



الجمهورية الجزائرية الديمقراطية الشعبية

REPUBLIQUE ALGERINNE DEMOCRATIQUE ET POPLAIRE

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جامعة أبي بكر بلقايد - تلمسان -

Université Abou bakr Belkaïd – Tlemcen –

Faculté de Technologie



MEMOIRE

Présenté pour l'obtention du diplôme de **MASTER**

En vertu de l'arrêté ministérielle n° **1275 (008)** (Institution économique)

En : Génie Industriel

Spécialité : Chaîne logistique / Ingénierie de la Production.

Par : BELAZREG Mohamed Elamine & BOUHALIS Sihem.

Sujet

Investigation autour de la planification et la prédiction en utilisant l'IA dans un système de production

Encadré par :

M. BETAOUAF Hichem	MCA	Université de Tlemcen
M. BENNEKROUF Mohammed	MCA	Université de Tlemcen

Devant le jury :

Président :

M. BELKAID Fayçal	Professeur	Université de Tlemcen	Présidente
M. BENSMAIN Yassir	MCB	Université de Tlemcen	Examinateur
Mme. BENABADJI	i2E	Université de Tlemcen	Examinateur
M. BOUHENNI Issam	Ing.	TAYAL SPA	Partenaire économique

Année universitaire 2024-2025

Remerciment

Alhamdulillah — Nous rendons grâce à Dieu, Le Très-Haut, qui nous a accordé la santé, la patience et la force morale nécessaires pour mener à bien ce travail de recherche.

Ce mémoire est le fruit d'un effort partagé, soutenu par l'accompagnement de nombreuses personnes à qui nous adressons nos remerciements les plus sincères.

Nous tenons à exprimer toute notre gratitude à **Monsieur BETAOUAF Hichem** et **Monsieur BENNEKROUF Mohammed**, qui ont accepté de nous encadrer avec bienveillance. Leur rigueur scientifique, leur écoute, leurs conseils éclairés et leur disponibilité constante ont grandement contribué à la progression de notre travail. Leur confiance a été une source réelle de motivation.

Nos remerciements vont également **aux ingénieurs de l'entreprise TAYAL SPA**, dont l'appui technique, les explications pratiques et la mise à disposition de données ont enrichi considérablement notre étude.

Nous remercions également **les membres du jury** pour l'intérêt qu'ils portent à notre travail, pour leur temps et pour les remarques constructives qu'ils formuleront à l'occasion de la soutenance.

Nos pensées reconnaissantes vont aussi à l'ensemble du corps **enseignant du département de Génie industriel** de l'Université de Tlemcen. Leur engagement pédagogique et leur professionnalisme tout au long de notre formation ont été essentiels à notre parcours académique.

Nous n'oublions pas **nos collègues de promotion**, pour les moments de collaboration, d'entraide et de réflexion partagée, qui ont été précieux sur le plan humain et intellectuel.

Enfin, nous adressons nos remerciements les plus chaleureux à **nos familles** et à **nos proches**, pour leur soutien inconditionnel, leur patience et leurs encouragements tout au long de cette aventure universitaire.

Dédicace

Je dédie ce modeste travail :

À mes parents,

Aucun mot ne saurait exprimer la profondeur de ma gratitude pour leur amour inconditionnel, leur patience et leurs prières. Que Dieu leur accorde santé, sérénité et longue vie.

À mes frères et leurs épouses et leurs enfants,

À mes sœurs et leurs époux et leurs enfants,

À toute ma famille, pour leur affection et leur soutien discret mais constant.

À mon binôme Mohamed Elamine,

Pour la collaboration harmonieuse, l'entraide mutuelle et l'engagement partagé tout au long de ce parcours.

À Ilham,

Proche de cœur, pour sa présence bienveillante et ses encouragements sincères.

À la famille BOUHALIS, sans exception,

Pour leur estime et leur soutien moral.

Et enfin, à tous **mes collègues** de la promotion **2025,**

Avec qui j'ai partagé cette aventure universitaire faite de défis, d'efforts et de réussites collectives.

Dédicace

Je dédie ce modeste travail :

À **mes parents**,

Pour leur amour inestimable, leur patience infinie et leur soutien inébranlable. Que Dieu leur accorde santé, sérénité et longue vie.

À **mes frères et mes sœurs**,

À toute ma **famille**, pour leur affection sincère et leur présence constante, même dans les moments les plus exigeants.

À **Sihem**, mon binôme,

Pour son engagement sans faille, son esprit de coopération et la belle synergie qui a animé notre travail commun. Travailler à ses côtés fut un véritable privilège.

À la **famille BELAZREG**, sans exception,

Pour leur bienveillance, leurs encouragements et leur soutien moral.

À **Islam, Ikbal, Imad, Moussa**

Présence fidèle, dont le soutien discret et les paroles justes ont toujours porté loin.

Et à tous mes **collègues de la promotion 2025**,

Avec qui j'ai partagé une aventure riche en apprentissages, en défis et en réussites partagées.

Résumé

Ce mémoire explore l'intégration de l'intelligence artificielle (IA) dans la prévision de la demande et la planification de la production au sein des systèmes industriels, avec une application concrète dans le secteur textile algérien à travers le cas de l'entreprise TAYAL SPA. Il répond aux limites des méthodes classiques de prévision, souvent inefficaces face aux comportements non linéaires, aux variations saisonnières et aux chocs exogènes.

Pour y remédier, une architecture hybride de prévision est proposée, combinant des modèles statistiques (SARIMA, Prophet) à des algorithmes de machine Learning (XGBoost), avec une optimisation intelligente des hyperparamètres via Optuna. Cette synergie vise à capter à la fois les dynamiques linéaires et les résidus non linéaires, pour améliorer la précision des prévisions.

Deux jeux de données ont été utilisés pour l'évaluation expérimentale : les données de production internes de TAYAL SPA et les données hebdomadaires d'exportation de coton de l'USDA. Le modèle hybride SARIMA–XGBoost–Optuna, mis à jour toutes les deux semaines et évalué selon des indicateurs tels que le WAPE, MAE, RMSE et MASE, a montré des performances nettement supérieures.

Les méthodes proposées ont été intégrées dans la solution intelligente VisionPlan, un outil d'aide à la décision qui automatise l'ingestion des données, les prévisions, la correction des erreurs et le suivi des performances. Sa structure modulaire et son assistant conversationnel IA facilitent son utilisation et renforcent l'interprétabilité des résultats (M2M).

Les résultats montrent que l'hybridation des approches IA et statistiques, déployée à travers VisionPlan, permet une planification industrielle plus agile, précise et robuste. Les perspectives futures incluent l'intégration de variables exogènes et de modèles de Deep Learning pour renforcer encore les capacités prédictives.

Mots-clés : Intelligence Artificielle (IA), Prévision de la demande, Modèles hybrides, SARIMA, Prophet, XGBoost, Optuna, Apprentissage automatique, Optimisation bayésienne.

Abstract

This thesis explores the integration of artificial intelligence (AI) into demand forecasting and production planning within industrial systems, with a concrete application in the Algerian textile sector through the case of TAYAL SPA. It addresses the limitations of classical forecasting methods, which often fail to handle nonlinear patterns, seasonal variations, and unexpected disruptions.

To overcome these challenges, a hybrid forecasting architecture is proposed, combining statistical models (SARIMA, Prophet) with machine learning algorithms (XGBoost), and optimized intelligently through Bayesian optimization using Optuna. This synergy captures both linear trends and nonlinear residuals, improving predictive accuracy.

Two datasets were used for experimental evaluation: TAYAL SPA's internal production data and USDA weekly cotton export data. The SARIMA–XGBoost–Optuna hybrid model, updated biweekly and evaluated using WAPE, MAE, RMSE, and MASE indicators, demonstrated significantly improved performance.

These methods are integrated into the intelligent solution VisionPlan — a decision-support tool that automates data ingestion, forecasting, error correction, and performance tracking. Its modular design and AI-driven conversational assistant make it user-friendly and improve the interpretability of results.

The findings show that combining AI with statistical approaches, as deployed in VisionPlan, enables more agile, accurate, and robust industrial planning. Future directions include the integration of exogenous variables and deep learning techniques to further enhance forecasting capabilities.

Keywords: Artificial Intelligence, Demand Forecasting, Hybrid Models, SARIMA, Prophet, XGBoost, Optuna, Machine Learning, Bayesian Optimization.

الملخص

يهدف هذا البحث إلى استكشاف دمج تقنيات الذكاء الاصطناعي في التنبؤ بالطلب وتخطيط الإنتاج داخل الأنظمة الصناعية، من خلال تطبيق عملي في قطاع النسيج الجزائري عبر دراسة حالة لشركة TAYAL. يعالج البحث أوجه القصور في أساليب التنبؤ التقليدية، التي غالبًا ما تعجز عن التعامل مع السلوكيات غير الخطية، والتغيرات الموسمية، والصدمات المفاجئة.

للتغلب على هذه التحديات، تم اقتراح نموذج تنبؤ هجين يجمع بين النماذج الإحصائية) مثل Prophet و SARIMA وخوارزميات التعلم الآلي) مثل XGBoost، مع تحسين ذكي للمعاملات باستخدام تقنيات التحسين البايزي (Optuna). يهدف هذا النموذج إلى دمج القدرة على التقاط الاتجاهات الخطية من النماذج الإحصائية مع التعامل مع البقايا غير الخطية عبر خوارزميات التعلم الآلي، مما يعزز دقة التنبؤ.

شملت الدراسة تقييمًا تجريبيًا باستخدام مجموعتي بيانات: بيانات الإنتاج الداخلي لشركة TAYAL وبيانات التصدير الأسبوعية للقطن من وزارة الزراعة الأمريكية (USDA). أظهرت نتائج التقييم أن النموذج الهجين SARIMA-XGBoost-Optuna، الذي يتم تحديثه بصفة دورية كل أسبوعين، قد تفوق بشكل واضح من حيث الأداء وفقًا لمؤشرات مثل MAE و RMSE و MASE.

تم دمج هذا النموذج ضمن منصة ذكية لدعم اتخاذ القرار تُعرف بـ VisionPlan، والتي تعمل على أتمتة عمليات جمع البيانات، والتنبؤ، وتصحيح الأخطاء، ومتابعة الأداء. يتميز النظام ببنية معيارية ومساعد تفاعلي قائم على الذكاء الاصطناعي، مما يسهل استخدامه ويعزز فهم وتفسير النتائج.

تؤكد النتائج أن دمج الذكاء الاصطناعي مع النماذج الإحصائية داخل إطار تطبيقي مثل VisionPlan يمكن أن يحقق تخطيطًا إنتاجيًا أكثر مرونة ودقة وموثوقية. وتشمل الآفاق المستقبلية للعمل تطوير النموذج من خلال إدماج متغيرات خارجية واستخدام تقنيات التعلم العميق لتعزيز القدرة التنبؤية بشكل أكبر.

الكلمات المفتاحية: الذكاء الاصطناعي، التنبؤ بالطلب، النماذج الهجينة، SARIMA، Prophet، XGBoost، Optuna، التعلم الآلي، التحسين البايزي.

Table des matières

Chapitre I : Vers une Industrie Textile Agile : Défis et Perspectives chez TAYAL

1.	INTRODUCTION.....	21
1.1	Historique et évolution de l'industrie textile.....	21
1.2	Importance économique de l'industrie textile	21
1.3	Défis Actuels de l'industrie Textile	22
1.3.1	Enjeux de Durabilité et Pressions Environnementales	22
1.3.2	Volatilité des Coûts et des Matières Premières	22
1.4	Opportunités et avenir de l'industrie textile	22
1.4.1	Innovation et Recherche dans le Textile.....	22
1.4.2	Digitalisation et Automatisation des Processus	23
1.4.3	Expansion et Internationalisation des Marchés	23
1.5	Aperçu de l'industrie textile en Algérie	23
1.5.1	Défis et Contraintes du Marché Algérien	24
1.6	Étude de cas : présentation de l'entreprise tayal spa	24
1.6.1	Historique et Rôle Stratégique de TAYAL SPA	24
1.6.2	Organisation et Structure de l'Entreprise	25
1.6.3	Les Gammes de Produits Fabriqués par TAYAL	27
1.6.4	Les unités de production de TAYAL SPA	28
1.7	Conclusion	38
2.	INTRODUCTION.....	40
2.1	Fondements généraux de la prévision.....	40
2.2	Modèles statistiques de séries temporelles	41
2.2.1	Lissage exponentiel (Holt–Winters).....	41
2.2.2	Modèles ARIMA et extension SARIMA	42
2.2.3	Modèle Prophet	49
2.3	Approches par apprentissage automatique	50
2.3.1	Principes généraux du Machine Learning pour la prévision	50
2.3.2	XGBoost appliqué aux séries temporelles	51
2.3.3	Optimisation des hyperparamètres avec Optuna	52
2.4	Analyse critique des approches existantes	55
2.4.1	Forces et limites des modèles classiques	55
2.4.2	Atouts et contraintes des méthodes ML « pures »	56
2.4.3	Motivation pour une hybridation.....	58
2.5	Conclusion	58

3.	INTRODUCTION.....	60
3.1	De la prévision classique à l'IA : une transition clé pour l'industrie textile	60
3.1.1	Contraintes opérationnelles et défis.....	61
3.2	Protocole expérimental	62
3.2.1	Analyse descriptive des données industrielles internes (TAYAL SPA).....	62
3.2.2	Données Hebdomadaires d'Exportation de Coton (USDA).....	66
3.2.3	Approches hybrides de prévision	68
3.2.4	Régression Non-Linéaire des Résidus avec XGBoost.....	69
3.3	Architecture Modulaire du Pipeline Hybride	69
3.3.1	Prétraitement des données temporelles.....	69
3.3.2	Modélisation statistique linéaire.....	69
3.3.3	Extraction et préparation des résidus.....	70
3.3.4	Vectorisation et enrichissement des données résiduelles	70
3.3.5	Modélisation non linéaire avec XGBoost.....	70
3.3.6	Génération de la prévision hybride.....	70
3.4	Optimisation des Paramètres et Apprentissage Adaptatif.....	70
3.4.1	Optimisation Systématique des Hyperparamètres	70
3.4.2	Apprentissage Adaptatif et Réinjection des Prédictions.....	71
3.5	Intégration de l'assistant IA pour l'analyse exploratoire	71
3.6	Indicateurs de performance	71
3.6.1	WAPE (Weighted Absolute Percentage Error)	71
3.6.2	MAE (Mean Absolute Error).....	72
3.6.3	MASE (Mean Absolute Scaled Error).....	73
3.6.4	RMSE (Root Mean Squared Error)	73
3.7	Résultats expérimentaux	74
3.7.1	Stratégie de prévision bi-hebdomadaire	74
3.8	Cas d'étude 1 : demande hebdomadaire de Tayal Spa	75
3.8.1	Visualisations SARIMA-OP-XGBoost-OP	75
3.8.2	Visualisations auto_ARIMA-XGBoost-OP	76
3.8.3	Visualisations Prophet-XGBoost-OP.....	78
3.9	Cas d'étude 2 : exportations hebdomadaires de coton (USDA)	79
3.9.1	Visualisations SARIMA-OP-XGBoost-OP	79
3.9.2	Visualisations auto_ARIMA-XGBoost-OP	81
3.9.3	Visualisations Prophet-XGBoost-OP.....	82
3.10	Analyse globale et discussion	84
3.10.1	Analyse des cas d'étude.....	84
3.10.2	Comparaison des pipelines hybrides	84
3.10.3	Robustesse et adaptation en ligne.....	85
3.10.4	Limites et perspectives d'amélioration.....	85
3.11	Implications opérationnelles	85
3.12	Conclusion.....	87

