité évalue la similarité statistique entre les données réelles et synthétiques. Ce rapport permet de découvrir quels motifs des données réelles ont été capturés par les données synthétiques. Voici les principales composantes de ce rapport, Pour générer le rapport de qualité, nous avons utilisé la fonction `evaluate_quality` de la bibliothèque SDMetrics. Nous pouvons voir la distribution des données synthétiques en comparaison avec les données réelle dans la figure suivante.

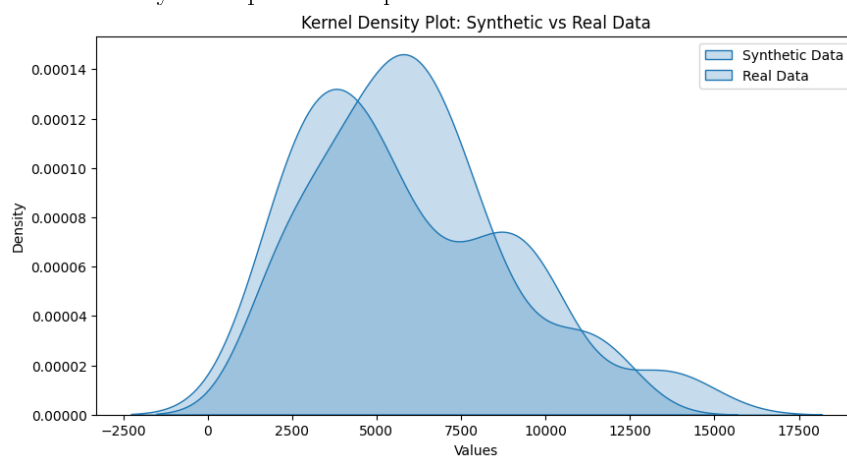


Figure 5.15 – Kernel Density plot entre les données réelles et synthétiques

Le score de 92.7% indique que les données synthétiques sont globalement similaires aux données réelles, environ 93% des tendances des données réelles étant capturées par les données synthétiques.

5.1.5 Comparaisons entre les modèles prévisionnels

Evaluation des résultats du génération de données

Après comparaisons entre les résultats de l'application des modèles sur la base de données originale et sur la base de données originale plus base de données synthétique, nous remarquons qu'il y avait en moyenne une amélioration de 8 % entre les résultats d'entraînement sur les données originales et

l'entraînement sur les données synthétiques plus les données originales.

Nous comparons les résultats de génération de données entre les trois derniers modèles de machine Learning, c'est-à-dire, XgBoost, RNN et LSTM, car les données synthétiques n'ont pas véritablement amélioré le déroulement des modèles de Fb Prophet (Les résultats se trouvent en annexe I).

		A	B	C	D	E	F
Xgboost	OG	48,47%	48,11%	32,97%	34,15%	26,09%	27,07%
	OG+Synth	29,39%	43,63%	31,43%	34,58%	17,94%	26,95%
RNN	OG	53,11%	27,32%	13,67%	38,72%	30,34%	16,85%
	OG+Synth	14,77%	26,28%	9,68%	28,73%	24%	13,33%
LSTM	Og	46,59%	27,14%	18,23%	18,05%	18%	17,48%
	Og+Synth	29,92%	18,63%	12,75%	12,16%	15,97%	12,12%

Tableau 5.10 – Tableau comparatif des erreurs MAPE

Pour les algorithmes XgBoost et LSTM, nous avons une amélioration d'environ 5 % des résultats en moyenne. Cependant, pour les RNNs, le modèle s'est amélioré de **14 %**. Il reste que pour la majorité des cas, LSTM a donné de meilleurs résultats.

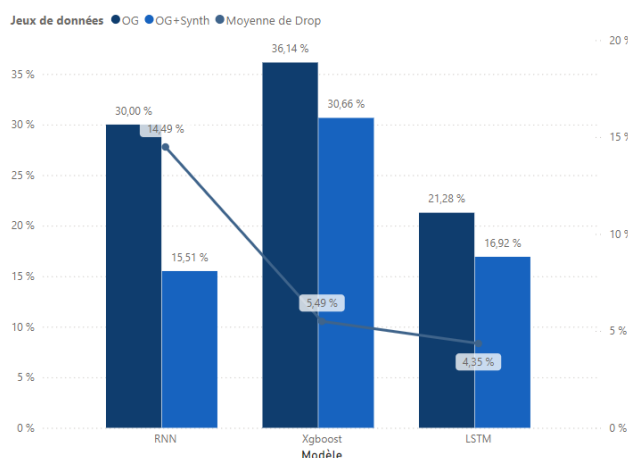


Figure 5.16 – Evolution des moyennes d'amélioration de l'erreur MAPE

Les résultats varient également d'une pièce à l'autre, car toutes les pièces n'ont pas eu le même changement en termes de MAPE. Par exemple, certaines pièces ont connu une réduction de 24 %, comme la pièce A, qui avait une tendance extrêmement difficile à capturer avec une moyenne de 49,39 % avec peu de données. En utilisant la génération de données, nous avons obtenu des résultats pertinents, avec une réduction de 25 % de la moyenne de MAPE, et le meilleur modèle a atteint un MAPE de 14,77 %. Pour la majorité des autres pièces, il y avait une réduction de 3 % à 5 % en MAPE, mais cela a également entraîné une réduction importante dans plusieurs cas, pouvant aller de 10 à 15 %.

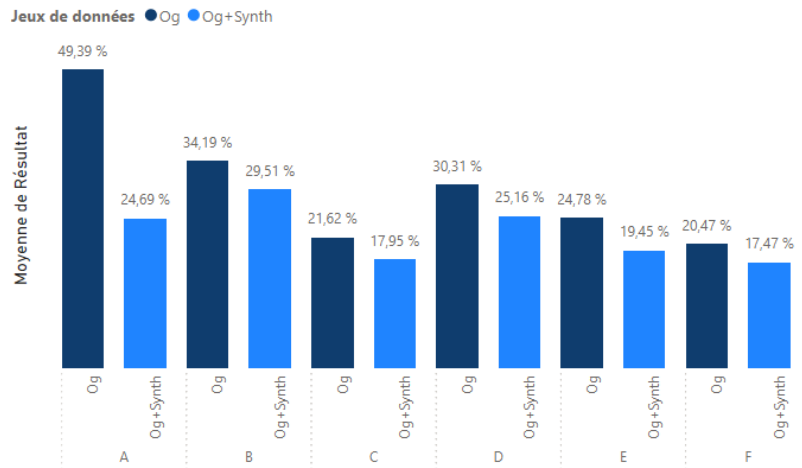


Figure 5.17 – Moyenne des résultats par rapport aux pièces étudiées

Comparatif des résultats de tous les modèles : Le tableau suivant illustre les résultats du pourcentage d'erreur moyenne absolue MAPE par rapport aux pièces étudiées ainsi qu'aux modèles utilisés pour notre étude prévisionnelle de la consommation des pièces du fournisseur sélectionnée.

Modèles	Modèles de lissage	ARIMA	Fb Prophet	Xgboost	RNN	LSTM
A	32.24%	/	43.55%	29,39%	14,77%	29,92%
B	20.07%	/	32.51%	43,63%	26,28%	18,63%
C	24.45%	28.04%	27.55%	31,43%	9,68%	12,75%
D	26.41%	23.6%	27.93%	34,58%	28,73%	12,16%
E	19.28%	19.48%	16.49%	17,94%	24%	15,97%
F	26.9%	27.9%	27.62%	26,95%	13,33%	12,12%

Tableau 5.11 – Comparatif des résultats de prévision par modèle étudié

Cette étude prévisionnelle nous a permis de mieux comprendre les tendances de consommation des pièces importantes du fournisseur en analysant les séries chronologiques, leurs tendances et leurs caractéristiques. Nous avons ensuite testé une série de modèles pour sélectionner celui le plus adapté à nos données. Constatant le faible taux de précision initial, nous avons conclu qu'il était nécessaire de générer davantage de données pour améliorer la fiabilité des prévisions, ce qui nous a orientés vers la génération de données synthétiques.

Cette génération de données a considérablement amélioré les taux d'erreur, avec une moyenne d'erreur de 13.9% pour les prévisions avec la LSTM et les RNN conjointement. Par conséquent, pour la suite de notre étude, notamment dans le but de proposer un engagement contractuel en volume, nous nous baserons sur les prévisions obtenues à travers les réseaux de neurones artificiels.

5.2 Estimation des gains en coûts à travers l'engagement contractuel en volume

Dans cette partie de la solution, nous évaluons le gain potentiel suite à l'application des engagements contractuels en volume, en utilisant les résultats des prévisions de consommation des pièces de rechange du fournisseur concerné. Notre objectif est d'estimer les réductions potentielles liées aux coûts de transport. Ne disposant pas des coûts de transport spécifiques à l'approvisionnement de ces pièces, nous évaluerons les réductions liées aux délais de livraison. Ces réductions serviront de base pour projeter les économies potentielles issues des engagements contractuels basés sur le volume.

Pour estimer ces résultats, nous avons choisi d'utiliser une estimation probabilistique via la méthode de Monte-Carlo. Cette approche fournira à l'équipe de Sourcing de SLB une base pour la négociation avec le fournisseur.

Estimation des Réductions de Coûts de Transport à l'Aide de la Méthode Monte-Carlo

Tout d'abord, nous allons calculer n_1 , qui représente le nombre de commandes effectuées l'année passée. Ensuite, nous supposerons que n_2 est le nombre de commandes prévues pour l'année suivante. Ce nombre est calculé en divisant les prévisions établies dans le deuxième chapitre par le stock de sécurité calculé dans le premier chapitre (car M&S Prime passe une commande lorsque le stock disponible atteint le niveau de stock de sécurité).

Nous ajouterons ensuite l_{q25} , l_m et l_{q75} au modèle. Ces valeurs représentent respectivement le 25ème percentile, la moyenne, et le 75ème percentile du délai de livraison pour des conditions similaires à celles du fournisseur.

Puis nous calculons :

$$R_1 = \left(1 - \frac{n_1}{n_2}\right) \times \frac{l_{q25}}{l_{q75}}$$

$$R_2 = \left(1 - \frac{n_1}{n_2}\right) \times \frac{l_m}{l_{q75}}$$

Ces ratios sont utilisés pour définir les bornes inférieure et supérieure des estimations de Monte-Carlo, offrant ainsi une base pour simuler les variations possibles des coûts de transport en fonction des variations de la demande et des délais de livraison.

La méthode Monte Carlo est utilisée pour simuler plusieurs trajectoires possibles des coûts de transport futurs. En utilisant des distributions probabilistes basées sur les données historiques, nous générons un grand nombre de scénarios pour estimer la variabilité des réductions de coûts de transport. Chaque simulation prend en compte les variations aléatoires des paramètres tels que les variations du délai de livraison. (Pour plus de détails, l'annexe contient la fonction utilisée pour estimer les réductions de coûts de transport.)

Après avoir exécuté les simulations Monte Carlo, nous analysons les résultats pour obtenir des estimations statistiques robustes. Notre modèle Monte Carlo prédit une réduction moyenne de 11,25% des coûts de transport, avec un écart-type de 1.05%. Cela signifie que, selon nos simulations, nous anticipons une réduction moyenne des coûts de transport de 11,25%, avec une variabilité possible allant jusqu'à $\pm 1.05\%$.

Pour conclure, après avoir augmenté le niveau d'automatisation de 21% des pièces commandées, et en estimant une réduction moyenne des coûts de transport pour le fournisseur d'environ 11%, on peut estimer les bénéfices pour ce fournisseur. Après avoir assuré une certaine stabilité de lancement des commandes grâce à l'automatisation et en stabilisant les prévisions à travers les modèles d'apprentissage profond, nous avons discuté avec les experts de SLB. Selon eux, ces résultats indiquent qu'il serait approprié de proposer une réduction des prix, en envisageant également un engagement sur volume pour les pièces à flux de demande stable. Cette réduction de prix garantirait également une stabilisation de la part de marché du fournisseur. SLB souhaite que ce fournisseur devienne le leader en termes de part de marché pour ces produits. Cela serait assuré par l'engagement sur volume, surtout que les deux piliers nécessaires pour un engagement sur volume fiable, à savoir la stabilisation de la commande et de la demande, sont assurés grâce au travail effectué au cours des deux derniers chapitres.