4.	INTRODUCTION.....	89
4.1	Identité de l'Application.....	89
4.2	Architecture logicielle	90
4.2.1	Recueil pragmatique des besoins du prototype	90
4.2.2	Environnement technique et outils utilisés	91
4.2.3	Conception modulaire des composants	91
4.2.4	Implémentation back-end	92
4.3	Interface utilisateur et flux de travail	94
4.3.1	Page de Connexion (Login Page)	94
4.3.2	Tableau de Bord (Dashboard)	95
4.3.3	Vue de Prévision (Forecasting View).....	97
4.3.4	Real Order Logs	97
4.3.5	Predicted Order Logs.....	99
4.3.6	Forecast Engine	100
4.3.7	Forecast Setting	102
4.3.8	Vue d'Analyse des Données (Data Analysis View).....	104
4.4	Procédures d'Installation et de Déploiement	106
4.4.1	Prérequis système :	106
4.4.2	Installation des dépendances :	106
4.5	Lignes Directrices pour la Maintenance et l'Évolution	107
4.5.1	Sauvegardes régulières	107
4.5.2	Surveillance des modèles	107
4.5.3	Documentation et mises à jour	107
4.6	Perspectives : Vers une Intégration Future de la Planification dans VisionPlan	107
4.7	Conclusion	109
5.	CONCLUSION GENERALE	110
6.	REFERENCES.....	112
6.	BMC	112
6.	SWOT.....	112

Table des figures

Figure 1-1: Partenariat Algero-turc	25
Figure 1-2: Plan de l'entreprise TAYAL.....	26
Figure 1-3: Capacité de production annuelle de diverses catégories de produits ³ textiles.	28
Figure 1-4: Schéma de l'enchaînement des opérations et des relations fonctionnelles entre les unités de production chez TAYAL	29
Figure 1-5 : Un schéma de l'organisation interne de l'Unité 2	30
Figure 1-6 : la classification des fibres textiles (fibres naturelles vs fibres synthétiques) 31	31
Figure 1-7: Dessin montrant une partie de l'organisation spatiale des unités 3 et 4.	34
Figure 1-8: Schéma du Processus de Filature et des Différents Types de Fils Produits ⁷ 35	35
Figure 2-1: Classification des méthodes de prévision	40
Figure 2-2: Séries Stationnaires Et Non-Stationnaires	42
Figure 2-3: Forecasts of monthly US leisure and hospitality employment using the ARIMA (Hyndman & Athanasopoulos, 2021).....	45
Figure 2-4 : Représentation graphique de l'ACF sur la série temporelle étudiée (Zheng & Zhu, 2020)	46
Figure 2-5: Représentation graphique de la PACF sur la série temporelle étudiée (Zheng & Zhu, 2020).	47
Figure 2-6:Prévisions sur deux jours du modèle Prophet pour la demande d'électricité de l'État de Victoria (Hyndman & Athanasopoulos, 2021).....	50
Figure 2-7: Schéma structurel du modèle XGBoost (Zhang & Li, 2023).	51
Figure 2-8: Représentation schématique de l'algorithme d'optimisation TPE (Kapočiūtė-Dzikienė, 2020).	55
Figure 3-1: Série journalière brute de production de coton	63
Figure 3-2: Histogramme de distribution des volumes journaliers (2024).....	63
Figure 3-3: Décomposition additive de la série journalière de production (tendance, saisonnalité, résidus)	63
Figure 3-4:Évolution de la production journalière avec moyenne mobile (fenêtre de 12 semaines).....	63
Figure 3-5: Série journalière brute de production de coton	65
Figure 3-6: Histogramme de distribution des volumes journaliers (2024).....	65
Figure 3-7: Décomposition additive de la série journalière de production (tendance, saisonnalité, résidus)	65
Figure 3-8: Évolution de la production journalière avec moyenne mobile (fenêtre de 12 semaines).....	65
Figure 3-9: Évolution quotidienne des commandes de coton – données USDA.....	67
Figure 3-10 : Distribution des volumes journaliers – données USDA (2021–2025)	67
Figure 3-11: Décomposition additive de la série journalière USDA (2019–2025).....	67
Figure 3-12: Lissage de la série USDA par moyenne mobile (fenêtre de 12 semaines) ..	67
Figure 3-13 : Comparaison des valeurs prévues et historiques du modèle SARIMA–OP–XGBoost-OP	75

Figure 3-14 : Évolution de l'erreur de prévision (%) du modèle SARIMA–OP–XGBoost-OP.....	75
Figure 3-15: Distribution des erreurs absolues (%) du modèle SARIMA–OP–XGBoost-OP.....	75
Figure 3-16 : Analyse 3d de l'erreur absolue en fonction de l'horizon temporel et du pourcentage d'erreur (APE).....	75
Figure 3-17 : Comparaison des valeurs prévues et historiques du modèle auto_ARIMA–XGBoost-OP	77
Figure 3-18 : Évolution de l'erreur de prévision (%) du modèle auto_ARIMA–XGBoost-OP.....	77
Figure 3-19: Distribution des erreurs absolues (%) du modèle auto_ARIMA–XGBoost-OP	77
Figure 3-20 : Analyse 3d de l'erreur absolue en fonction de l'horizon temporel et du pourcentage d'erreur (APE).....	77
Figure 3-21 : Comparaison des valeurs prévues et historiques du modèle Prophet–XGBoost-OP	78
Figure 3-22 : Évolution de l'erreur de prévision (%) du modèle Prophet–XGBoost-OP .	78
Figure 3-23: Distribution des erreurs absolues (%) du modèle Prophet–XGBoost-OP ...	78
Figure 3-24 : Analyse 3d de l'erreur absolue en fonction de l'horizon temporel et du pourcentage d'erreur (APE).....	78
Figure 3-25 : Comparaison des valeurs prévues et historiques du modèle Prophet–XGBoost-OP	80
Figure 3-26 : Évolution de l'erreur de prévision (%) du modèle Prophet–XGBoost-OP .	80
Figure 3-27: Distribution des erreurs absolues (%) du modèle Prophet–XGBoost-OP ...	80
Figure 3-28 : Analyse 3d de l'erreur absolue en fonction de l'horizon temporel et du pourcentage d'erreur (APE).....	80
Figure 3-29 : Comparaison des valeurs prévues et historiques du modèle auto_ARIMA–XGBoost.....	81
Figure 3-30 : Évolution de l'erreur de prévision (%) du modèle auto_ARIMA–XGBoost	81
Figure 3-31: Distribution des erreurs absolues (%) du modèle auto_ARIMA–XGBoost	81
Figure 3-32 : Analyse 3d de l'erreur absolue en fonction de l'horizon temporel et du pourcentage d'erreur (APE).....	81
Figure 3-33 : Comparaison des valeurs prévues et historiques du modèle Prophet–XGBoost-OP	83
Figure 3-34 : Évolution de l'erreur de prévision (%) du modèle Prophet–XGBoost-OP .	83
Figure 3-35: Distribution des erreurs absolues (%) du modèle Prophet–XGBoost-OP ...	83
Figure 3-36 : Analyse 3d de l'erreur absolue en fonction de l'horizon temporel et du pourcentage d'erreur (APE).....	83
Figure 3-37 : Assistance IA pour l'analyse des prévisions.....	86
Figure 4-1: Écran de démarrage de l'application VisionPlan	90
Figure 4-2: Chaîne de Traitement Hybride pour l'Amélioration des Prévisions.....	92
Figure 4-3: Diagramme entité-relation de la base MySQL	93
Figure 4-4: l'interface de connexion	94

Figure 4-5 : Interface de visualisation des performances de prévision dans VisionPlan .	95
Figure 4-6 : Vue de Prévision (Forecasting View) dans VisionPlan.....	97
Figure 4-7: Vue des données historiques dans VisionPlan.....	98
Figure 4-8 : Vue des prévisions enregistrées (Predicted Logs) dans VisionPlan.....	99
Figure 4-9: Interface Forecast Engine dans VisionPlan	101
Figure 4-10: Interface Forecast Settings dans VisionPlan.....	103
Figure 4-11:Vue d'Analyse Temporelle des Données dans VisionPlan	105
Figure 4-12:Fenêtre de confirmation de sauvegarde dans VisionPlan	107

List des tables :

Tableau 1-1: Evolution de l'Indice de la production industrielle des textiles, bonneterie et confection (APS, 2022).....	23
Tableau 1-2: Types de produits fabriqués par TAYAL.....	27
Tableau 1-3 : Répartition des Espaces de l'unité 2.....	31
Tableau 1-4: Tableau récapitulatif des matières premières stockées dans l'Unité 2.....	32
Tableau 1-5: Un tableau illustratif détaillant l'organisation spatiale des unités 3 et 4.....	34
Tableau 1-6: Répartition des Espaces et Zones Fonctionnelles de l'Unité 5.....	36
Tableau 1-8 : Types d'Emballages et Leur Utilité pour la Conservation des Fils ⁷	37
Tableau 2-1 : Comparaison des valeurs AIC et BIC pour les modèles ARIMA testés. ...	49
Tableau 3-1: Statistiques descriptives de la série journalière (production de coton, 2024)	63
Tableau 3-2: Résumé statistique de la série hebdomadaire agrégée (production de coton, 2024).....	64
Tableau 3-3: Statistiques descriptives de la série journalière USDA (janvier 2019 – janvier 2025).....	66
Tableau 3-4 : Évaluation des modèles hybrides de prévision (Tayal Spa).....	75
Tableau 3-5 : Évaluation des modèles hybrides de prévision (USDA).....	79
Tableau 4-1: Dictionnaire des tables de la base de données MySQL.....	93
Tableau 2: Détail de la Masse Salariale Mensuelle et Annuelle par Poste (Salaires Chargés)	121
Tableau 3: Projection du Chiffre d'Affaires par Produit/Service.....	123

Liste des abréviations

ACF	Autocorrelation Function
AIC	Akaike Information Criterion
APE	Absolute Percentage Error
API	Application Programming Interface
AR	AutoRegressive (Autorégressif)
ARIMA	AutoRegressive Integrated Moving Average
B	Backshift Operator (Opérateur de décalage)
BIC	Bayesian Information Criterion
CO	Coton
CRUD	Create, Read, Update, Delete
DEN	Denier (unité de finesse des fibres textiles)
Docker	Conteneurisation logicielle
EI	Expected Improvement
ERP	Enterprise Resource Planning
GUI	Graphical User Interface
IA	Intelligence Artificielle
I	Integrated (Intégré)
LSTM	Long Short-Term Memory (réseau de neurones récurrent)
MA	Moving Average (Moyenne mobile)
MAE	Mean Absolute Error
MASE	Mean Absolute Scaled Error
ML	Machine Learning (Apprentissage automatique)
MLOps	Machine Learning Operations
MVVM	Model-View-ViewModel
NumPy	Numerical Python
OP	Optuna (Optimisation Bayésienne)
PACF	Partial Autocorrelation Function
Pandas	Python Data Analysis Library
PES	Polyester
PIB	Produit Intérieur Brut
PyQt5	Python Qt version 5
RMSE	Root Mean Squared Error
SARIMA	Seasonal AutoRegressive Integrated Moving Average

SPA	Société par Actions
SQL	Structured Query Language
Statsmodels	Python module for statistical modeling
TPE	Tree-structured Parzen Estimator
T.U.C	Taux d'Utilisation des Capacités
USDA	United States Department of Agriculture
UoM	Unité de Mesure (Unit of Measurement)
WAPE	Weighted Absolute Percentage Error
XGBoost	Extreme Gradient Boosting

INTRODUCTION GÉNÉRALE

Contexte

L'industrie textile algérienne, bien que soutenue par des projets d'envergure tels que le complexe intégré TAYAL SPA, fait face à une pression croissante pour moderniser ses outils de gestion et améliorer la performance de ses processus de production. L'un des leviers majeurs de cette modernisation est la digitalisation des systèmes industriels, qui s'accompagne d'une intégration croissante des technologies intelligentes au cœur de la chaîne de valeur.

Dans cette logique, l'intelligence artificielle (IA) s'impose comme un catalyseur de transformation. Elle permet de traiter de grands volumes de données en temps réel, de détecter des patterns complexes et de générer des prédictions précises, ouvrant la voie à des systèmes de production plus autonomes, plus réactifs et plus efficaces. Parmi les nombreuses applications de l'IA dans l'industrie, la prédiction occupe une place centrale : elle permet d'anticiper les comportements futurs du système, de planifier les actions à venir et d'optimiser l'allocation des ressources de manière proactive.

Dans le cas spécifique de TAYAL SPA, dont les lignes de production sont organisées de manière séquentielle et fortement interdépendante, le besoin d'un système de prévision fiable est particulièrement critique. Les déséquilibres entre les ordres de fabrication et les besoins réels du marché entraînent des conséquences opérationnelles lourdes : pertes de temps machine, accumulation de stocks inutiles, ruptures d'approvisionnement et gaspillage de matières. C'est dans ce contexte que la mise en œuvre de solutions prédictives basées sur l'IA devient non seulement pertinente, mais nécessaire.

Problématique et objectifs

Les méthodes classiques de prévision, fondées sur des modèles statistiques comme ARIMA ou SARIMA et prophet, sont bien adaptées aux séries temporelles linéaires et stationnaires. Cependant, elles atteignent leurs limites dès lors que les données présentent des comportements non linéaires, des discontinuités ou des facteurs exogènes imprévus. Cela limite leur capacité à répondre aux besoins complexes et dynamiques d'un système industriel réel.

Afin de dépasser ces limites, ce mémoire explore une solution basée sur une architecture hybride combinant des modèles statistiques traditionnels avec des algorithmes d'intelligence artificielle, notamment XGBoost pour la modélisation des résidus et Optuna pour l'optimisation intelligente des hyperparamètres. L'objectif est d'obtenir un modèle prédictif à la fois robuste, explicable et capable de s'adapter aux variations du système de production.

Les objectifs de cette étude sont les suivants :

- Évaluer la performance d'une solution prédictive hybride, sur deux sources de données distinctes : les historiques d'ordres de production internes de TAYAL SPA, et les données hebdomadaires d'exportation de coton fournies par l'USDA.
- Implémenter un mécanisme d'apprentissage adaptatif en ligne, mis à jour toutes les deux semaines, qui intègre les erreurs passées pour améliorer la précision prédictive au fil du temps.
- Développer un outil applicatif, VisionPlan, intégrant l'ensemble de la chaîne de traitement — collecte, traitement, prévision, visualisation — avec restitution des

indicateurs de performance (WAPE, MAE, MASE, RMSE) pour une utilisation directe par les équipes opérationnelles.

Structure du mémoire

Ce travail s'articule autour de quatre grands axes :

La première partie pose les bases contextuelles et industrielles, avec une attention particulière portée à l'organisation, aux contraintes et aux besoins spécifiques du complexe TAYAL SPA.

La deuxième partie traite des fondements théoriques et des outils de modélisation : elle présente et compare les approches statistiques classiques et les techniques d'apprentissage automatique utilisées en prévision.

La troisième partie est consacrée à la mise en œuvre du protocole expérimental, en décrivant la préparation des données, la modélisation hybride, l'optimisation bayésienne et les résultats des prédictions sur les deux cas d'usage.

La quatrième partie porte sur la conception de l'outil *VisionPlan*, en détaillant son architecture logicielle, ses fonctionnalités, son interface utilisateur et les modalités de déploiement dans un environnement industriel réel.