CONCLUSION GENERALE

Dans une industrie parapétrolière marquée par une concurrence intense, l'optimisation du trio coût, qualité et délais représente un levier essentiel et une réelle problématique, notamment dans la partie amont de la supply chain. Notre analyse nous a permis d'identifier que la variabilité des parts de marché des fournisseurs constitue un axe stratégique d'amélioration. Stabiliser ces parts de marché est primordial pour assurer une bonne maîtrise des coûts et renforcer la résilience de la Supply Chain.

Nous avons mené notre projet au sein du département Sourcing de SLB afin de proposer une solution à cette volatilité en stabilisant les relations contractuelles avec les fournisseurs par des engagements par volume. Après avoir analysé le marché concurrentiel des fournisseurs de l'entreprise, ce qui nous a donné une vue éclairée sur les variations des parts de marché, nous nous sommes orientés vers la catégorie des produits M&S pour illustrer notre approche, en sélectionnant un contrat particulier pour appliquer notre solution. De plus, nous avons audité les processus clés liés à la gestion des produits de cette catégorie en nous focalisant sur les processus amont afin d'identifier les dysfonctionnements majeurs et de cibler notre problématique.

Notre deuxième niveau d'analyse s'est articulé autour de la sélection des produits pour l'automatisation et la stabilisation des commandes. Cette étape était essentielle pour remédier au problème d'approvisionnement de pièces non nécessaires, en raison d'une classification inadéquate. À cette fin, nous avons :

- Appliqué des modèles de clustering pour proposer une nouvelle classification basée sur la consommation actuelle, suivant une méthodologie CRISP-DM.
- Proposé une nouvelle mesure des stocks de sécurité pour les pièces incluses dans l'automatisation des commandes.
- Développé un tableau de bord illustratif des pièces concernées, pour une meilleure visibilité de leur consommation et d'autres caractéristiques essentielles à la prise de décision.

Notre troisième contribution concerne l'amélioration du contrat avec un fournisseur spécifique pour stabiliser la relation et réaliser des économies potentielles. La solution consistait en la proposition d'un engagement en volume basé sur l'estimation des besoins futurs de l'entreprise. Pour ce faire, nous avons :

- Sélectionné les pièces les plus importantes du fournisseur à travers une classification ABC combinée à une classification XYZ.
- Réalisé une étude prévisionnelle des pièces sélectionnées en utilisant des modèles statistiques, puis des modèles de machine learning.
- Utilisé des modèles de génération de données synthétiques pour améliorer la précision de nos prévisions et comparé les résultats obtenus.

Enfin, nous avons estimé les gains potentiels suite à l'application de ces engagements en volume et à l'automatisation des commandes des pièces sélectionnées, en évaluant la réduction des délais de livraison liés à l'approvisionnement des pièces. Nous avons également inclus une vue probabiliste en utilisant le modèle Monte-Carlo pour fournir une marge d'erreur. Ces estimations représentent une base de négociation pour SLB avec le fournisseur afin de réaliser d'éventuelles économies.

Pistes d'amélioration et perspectives futures

Pour garantir une meilleure optimisation des coûts liés aux approvisionnements et à la supply chain amont de SLB, notre approche de résolution et nos propositions constituent un réel avantage et une base de référence pour la réalisation des futurs contrats avec tous les fournisseurs dont SLB NAF est responsable. Cette approche vise à stabiliser les relations avec les fournisseurs, mais aussi à stabiliser les commandes des pièces les plus importantes grâce à l'automatisation. Étendre l'utilisation de nos propositions aux autres catégories de produits et aux autres relations semble être la solution la plus adéquate pour répondre aux besoins stratégiques de l'entreprise, qui cherche à améliorer sa résilience face aux perturbations d'approvisionnement, garantir la disponibilité des matières premières et équipements, et optimiser le rapport prix-qualité-délais.

En conclusion, notre solution, composée de plusieurs propositions généralisables à l'ensemble du service d'approvisionnement de l'entreprise, est fidèle à notre analyse et aux constats émis lors de notre stage et visites sur le terrain. Bien qu'elle permette de résoudre en partie les problématiques identifiées, elle peut être enrichie et améliorée, notamment en intégrant les opérations et missions de l'entreprise sur le terrain ainsi que d'autres facteurs. Elle doit également faire l'objet d'une amélioration continue afin de refléter fidèlement la réalité du terrain.

Références

- Acholi, M. (2021). *LA DÉMARCHE DE CONCEPTION D'UN RÉFÉRENTIEL DE MESURE DE LA PERFORMANCE LOGISTIQUE*. *Revue D'Études en Management et Finance D'Organisation*.
- AMIDI, A. (s. d.). *CS 230—Recurrent Neural Networks Cheatsheet*. Consulté 1 juillet 2024, à l'adresse <https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks>
- Bandara, K., Hewamalage, H., Liu, Y.-H., Kang, Y., & Bergmeir, C. (2020). *Improving the Accuracy of Global Forecasting Models using Time Series Data Augmentation* (arXiv:2008.02663). arXiv. <https://doi.org/10.48550/arXiv.2008.02663>
- Benoit (née Moeller), S., S./Fassnacht, & Klose, S. (2006). A Framework for Supplier Relationship Management (SRM). *Journal of Business-to-Business Marketing*, 13. https://doi.org/10.1300/J033v13n04_03
- Bothin, J. (2024). Relation fournisseurs en entreprise. *Officéo*. <https://officeopro.com/services/gestion-des-achats/relation-fournisseur>
- Chaudhari, N. (2019). Impact of Automation Technology on Logistics and Supply Chain Management. *American Journal of Theoretical and Applied Business*, 5, 53. <https://doi.org/10.11648/j.ajtab.20190503.12>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. <https://doi.org/10.1145/2939672.2939785>
- Chima, C. M. (2007). Supply-Chain Management Issues In The Oil And Gas Industry. *Journal of Business & Economics Research (JBER)*, 5(6), Article 6. <https://doi.org/10.19030/jber.v5i6.2552>
- Chopra, S., & Meindl, P. (2016). *Supply Chain Management: Strategy, Planning, and Operation*. Pearson Education.
- Comité Professionnel du Pétrole. (2023). *ESTIMATIONS DE LA PRODUCTION MONDIALE DE PÉTROLE BRUT (975)*. Comité Professionnel du Pétrole. https://www.cpdp.org/modules/custom/gc_minipdf/pdf/NIE-975-Prod-mondiale-petrole-brut-et-reserves.pdf
- Cooper, M. C., Lambert, D. M., & Pagh, J. D. (1997). Supply Chain Management: More Than a New Name for Logistics. *The International Journal of Logistics Management*, 8(1), 1-14. <https://doi.org/10.1108/09574099710805556>
- Cornuéjols, A., & Miclet, L. (2010). *Apprentissage artificiel* (2e éd). Eyrolles.
- CSI Market. (2024). *Schlumberger Limited Comparisons to its Competitors and Market Share—CSIMarket*. CSI Market. https://csimarket.com/stocks/compet_glance.php?code=SLB
- Dbich, I. (2018). *AUDIT DES PERFORMANCES LOGISTIQUES*. UAE Tanger.
- Dertevelle, A. (2016). *L'Audit Logistique*. Librairie MatheO.
- Donyavi, Z., & Asadi, S. (2020). Diverse training dataset generation based on a multi-objective optimization for semi-Supervised classification. *Pattern Recognition*, 108, 107543. <https://doi.org/10.1016/j.patcog.2020.107543>
- Durango-Cohen, E., & Yano, C. (2006). Supplier Commitment and Production Decisions Under a Forecast-Commitment Contract. *Management Science*, 52, 54-67. <https://doi.org/10.1287/mnsc.1050.0471>
- Dwyer, F. R., Schurr, P. H., & Oh, S. (1987). Developing Buyer-Seller Relationships. *Journal of Marketing*, 51(2), 11-27. <https://doi.org/10.2307/1251126>
- Energy Information Administration. (2024, mai). *Short-Term Energy Outlook—U.S. Energy Information Administration (EIA)*. eia.org. https://www.eia.gov/outlooks/steo/report/global_oil.php
- Festa, F. (2021). *The Benefits & Downfalls of Volume Commitments*. <https://www.persuit.com/blog/legal-finance/benefits-downfalls-of-volume-commitments>
- France-Anne, G. L. F.-C. (2007). *Référentiel d'évaluation de la performance d'une chaîne logistique*.

- Gacogne, L. (2015). *Intelligence artificielle* Louis Gacogne—Librairie Mollat Bordeaux. <https://www.mollat.com/livres/1589615/louis-gacogne-intelligence-artificielle-cours-exercices-et-projets>
- Ganeshan, R., & Harrison, T. P. (1995). *An Introduction to Supply Chain Management*. https://gtl.csa.iisc.ac.in/scm/supply_chain_intro.html
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Networks* (arXiv:1406.2661). arXiv. <https://doi.org/10.48550/arXiv.1406.2661>
- Gourieroux, C., & Monfort, A. (1983). *Séries temporelles et modèles dynamiques*.
- Gourieroux, C., & Monfort, A. (1995). *Séries temporelles et modèles dynamiques*. *ECONOMICA*.
- Gundlach, G., Achrol, R., & Mentzer, J. (1995). The Structure of Commitment in Exchange. *Journal of Marketing*, 59. <https://doi.org/10.2307/1252016>
- Hamisultane, H. (2016). *ECONOMETRIE DES SERIES TEMPORELLES*. cel-01261174. <https://shs.hal.science/cel-01261174>
- Hubert, T. (2013). *Prévision de la demande et pilotage des flux en approvisionnement lointain*. tel-00879853. <https://theses.hal.science/tel-00879853>
- Hubert, & Thibault. (2013). *Prévision de la demande et pilotage des flux en approvisionnement lointain* [These de doctorat, Châtenay-Malabry, Ecole centrale de Paris]. <https://theses.fr/2013ECAP0012>
- Hugos, M. (2018). *Essentials of Supply Chain Management*. <https://doi.org/10.1002/9781119464495>
- Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., Cohen, S. N., & Weller, A. (2022). *Synthetic Data—What, why and how?* (arXiv:2205.03257). arXiv. <http://arxiv.org/abs/2205.03257>
- Kang, Y., Hyndman, R. J., & Li, F. (2020). GRATIS: GeneRAting Time Series with diverse and controllable characteristics. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 13(4), 354-376. <https://doi.org/10.1002/sam.11461>
- Kramar, V., & Alchakov, V. (2023). Evaluation of the Machine Learning Techniques for Forecasting the Seasonal Time Series. *2023 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT)*, 262-265. <https://doi.org/10.1109/USBREIT58508.2023.10158869>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25. https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html
- Krzyk, K. (2018, décembre 4). *Coding Deep Learning For Beginners*. Medium. <https://towardsdatascience.com/coding-deep-learning-for-beginners-types-of-machine-learning-b9e651e1ed9d>
- La finance pour tous. (2024, juin 14). *Le marché du pétrole*. La finance pour tous. <https://www.lafinancepourtous.com/decryptages/entreprise/secteurs-dactivites/le-prix-de-l-energie/le-prix-du-petrole/le-marche-du-petrole/>
- Lambert, D. M., Ellram, L. M., & Stock, J. R. (1998). *Fundamentals of logistics management*. Irwin/McGraw-Hill.
- Lisitsa, S., Levina, A., & Lepekhin, A. (2019). Supply-chain management in the oil industry. *E3S Web of Conferences*, 110, 02061. <https://doi.org/10.1051/e3sconf/201911002061>
- Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N. J. A. H., Ramírez-Quintana, M. J., & Flach, P. A. (2019). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*. <https://doi.org/10.1109/TKDE.2019.2962680>

- McKinsey. (2024). *Apparel brands deepen supplier relationships* | McKinsey. <https://www.mckinsey.com/industries/retail/our-insights/apparel-brands-finally-shift-to-deeper-supplier-relationships>
- Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining Supply Chain Management. *Journal of Business Logistics*, 22(2), 1-25. <https://doi.org/10.1002/j.2158-1592.2001.tb00001.x>
- Mialhe, N., & Hodes, C. (s. d.). *LA TROISIÈME ÈRE DE L'INTELLIGENCE ARTIFICIELLE*.
- Ministère de l'Énergie et des Mines. (2019). *Produits Pétroliers*. energy.gov. <https://energy.gov.dz>
- Monczka, R., Handfield, R., Giunipero, L., & Patterson, J. (2008). *Purchasing and Supply Chain Management*. Cengage Learning.
- Monczka, R. M., Trent, R. J., & Handfield, R. B. (1998). *Purchasing and Supply Chain Management*. South-Western College Pub.
- Mordor Intelligence. (2023). *Oilfield Services Market—OFS - Size, Companies & Share*. Mordor Intelligence. <https://www.mordorintelligence.com/industry-reports/global-oil-field-services-market-outlook-industry>
- Mordor Intelligence. (2024). *Analyse de la taille et de la part du marché du pétrole et du gaz en Algérie—Rapport de recherche de l'industrie—Tendances de croissance*. Mordor Intelligence. <https://www.mordorintelligence.com/fr/industry-reports/algeria-oil-and-gas-market>
- Oil Price. (2024). *Oil Price Charts*. OilPrice.Com. <https://oilprice.com/oil-price-charts/>
- OPEC. (2024). *OPEC: OPEC Basket Price*. Opec.org. https://www.opec.org/opec_web/en/data_graphs/40.htm
- Plambeck, E. L., & Taylor, T. A. (2006). *Simple Relational Contracts for Capacity Investment: Price-Only vs. Quantity-Commitment*. Stanford Graduate School of Business. <https://www.gsb.stanford.edu/faculty-research/publications/simple-relational-contracts-capacity-investment-price-only-vs-quantity>
- PWC. (2013). *Supplier Relationship Management How key suppliers drive your company's competitive advantage*.
- Pyke, D., Johnson, M., Harrison, T., Lee, H., & Neale, J. (2004). *Sourcing Strategy and Supplier Relationships: Alliances Versus eProcurement* (p. 77-89). https://doi.org/10.1007/0-387-27275-5_5
- Ren, Z., Cohen, M., Ho, T., & Terwiesch, C. (2006). Sharing Forecast Information in a Long-term Supply Chain Relationship. *Operations Research*.
- Saxena, A. (2008). *Enterprise Contract Management: A Practical Guide to Successfully Implementing an ECM Solution*. J. Ross Publishing.
- Scikit-learn: Machine learning in Python—Scikit-learn 1.5.0 documentation*. (2024). <https://scikit-learn.org/stable/>
- SLB. (2024). *SLB History*. SLB. <https://www.slb.com/>
- Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. (2015). *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*. <https://doi.org/10.48550/arXiv.1503.03585>
- SONATRACH. (2022). *Rapport 2022 | SONATRACH*. sonatrach.com. <https://sonatrach.com/rapports>
- Stanford—Unsupervised Learning*. (2024). [Stanford.edu]. Stanford. <https://stanford.edu/~shervine/teaching/cs-229/cheatsheet-unsupervised-learning>
- Statista. (2023). *Graphique: Pétrole: Trois pays assurent plus de 40 % de la production mondiale* | Statista. <https://fr.statista.com/infographie/19382/plus-gros-producteurs-de-petrole-brut-dans-le-monde/>
- Statista. (2024). *Brent crude oil price annually 1976-2024*. Statista. <https://www.statista.com/statistics/262860/uk-brent-crude-oil-price-changes-since-1976/>

- Talagala, T. S., Li, F., & Kang, Y. (2022). FFORMPP: Feature-based forecast model performance prediction. *International Journal of Forecasting*, 38(3), 920-943. <https://doi.org/10.1016/j.ijforecast.2021.07.002>
- Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *The American Statistician*, 72(1), 37-45.
- The Business Research Company. (2024). *Oilfield Services Market Size, Share, Trends, Report By 2033*. The Business Research Company. <https://www.thebusinessresearchcompany.com/report/oilfield-services-global-market-report>
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the Number of Clusters in a Data Set Via the Gap Statistic. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 63(2), 411-423. <https://doi.org/10.1111/1467-9868.00293>
- Wang, Z., Zhou, Z., Lu, H., & Jiang, J. (2020). Global and local sensitivity guided key salient object re-augmentation for video saliency detection. *Pattern Recognition*, 103, 107275. <https://doi.org/10.1016/j.patcog.2020.107275>
- Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). *Modeling Tabular data using Conditional GAN* (arXiv:1907.00503). arXiv. <http://arxiv.org/abs/1907.00503>
- Zone Bourse. (2024). *Schlumberger Limited: Actionnaires Dirigeants et Profil Société | SLB*. Zone Bourse. <https://www.zonebourse.com/cours/action/SCHLUMBERGER-LIMITED-14411/societe/>

Annexes :

Annexe A : Marché Pétrolier et Présentation d'entreprise

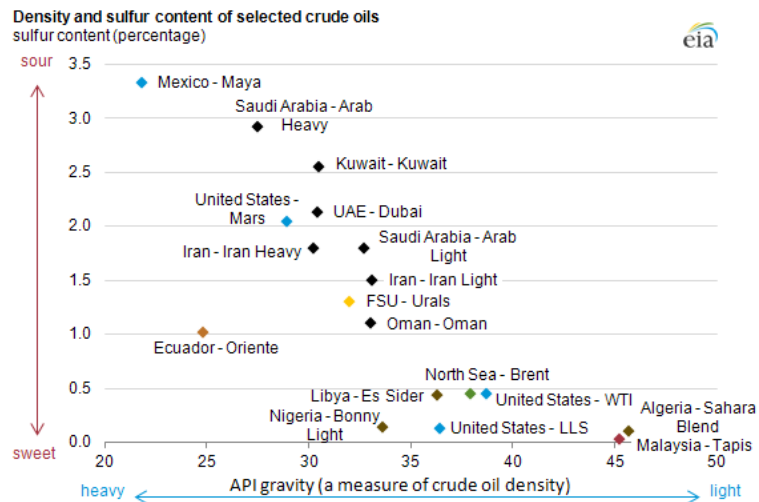


Figure A.1 – Caractéristiques de qualité des différents pétroles bruts (Energy Information Administration, 2024)

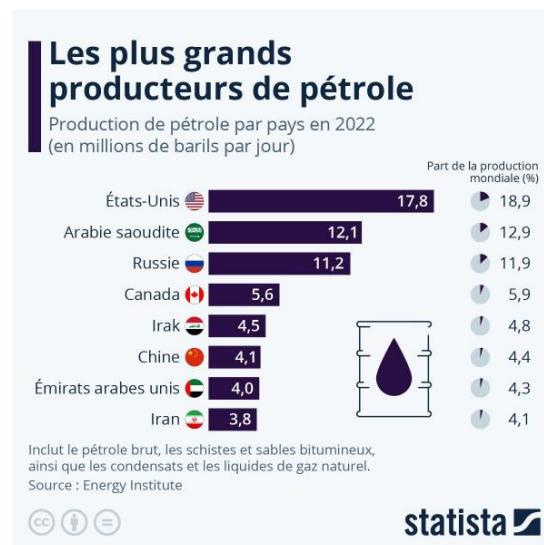


Figure A.2 – Les plus grands producteurs de pétrole (Statista, 2023)

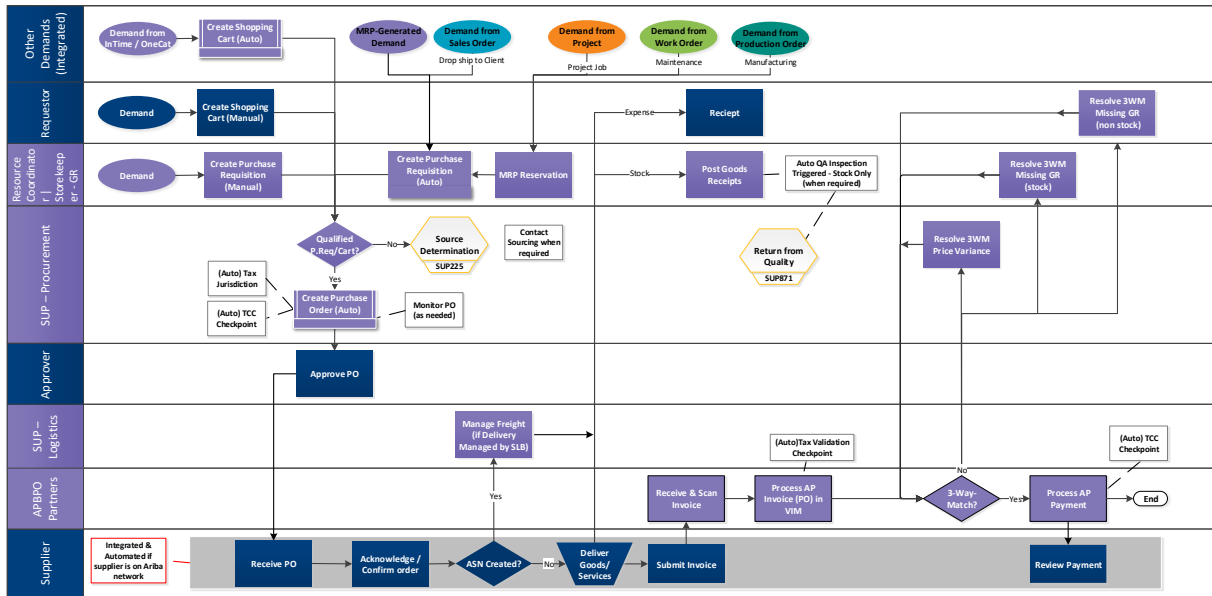


Figure A.3 – Processus Procure to Pay

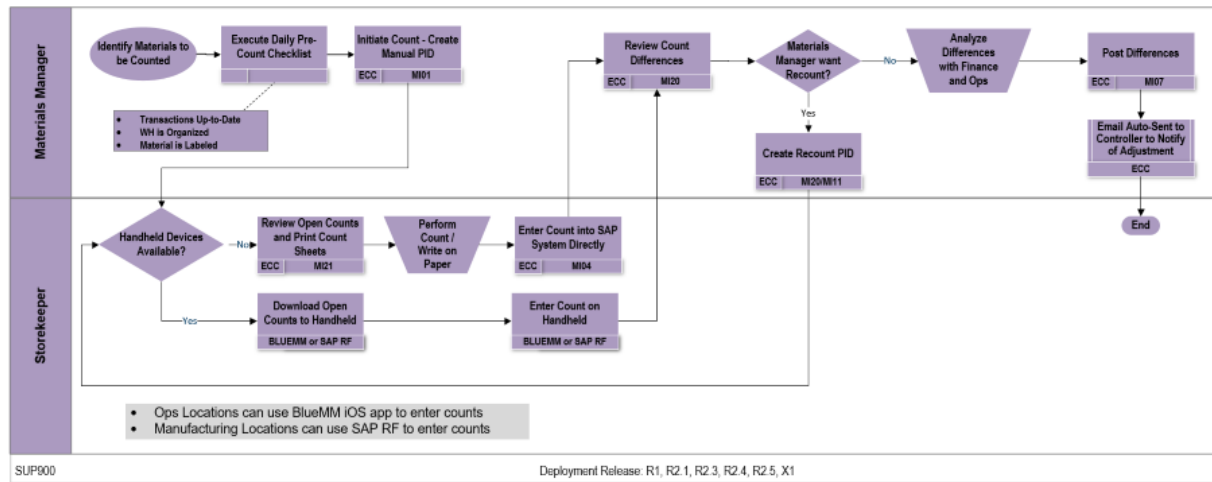


Figure A.4 – Processus Material Management (Physical Inventory)