Chapitre I :
Vers une Industrie Textile Agile : Défis et
Perspectives chez TAYAL SPA

1. Introduction

Ce chapitre donne un aperçu de l'industrie textile pour présenter le contexte général de nos travaux. Nous aborderons dans un premier temps la spécificité et l'importance de cette industrie à l'échelle mondiale et nationale. Nous nous concentrerons ensuite sur la principale entreprise textile algérienne, TAYAL SPA, pour comprendre son rôle et sa contribution. Enfin, nous présentons les problématiques qui ont constitué le point de départ de notre projet, en nous concentrant sur les innovations en matière de planification, de prévision dans l'entreprise.

1.1 Historique et évolution de l'industrie textile

L'habillement a toujours répondu à l'un des besoins fondamentaux de l'homme. Ce secteur, autrefois basé sur un artisanat manuel, s'est transformé au fil du temps en une industrie mondiale dynamique. Les premiers textiles, fabriqués à partir de fibres naturelles comme le lin, la laine et le coton, témoignaient déjà de la diversité et de l'ingéniosité humaine. Parmi ces matières, le coton, notamment en Inde, s'est imposé comme un pilier de la production et du commerce mondial grâce à la renommée de ses tissus fins (Corbman, 1985).

La révolution industrielle, au XVIIIe siècle, a marqué un tournant décisif. Des inventions comme la filature Jenny (Spinning Jenny), le cadre à eau (Water Frame) et le métier à tisser mécanique (Power Loom) ont permis la mécanisation des procédés, augmentant considérablement la vitesse et la capacité de production. Ces avancées technologiques ont non seulement réduit les coûts, mais ont également démocratisé l'accès aux textiles (Chandrakant & Gupta, 2019).

Au XXe siècle, l'introduction des fibres synthétiques, telles que le nylon, a bouleversé le paysage textile. Issues de l'industrie pétrochimique, ces fibres ont offert des avantages significatifs : durabilité, élasticité et coût réduit. Elles ont élargi les possibilités tant dans la mode que dans les textiles fonctionnels.

Aujourd'hui, l'industrie textile fait face à un défi majeur : la durabilité. La production doit désormais intégrer des pratiques respectueuses de l'environnement, avec des efforts pour développer des matériaux écologiques, adopter des technologies de recyclage et minimiser l'impact environnemental. L'objectif est de concilier innovation et préservation pour garantir un avenir durable aux générations futures (Christopher, 2016).

1.2 Importance économique de l'industrie textile

Le secteur textile occupe une place importante dans l'économie mondiale en raison de son impact sur la production industrielle et le commerce international ainsi que sur l'emploi au niveau mondial. Présent dans plusieurs nations à travers le monde entier et contribuant considérablement au progrès industriel tout en se conformant aux avancées technologiques et aux impératifs environnementaux.

Dans certaines nations du monde entier et notamment dans les économies en développement, l'industrie textile joue un rôle significatif dans le produit intérieur brut (PIB). Par exemple au Bangladesh, cette industrie représente environ 12 % du PIB et contribue à hauteur de 80 % des exportations totales du pays. En Inde, ce secteur contribue à environ 7 %

du PIB national, avec une forte implication dans les exportations textiles et vestimentaires. De manière similaire, au Vietnam, la production textile joue un rôle crucial dans le commerce extérieur ainsi que dans les investissements directs étrangers (Kilduff, 2005).

Dans les pays industrialisés comme la France et l'Italie en particulier, le secteur du textile se démarque par son expertise dans des créneaux à haute valeur ajoutée. C'est en effet en produisant des textiles luxueux et techniques que ces nations se distinguent, étant fortement ancrées dans les domaines de la mode et des technologies textiles. Quant à la Turquie laquelle jouit d'un emplacement géographique privilégié aux portails de l'Europe et offre flexibilité dans sa capacité productive, elle occupe une position clé dans l'approvisionnement pour diverse grands marques mondiales (INSEE, 2019).

1.3 Défis Actuels de l'industrie Textile

1.3.1 Enjeux de Durabilité et Pressions Environnementales

L'industrie textile est l'une des industries les plus polluantes, produisant 10 % des émissions mondiales de CO₂ et consommant de grandes quantités d'eau. La contamination de l'eau due aux teintures et aux traitements chimiques, la mauvaise gestion des déchets textiles (des millions de tonnes sont jetées chaque année) et l'empreinte carbone élevée (notamment associée à la fast fashion) constituent des défis majeurs.

Face à ces défis, les entreprises doivent adopter des pratiques plus durables, encouragées par la réglementation et la demande des consommateurs. L'utilisation de matériaux recyclés, les procédés écologiques et l'économie circulaire s'imposent comme des solutions pour réduire l'impact environnemental de l'industrie (CAWTAR, 2020).

1.3.2 Volatilité des Coûts et des Matières Premières

L'industrie textile s'appuie sur des matières premières comme le coton et les fibres synthétiques, dont les prix fluctuent en raison de facteurs économiques et climatiques. Les sécheresses et les inondations affectent les récoltes, entraînant une augmentation des coûts et des ruptures d'approvisionnement.

Les tensions géopolitiques et la pandémie de COVID-19 ont mis en évidence les vulnérabilités de la chaîne d'approvisionnement, entraînant des retards et une hausse des coûts. De plus, la hausse des prix de l'énergie exerce une pression sur le secteur manufacturier, en particulier pour les processus qui nécessitent une forte consommation d'énergie.

Pour relever ces défis, les entreprises doivent diversifier leurs sources d'approvisionnement, explorer des alternatives durables et optimiser la gestion des stocks afin de limiter leur vulnérabilité aux fluctuations du marché (Expo, n.d.).

1.4 Opportunités et avenir de l'industrie textile

1.4.1 Innovation et Recherche dans le Textile

L'industrie textile évolue rapidement grâce aux avancées technologiques. Les textiles intelligents intègrent désormais des capteurs et des matériaux réactifs capables de s'adapter aux conditions environnementales ou de surveiller des paramètres biométriques. Par ailleurs,

l'impression 3D textile permet une production personnalisée et à la demande, réduisant ainsi les déchets et les coûts de stockage. Les biomatériaux, comme les fibres à base d'algues ou le cuir végétal, ouvrent de nouvelles perspectives pour une mode plus écologique et durable (APS, 2022).

1.4.2 Digitalisation et Automatisation des Processus

L'outil d'aide à la décision basé sur l'intelligence artificielle transforme le secteur textile en optimisant la gestion des stocks et en améliorant la prévision de la demande. Grâce aux algorithmes avancés, les marques peuvent ajuster leur production en fonction des tendances du marché, évitant ainsi le gaspillage. La personnalisation des vêtements est également facilitée par la digitalisation, offrant aux consommateurs des produits sur mesure. Enfin, l'automatisation des usines réduit la dépendance à la main-d'œuvre et améliore la rentabilité des entreprises (Expo, n.d.).

1.4.3 Expansion et Internationalisation des Marchés

L'essor des textiles techniques ouvre des débouchés dans des secteurs variés tels que l'aéronautique, la santé et le sport. Par ailleurs, la croissance des classes moyennes en Afrique et en Asie stimule la demande pour des produits textiles abordables et de qualité. Les investissements dans ces régions permettent aux entreprises d'accéder à de nouveaux marchés à fort potentiel, renforçant ainsi leur compétitivité à l'échelle mondiale (CAWTAR, 2020).

1.5 Aperçu de l'industrie textile en Algérie

L'évolution récente de la production industrielle dans le secteur des textiles, bonneterie et confection en Algérie met en lumière les difficultés structurelles que connaît cette filière. Le tableau ci-dessous résume les principales tendances observées entre 2019 et 2021.

Tableau 1-1: Evolution de l'Indice de la production industrielle des textiles, bonneterie et confection (APS, 2022)

Code	Intitulé/ Agrégation	Indices			Variations			T.U.C.		
		2019	2020	2021	2019	2020	2021	2019	2020	2021
NSA 11	Ind. Text. Bonn. Confection	10	8,7	7,7	-0,9	-13,1	-11,5	31,7	31,3	27,3

Selon l'Office National des Statistiques (ONS, 2021), le secteur textile, habillement et cuir en Algérie a connu une baisse continue de sa production entre 2019 et 2021, avec un indice passant de 10 à 7,7 et un taux d'utilisation des capacités tombé à 27,3 %. Malgré des partenariats avec des entreprises étrangères pour améliorer la production locale, l'industrie reste peu performante, couvrant moins de 6 % du marché national, tandis que les importations représentent près de 90 % de la demande (APS, 2023). L'objectif d'atteindre 2 milliards d'euros d'exportations à l'horizon 2022 n'a pas été atteint. La baisse des importations observée en 2022 semble davantage liée à des facteurs conjoncturels qu'à une réelle relance de la production nationale (APS, 2023).

1.5.1 Défis et Contraintes du Marché Algérien

L'industrie textile en Algérie est confrontée à des difficultés liées à une planification imprécise de la demande et à une gestion inefficace de l'offre. Le manque d'outils adéquats entraîne des écarts entre l'offre et la demande, ce qui entraîne soit des excédents de stocks coûteux, soit des ruptures de stock qui impactent la satisfaction des clients. Le manque d'analyses et de prévisions de marché fiables fondées sur des données entrave la capacité des entreprises à s'adapter aux tendances et aux fluctuations saisonnières (APS, 2022).

De plus, la chaîne d'approvisionnement manque de coordination et de liquidité. Les retards dans l'importation des matières premières allongent les cycles de production et augmentent les coûts logistiques. Contrairement à d'autres pays qui automatisent leurs processus, l'Algérie reste dépendante des méthodes traditionnelles, ce qui réduit sa compétitivité (CAWTAR, 2020).

1.6 Étude de cas : présentation de l'entreprise tayal spa

Dans le cadre de notre projet de fin d'études, nous avons eu l'opportunité d'effectuer un stage au sein de TAYAL SPA, l'une des plus grandes entreprises textiles en Algérie. Cette immersion nous a permis d'acquérir une compréhension approfondie du fonctionnement de l'entreprise et des défis auxquels elle est confrontée. Nous avons ainsi exploré le secteur textile algérien, son organisation industrielle et son impact économique. Ce stage nous a offert une vision concrète de l'importance de l'intégration de nouvelles technologies et de stratégies de gestion efficaces dans le domaine textile.

1.6.1 Historique et Rôle Stratégique de TAYAL SPA

TAYAL SPA est une entreprise textile intégrée, située dans la zone industrielle de Sidi Khettab, Relizane, en Algérie. Issue d'un partenariat entre INTERTAY (filiale du groupe turc TAYPA, avec plus de 40 ans d'expérience dans l'industrie textile et de l'habillement) et les entreprises publiques algériennes C&H Group, TEXALG et MADAR Holding S.P.A, elle incarne une ambition stratégique pour le développement du secteur textile national¹.

Le complexe industriel s'étend sur 250 hectares, ce qui en fait la plus grande installation de production textile en Afrique². Conçu pour couvrir l'ensemble du cycle de production textile, il intègre toutes les étapes, de la transformation des matières premières à la confection des produits finis. Cette approche vise à réduire la dépendance de l'Algérie aux importations textiles tout en renforçant son positionnement sur le marché international.

La production a démarré progressivement en mars 2018, et l'usine a atteint une capacité de 30 millions d'articles de prêt-à-porter par an, comprenant 12 millions de jeans, 12 millions de pulls et 6 millions de chemises³. Une fois à pleine capacité, TAYAL SPA pourra couvrir près de la moitié des besoins nationaux en prêt-à-porter ainsi qu'une part importante de la demande en tissus et fils.

¹ Présentation Orientation - TAYAL SPA, 2021. Disponible sur <https://tayal.dz/>

² Présentation Orientation - TAYAL SPA, 2021. Disponible sur <https://tayal.dz/wp-content/uploads/2021/01/TAYAL-Inc.-Company-Profile-2021.pdf>

³ Présentation Orientation - TAYAL SPA, 2021

Ce projet s'inscrit dans la stratégie de diversification industrielle de l'Algérie, visant à stimuler l'emploi, favoriser le transfert de savoir-faire et renforcer la compétitivité du secteur textile. Grâce à ses infrastructures modernes et à son intégration verticale, TAYAL SPA aspire à faire de l'Algérie un acteur clé du textile sur la scène africaine et internationale.

Le groupe TAYPA :

TAYPA créé en 1977 commencé en 1977 lorsque Mesut Toprak, le fondateur de la société, a échangé du tissu. Aujourd'hui TAYPA fourni les plus grandes marques du monde, Elle fait partie des 500 plus grandes organisations industrielles de Turquie⁴, Elle entend maintenir sa croissance en jouant un rôle plus actif dans les principaux pays européens d'ici 2020. Elle exerce activité dans plusieurs payé comme l'Egypte, l'Algérie, la Turquie. etc.

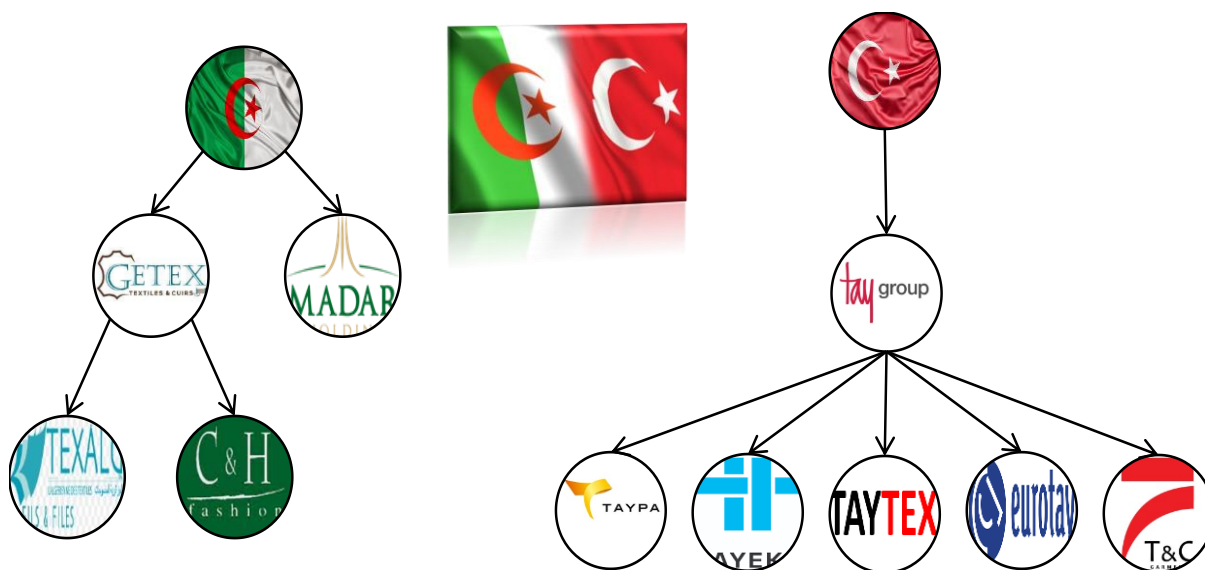


Figure 1-1: Partenariat Algero-turc

1.6.2 Organisation et Structure de l'Entreprise

TAYAL SPA est structurée en plusieurs divisions spécialisées, chacune jouant un rôle clé dans l'efficacité et la gestion optimale des opérations industrielles. Son organisation repose sur un modèle intégré favorisant une coordination efficace entre les différentes unités pour assurer une production fluide et performante.