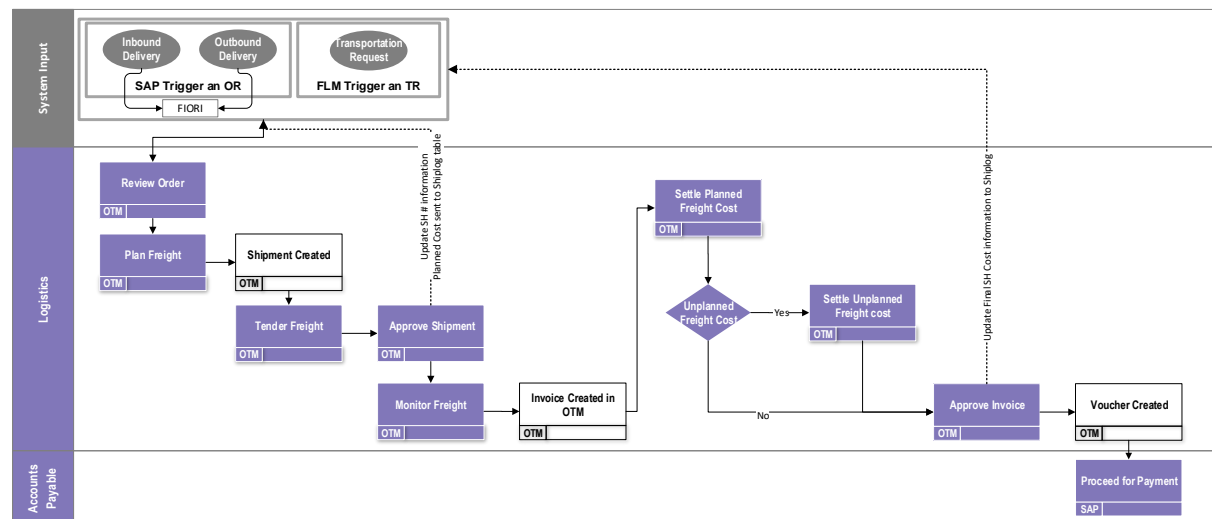


Figure A.5 – Processus Logistics Management (High level process)

Annexe B : Visualisations des sous catégories des fournisseurs

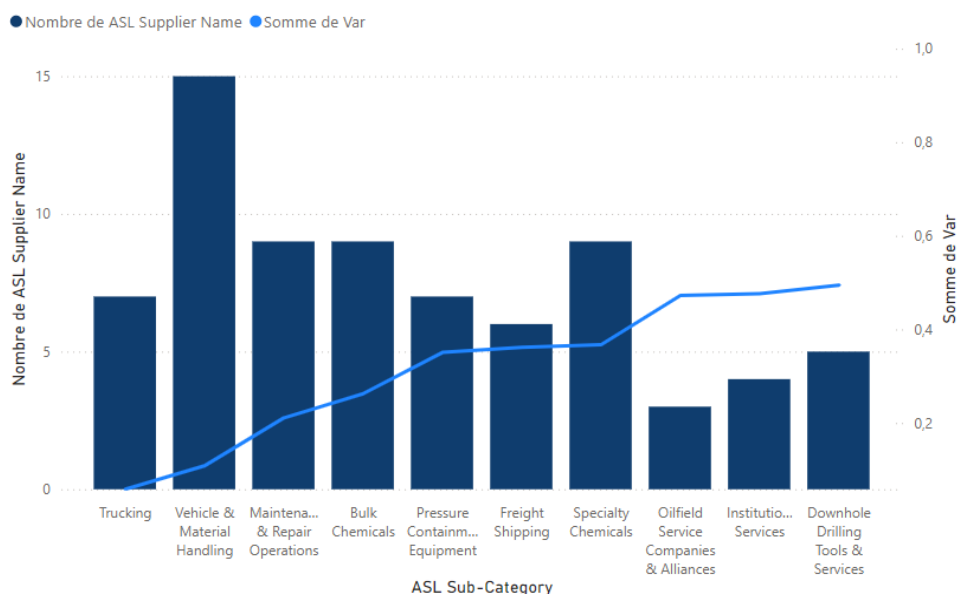


Figure B.1 – Nombre des fournisseurs et somme des variances par sous-catégorie de fournisseur

Variation des parts de marché des fournisseurs de produits de la catégorie M&S de la fonction TLM :
 Cette visualisation illustre bien la variation des parts de marché des fournisseurs de la catégorie de produits M&S en fonction du temps.

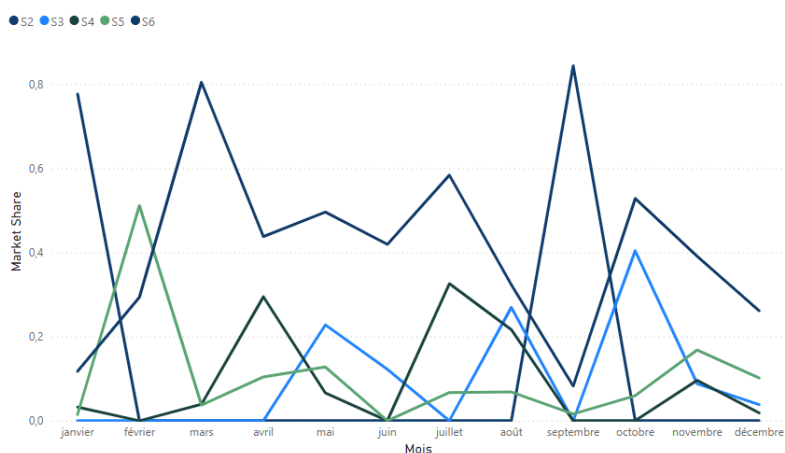


Figure B.2 – Variation des parts de marché des fournisseurs M&S

Annexe C : Référentiels d'audit logistique et comparaison

1) Présentation des différents référentiels d'audit logistique :

Pour auditer les supply chains, plusieurs référentiels sont couramment utilisés, parmi lesquels :

- Référentiel ASLOG :

Le guide d'audit logistique de l'ASLOG (Association française pour la Logistique) est reconnu comme un référentiel standard visant l'excellence logistique. Il se base sur huit processus de la Supply Chain : conception produit, achat, approvisionnement, production, livraison, stockage, ventes, maintenance, et

retour. Ce modèle permet à une entreprise de définir sa situation actuelle (« as is ») et d'évaluer sa performance logistique. Les auditeurs de l'ASLOG analysent la position de l'entreprise dans sa Supply Chain et proposent des recommandations pour des améliorations futures (« to be ») (France-Anne, 2007).

- Référentiel EVALOG :

Développé par les constructeurs et équipementiers de l'industrie automobile, EVALOG est un référentiel mondial standard pour évaluer et améliorer la performance logistique. Il couvre six thématiques principales : relation client, relation fournisseur, production, développement produit, stratégie d'entreprise et organisation. Utilisant un questionnaire pondéré, EVALOG permet l'auto-évaluation et l'audit logistique des partenaires, identifiant les priorités d'amélioration pour fiabiliser les flux physiques et logistiques. Son approche commune aux fournisseurs et clients facilite la collaboration et l'optimisation logistique (Acholi, 2021).

- Référentiel Supply Chain Master (SCM) :

Créé en 2007, le Référentiel Supply Chain Master évalue la maturité logistique des PME de 10 à 250 salariés dans divers secteurs tels que les produits de grande consommation et la mécanique. Il analyse les pratiques logistiques de l'entreprise, identifie ses forces et faiblesses sur une échelle de 0 à 4, et définit un plan d'amélioration prioritaire. Conçu par des experts de la supply chain, il s'appuie sur les normes et bonnes pratiques de grands organismes tels que AFNOR, APICS et ISO. Les niveaux de maturité vont de l'absence de structure logistique (0) à une supply chain collaborative avancée (4), avec des étapes intermédiaires de logistique fragmentée (1), d'excellence interne (2), et d'intégration externe (3) (Dbich, 2018).

- Référentiel WCL (World Class Logistics) :

Développé par la Michigan State University, le Référentiel WCL évalue la performance logistique des entreprises à partir d'un questionnaire de 68 questions, générant des indicateurs pour des comparaisons inter-entreprises. Appliqué annuellement à différents secteurs tels que l'automobile et l'électronique, ce modèle inclut des compétences comme le positionnement stratégique, la stratégie globale, la synchronisation de la Supply Chain, la structuration des infrastructures et des réseaux, ainsi que l'organisation et l'engagement des ressources humaines (Dbich, 2018).

- Référentiel SCOR :

Le modèle SCOR (Supply Chain Operations Reference) est une méthodologie standardisée pour décrire et évaluer les flux au sein de la Supply Chain. Construit par des industriels, il structure des processus types et identifie les meilleures pratiques. Il s'articule autour de cinq processus principaux : planification, approvisionnement, fabrication, livraison, et gestion des retours. SCOR relie la stratégie organisationnelle à la gestion opérationnelle, couvrant l'ensemble des interactions de la Supply Chain, facilitant l'analyse des performances et l'optimisation des processus à travers des démarches de benchmarking et une mise en œuvre structurée en quatre étapes : stratégique, opérationnelle, systémique, et mise en production (France-Anne, 2007).

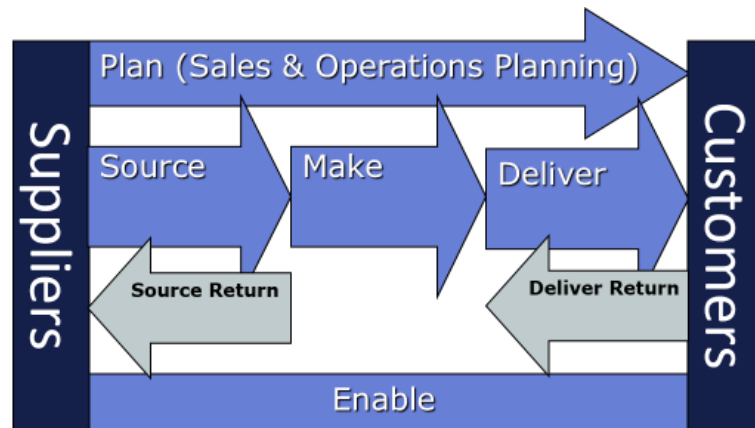


Figure C.1 – Schématisation des processus SCOR

2) Comparaison des différents référentiels :

Critères	SCOR	EVALOG	ASLOG	SCM
Domaine d'application	Général	Industrie automobile	Général	Général
Périmètre d'intervention	Supply Chain	Supply Chain	Supply Chain	Supply Chain
Envergure d'entreprise	Très grandes entreprises	Très grandes entreprises	Grandes entreprises	Petites et moyennes entreprises
Niveau de détail des processus	Description détaillée des activités	Description détaillée des activités	Description détaillée des activités	Processus
Outils supports	- Grilles d'évaluations - Indicateurs de performance	Questionnaire des bonnes pratiques par activité	Questionnaire des bonnes pratiques par activité	Indicateurs représentés sous quatre leviers
Meilleures pratiques identifiées	Diagnostic de la performance logistique Utilisation d'une notation	Plan d'amélioration continue, référentiel unique pour fournisseur et client	Diagnostic de la performance logistique Utilisation d'une notation	Identification des indicateurs selon 4 leviers
Intégration de la dimension environnementale dans la performance	Oui	Oui	Oui	Oui

Tableau C.1 – Synthèse et comparaison des principaux référentiels existants (Acholi, 2021)

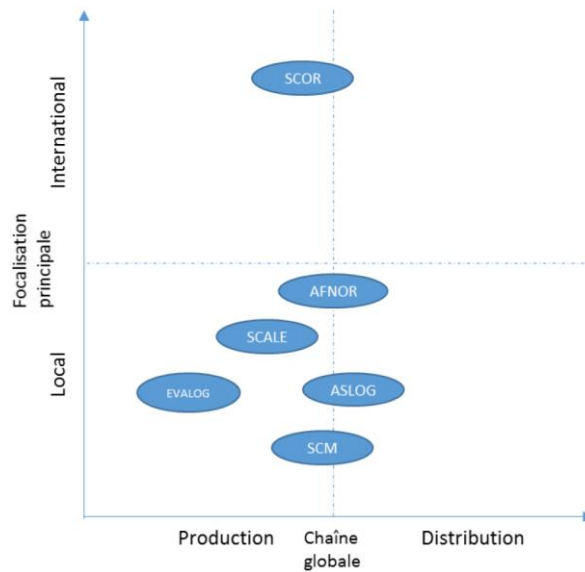


Figure C.2 – Positionnement des référentiels (Dertevelle, 2016)

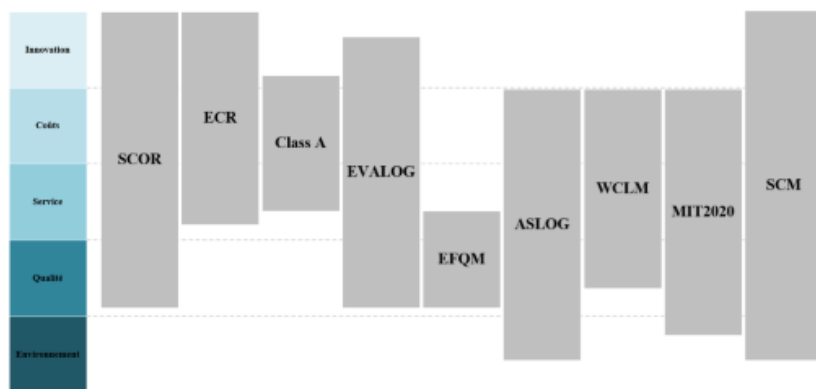


Figure C.3 – Domaines couverts par les référentiels logistiques (Dbich, 2018)

En prenant en compte ces critères, il est clair que le référentiel SCOR est le plus approprié pour l'audit de l'entreprise Schlumberger. Voici pourquoi :

- SLB est une entreprise de très grande taille avec une orientation internationale, ce qui correspond bien à l'application du modèle SCOR.
- Comme illustré dans la figure précédente, SCOR s'applique à l'ensemble de la Supply Chain.
- SCOR couvre le domaine de l'innovation, un élément clé de la philosophie de SLB, en plus des autres domaines pertinents (Figure C.3).