⁴ Présentation Orientation - TAYAL SPA, 2021. Disponible sur <https://tayal.dz/>

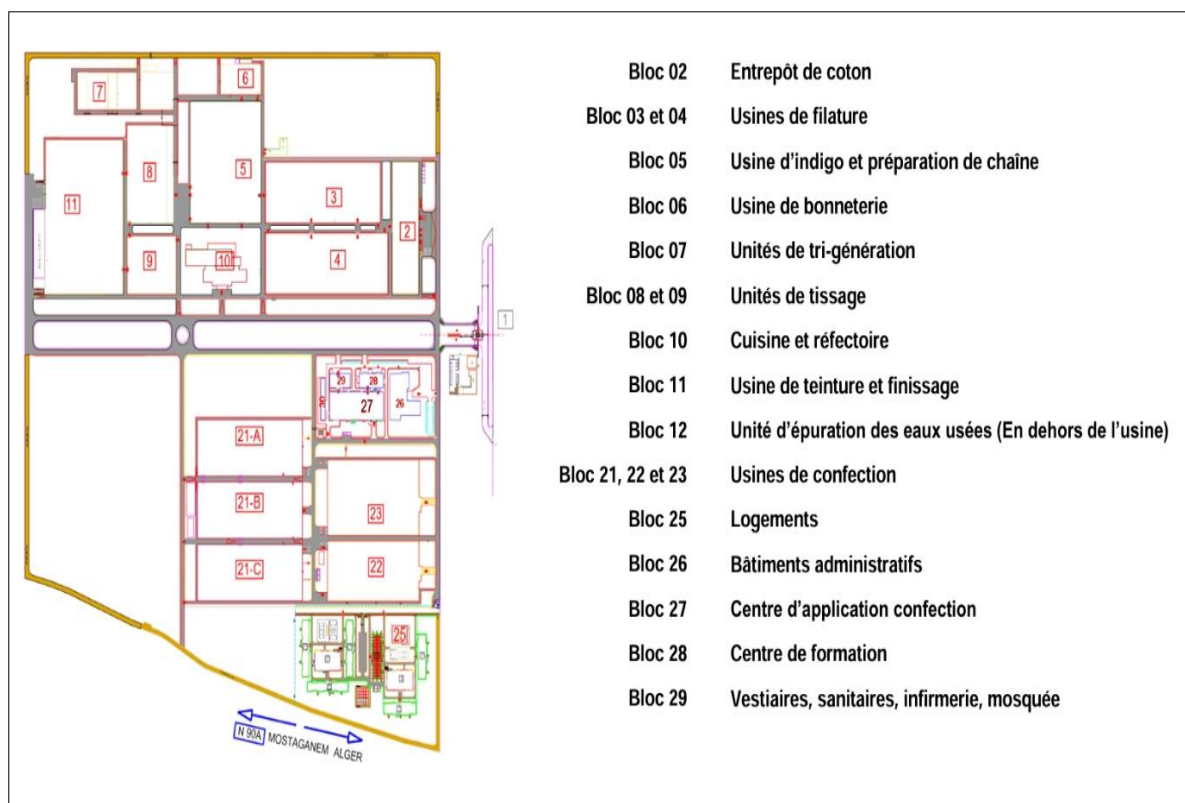


Figure 1-2: Plan de l'entreprise TAYAL⁵

- **Bloc 02 – Entrepôt de coton** : Espace dédié au stockage du coton brut avant son passage en filature, assurant ainsi une alimentation continue des lignes de production.
- **Blocs 03 et 04 – Usines de filature** : Ces unités assurent la transformation du coton brut en fils grâce à des procédés de cardage, étirage et filage, constituant la première étape industrielle du processus textile.
- **Bloc 05 – Usine d'indigo et préparation de chaîne** : Ce bloc est spécialisé dans la teinture des fils à l'indigo et leur préparation pour le tissage. Il assure l'enroulement et l'alignement des fils de chaîne.
- **Bloc 06 – Usine de bonneterie** : Elle produit des textiles en maille, utilisés notamment pour les vêtements souples comme les t-shirts ou les sous-vêtements.
- **Bloc 07 – Unités de tri-génération** : Installations énergétiques permettant de produire simultanément de l'électricité, de la chaleur et du froid pour alimenter l'ensemble du complexe en énergie.
- **Blocs 08 et 09 – Unités de tissage** : Ces ateliers transforment les fils en tissus grâce à des métiers à tisser, en croisant les fils de chaîne et de trame selon différents motifs.
- **Bloc 10 – Cuisine et réfectoire** : Infrastructure destinée à la restauration du personnel, contribuant à leur bien-être quotidien.
- **Bloc 11 – Usine de teinture et de finissage** : Cette unité assure les traitements finaux des textiles, notamment la coloration, le lavage, le séchage et d'éventuelles finitions spécifiques (douceur, imperméabilité, etc.).

⁵ Rapport de stage, TAYAL SPA, 2024.

- **Bloc 12 – Unité d'épuration des eaux usées** : Installée en dehors du site principal, elle permet le traitement des eaux industrielles avant rejet, en conformité avec les normes environnementales.
- **Blocs 21, 22 et 23 – Usines de confection** : Ces blocs sont consacrés à la coupe, à l'assemblage et à la couture des vêtements à partir des tissus produits.
- **Bloc 25 – Logements** : Hébergements réservés au personnel résident ou aux travailleurs en déplacement.
- **Bloc 26 – Bâtiments administratifs** : Ce bloc regroupe les bureaux des services de gestion, des ressources humaines et de la direction de l'usine.
- **Bloc 27 – Centre d'application confection** : Espace dédié à l'innovation, aux essais de prototypes et à l'amélioration des techniques de confection.
- **Bloc 28 – Centre de formation** : Structure pédagogique destinée à former les ouvriers et techniciens dans les métiers du textile et de la confection.
- **Bloc 29 – Vestiaires, sanitaires, infirmerie, mosquée** : Infrastructures de soutien pour le personnel, incluant des équipements d'hygiène, de santé et un espace de culte.

1.6.3 Les Gammes de Produits Fabriqués par TAYAL

TAYAL se spécialise dans la production de trois types de produits, en partant de la matière première, le coton, qui est transformé en fils, puis en tissu, avant d'être confectionné en vêtements prêts à porter.⁵

Tableau 1-2: Types de produits fabriqués par TAYAL

Catégorie de Produits	Types de Produits
Le Fil	- Open End - Compact - Carded - Combed
Le Tissu	- Denim - Gabardine et tissus de chemises
Le Prêt-à-Porter	- Vêtements Denim & Non-Denim - Tricot - Chemises

Le graphique ci-dessous montre la capacité de production annuelle par catégorie dans l'unité de fabrication textile de Tayal, exprimée en unités produites par an. Chaque catégorie représente un segment majeur de la chaîne de production, comprenant le fil, le tissu tissé, le tissu tricoté, le denim, la gabardine et les chemises, ainsi que les vêtements prêts à l'emploi. Ces capacités reflètent une organisation diversifiée optimisée pour répondre aux divers besoins du marché, depuis les matières premières (fils et tissus) jusqu'aux produits finis (vêtements prêts à porter).

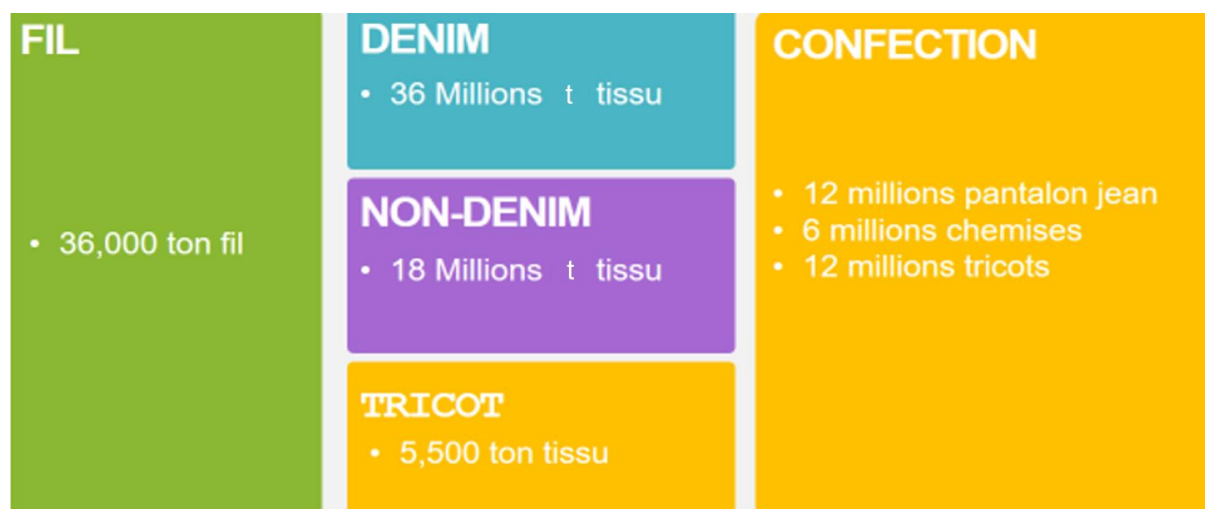


Figure 1-3: Capacité de production annuelle de diverses catégories de produits³ textiles.

1.6.4 Les unités de production de TAYAL SPA

TAYAL SPA est structurée en plusieurs unités de production, chacune ayant un rôle spécifique dans le processus de fabrication textile. Grâce à cette organisation intégrée, l'entreprise assure une production optimisée allant du stockage des matières premières jusqu'à la confection des produits finis. Cette structuration permet non seulement une meilleure gestion des ressources, mais aussi un contrôle accru sur la qualité et l'efficacité de la chaîne de production.

Toutefois, dans le cadre de cette étude, l'attention sera portée sur quatre unités spécifiques, directement impliquées dans le processus de filature. Cette sélection repose sur leur impact stratégique au sein du cycle de production, notamment en ce qui concerne la transformation des matières premières en fils, leur stockage et leur préparation pour les étapes suivantes. L'analyse de ces unités vise à identifier les défis opérationnels, à évaluer les méthodes de gestion en place et à explorer d'éventuelles opportunités d'amélioration en matière de planification et d'optimisation des flux de production.

La répartition et le fonctionnement de ces unités sont illustrés dans la figure 1.5 ci-dessous.

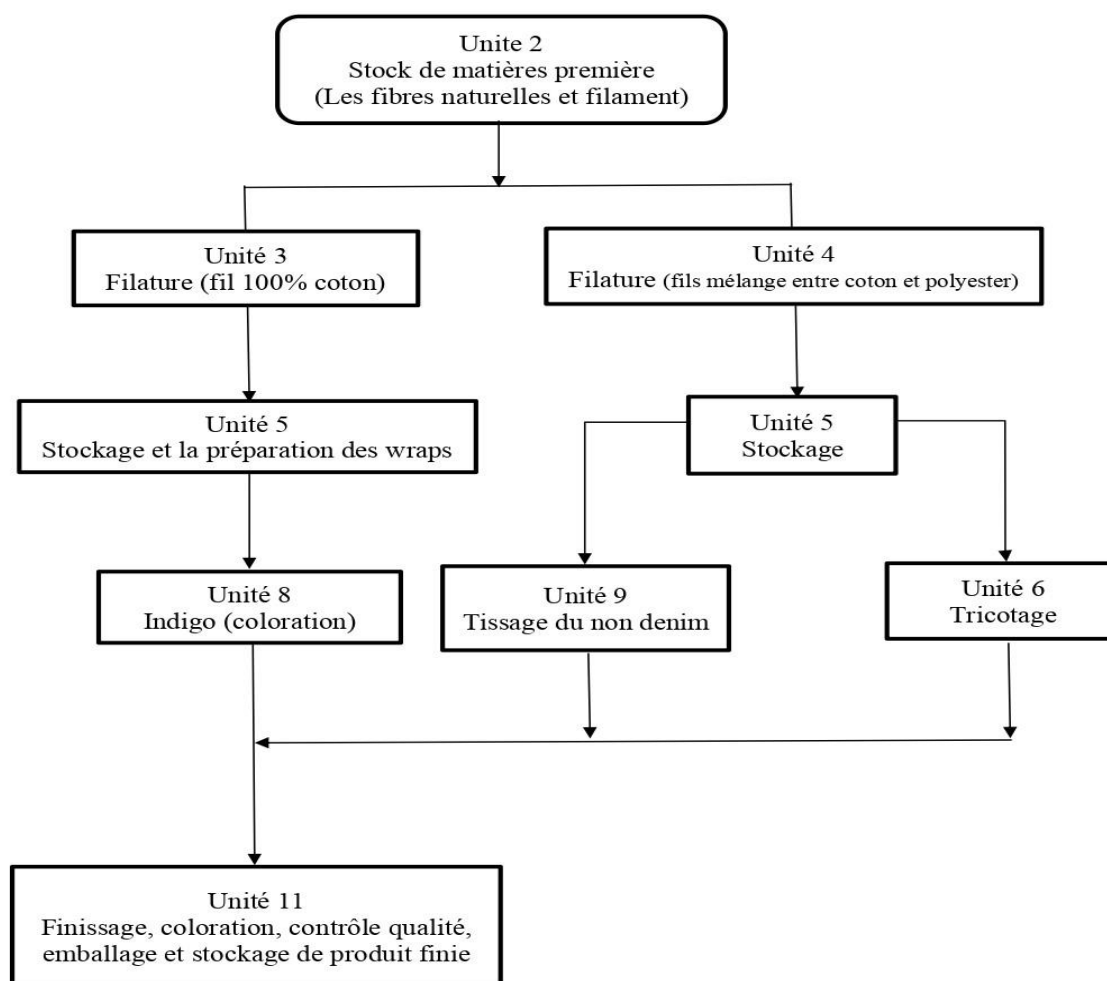


Figure 1-4: Schéma de l'enchaînement des opérations et des relations fonctionnelles entre les unités de production chez TAYAL

1.6.4.1 Unité 2 (Stockage des Matières Premières)

L'Unité 2 est une composante essentielle du processus de production textile de TAYAL SPA. Elle constitue le point de départ de la chaîne de fabrication, où sont stockées les matières premières nécessaires à la production de fils et de tissus. Cette unité joue un rôle crucial dans l'approvisionnement des unités de filature, garantissant la disponibilité et la gestion efficace des fibres utilisées dans les différentes étapes de transformation.

A. Organisation et Fonctionnement de l'Unité 2

L'Unité 2 est un entrepôt de grande capacité, couvrant une superficie totale de 9 072 m², conçu pour stocker plusieurs types de matières premières, notamment :

- **Coton brut** : Principalement sous forme de balles compactées pour une conservation optimale.
- **Waste Cotton (coton recyclé)** : Déchets issus des processus de cardage et de peignage, réutilisés pour minimiser le gaspillage.

- **Polyester** : Stocké en bobines ou en balles, destiné aux mélanges avec du coton.
- **Viscose** : Une fibre semi-synthétique dérivée de la cellulose, utilisée pour améliorer la qualité des fils.
- **Modal et Tencel** : Fibres cellulosiques haut de gamme, reconnues pour leur douceur et leur résistance.
- **Paraffine** : Utilisée comme lubrifiant dans la filature pour améliorer la glisse des fils et réduire les ruptures⁶.