De plus, l'approche SCOR permet de définir clairement les activités appartenant à la Supply Chain de manière exhaustive et structurée. Enfin, SCOR est le seul modèle de référence complet conçu pour l'ensemble de la chaîne, applicable à toutes les entreprises, qu'elles soient industrielles ou prestataires de services.

3) Décomposition en troisième niveau des processus audités :

Niveau 1	Processus de planification		
Niveau 2	Planification de la Supply Chain	Planification des commandes	Planification des approvisionnement
Niveau 3	<p>P1.1 Capturer les signaux du marché externe</p> <p>P1.2 Agréger les besoins de la Supply Chain</p> <p>P1.3 Évaluer et créer la réponse en matière d'approvisionnement</p> <p>P1.4 Équilibrer les signaux du marché externe, les besoins de la SC et sa réponse en utilisant des techniques de modélisation</p> <p>P1.5 Replanifier, analyser et sélectionner la réponse optimale de la Supply Chain</p> <p>P1.6 Communiquer le plan aux fonctions d'exécution et évaluer</p>	<p>2.1 Agréger les besoins en commandes clients</p> <p>P2.2 Évaluer et créer la réponse initiale aux commandes</p> <p>P2.3 Équilibrer les besoins et la réponse aux commandes</p> <p>P2.4 Replanifier, analyser et sélectionner la réponse optimale aux commandes</p> <p>P2.5 Communiquer la réponse finalisée aux commandes à P1.3</p>	<p>P3.1 Agréger les besoins en approvisionnement</p> <p>P3.2 Évaluer et créer la réponse initiale en approvisionnement</p> <p>P3.3 Équilibrer les besoins et la réponse en approvisionnement</p> <p>P2.4 Replanifier, analyser et sélectionner la réponse optimale aux approvisionnements</p> <p>P3.5 Communiquer la réponse finale en approvisionnement à P1.3</p>

Tableau C.2 – Décomposition en niveau 3 du processus de planification

Niveau 1	Processus de planification		
Niveau 2	Approvisionnement stratégique	Approvisionnement direct	Approvisionnement indirect
Niveau 3	<p>S1.1 Définir le besoin commercial</p> <p>S1.2 Effectuer une analyse du marché de l'approvisionnement</p> <p>S1.3 Développer la stratégie d'approvisionnement</p> <p>S1.4 Tester le marché en pré-procurement</p> <p>S1.5 Approvisionner le marché</p> <p>S1.6 Préqualifier les fournisseurs</p> <p>S1.7 Déterminer le niveau</p>	<p>S2.1 Établir le signal de commande</p> <p>S2.2 Programmer la livraison des produits</p> <p>S2.3 Gérer le transport entrant</p> <p>S2.4 Recevoir le produit</p> <p>S2.5 Inspecter et vérifier</p> <p>S2.6 Transférer le produit</p> <p>S2.7 Autoriser le paiement au fournisseur</p>	<p>S3.1 Établir le signal de commande</p> <p>S3.2 Programmer la livraison des produits</p> <p>S3.3 Gérer le transport entrant</p> <p>S3.4 Recevoir le produit</p> <p>S3.5 Inspecter et vérifier</p> <p>S3.6 Transférer le produit</p> <p>S3.7 Autoriser le paiement au fournisseur</p>

	de collaboration S1.8 Inviter à soumissionner / Demande de devis S1.9 Analyser les offres et sélectionner les fournisseurs S1.10 Négocier et attribuer le contrat		
--	--	--	--

Tableau C.3 – Décomposition en niveau 3 du processus d'approvisionnement

4) Grille d'évaluation de l'audit SCOR :

a) Processus de planification :

- Planification de la supply chain :

Métrique/ Pratiques	Intitulés	Réponse
Métrique	RL.3.47 Précision des prévisions	Non
	RL.3.55 Variabilité du rendement	Non
	RS.3.62 Identifier, hiérarchiser et regrouper les exigences de la Supply Chain Temps de cycle	Oui
	AM.2.2 Jours d'inventaire	Oui
	RS.3.39 Équilibrer les ressources en produits avec le temps de cycle des exigences des produits	Oui
	RL.3.46 Taux de remplissage	Oui
Pratiques	BP.002 Stratégies de gestion des risques	Non
	BP.014 Planification et prévision de la demande	Non
	BP.016 Planification du réseau d'approvisionnement	Non
	BP.021 Planification des ventes et des opérations (S&OP)	Oui
	BP.024 Optimisation de la supply chain (SCO)	Oui
	BP.030 Exactitude des enregistrements d'inventaire	Oui
	BP.033 Amélioration de la prévision de la demande traditionnelle	Non
	BP.056 Assurance qualité des matières premières fournies par les	Oui

	fournisseurs	
	BP.061 Réduire le niveau d'inventaire non stratégique	Non
	BP.063 Optimiser les décisions d'approvisionnement vers le point d'approvisionnement local	Oui
	BP.086 Planification du réseau d'approvisionnement	Non
	BP.082 Amélioration continue	Non

Tableau C.4 – Grille d'évaluation du sous processus Planification de la Supply Chain

- Planification des approvisionnements :

Métrique/ Pratiques	Intitulés	Réponse
Métrique	AG.3.12 Volume d'approvisionnement actuelle	Oui
	AG.3.14 Approvisionnement de la demande - Contraintes des fournisseurs	Non
	CO.2.001 Coût de la planification	Oui
	AM.2.2 Jours d'inventaire	Oui
	RS.3.52 Établir le temps de cycle des plans d'approvisionnement	Non
Pratiques	BP.001 Gestion des risques de Supply Chain (SCRM)	Non
	BP.024 Optimisation de la Supply Chain (SCO)	Oui
	BP.027 Réapprovisionnement des stocks basé sur la demande	Oui
	BP.029 Gestion des stocks à l'aide de l'optimisation du réseau de la Supply Chain	Non
	BP.042 Révision des termes et conditions d'approvisionnement	Oui
	BP.056 Assurance qualité des matières premières fournies par les fournisseurs	Oui
	BP.068 Rapport sur la performance des fournisseurs	Oui
	BP.096 Planification logistique et d'entrepôt	Oui
	BP.100 Approvisionnement stratégique	Oui
	BP.101 Stratégie d'achat et d'approvisionnement	Non
BP.134 Évaluation des fournisseurs	Oui	

	BP.161 Analyse des dépenses	Oui
	BP.162 Accord ou partenariat à long terme avec les fournisseurs	Oui
	BP.163 Rationalisation de la base de fournisseurs	Oui
	BP.279 Segmentation des fournisseurs	Oui
	BP.294 Rapport sur la durabilité	N/A

Tableau C.5 – Grille d'évaluation du sous processus Planification des approvisionnements

- Planification des commandes :

Métrique/ Pratiques	Intitulés	Réponse
Métriques	AG.2.1 Agilité des commandes	Oui
	RS.1.1 Temps de cycle d'exécution des commandes	Oui
	RL.3.37 La précision des prévisions	Non
Pratiques	BP.017 Planification de la distribution	Oui
	BP.027 Réapprovisionnement des stocks basé sur la demande	Oui
	BP.082 Amélioration continue	Non
	BP.122 Inventaire géré par le fournisseur (VMI)	Non
	BP.201 Gestion dynamique des stocks	Oui

Tableau C.6 – Grille d'évaluation du sous processus Planification des commandes

- b) Processus d'approvisionnement :

- Approvisionnements stratégiques :

Métrique/ Pratiques	Intitulés	Réponse
Métriques	AG.3.12 Volume d'approvisionnement actuel	Oui
	RS.3.52 Temps de cycle d'établissement des plans d'approvisionnement	Oui
	AM.3.7 Pourcentage de matériaux recyclables ou réutilisables	Non
	RL.3.34 Pourcentage de fournisseurs respectant les critères et les métriques environnementaux	Oui

	RL.3.35 Pourcentage de fournisseurs certifiés ISO 14001	N/A
	RS.2.2 Temps de cycle d'approvisionnement	Oui
	RS.3.10 Temps de cycle de sélection des fournisseurs et de négociation	Oui
Pratiques	BP.049 Planification lean	Non
	BP.061 Réduction du niveau d'inventaire non stratégique	Oui
	BP.149 Mise en œuvre et analyse de la stratégie de gestion des risques de la Supply Chain (SCRM)	Oui
	BP.173 Surveillance des risques de la Supply Chain	Non
	BP.184 Planification de scénarios	Oui
	BP.194 Analyse prédictive	Non
	BP.285 Passeport matériel	N/A

Tableau C.7 – Grille d'évaluation du sous processus Approvisionnement Stratégique

- Approvisionnement direct :

Métrique/ Pratiques	Intitulés	Réponse
Métriques	AG.3.12 Volume d'approvisionnement actuel	Oui
	CO.2.6 Coût des matières directes	Oui
	RL.3.44 Pourcentage de plannings modifiés dans le délai de préavis du fournisseur	Non
	RS.3.9 Cycle de temps de retour de la commande complète	Non
	RS.3.36 Nombre moyen de jours par modification d'ingénierie	Non
	CO.3.13 Frais de transport entrant et droits de douane	Oui
	RL.2.5 Pourcentage de commandes livrées en totalité par le fournisseur	Oui
	RL.3.15 Réalisation du fournisseur à la date d'engagement initiale de l'organisation	Non
	RL.3.37 Pourcentage de commandes ou de lignes reçues sans défaut	Oui
	RL.3.38 Pourcentage de commandes ou de lignes reçues à temps selon les exigences de la demande	Oui

	RS.3.101 Recevoir et vérifier le cycle de temps de produit	Non
	CO.3.14 Réception et stockage de matériel	Non
	RL.2.8 Commande de fournisseur en parfait état	Oui
	AM.2.2 Jours d'inventaire	Oui
Pratiques	BP.035 Examen des règles métier	Oui
	BP.056 Assurance qualité des matières premières fournies par le fournisseur	Oui
	BP.100 Approvisionnement stratégique	Oui
	BP.131 Demande d'information (RFI)	Oui
	BP.132 Demande de devis (RFQ)	Oui
	BP.134 Évaluation du fournisseur	Oui
	BP.145 Collaboration avec les fournisseurs	Non
	BP.147 Inspection de la réception des marchandises	Oui
	BP.161 Analyse des dépenses	Oui
	BP.163 Rationalisation de la base d'approvisionnement	Oui
	BP.197 Visualisation et visibilité avancées des données	Non
	BP.286 Plateformes de partage	Oui
	BP.300 Traçabilité	Oui

Tableau C.8 – Grille d'évaluation du sous processus Approvisionnement direct

- Approvisionnement indirect :

Métrique/ Pratiques	Intitulés	Réponse
Métriques	AG.3.12 Volume d'approvisionnement actuel	Oui
	AG.3.6 Volumes d'approvisionnement supplémentaires obtenus en 30 jours	Oui
	CO.3.13 Fret entrant et droits	Oui
	RL.2.5 Pourcentage de commandes livrées en totalité par le fournisseur	Non
	RL.3.30 Pourcentage de factures sans erreur	Oui

	RL.3.37 Pourcentage de commandes ou de lignes reçues sans défaut	Oui
	RL.2.8 Commande fournisseur en parfait état	Oui
	RS.3.6 Délai de cycle d'autorisation de paiement du fournisseur	Oui
Pratiques	BP.041 Optimisation du transport	Non
	BP.155 Procédures opérationnelles standard	Oui
	BP.183 Planification intégrée des opérations (IBP)	Oui
	BP.283 Produit en tant que service	N/A

Tableau C.9 – Grille d'évaluation du sous processus Approvisionnement indirect

Annexe D : Modèles prévisionnels

1) La définition des autres concepts de prévision :

Nous distinguons généralement deux grandes familles de prévision pour estimer les besoins futurs de l'entreprise : la famille des méthodes qualitatives, basées sur l'expertise et le jugement des managers, et la famille des méthodes quantitatives, fondées sur des bases statistiques et mathématiques (Hubert, 2013).

- **Méthodes qualitatives :**

Les méthodes qualitatives utilisent des données subjectives. Le résultat dépend donc fortement de l'expérience et de l'expertise de ceux en charge de la réalisation de ces prévisions. Parmi les méthodes les plus connues, on cite :

- La méthode Delphi
- Les enquêtes auprès des consommateurs
- Les analogies historiques
- Les panels d'experts

Bien que les méthodes qualitatives permettent de prendre en compte des facteurs intangibles et d'obtenir des résultats avec peu de données, comme dans le cas de l'introduction d'un nouveau produit, elles restent cependant assez longues et coûteuses à mettre en œuvre lorsqu'on sollicite l'avis d'experts ou de consommateurs. De plus, il existe un risque d'obtenir des prévisions biaisées ou arbitraires.

- **Méthodes quantitatives :**

Contrairement aux méthodes qualitatives qui sont basées sur l'expérience et l'opinion des managers, les méthodes quantitatives sont fondées sur l'analyse statistique du comportement et de l'évolution des données historiques, appelées séries temporelles, ou sur la détermination d'une relation cause à effet entre certaines variables explicatives, dites exogènes, pour expliquer une variable à prédire, dite endogène.

Les méthodes quantitatives de prévision peuvent être classées en deux grandes catégories :

- **Les modèles causaux**

Ces méthodes permettent de déterminer les liens de causalité entre certaines variables indépendantes (variables exogènes) et l'objet de la prévision (variable endogène) à travers une théorie économique. Elles développent des modèles de régression à partir des séries chronologiques relevées de ces variables pour prévoir le comportement de la variable endogène selon l'évolution des variables exogènes.

Cependant, en raison de la complexité de ces méthodes, notamment en termes d'identification des variables explicatives, elles sont généralement difficiles à appliquer et nécessitent plus de temps et d'énergie comparativement aux autres approches.

- **Les modèles auto-projectifs**

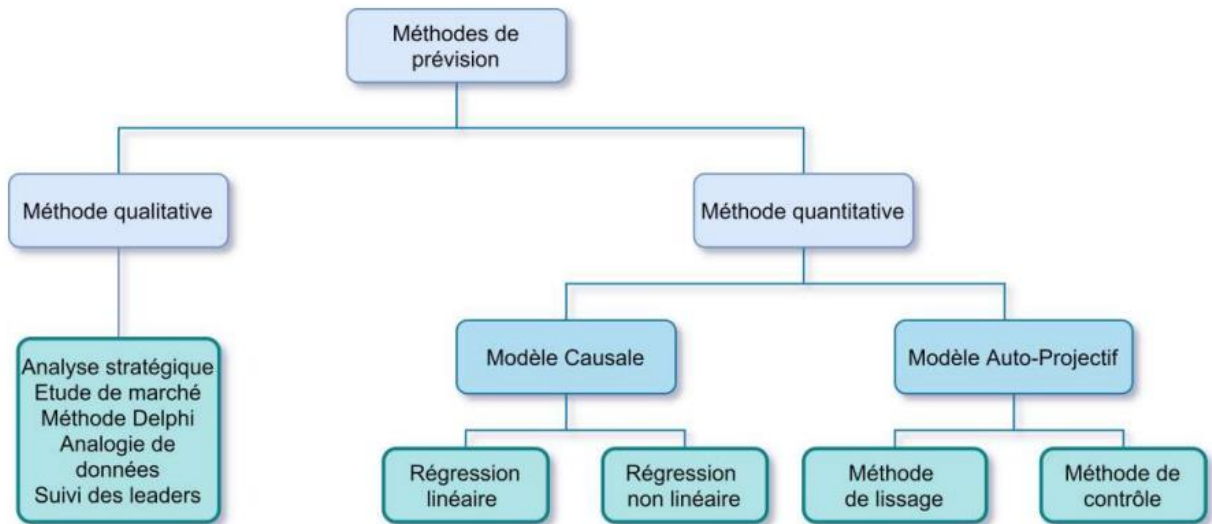


Figure D.1 – Schéma des méthodes de prévision

2) Schémas de la méthodologie Box Jenkins :

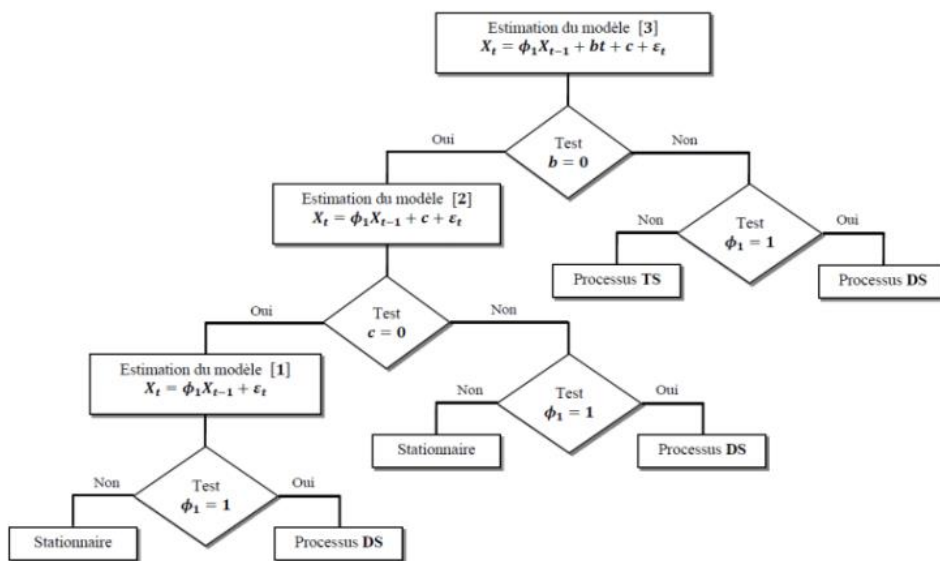


Figure D.2 – Stratégie simplifiée de Dicky Fuller

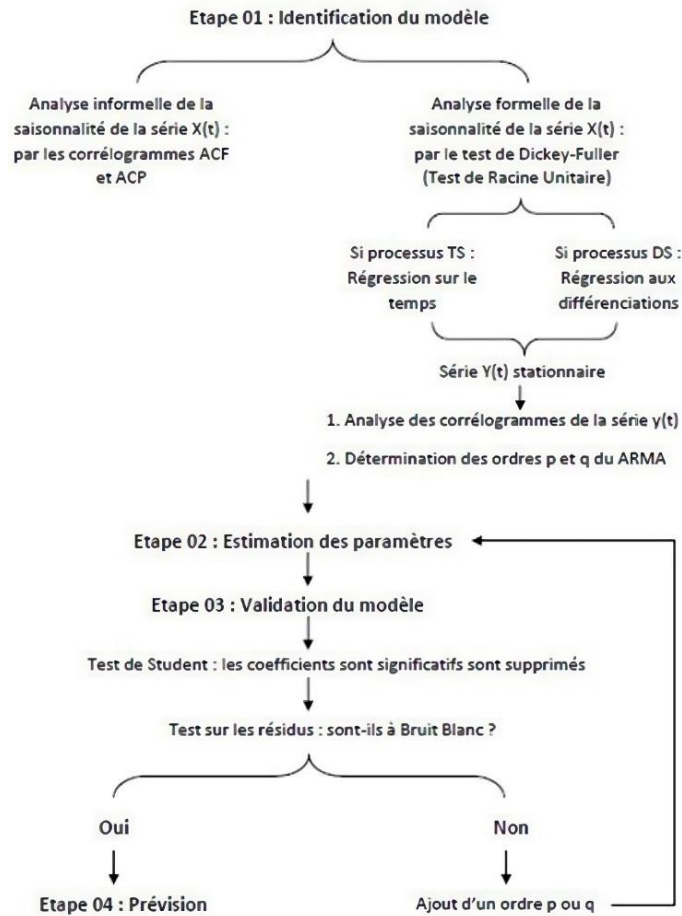


Figure D.3 – Schéma de la méthodologie Box Jenkins

Annexe E : Les fonctions d'activation des réseaux de neurones

La fonction Sigmoidale : La fonction sigmoïde ou la fonction logistique est l'une des fonctions les plus utilisées en apprentissage profond notamment dans la couche de sortie pour un problème de 25 classification binaire. Elle est bornée entre 0 et 1 et elle peut être interprétée comme la probabilité que le neurone s'active. La sortie de la fonction sigmoïde est donnée par l'équation et la figure suivante :

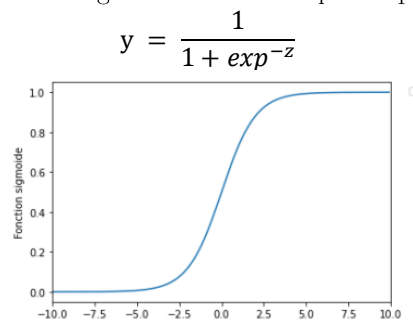


Figure E.1 – Illustration de la fonction Sigmoidale

La fonction ReLu : La fonction RELU (en anglais rectified linear unit) est récemment devenue un choix incontournable pour les réseaux de neurones. Cette fonction nous renvoie 0 si l'entrée est inférieure à 0 et retourne l'entrée elle-même si elle est plus grande que 0. La sortie de la fonction relue est donnée par l'équation et la figure suivante : $y = \max(0, z)$