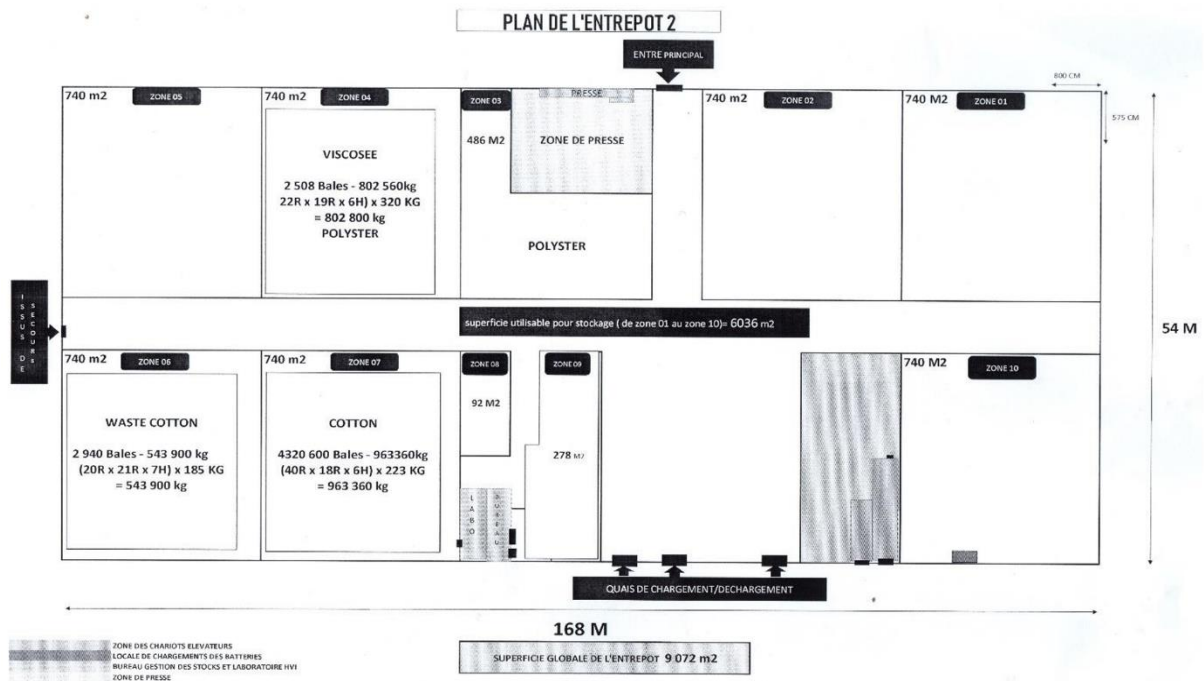


Figure 1-5 : Un schéma de l'organisation interne de l'Unité 2⁷

B. Répartition des Espaces et Zones Fonctionnelles

L'Unité 2 est divisée en plusieurs zones, chacune dédiée à une fonction précise :

⁶ Observation personnelle réalisée lors du stage à TAYAL SPA (Décembre 2024 - Janvier 2025).

⁷ Planification de Production, Rapport interne, 2019

Tableau 1-3 : Répartition des Espaces de l'unité 2

Zone	Description
Zone de Stockage du Coton	Stocke les balles de coton brut et recyclé avant transformation.
Zone de Stockage du Polyester	Contient les fibres synthétiques en bobines ou en balles.
Zone de Stockage de la Viscose	Stocke la viscose sous forme de fibres prêtes à être filées.
Zone de Presse	Dédiée à la compression et au conditionnement des matières premières.
Quais de Chargement et Déchargement	Point logistique pour l'entrée et la sortie des matières premières.
Zone de Contrôle Qualité	Vérification des matières premières avant leur acheminement vers la filature.

C. Répartition des Matières Premières dans l'Unité 2

L'Unité 2 de TAYAL SPA constitue un espace stratégique où sont entreposées les matières premières indispensables au processus de production textile. Son organisation repose sur une classification rigoureuse des matériaux en fonction de leur nature et de leur usage dans la chaîne de fabrication. Cette structuration permet d'optimiser la gestion des stocks, de faciliter les flux logistiques et d'assurer un contrôle qualité efficace.

Les fibres naturelles, telles que le coton grec et turc, sont conservées sous forme de balles compactées afin de garantir leur intégrité avant leur transformation. De même, les fibres synthétiques et cellulosiques, notamment le polyester, la viscose et le Tencel, sont stockées sous forme de bobines ou de balles selon leurs propriétés spécifiques.

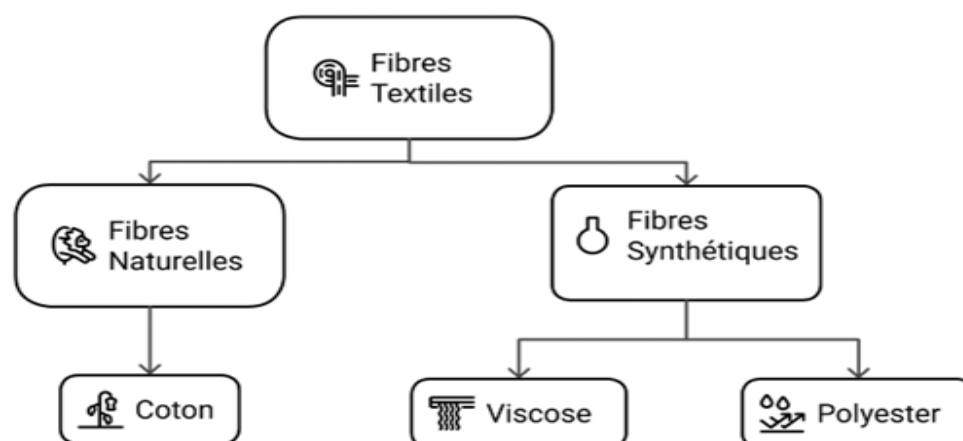


Figure 1-6 : la classification des fibres textiles (fibres naturelles vs fibres synthétiques)

Par ailleurs, la gestion des déchets de coton (Waste Cotton) représente un axe majeur de la politique de durabilité de l'entreprise. Ces matières résiduelles issues des étapes de cardage et de peignage sont soigneusement collectées, recyclées et réintégrées dans certains processus de production afin de réduire le gaspillage et d'optimiser l'utilisation des ressources.

En complément, divers produits auxiliaires comme la paraffine et le fil métallique sont également stockés au sein de l'unité. Ces éléments, bien que secondaires, jouent un rôle essentiel dans le bon déroulement des opérations de filature et de tissage, garantissant ainsi la qualité et la performance des fils produits.

Tableau 1-4: Tableau récapitulatif des matières premières stockées dans l'Unité 2

Famille de matières	Code article	Description du produit	Unité de mesure (UoM)
Lycra Elastane	000001-EL-CREORA-20DN-WHITE-R	Elastane 20 DEN	Box
	000002-EL-CREORA-30DN-WHITE-R	Elastane 30 DEN	Box
	000003-EL-CREORA-40DN-WHITE-R	Elastane 40 DEN	Box
	000004-EL-CREORA-70DN-WHITE-R	Elastane 70 DEN	Box
Fil Polyester	000005-PL-RBKARESI-50DN-WHITE-R	PES-RBKARESI-50/36-WHT-R	PCS
Coton	000002-CO-GREEK-SW-MLS-NA-STD	Coton grec (Medium Long Stable)	Bale
	000003-CO-SPAIN-SW-MLS-NA-STD	Coton espagnol (Medium Long Stable)	Bale
	000004-CO-USASHORT-SW-SS-NA-STD	Coton USA Short Cotton	Bale
	000009-CO-USAMEDIUM-SW-MLS-NA-STD	Coton USA Medium Long Stable	Bale
Waste Cotton (Coton recyclé)	000001-04-YNRHARDWST	Déchet de fil (Murata Ustupu)	Bale
	000004-02-CROVINGSTR	Roving Stripper waste (100% CO)	Bale
	000005-02-CROVINGSTR	Roving Stripper waste (Blend)	Bale
	000006-02-FLOORWST	Floor Waste	Bale
	000009-02-BCARDWASTE	Card Waste (Blend)	Bale
	000015-02-CSLIVERWST	Silver Waste (100% CO)	Bale
	000018-02-COMBERLONG	Comber Piecing Waste (Long Stable)	Bale
	000020-02-CCARDWST	Card Waste (100% CO)	Bale

	000028-05-WFBWASTE	Waste Clipped Fabric	Bale
	000056-02-BLOWROMPW	Briseur (Blowroom)	Bale
Viscose	000005-VI-LENZING-NA-32MM-1.3DTEX-OPW	Fibre de viscose 1,3dtex/32mm	Bale
Polyester	000006-PL-SASA-NA-38MM-1.6DTEX-OPW	Polyester 38 mm 1.6dtex Sasa optic white	Bale
Modal Fiber	000007-MD-LENZING-NA-38MM-1.3DTEX-ECR	Modal fiber 38mm-1,3dtex Lenzing	Bale
Tencel Fiber	000008-TN-LENZING-NA-38MM-1.3DTEX-ECR	Tencel fiber 38mm-1,3dtex Lenzing	Bale
Paraffine (Wax)	000243-PROD-WAX-MURATAWAX	Paraffine (Open End) 20,57543	KG
	000244-PROD-WAX-RIETERAWAX	Paraffine (Murata) 15,43843	KG
Fil Métallique	000289-PROD-FIL-FIL	Fil métallique 4,5MM	Wire

D. Le Rôle de la Sous-Douane dans l'Unité 2

Une particularité importante de l'Unité 2 est l'existence d'un espace sous-douane, qui fonctionne comme une extension du système douanier. Lorsque TAYAL SPA ne dispose pas de fonds suffisants pour l'achat immédiat de coton, l'État algérien peut intervenir en acquérant ces matières premières, qui sont ensuite stockées dans la sous-douane de l'usine. Cette zone permet à TAYAL SPA d'accéder progressivement aux quantités nécessaires à la production, en procédant à un dédouanement progressif selon les besoins de l'entreprise. Cette approche permet une meilleure gestion des flux financiers et évite les interruptions de production en assurant une réserve stratégique de coton accessible à tout moment.

L'ensemble de cette organisation témoigne de l'importance stratégique de l'Unité 2 dans la gestion des matières premières chez TAYAL SPA. Une structuration optimale de l'espace et une logistique bien maîtrisée permettent d'assurer un approvisionnement fluide des unités de production, contribuant ainsi à la stabilité et à l'efficacité du processus industriel.

1.6.4.2 Unités 03 et 04 (Processus de Filature)

Les unités 3 et 4 de TAYAL SPA constituent le cœur de la filature, garantissant la transformation des matières premières en fils destinés au tissage et au tricotage. Ces deux unités, bien que similaires dans leur fonctionnement, se distinguent par la nature des matières premières traitées :

- **Unité 3** : Spécialisée dans la production de fils 100 % coton, utilisés pour des textiles nécessitant une qualité et une résistance accrues.
- **Unité 4** : Dédiée à la production de fils mélangés coton-polyester, adaptés aux besoins de performance et de durabilité dans la confection textile.

La gestion efficace de ces unités est essentielle pour assurer un approvisionnement continu aux unités de tissage et de tricotage, tout en garantissant une production optimisée en termes de qualité et de rendement.

A. Organisation et Répartition des Espaces dans les Unités 3 et 4

L'optimisation des espaces et la fluidité des opérations sont des enjeux majeurs dans les unités de filature. Chaque zone est conçue pour répondre à un besoin spécifique et faciliter la transition des matières premières à l'étape de production. Voici la répartition des zones fonctionnelles :

Tableau 1-5: Un tableau illustratif détaillant l'organisation spatiale des unités 3 et 4

Zone	Description
Zone de réception des fibres	Assure le stockage temporaire des matières premières et leur inspection avant utilisation.
Zone de préparation des fils	Mélange et ajustement des fibres pour garantir une qualité homogène du fil.
Zone de filature principale	Accueille les machines de filature responsables de la transformation des fibres en fils.
Zone de contrôle qualité	Vérifie la conformité des fils produits selon les normes de TAYAL SPA.
Zone de stockage des fils	Stocke temporairement les bobines avant leur transfert vers les unités de tissage ou de tricotage.

Ce tableau synthétise les principales zones fonctionnelles impliquées dans le processus de filature au sein des unités 3 et 4. Afin de mieux visualiser cette organisation spatiale, (la figure 1-8) suivante en présente une représentation graphique.

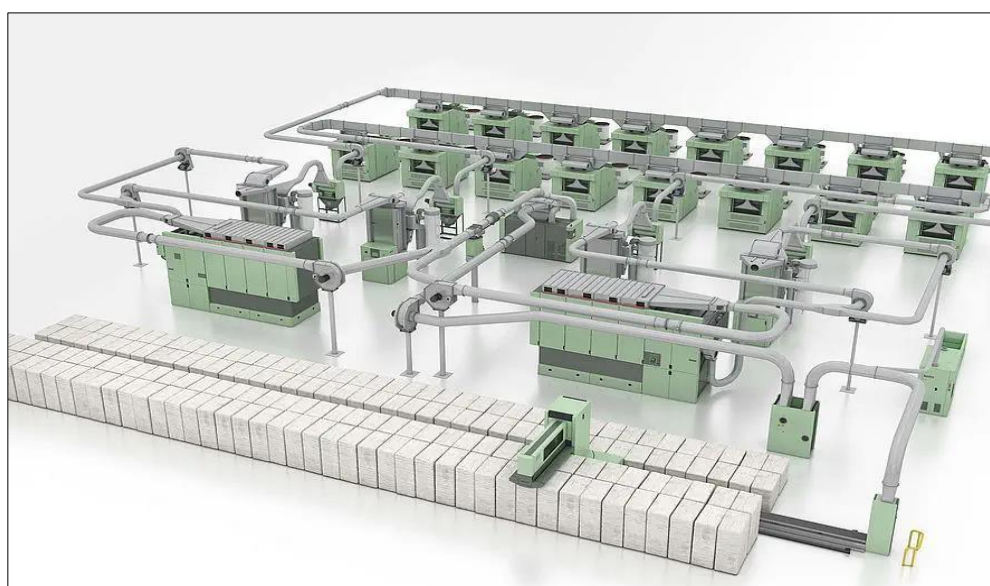


Figure 1-7: Dessin montrant une partie de l'organisation spatiale des unités 3 et 4.

B. Classification des Types de Fils Produits

La qualité du fil dépend du procédé de filage utilisé. Chaque type de fil suit un chemin spécifique avec les machines appropriées, ce qui affecte ses propriétés finales telles que la résistance, l'uniformité et l'utilisation prévue. Vous trouverez ci-dessous une explication détaillée des trois principales catégories de fils produites.