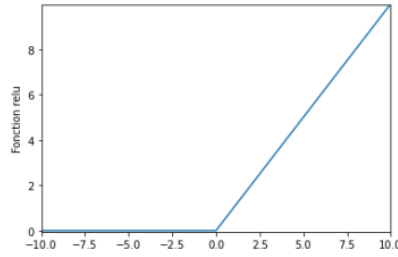


Figure E.2 – Illustration de la fonction ReLu

La fonction Softmax : La fonction softmax ou régression Softmax est une généralisation de la fonction sigmoïde que nous pouvons l'utiliser en couche de sortie dans le cas d'une classification multi-classes. La sortie de la fonction softmax est donnée par l'équation suivante, avec C le nombre de classe de sortie :

$$y_i = \frac{\exp^{-z_i}}{\sum_{j=1}^C \exp^{-z_j}}$$

La fonction Tanh (Hyperbolic Tangent Function) : La fonction tangente hyperbolique connue sous le nom de fonction tanh, est une fonction centrée sur le zéro, plus douce que la fonction sigmoïde et dont la plage se situe entre -1 et 1. La sortie de la fonction tanh est donnée par l'équation et la figure suivante :

$$y_i = \frac{\exp^z - \exp^{-z}}{\exp^z + \exp^{-z}}$$

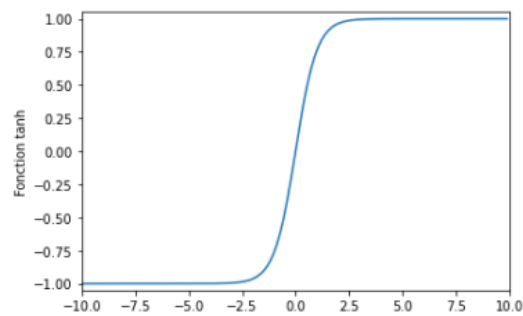


Figure E.3 – Illustration de la fonction Tanh

Annexe F : Partitionnement de données

```
k_values_silhouette = range(2, 9)
table_silhouette = pd.DataFrame(columns=['k', 'silhouette_score'])

for k in k_values_silhouette:
    kmeans = KMeans(n_clusters=k, random_state=123).fit(df_scaled)
    silhouette_avg = silhouette_score(df_scaled, kmeans.labels_)
    table_silhouette = table_silhouette.append({'k': k, 'silhouette_score': silhouette_avg}, ignore_index=True)

print("silhouette Scores:")
print(table_silhouette)
```

Figure F.1 – Capture du code utilisé pour avoir table du silhouette score


```
kmeans = KMeans(n_clusters = 3, random_state = 111)
y_means = kmeans.fit(df_scaled)
labels = kmeans.labels_
```

utputs are collapsed ...

```
pd.Series(kmeans.labels_).value_counts()
```

Figure F.2 – Capture de l’algorithme kmeans pour créations de 3 clusters et nombre de pièces par cluster

```
distance_matrix = linkage(X, method = 'ward', metric = 'euclidean')

labels = fcluster(distance_matrix, 5, criterion='maxclust')
print(len(set(labels)) )
print(f"Silhouette Coefficient: {metrics.silhouette_score(X, labels):.3f}")
print(f"Calinski-harabasz : {metrics.calinski_harabasz_score(X, labels):.3f}")
print(f"Indice de Davies-Bouldin : {metrics.davies_bouldin_score(X, labels):.3f}")

dn = dendrogram(distance_matrix)
plt.show()
```

Figure F.3 – Capture d’écran de l’algorithme CAH

```
db = DBSCAN(eps=0.2, min_samples=30).fit(X)
labels = db.labels_

n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)
n_noise_ = list(labels).count(-1)

print("Estimated number of clusters: %d" % n_clusters_)
print("Estimated number of noise points: %d" % n_noise_)
print(f"Silhouette Coefficient: {metrics.silhouette_score(X, labels):.3f}")
print(f"Calinski-harabasz : {metrics.calinski_harabasz_score(X, labels):.3f}")
print(f"Indice de Davies-Bouldin : {metrics.davies_bouldin_score(X, labels):.3f}")
```

Figure F.4 – Capture de l’algorithme DbSCAN

```
kmeans = MiniBatchKMeans(n_clusters=3, random_state=0, batch_size=32,
                          max_iter=10, n_init="auto").fit(df_scaled)
labels = kmeans.labels_
print(f"Silhouette Coefficient: {metrics.silhouette_score(X, labels):.3f}")
print(f"Calinski-harabasz : {metrics.calinski_harabasz_score(X, labels):.3f}")
print(f"Indice de Davies-Bouldin : {metrics.davies_bouldin_score(X, labels):.3f}")
```

Figure F.5 – Capture de l’algorithme Mini Batch kmeans

```
bandwidth = estimate_bandwidth(X, quantile=0.9, n_samples=500)
ms = MeanShift(bandwidth=bandwidth, bin_seeding=True)
ms.fit(X)
labels = ms.labels_
cluster_centers = ms.cluster_centers_

labels_unique = np.unique(labels)
n_clusters_ = len(labels_unique)

print("number of estimated clusters : %d" % n_clusters_)
print(f"Silhouette Coefficient: {metrics.silhouette_score(X, labels):.3f}")
print(f"Calinski-harabasz : {metrics.calinski_harabasz_score(X, labels):.3f}")
print(f"Indice de Davies-Bouldin : {metrics.davies_bouldin_score(X, labels):.3f}")
```

Figure F.6 – Capture de l’algorithme Meanshift pour créations des clusters

Annexe G : Résultats de la classification XYZ

N°	Numéro de la pièce	Critère de variabilité (CV)	Pourcentage du CV	Pourcentage du CV cumulé	Classe XYZ
1	B021698	0,7838	0,0107	0,0107	X
2	T6016548	0,8106	0,0110	0,0217	X
...
8	B013366	0,9991	0,0136	0,1004	Y
9	B015518	1,0072	0,0137	0,1141	Y
...
55	T6017303	1,6037	0,0219	0,9551	Z
57	H712973	1,6810	0,0229	1,0000	Z

Tableau G.1 – Résultats de la classification XYZ

Annexe H : Résultat des statistiques

Forecast Evaluation						
Date: 06/22/24 Time: 01:19						
Sample: 2021M10 2024M08						
Included observations: 35						
Evaluation sample: 2021M10 2024M08						
Number of forecasts: 1						
Evaluation statistics						
Forecast	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
PART_A	2157.103	1665.355	35.39781	32.94465	0.159531	0.898089

Figure H.1 – Résultats des prévisions par le modèle LES

Forecast Evaluation						
Date: 06/22/24 Time: 01:23						
Sample: 2021M10 2024M08						
Included observations: 35						
Evaluation sample: 2021M10 2024M08						
Number of forecasts: 1						
Evaluation statistics						
Forecast	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
PART_A	2587.691	2112.933	61.07743	44.70717	0.188051	1.635023

Figure H.2 – Résultats des prévisions par le modèle LED

Forecast Evaluation						
Date: 06/22/24 Time: 01:06						
Sample: 2021M10 2024M08						
Included observations: 35						
Evaluation sample: 2021M10 2024M08						
Number of forecasts: 1						
Evaluation statistics						
Forecast	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
PART_A	2195.287	1664.626	32.24475	32.37564	0.158900	0.855366

Figure H.3 - Résultats des prévisions par le modèle HWNS

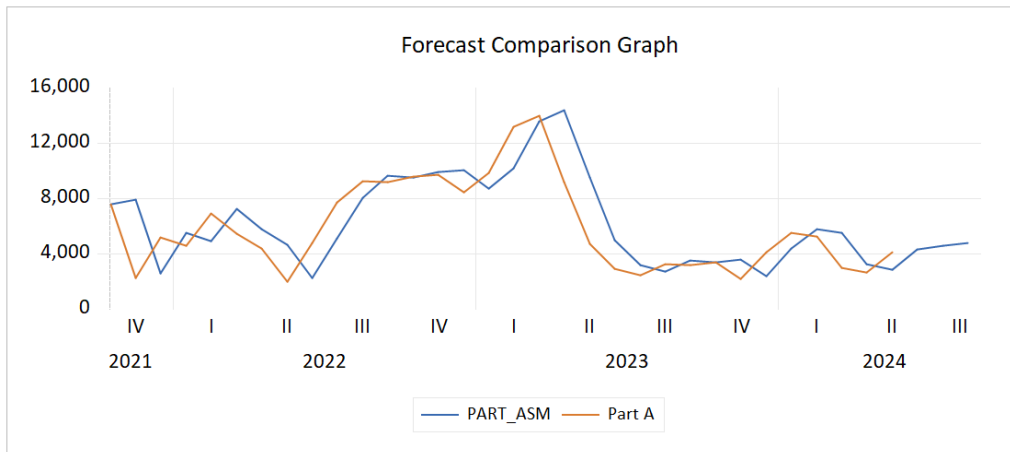


Figure H.4 – Résultats des prévisions par HWNS

Résultats de la méthodologie Box Jenkins

1) Test de Dickey Fuller

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(PART_A)
 Method: Least Squares
 Date: 06/21/24 Time: 22:38
 Sample (adjusted): 2021M12 2024M05
 Included observations: 30 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PART_A(-1)	-0.271690	0.112991	-2.404524	0.0233
D(PART_A(-1))	0.285127	0.167937	1.697820	0.1010
C	1719.069	762.7617	2.253743	0.0325

Figure H.5 - Résultat du test DF avec contrainte

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(PART_A,2)
 Method: Least Squares
 Date: 06/21/24 Time: 23:01
 Sample (adjusted): 2022M01 2024M05
 Included observations: 29 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(PART_A(-1))	-0.823720	0.222783	-3.697404	0.0010
D(PART_A(-1),2)	0.153813	0.162509	0.946487	0.3526
C	-66.47640	359.9212	-0.184697	0.8549

Figure H.7 - Résultat du test DF avec contrainte sur la série différencié

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(PART_A,2)
 Method: Least Squares
 Date: 06/21/24 Time: 22:56
 Sample (adjusted): 2022M01 2024M05
 Included observations: 29 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(PART_A(-1))	-0.828484	0.230902	-3.588038	0.0014
D(PART_A(-1),2)	0.154049	0.165697	0.929706	0.3614
C	20.88064	845.4194	0.024699	0.9805
@TREND("2021M10")	-5.137249	44.78953	-0.114698	0.9096

Figure H.6 - Résultat du test DF avec tendance et contrainte sur la série différencié

Null Hypothesis: D(PART_A) has a unit root
 Exogenous: None
 Lag Length: 1 (Automatic - based on SIC, maxlag=7)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.760942	0.0005
Test critical values:		
1% level	-2.647120	
5% level	-1.952910	
10% level	-1.610011	

Figure H.8 - Résultat du test DF modèle [3] sur la série différencié

2) Résultat du modèle Automatic ARIMA :

Automatic ARIMA Forecasting
 Selected dependent variable: DPART_A
 Date: 06/21/24 Time: 23:27
 Sample: 2021M10 2024M05
 Included observations: 31
 Forecast length: 0
 Model maximums: (4,4)2(0,0)
 Regressors: C

Number of estimated ARMA models: 25
 Number of non-converged estimations: 0
 Selected ARMA model: (2,3)(0,0)
 AIC value: 18.093820541

Figure H.9 : Résultat de l'automatic ARIMA

Dependent Variable: D(PART_A)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 06/22/24 Time: 00:33
 Sample: 2021M11 2024M05
 Included observations: 31
 Failure to improve objective (non-zero gradients) after 11 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-100.0940	134.2478	-0.745592	0.4624
AR(2)	0.084198	0.220072	0.382595	0.7050
MA(3)	-1.000000	2611.013	-0.000383	0.9997
SIGMASQ	2613343.	3.41E+09	0.000766	0.9994
R-squared	0.451598	Mean dependent var	-111.3226	
Adjusted R-squared	0.390664	S.D. dependent var	2219.058	
S.E. of regression	1732.196	Akaike info criterion	18.10737	
Sum squared resid	81013620	Schwarz criterion	18.29240	
Log likelihood	-276.6642	Hannan-Quinn criter.	18.16768	
F-statistic	7.411307	Durbin-Watson stat	1.484324	
Prob(F-statistic)	0.000895			
Inverted AR Roots	.29	-.29		
Inverted MA Roots	1.00	-.50+.87i	-.50-.87i	

Figure H.10 : Résultat du test de student pour l'automatic ARIMA

3) Résultats du test de student sur les modèles ARMA pour la série Part_A :

Dependent Variable: D(PART_A)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 06/22/24 Time: 00:30
 Sample: 2021M11 2024M05
 Included observations: 31
 Failure to improve objective (non-zero gradients) after 14 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-101.0207	139.9195	-0.721992	0.4765
AR(3)	0.078095	0.270396	0.288817	0.7749
MA(3)	-0.999999	3202.070	-0.000312	0.9998
SIGMASQ	2653836.	4.27E+09	0.000622	0.9995
R-squared	0.443100	Mean dependent var	-111.3226	
Adjusted R-squared	0.381222	S.D. dependent var	2219.058	
S.E. of regression	1745.565	Akaike info criterion	18.10861	
Sum squared resid	82268927	Schwarz criterion	18.29364	
Log likelihood	-276.6835	Hannan-Quinn criter.	18.16893	
F-statistic	7.160894	Durbin-Watson stat	1.483938	
Prob(F-statistic)	0.001092			
Inverted AR Roots	.43	-.21-.37i	-.21+.37i	
Inverted MA Roots	1.00	-.50+.87i	-.50-.87i	