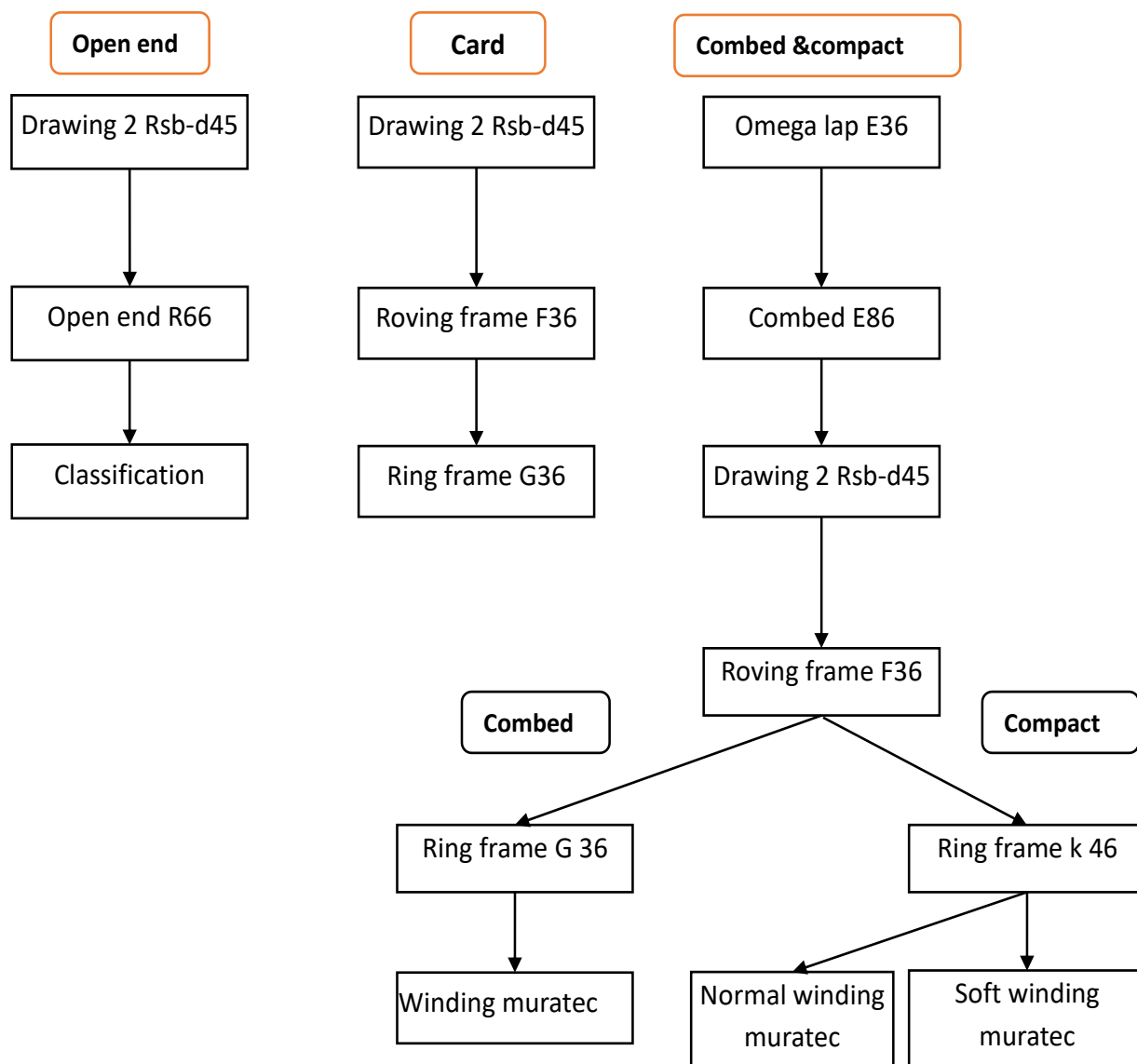


Figure 1-8: Schéma du Processus de Filature et des Différents Types de Fils Produits ⁷

Le fil open-end est fabriqué à l'aide d'un processus de filage sans boucle, ce qui le rend plus rapide et plus économique que les autres méthodes. Après un premier étirage à l'aide de la Drawing 2 RSB-D 45, le fil est envoyé vers la machine Open End R66, où les fibres sont filées dans une chambre rotative sans nécessiter de torsion mécanique conventionnelle. Ce procédé produit un fil moins cher mais aussi moins résistant et moins uniforme. Ce type de fil est généralement utilisé dans les textiles de faible qualité. Une fois le fil produit, il subit un classement pour garantir un certain degré d'uniformité avant utilisation.

Le fil cardé est un compromis entre qualité et coût, offrant une meilleure régularité et résistance par rapport à l'Open End. Le processus débute avec la Drawing 2 RSB-D 45, qui homogénéise les mèches avant qu'elles ne passent par le Roving Frame F36, un banc d'étirage qui réduit progressivement la section du fil tout en lui appliquant une légère torsion. Ensuite, la

machine Ring Frame G36 finalise la filature en appliquant une torsion plus stable et contrôlée, garantissant un fil de meilleure qualité.

La filature peignée et compacte est destinée aux fils de haute qualité, utilisés dans les textiles haut de gamme et techniques. Le processus commence par la machine Omega Lap E36, qui prépare des nappes homogènes avant d'être envoyées à la Combed E86, où les fibres courtes sont éliminées pour obtenir un fil plus fin et plus solide. La Drawing 2 RSB-D 45 intervient ensuite pour assurer une uniformité parfaite avant le passage au Roving Frame F36⁷. À partir de cette étape, le processus se divise en deux variantes :

- **Fil Peigné (Combed) :** Il passe par la Ring Frame G36, qui assure une structure régulière et résistante, avant d'être enroulé sur la machine Winding Muratec.
- **Fil Compact (Compact) :** Il est traité par la Ring Frame K46, une machine spécialisée qui réduit la pilosité du fil, améliorant ainsi sa finesse et sa résistance. Le fil est ensuite bobiné soit avec un enroulement classique (Normal Winding Muratec) soit avec un enroulement souple (Soft Winding Muratec) selon l'application prévue.

1.6.4.3 Unité 5 (Stockage et Préparation des Fils)

L'Unité 5 de TAYAL SPA est une entité essentielle dans la gestion des fils après la phase de filature. Elle assure le stockage, la préparation et l'expédition des fils vers leurs destinations finales, que ce soit pour l'exportation, la teinture indigo, le tissage, le tricotage ou les commandes locales. Une gestion efficace des stocks dans cette unité est primordiale pour éviter les ruptures de production et garantir un approvisionnement fluide des unités en aval.

A. Répartition des Espaces et Zones Fonctionnelles

L'Unité 5 est organisée en plusieurs zones fonctionnelles afin de faciliter la gestion des flux de fils et d'optimiser les opérations de stockage et de préparation.

Tableau 1-6: Répartition des Espaces et Zones Fonctionnelles de l'Unité 5

Zone / Type de Fil	Description	Destination Finale
Zone de stockage des fils pour tissage	Contient les fils destinés à l'Unité 8 pour le tissage.	Envoyés vers l'Unité 8 pour la fabrication de textiles.
Zone de stockage des fils pour tricotage	Stocke les fils utilisés pour le tricotage.	Alimentent les machines de tricotage de l'usine.
Zone des retours de fils	Regroupe les fils renvoyés par la production pour réutilisation ou correction.	Réintégrés après contrôle qualité ou reconditionnés pour d'autres usages.
Zone des fils destinés à l'exportation	Contient les fils emballés et prêts à être expédiés vers les marchés internationaux.	Emballés et stockés avant expédition vers les marchés étrangers.

Zone des commandes locales	Répond aux besoins des clients nationaux selon leurs commandes spécifiques.	Livrés directement aux clients nationaux.
Zone des équipements d'emballage	Espace dédié au stockage des cartons, cornes et autres emballages pour la protection des fils.	

B. Préparation des Wraps pour l'Indigo

Une partie des fils stockés dans l'Unité 5 est destinée à la coloration indigo avant leur transformation en textile. Cette phase est essentielle, car elle conditionne l'apparence et la qualité des tissus finaux.

Étapes du processus de préparation des wraps :

- Tri des fils selon leur compatibilité avec la teinture.
- Conditionnement et emballage des fils pour éviter tout dommage avant teinture.
- Envoi vers l'Unité de teinture Indigo avant leur transfert en Unité 8 pour le tissage.

Une bonne gestion de cette étape permet d'optimiser l'absorption de la teinture et d'améliorer la régularité des couleurs sur les textiles finis.

C. Emballage et Équipements de Protection des Fils

Afin de protéger et conserver la qualité des fils stockés, l'Unité 5 utilise plusieurs types d'emballages et équipements logistiques⁸.

Tableau 1-7 : Types d'Emballages et Leur Utilité pour la Conservation des Fils ⁷

Équipement d'Emballage	Utilité
Cartons	Protègent les bobines de fil lors du stockage et du transport.
Cornes	Maintiennent la structure des bobines et facilitent leur manipulation.
Film plastique	Assure une protection contre l'humidité et la poussière.
Palettes en bois	Facilitent l'organisation et le transport des bobines.

⁸ Rapport de stage, TAYAL SPA

1.7 Conclusion

L'industrie textile repose sur des chaînes de production complexes où la coordination entre les différentes unités est essentielle. Malgré l'utilisation de technologies modernes, nous avons constaté que TAYAL SPA rencontre des difficultés en matière de planification de la production et de gestion des stocks, ce qui impacte directement son efficacité opérationnelle et sa compétitivité.

L'un des principaux problèmes que nous avons identifiés réside dans le manque de synchronisation entre la production et la demande réelle. L'entreprise adopte une approche de production de masse, sans outils de prévision avancés, ce qui entraîne des déséquilibres majeurs : une surproduction en période de faible demande, immobilisant des ressources, et une incapacité à répondre efficacement aux pics de commandes, causant des retards et une insatisfaction des clients.

En parallèle, la gestion des stocks demeure inefficace. Nous avons observé une accumulation excessive de matières premières et de fils, générant des coûts de stockage élevés, tandis que certaines pénuries perturbent la production et ralentissent le processus industriel. L'absence d'un système dynamique de planification, basé sur des analyses précises, empêche d'adapter rapidement les niveaux de production aux fluctuations du marché, aggravant ainsi les inefficacités.

Ces lacunes entraînent des surcoûts, des retards et une instabilité dans l'approvisionnement, compromettant la rentabilité et la satisfaction des clients. Il est donc primordial de proposer des solutions innovantes pour améliorer la gestion de ces aspects stratégiques.

À travers notre étude, nous cherchons à apporter des réponses concrètes à ces problématiques. L'intégration de technologies avancées et d'outils analytiques permettrait d'optimiser la planification de la production, de réduire les coûts liés aux stocks et d'améliorer la compétitivité de l'entreprise. Notre approche vise ainsi à repenser les méthodes actuelles en s'appuyant sur des modèles prédictifs et des systèmes intelligents, afin d'assurer une production plus fluide, réactive et alignée avec les besoins du marché.

Chapitre II :

**Prévision Temporelle : Synergie entre Modélisation
Statistique et Machine Learning**

2. Introduction

Dans un environnement industriel marqué par une forte variabilité de la demande, la prévision constitue un levier stratégique essentiel pour optimiser les opérations et renforcer la compétitivité. Ce besoin est particulièrement crucial pour TAYAL SPA, acteur majeur du secteur textile en Algérie, où les fluctuations saisonnières nécessitent une planification rigoureuse.

Ce chapitre propose un cadre théorique axé sur l'harmonisation entre les modèles statistiques classiques (ARIMA, SARIMA, Prophet) et les approches modernes de Machine Learning (XGBoost, Optuna). Cette synergie vise à tirer parti des dynamiques linéaires capturées par les méthodes traditionnelles tout en intégrant les capacités prédictives des techniques d'apprentissage automatique pour modéliser les non-linéarités complexes. L'objectif est de justifier le recours à des modèles hybrides pour optimiser la précision des prévisions dans un contexte industriel exigeant.

2.1 Fondements généraux de la prévision

La prévision représente un processus analytique essentiel permettant d'anticiper l'évolution de variables stratégiques au sein des flux industriels, qu'il s'agisse de flux de production, de matières premières, de ressources énergétiques ou de capacités logistiques. Elle repose sur l'exploitation rigoureuse de données historiques, l'analyse de tendances cycliques et l'intégration de facteurs exogènes tels que les fluctuations du marché ou les aléas climatiques (Hyndman & Athanasopoulos, 2018). En offrant une visibilité accrue sur les besoins futurs, la prévision constitue un levier stratégique majeur pour l'optimisation des ressources, le pilotage des opérations industrielles et l'amélioration de la réactivité organisationnelle. Sur le plan opérationnel (Chopra & Meindl, 2016), elle permet une synchronisation fine entre la planification de la production, l'approvisionnement en matières premières et la gestion des capacités de stockage, minimisant ainsi les coûts d'entreposage et les risques de congestion logistique (Simchi-Levi, 2014). Dans un environnement incertain, elle offre également une flexibilité accrue pour s'adapter rapidement aux variations de flux, optimisant ainsi la chaîne de valeur et renforçant la compétitivité industrielle.

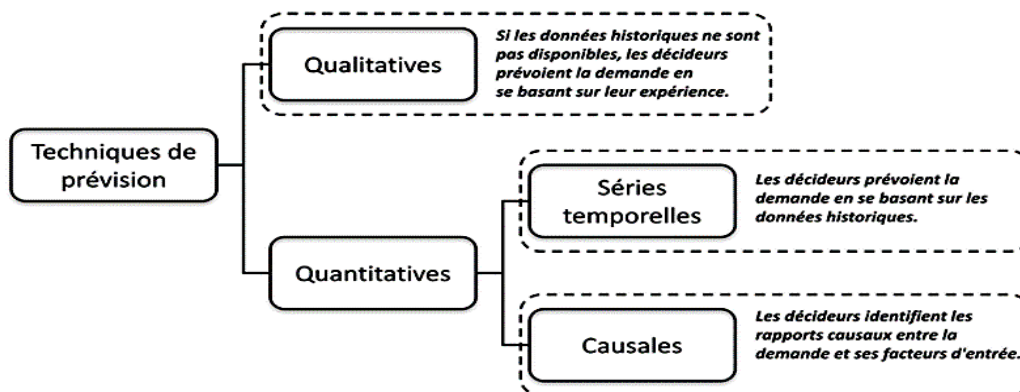


Figure 2-1: Classification des méthodes de prévision

2.2 Modèles statistiques de séries temporelles

Les modèles statistiques de séries temporelles constituent un cadre analytique essentiel pour la modélisation prédictive des données chronologiques. Depuis les travaux fondateurs de Box et Jenkins, ces approches ont démontré leur efficacité dans l'anticipation de phénomènes cycliques et saisonniers (Box & Jenkins, 1976). Leur robustesse méthodologique repose sur l'identification des schémas temporels récurrents, permettant ainsi d'améliorer la prise de décision dans divers contextes industriels.

2.2.1 Lissage exponentiel (Holt–Winters)

La méthode de Holt–Winters est une technique de lissage exponentiel développée pour prévoir des séries temporelles présentant simultanément une tendance et une saisonnalité (Hyndman & Athanasopoulos, 2018). Elle repose sur la décomposition de la série en trois composantes fondamentales :

- **Niveau** (L_t) : Représente la moyenne estimée de la série à l'instant t .
- **Tendance** (T_t) : Indique l'évolution (croissante ou décroissante) de la série.
- **Saisonnalité** (S_t) : Traduit les fluctuations périodiques.

Chaque composante est ajustée au fil du temps à l'aide de coefficients de lissage notés α (niveau), β (tendance) et γ (saisonnalité), tous compris dans l'intervalle $[0,1]$. Le modèle Holt Winters se décline en deux variantes principales : additive ou multiplicative, en fonction du comportement de la saisonnalité.

Dans le cas du modèle additif, adapté aux séries où la saisonnalité a une amplitude constante, les mises à jour s'effectuent selon les équations suivantes pour une périodicité saisonnière m (par ex., $m = 12$ pour des données mensuelles) :

- **Mise à jour du niveau**

$$L_t = \alpha(y_t - S_{t-m}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad 2-1$$

- **Mise à jour de la tendance**

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad 2-2$$

- **Mise à jour de la saison**

$$S_t = \gamma(y_t - L_t) + (1 - \gamma)S_{t-m} \quad 2-3$$

- **Prévision à k pas**

$$\hat{y}_{t+k} = L_t + k T_t + S_{t+k-m} \quad 2-4$$

Dans ces expressions, y_t désigne l'observation de la série à l'instant t . Les paramètres α , β et γ sont généralement calibrés par minimisation de l'erreur quadratique moyenne sur un échantillon d'apprentissage.