Figure H.11 : Résultat du test de student pour ARMA(3,3)

Dependent Variable: D(PART_A)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 06/22/24 Time: 00:31
 Sample: 2021M11 2024M05
 Included observations: 31
 Convergence achieved after 8 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-70.67304	252.3432	-0.280067	0.7815
AR(3)	-0.519253	0.166284	-3.122693	0.0041
SIGMASQ	3605994.	1107257.	3.256691	0.0029
R-squared	0.243293	Mean dependent var	-111.3226	
Adjusted R-squared	0.189242	S.D. dependent var	2219.058	
S.E. of regression	1998.087	Akaike info criterion	18.15994	
Sum squared resid	1.12E+08	Schwarz criterion	18.29871	
Log likelihood	-278.4791	Hannan-Quinn criter.	18.20518	
F-statistic	4.501207	Durbin-Watson stat	1.351483	
Prob(F-statistic)	0.020183			
Inverted AR Roots	.40+.70i	.40-.70i	-.80	

Figure H.12 : Résultat du test de student pour AR(3)

Dependent Variable: D(PART_A)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 06/22/24 Time: 00:32
Sample: 2021M11 2024M05
Included observations: 31
Failure to improve objective (non-zero gradients) after 36 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-100.2259	131.3421	-0.763090	0.4518
MA(3)	-0.999998	2638.825	-0.000379	0.9997
SIGMASQ	2629156.	3.53E+09	0.000744	0.9994

R-squared	0.448279	Mean dependent var	-111.3226
Adjusted R-squared	0.408871	S.D. dependent var	2219.058
S.E. of regression	1706.122	Akaike info criterion	18.04846
Sum squared resid	81503838	Schwarz criterion	18.18723
Log likelihood	-276.7511	Hannan-Quinn criter.	18.09369
F-statistic	11.37515	Durbin-Watson stat	1.458829
Prob(F-statistic)	0.000242		

Inverted MA Roots	1.00	-50-.87i	-50+ .87i
-------------------	------	----------	-----------

Figure H.13 : Résultat du test de student pour MA(3)

Forecast Evaluation
Date: 06/22/24 Time: 01:00
Sample: 2021M10 2024M08
Included observations: 35
Evaluation sample: 2021M10 2024M08
Number of forecasts: 1

Evaluation statistics						
Forecast	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
PART_A	3459.758	2507.523	54.65707	41.94273	0.299679	8.800794

Figure H.14 : Résultats des prévisions par le modèle AR(3)

4) Test de normalité :

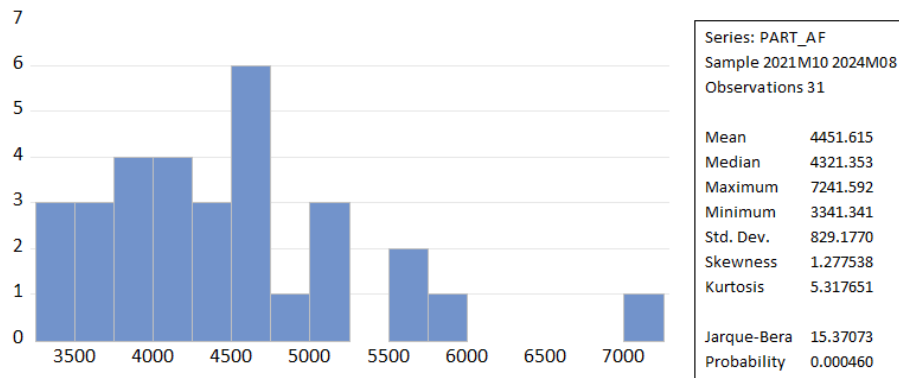


Figure H.15 : Test de normalité du modèle AR(3)

Annexe I : Algorithmes de Machine Learning pour les prévisions

1) Script des modèles :

```

file_path = '/content/Data to forecast FB.xlsx'
df = pd.read_excel(file_path)
df['Date'] = pd.to_datetime(df['Date'], format='%b-%y')

def evaluate_model(y_true, y_pred):
    r2 = r2_score(y_true, y_pred)
    mae = mean_absolute_error(y_true, y_pred)
    mse = mean_squared_error(y_true, y_pred)
    rmse = mean_squared_error(y_true, y_pred, squared=False)
    mape = np.mean(np.abs((y_true - y_pred) / y_true)) * 100
    return {'R2': r2, 'MAE': mae, 'MSE': mse, 'RMSE': rmse, 'MAPE': mape}

parts = ['Part A', 'Part B', 'Part C', 'Part D', 'Part E', 'Part F']
results = {}

for part in parts:
    df_prophet = df[['Date', part]].rename(columns={'Date': 'ds', part: 'y'})

    model_prophet = Prophet()
    model_prophet.fit(df_prophet)
    forecast = model_prophet.predict(df_prophet[['ds']])
    model_prophet.plot(forecast)
    y_pred_prophet = forecast['yhat'].values
    y_true = df_prophet['y'].values

    results[part] = {}
    results[part]['Prophet'] = evaluate_model(y_true, y_pred_prophet)
for part in results:
    print(f"Résultats pour {part}:")
    for model in results[part]:
        print(f" {model}: {results[part][model]}")

```

Figure I.1 – Capture d'écran du code du modèle Facebook Prophet

```

df = pd.read_excel(relative_path, sheet_name='B')
df['month'] = df['ds'].dt.month
df['year'] = df['ds'].dt.year
df['quarter'] = df['ds'].dt.quarter
df = df.drop(columns=['ds'])
test_size=0.2

```

```

X = df.drop(columns=['qte']) # Include all relevant features except the target
y = df['qte']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=test_size, shuffle=False)
model = xgb.XGBRegressor(
    learning_rate=0.1,
    n_estimators=1000,
    objective='reg:squarederror',
    random_state=42
)
model.fit(X_train, y_train, eval_set=[(X_test, y_test)], early_stopping_rounds=10, verbose=False)
y_pred = model.predict(X_test)

```

Figure I.2 – Capture d'écran sur le code utilisé pour les prévisions en utilisant Xgboost

```

scaled_data = scaler.fit_transform(df)
seq_length = 32
def create_sequences(data, seq_length):
    X, y = [], []
    for i in range(len(data) - seq_length):
        X.append(data[i:i+seq_length])
        y.append(data[i+seq_length, 0])
    return np.array(X), np.array(y)

X, y = create_sequences(scaled_data, seq_length)
X = X.reshape((X.shape[0], X.shape[1], X.shape[2]))
input_shape = (seq_length, X.shape[2])

test_size = int(len(df) * 0.2)

X_train = X[:-test_size]
X_test = X[-test_size:]
y_train = y[:-test_size]
y_test = y[-test_size:]

def build_model(input_shape):
    model = Sequential()
    model.add(SimpleRNN(100, activation='relu', input_shape=input_shape, return_sequences=True))
    model.add(Dropout(0.2))
    model.add(SimpleRNN(50, activation='relu'))
    model.add(Dropout(0.2))
    model.add(Dense(1))
    model.compile(optimizer='adam', loss='mse')
    return model

model = build_model(input_shape)
model.fit(X_train, y_train, epochs=1000, batch_size=8, verbose=1)
predictions = model.predict(X_test)

```

Figure I.3 – Capture d'écran sur le code utilisé pour les prévisions en utilisant les RNN

```

sc = MinMaxScaler(feature_range=(0, 1))
training_scaled = sc.fit_transform(trainset)
# Window size for the sliding window
window = 4
x_train = []
y_train = []
for i in range(window, len(df_train)):
    x_train.append(training_scaled[i-window:i, 0])
    y_train.append(training_scaled[i, 0])
x_train, y_train = np.array(x_train), np.array(y_train)
x_train = np.reshape(x_train, (x_train.shape[0], x_train.shape[1], 1))
# Clear any previous session
tf.keras.backend.clear_session()
# Build the model
regressor = Sequential()
regressor.add(LSTM(units=100, return_sequences=True, input_shape=(x_train.shape[1], 1)))
regressor.add(Dropout(0.2))
regressor.add(LSTM(units=50))
regressor.add(Dropout(0.2))
regressor.add(Dense(units=1))

regressor.compile(optimizer='adam', loss='mse', metrics=['mae'])

early_stopping = tf.keras.callbacks.EarlyStopping(monitor='loss', mode='min', patience=20)
mc = tf.keras.callbacks.ModelCheckpoint('best_model.h5', monitor='loss', mode='min', verbose=0, save_best_only=True)

hist = regressor.fit(x_train, y_train, epochs=1000, batch_size=32, callbacks=[mc, early_stopping])

regressor = tf.keras.models.load_model('best_model.h5')

prediction_train = regressor.predict(x_train)
prediction_train = sc.inverse_transform(prediction_train)

```

Figure I.4 – Capture d'écran sur le code utilisé pour les prévisions en utilisant LSTM


```

sc = MinMaxScaler(feature_range=(0, 1))
training_scaled = sc.fit_transform(trainset)
# Window size for the sliding window
window = 4
x_train = []
y_train = []
for i in range(window, len(df_train)):
    x_train.append(training_scaled[i-window:i, 0])
    y_train.append(training_scaled[i, 0])
x_train, y_train = np.array(x_train), np.array(y_train)
x_train = np.reshape(x_train, (x_train.shape[0], x_train.shape[1], 1))
# Clear any previous session
tf.keras.backend.clear_session()
# Build the model
regressor = Sequential()
regressor.add(LSTM(units=100, return_sequences=True, input_shape=(x_train.shape[1], 1)))
regressor.add(Dropout(0.2))
regressor.add(LSTM(units=50))
regressor.add(Dropout(0.2))
regressor.add(Dense(units=1))

regressor.compile(optimizer='adam', loss='mse', metrics=['mae'])

early_stopping = tf.keras.callbacks.EarlyStopping(monitor='loss', mode='min', patience=20)
mc = tf.keras.callbacks.ModelCheckpoint('best_model.h5', monitor='loss', mode='min', verbose=0, save_best_only=True)

hist = regressor.fit(x_train, y_train, epochs=1000, batch_size=32, callbacks=[mc, early_stopping])

regressor = tf.keras.models.load_model('best_model.h5')

prediction_train = regressor.predict(x_train)
prediction_train = sc.inverse_transform(prediction_train)

```

Figure I.5 – Capture d’écran sur le code utilisé pour les prévisions en utilisant LSTM

Modèle	Part A	Part B	Part C	Part D	Part E	Part F
RMSE	2.4e+3	1.52e+3	1.26e+3	717.96	9.56	13.93
MAE	2.01e+3	1.26e+3	1.07e+3	572.92	7.70	11.55
MAPE	43.39%	41.89%	83.23%	32.97%	17.91%	45.95%

Tableau I.1 – Résultats des prévisions par FB Prophet après génération des données

```

metadata = SingleTableMetadata()
metadata.detect_from_dataframe(df)
metadata.update_column(column_name='entity', sdtype='id')
metadata.set_sequence_index(column_name='ds')
metadata.set_sequence_key(column_name='entity')
print(metadata.to_dict())

metadata1 = {
  "METADATA_SPEC_VERSION": "SINGLE_TABLE_V1",
  "columns": {
    "ds": {
      "sdtype": "datetime",
      "datetime_format": "%Y-%m-%d"
    },
    "qte": {
      "sdtype": "numerical"
    },
    "entity": {
      "sdtype": "id",
      "regex_format": "sdv-id-[A-Za-z0-9]{7}"
    }
  }
}

```

Figure I.6 – Code de la génération de données

Generating report ...

(1/2) Evaluating Column Shapes: ██████████ 3/3 [00:00<00:00, 299.64it/s]
Column Shapes Score: 89.06%

(2/2) Evaluating Column Pair Trends: ██████████ 3/3 [00:00<00:00, 230.80it/s]
Column Pair Trends Score: 95.49%

Overall Score (Average): 92.27%

Figure I.7 – Performance de la génération de données