La méthode Holt–Winters est particulièrement efficace pour des séries présentant une croissance stable et des effets saisonniers réguliers. Elle offre une approche simple d'implémentation tout en fournissant une interprétation intuitive des différentes dynamiques sous-jacentes à la série. Toutefois, ses performances peuvent être altérées en présence de changements structurels soudains ou de saisonnalités irrégulières (Hyndman & Athanasopoulos, 2018)

2.2.2 Modèles ARIMA et extension SARIMA

2.2.2.1 Les séries stationnaires et non-stationnaires

L'application des modèles prédictifs dans le cadre des séries temporelles repose sur une distinction fondamentale entre les séries stationnaires et non stationnaires. Cette classification conditionne directement le choix et l'efficacité des modèles statistiques utilisés, notamment dans le cadre d'ARIMA et de ses extensions. En effet, les méthodes traditionnelles de modélisation exigent souvent une transformation préalable des séries non stationnaires en séries stationnaires, permettant ainsi de capturer les dynamiques sous-jacentes de manière plus précise (Box, Jenkins, & Reinsel, 2008). Cette étape d'ajustement, bien que technique, optimise l'interprétation des relations temporelles et renforce la robustesse des prédictions.

Dans un contexte industriel, cette distinction s'avère cruciale pour anticiper les fluctuations de la demande, l'évolution des coûts de production ou encore les variations des capacités logistiques. Par conséquent, l'intégration de processus de stationnarisation constitue une étape préparatoire indispensable pour garantir la fiabilité des prévisions réalisées à l'aide de modèles statistiques (Hyndman & Athanasopoulos, 2018). Cette approche permet non seulement de stabiliser les composantes volatiles, mais également d'améliorer la convergence des modèles lors des phases d'entraînement et de validation (Tsay, 2010)

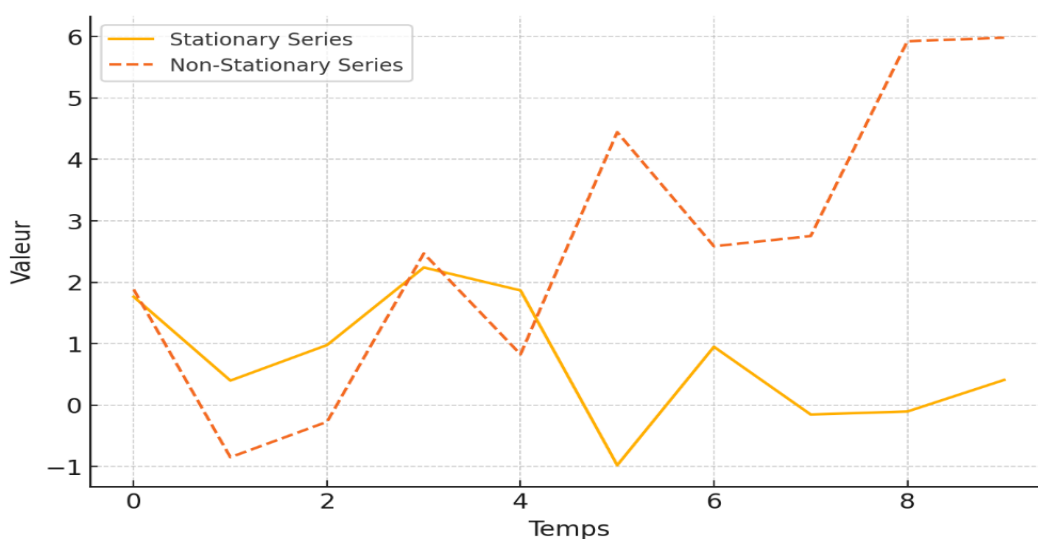


Figure 2-2: Séries Stationnaires Et Non-Stationnaires

Le graphique ci-dessous illustre la distinction entre les séries stationnaires et non-stationnaires. La série stationnaire oscille autour d'une moyenne constante, tandis que la série non-stationnaire présente une tendance croissante au fil du temps. Cette représentation visuelle permet de mieux comprendre l'importance de la stationnarisation avant l'application de modèles prédictifs tels qu'ARIMA.

2.2.2.2 Le modèle ARIMA

Le modèle ARIMA (AutoRegressive Integrated Moving Average) constitue une méthode largement répandue pour l'analyse et la prévision des séries chronologiques. Il est particulièrement adapté aux données non stationnaires, c'est-à-dire dont la moyenne, la variance ou la covariance varient dans le temps (Charpentier, 2012). Ce modèle combine trois composantes essentielles : l'auto-régression (AR), l'intégration (I), et la moyenne mobile (MA).

Le modèle ARIMA est défini par trois paramètres notés (p, d, q) , où :

- p : représente l'ordre du processus autorégressif, c'est-à-dire le nombre de valeurs passées utilisées comme prédicteurs.
- d : indique le degré de différenciation nécessaire pour rendre la série stationnaire.
- q : correspond à l'ordre du modèle à moyenne mobile, reliant la valeur actuelle à une combinaison linéaire des erreurs passées.

L'équation générale du modèle $ARIMA(p, d, q)$ s'écrit comme suit :

$$\Phi_p(L)(1 - L)^d Y_t = \Theta_q(L)\varepsilon_t \quad 2-5$$

Avec :

- $\Phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$: le polynôme autorégressif,
- $\Theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$: le polynôme de moyenne mobile,
- L : l'opérateur de retard ($LY_t = Y_{t-1}$),
- ε_t : le terme d'erreur blanche (bruit aléatoire).

Le modèle opère en trois étapes principales :

- La **différenciation** (composante I) est appliquée pour supprimer la tendance ou la saisonnalité et rendre la série stationnaire.
- La **composante autorégressive** (AR) modélise la relation entre les observations passées et la valeur actuelle.
- La **composante moyenne mobile** (MA) prend en compte les erreurs de prédiction passées pour améliorer la prévision (Wasi, 2022).

Le modèle ARIMA est un outil robuste et polyvalent pour la prévision des séries chronologiques, à condition que la série soit rendue stationnaire et que les paramètres soient

correctement estimés. Il constitue une référence incontournable dans les approches quantitatives de prévision (Charpentier, 2012) .

2.2.2.3 La méthode SARIMA (Seasonal ARIMA)

La méthode SARIMA (Seasonal Autoregressive Integrated Moving Average) est une extension du modèle ARIMA, spécifiquement conçue pour les séries chronologiques présentant une saisonnalité (Charpentier, 2012). Elle combine les composants non saisonniers (ARIMA classique) et saisonniers, permettant ainsi de modéliser une large variété de comportements temporels.

Le modèle SARIMA est noté sous la forme :

$$\text{SARIMA}(p, d, q)(P, D, Q)_s$$

Où :

- **p, d, q**: ordres respectifs de l'auto-régression (AR), de la différenciation (I) et de la moyenne mobile (MA) pour la composante **non saisonnière** ;
- **P, D, Q**: ordres analogues mais pour la composante **saisonnaire** ;
- **s** : période de la saisonnalité (par ex. 12 pour des données mensuelles avec une saisonnalité annuelle).

Le modèle SARIMA est défini à partir de l'application successive des différenciations non saisonnière et saisonnière sur la série temporelle y_t . Il peut être formulé comme suit :

$$\Phi_P(B^s) \cdot \phi_p(B) \cdot \nabla^d \nabla_s^D y_t = \Theta_Q(B^s) \cdot \theta_q(B) \cdot \varepsilon_t \quad 2-6$$

- B : l'opérateur de décalage (backshift), tel que ; $B^k y_t = y_{t-k}$
- $\nabla = 1 - B$: opérateur de différenciation non saisonnière d'ordre d ;
- $\nabla_s = 1 - B^s$: opérateur de différenciation saisonnière d'ordre D ;
- $\phi_p(B)$: polynôme autorégressif non saisonnier d'ordre p ;
- $\Phi_P(B^s)$: polynôme autorégressif saisonnier d'ordre P , avec des retards multiples de s ;
- $\theta_q(B)$: polynôme de moyenne mobile non saisonnière d'ordre q ;
- $\Theta_Q(B^s)$: polynôme de moyenne mobile saisonnière d'ordre Q ;
- ε_t : un bruit blanc, supposé centré et à variance constante (homoscédastique).

Les principales composantes du modèle SARIMA sont :

- **Différenciation non saisonnière (d)** : suppression de la tendance générale,
- **Différenciation saisonnière (D)** : élimination des effets saisonniers récurrents,
- **Composantes AR et MA** : modélisation des dépendances à court et à long terme (Wasi, 2022).

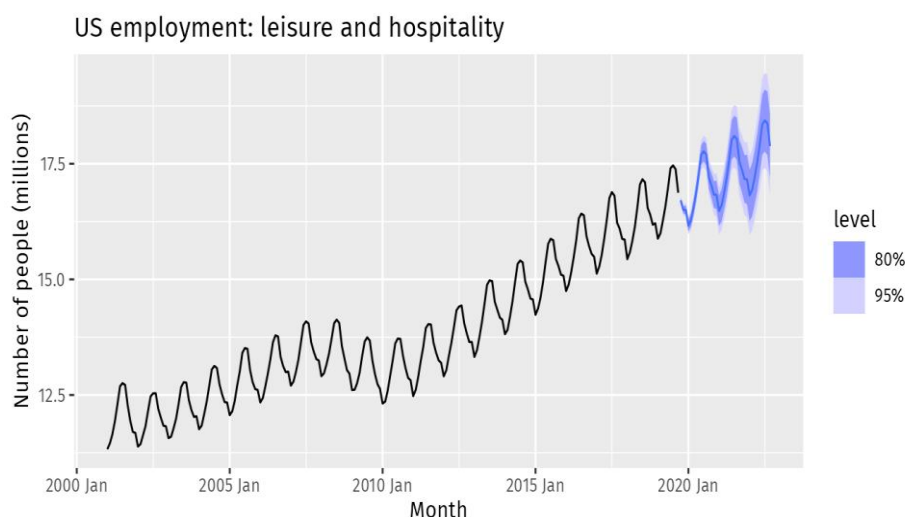


Figure 2-3: Forecasts of monthly US leisure and hospitality employment using the ARIMA (Hyndman & Athanasopoulos, 2021)

Le graphique illustre les prévisions de l'emploi mensuel dans le secteur des loisirs et de l'hôtellerie aux États-Unis, générées par un modèle $ARIMA(2,1,0)(1,1,1)$. La courbe noire représente les données historiques de l'emploi, marquées par une tendance haussière continue ainsi qu'une forte saisonnalité visible à chaque cycle annuel. À partir de 2020, les projections du modèle sont visualisées, accompagnées de deux intervalles de confiance : la zone bleu foncé indique un niveau de confiance de 80 %, tandis que la zone bleu clair représente un niveau de confiance de 95 %. Ces intervalles traduisent l'incertitude croissante liée aux prédictions futures. Le modèle capture efficacement les variations saisonnières et les tendances à long terme, tout en fournissant des marges d'erreur pour évaluer la fiabilité des prévisions.

2.2.2.4 Identification et sélection des paramètres par ACF, PACF, AIC/BIC

L'identification des paramètres dans les modèles ARIMA et SARIMA représente une étape clé pour assurer la pertinence des prévisions. Elle repose principalement sur l'analyse des corrélations temporelles, à travers les fonctions ACF (Autocorrelation Function) et PACF (Partial Autocorrelation Function), permettant d'estimer les ordres autorégressifs et de moyenne mobile. Par ailleurs, les critères d'information tels que l'AIC (Akaike Information Criterion) et le BIC (Bayesian Information Criterion) facilitent la sélection optimale des modèles en équilibrant précision et complexité (Montgomery, 2015). Ces outils analytiques jouent un rôle fondamental dans la structuration des modèles prédictifs adaptés aux dynamiques observées.

A. Analyse de l'ACF (Autocorrelation Function)

L'ACF permet de mesurer la corrélation entre une valeur actuelle de la série temporelle et ses valeurs passées. En d'autres termes, elle observe dans quelle mesure les variations présentes sont liées à celles observées à des instants antérieurs. Lorsque l'ACF décroît lentement ou oscille sans s'annuler, cela suggère une mémoire persistante dans le processus, typique des processus de type MA (Moving Average) (Box, Jenkins, Reinsel, & Ljung, 2015). Cette persistance traduit une dépendance prolongée avec les données passées, laissant des empreintes visibles même après plusieurs lags.

L'analyse graphique de la fonction d'autocorrélation (ACF) révèle des pics significatifs aux premiers lags, suggérant l'existence d'une structure de dépendance temporelle non négligeable. Cette configuration indique la présence probable d'une composante de moyenne mobile (MA) à intégrer dans le modèle. Ces observations sont cohérentes avec les principes formulés par (Shumway & Stoffer, 2017), qui mettent en évidence l'importance d'une décroissance rapide ou lente de l'ACF comme indicateur clé pour caractériser les processus de type MA.

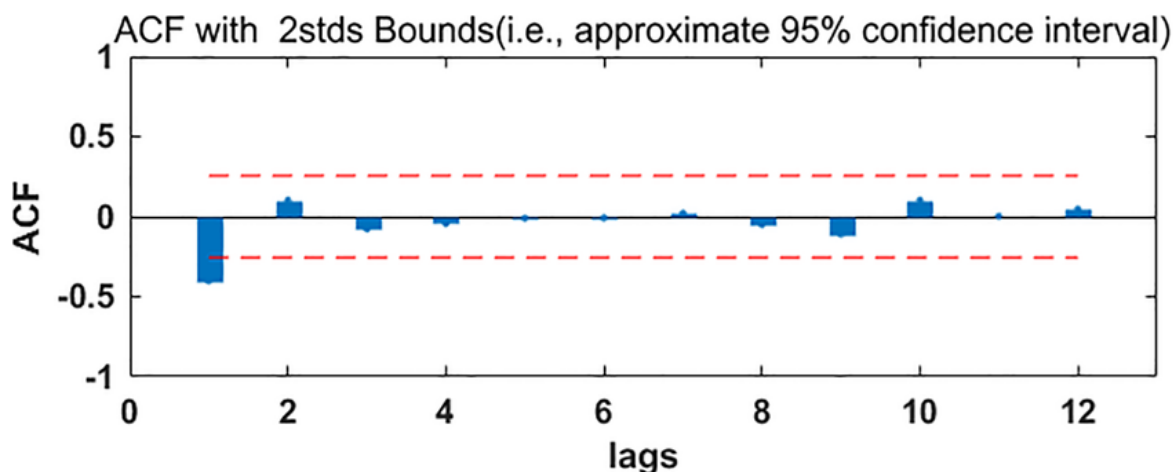


Figure 2-4 : Représentation graphique de l'ACF sur la série temporelle étudiée (Zheng & Zhu, 2020) .

Le graphique de la fonction d'autocorrélation (ACF) permet d'évaluer la dépendance temporelle d'une série en mesurant la corrélation entre ses valeurs présentes et passées, appelées lags ou décalages. Dans ce graphique, les barres bleues représentent les coefficients d'autocorrélation pour chaque lag, tandis que les lignes rouges pointillées délimitent les intervalles de confiance à 95 %. L'analyse montre que seule l'autocorrélation au premier lag dépasse légèrement la borne inférieure de l'intervalle, indiquant une corrélation faiblement significative à ce niveau. Les autres lags restent à l'intérieur des bornes, suggérant une absence de dépendance au-delà du premier retard. Cette configuration est caractéristique d'un processus de moyenne mobile (MA) d'ordre faible, reflétant une dynamique où l'effet des chocs ou perturbations s'atténue rapidement après une seule période.

B. Analyse de la PACF (Partial Autocorrelation Function)

Contrairement à la fonction d'autocorrélation (ACF), la fonction d'autocorrélation partielle (PACF) se concentre uniquement sur les corrélations directes, en éliminant l'influence des lags intermédiaires. Elle permet ainsi d'identifier plus clairement l'ordre autorégressif (AR) d'un modèle. Lorsque la PACF chute brutalement après un certain lag, cela suggère un processus AR d'ordre équivalent (Brockwell & Davis, 2002).

Le graphique explicatif permet de visualiser cette distinction fondamentale : alors que l'ACF tient compte à la fois des corrélations directes et indirectes, la PACF agit comme un filtre en ne conservant que les interactions directes. L'observation de décroissances marquées dans la PACF valide l'hypothèse d'un processus AR dominant. Ces résultats s'inscrivent

pleinement dans le cadre théorique proposé par (Brockwell & Davis, 2002), qui soulignent la pertinence de l'analyse de la PACF pour l'estimation de l'ordre AR.

Le graphe explicatif que nous avons généré permet de visualiser cette différence fondamentale : là où l'ACF considère toutes les connexions possibles (directes et indirectes), la PACF ne conserve que les interactions directes, agissant ainsi comme un filtre. L'analyse de la PACF montre des décroissances marquées, ce qui valide l'hypothèse d'un processus AR dominant. Ces résultats confirment les théories développées par (Brockwell & Davis, 2002) sur la pertinence de l'interprétation de la PACF pour identifier l'ordre AR.

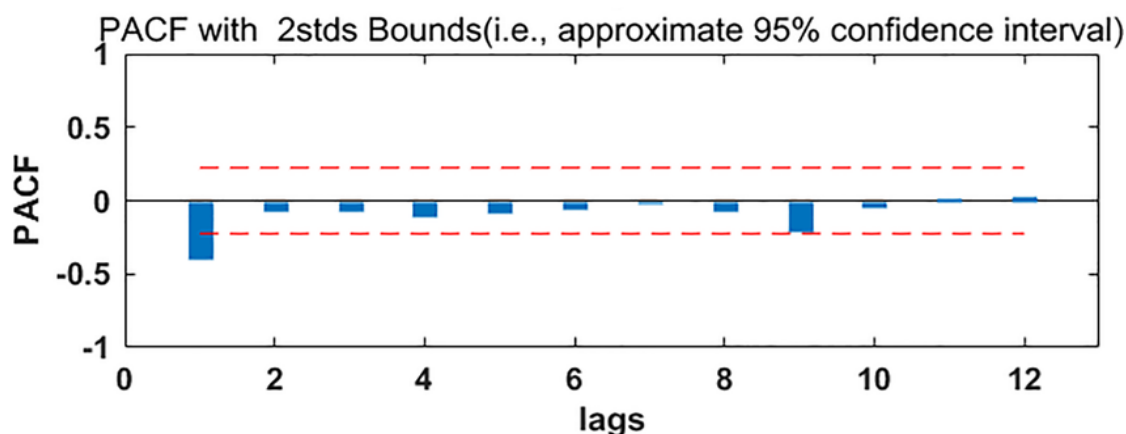


Figure 2-5: Représentation graphique de la PACF sur la série temporelle étudiée (Zheng & Zhu, 2020).

Le graphique PACF montre les corrélations partielles, c'est-à-dire les corrélations entre les valeurs actuelles et les valeurs passées, tout en éliminant l'effet des intervalles intermédiaires. Sur le graphique PACF, le seul pic significatif est visible au premier lag. Les autres lags ne franchissent pas les seuils de signification. Cela indique un processus AR (AutoRegressive) d'ordre 1. La chute brutale après le premier lag valide l'hypothèse d'un AR(1), c'est-à-dire que l'observation actuelle dépend uniquement de la valeur précédente, sans effet notable des valeurs antérieures.

C. Sélection de modèle par AIC et BIC

Une fois les ordres initiaux identifiés, il est nécessaire de comparer plusieurs modèles afin de sélectionner celui qui fournit le meilleur compromis entre qualité d'ajustement et simplicité. Deux critères majeurs guident cette sélection :

- **Akaike Information Criterion (AIC)** : Favorise les modèles qui équilibrent précision d'ajustement et nombre de paramètres, en minimisant la perte d'information (Montgomery, 2015) .
- **Bayesian Information Criterion (BIC)** : Applique une pénalisation plus forte pour la complexité du modèle, encourageant des structures plus simples (Charpentier, 2012)

Dans la pratique, comme le montre Wasi (2022) à travers des implémentations sur Kaggle, on ajuste plusieurs modèles aux données, puis on retient celui qui présente l'AIC ou le BIC le plus bas (Wasi, 2022). Cette méthode garantit une sélection objective, limitant le risque de surapprentissage et améliorant la généralisation du modèle.

Tableau 2-1 : Comparaison des valeurs AIC et BIC pour les modèles ARIMA testés.

(Mian & Khan, 2021)

Modèle	AIC	BIC
<i>ARIMA(0, 1, 0)</i>	3153.97	3161.47
<i>ARIMA(1, 1, 0)</i>	3145.82	3157.17
<i>ARIMA(0, 1, 1)</i>	3144.90	3156.24
<i>ARIMA(1, 1, 1)</i>	3145.55	3160.67
<i>ARIMA(0, 1, 2)</i>	3146.49	3161.62
<i>ARIMA(1, 1, 2)</i>	3146.13	3165.03

Le tableau comparatif affiché montre clairement les valeurs obtenues pour plusieurs configurations ARIMA. Le modèle optimal est celui qui minimise à la fois l'AIC et le BIC, indiquant un compromis optimal entre complexité et performance prédictive. Dans ce cas, le modèle *ARIMA(1,1,1)* semble être le plus approprié, affichant les scores les plus faibles.

2.2.3 Modèle Prophet

Le modèle Prophet, développé par Facebook en 2017, représente une avancée significative dans le domaine de la prévision des séries temporelles. Conçu pour modéliser des données présentant des variations complexes et des changements soudains, Prophet s'appuie sur une décomposition additive des séries temporelles. Cette méthodologie permet de capturer les dynamiques évolutives d'une série en trois composantes distinctes : la tendance, la saisonnalité et les effets ponctuels liés aux jours fériés (Taylor & Letham, 2017)

2.2.3.1 Structure et décomposition du modèle

Le fonctionnement de Prophet se distingue par sa capacité à segmenter le signal temporel en trois éléments principaux. La première composante, appelée tendance, représente l'évolution structurelle de la série sur le long terme. Elle capte les augmentations ou diminutions progressives, tout en permettant l'identification automatique des changements brusques appelés changepoints, qui modifient la trajectoire initiale de la série.

La seconde composante, la saisonnalité, modélise les variations cycliques observées à intervalles réguliers, tels que les fluctuations annuelles, mensuelles ou hebdomadaires. Prophet se distingue par sa capacité à détecter ces cycles récurrents sans nécessiter d'ajustement manuel, optimisant ainsi la précision des prévisions.

Enfin, les effets des jours fériés permettent d'intégrer les perturbations liées à des événements exceptionnels, comme les fêtes nationales, les événements commerciaux ou les crises économiques. Cette modélisation fine capture les pics de demande et les baisses d'activité associés à ces périodes spécifiques, renforçant ainsi la capacité prédictive du modèle (Taylor & Letham, 2017).

2.2.3.2 Intérêt stratégique

Le modèle Prophet présente plusieurs caractéristiques qui le rendent particulièrement pertinent pour les applications industrielles. Sa robustesse face aux ruptures de tendance permet de détecter automatiquement les changements structurels dans les séries temporelles, assurant ainsi une précision accrue dans les prévisions. Prophet se distingue également par sa tolérance aux données manquantes, garantissant une continuité des prédictions même en cas d'informations partielles. De plus, l'intégration de variables exogènes optimise la prise en compte des variations dues à des facteurs externes, renforçant ainsi la fiabilité des projections dans des environnements dynamiques. Cette flexibilité et cette modularité en font un outil performant pour l'analyse des cycles saisonniers complexes et les fluctuations irrégulières (Taylor & Letham, 2017).

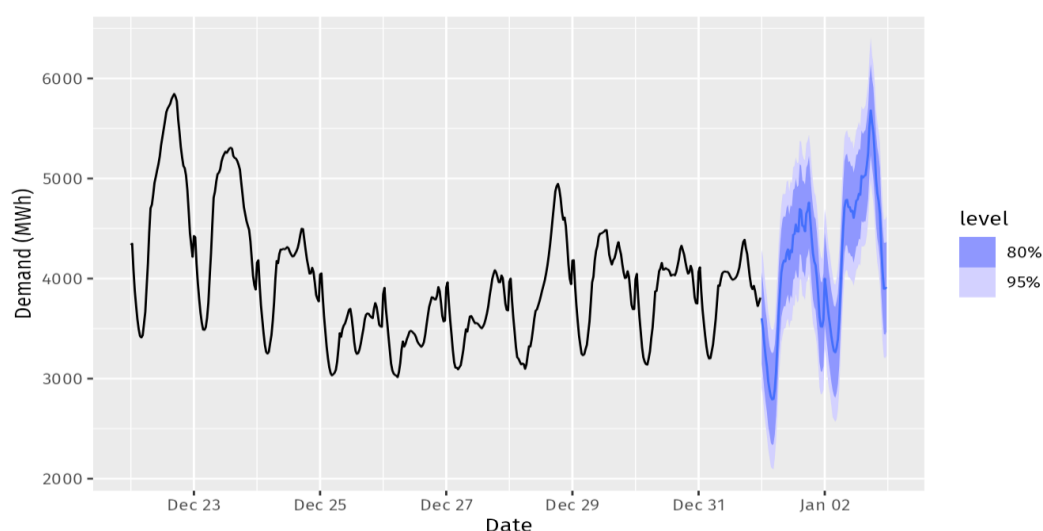


Figure 2-6:Prévisions sur deux jours du modèle Prophet pour la demande d'électricité de l'État de Victoria (Hyndman & Athanasopoulos, 2021).

Le graphique présente les prévisions de la demande en énergie (MWh) réalisées par le modèle Prophet. La courbe noire représente les données historiques, tandis que les zones bleues indiquent les intervalles de confiance à 80 % (bleu foncé) et à 95 % (bleu clair). Ces intervalles traduisent l'incertitude associée aux prévisions pour les premiers jours de janvier. Approches par apprentissage automatique

2.3 Approches par apprentissage automatique

2.3.1 Principes généraux du Machine Learning pour la prévision

Le Machine Learning (apprentissage automatique) désigne un ensemble de techniques permettant à des algorithmes d'apprendre à partir de données historiques afin d'effectuer des prédictions ou de prendre des décisions sans être explicitement programmés pour chaque situation (Nielsen, 2019). Dans le domaine de la prévision, ces méthodes représentent une avancée majeure par rapport aux approches statistiques classiques. Alors que les modèles traditionnels reposent souvent sur des hypothèses de linéarité et de stationnarité, les algorithmes

de Machine Learning sont capables de capturer des relations complexes, non linéaires et évolutives au sein des données (Brownlee, 2018).

L'application du Machine Learning à la prévision permet notamment d'améliorer la précision des estimations dans des contextes caractérisés par une grande variabilité et une forte incertitude, tels que les environnements industriels. De nombreux algorithmes sont utilisés dans ce cadre, notamment les réseaux de neurones et les méthodes de boosting.

Parmi ces méthodes, XGBoost (Extreme Gradient Boosting) s'est imposé comme l'un des algorithmes les plus performants pour les tâches de régression et de classification, et connaît un succès croissant dans le domaine de la prévision. Grâce à sa capacité à modéliser efficacement des relations complexes et à intégrer différentes sources de données, XGBoost constitue aujourd'hui une approche de référence pour les projets industriels de prévision avancée (Nielsen, 2019).

2.3.2 XGBoost appliqué aux séries temporelles

XGBoost (Extreme Gradient Boosting) est un algorithme d'apprentissage supervisé basé sur l'optimisation par gradient, développé par (Chen & Guestrin, 2016) Il s'inscrit dans la famille des méthodes d'ensemble boosting, dont le principe est de combiner plusieurs modèles faibles (généralement des arbres de décision) pour créer un modèle robuste et performant. XGBoost est reconnu pour sa précision prédictive, sa vitesse d'exécution et sa capacité à traiter des données de qualité variable (valeurs manquantes, bruit, etc.), ce qui en fait un choix populaire dans de nombreux domaines, notamment la finance, la santé, l'énergie, et les prévisions industrielles.

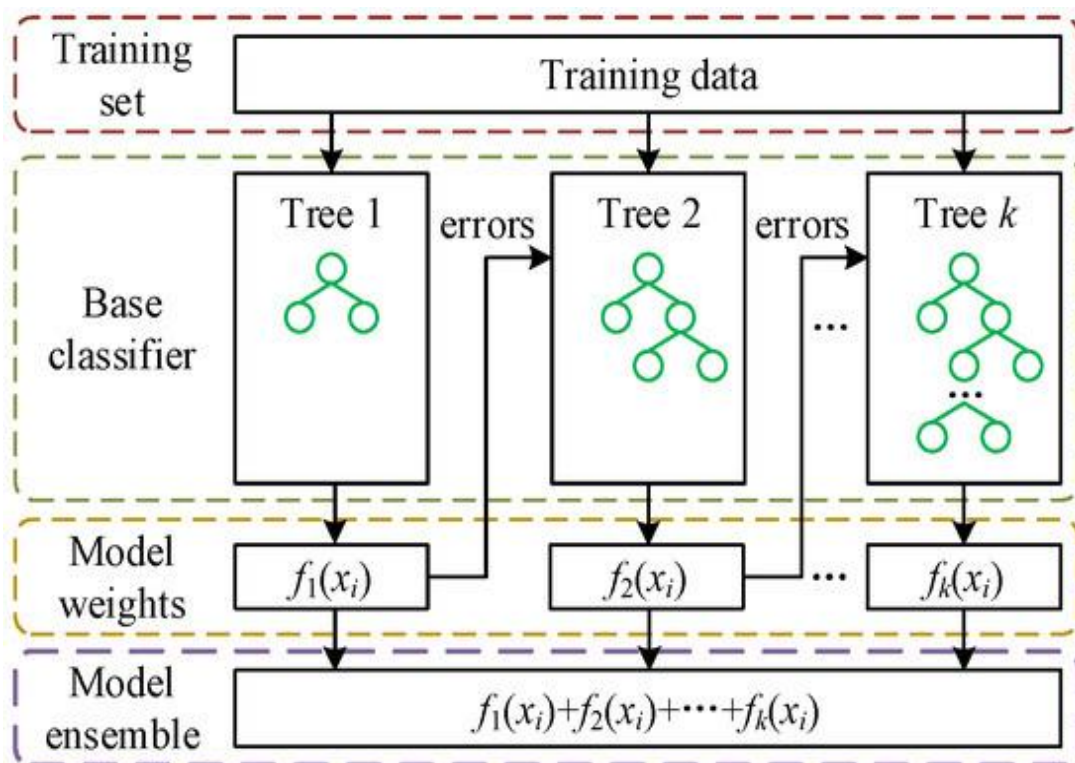


Figure 2-7: Schéma structurel du modèle XGBoost (Zhang & Li, 2023).

XGBoost repose sur une stratégie itérative où, à chaque étape, un nouvel arbre de décision est ajouté pour corriger les erreurs commises par la somme des prédictions précédentes. L'objectif global du modèle est de minimiser une fonction de coût composée de deux éléments :

- **Un terme de perte** $l(y_i, \hat{y}_i)$, mesurant l'écart entre les prédictions et les observations réelles ;
- **Un terme de régularisation** $\Omega(f)$, pénalisant la complexité du modèle afin de limiter le surapprentissage.

La formule de fonction objective :

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad 2-7$$

Avec :

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad 2-8$$

- l : fonction de perte (ex. : erreur quadratique pour les régressions),
- $\Omega(f)$: terme de régularisation,
- T : nombre de feuilles dans l'arbre,
- w : scores des feuilles.

XGBoost se distingue par sa grande précision et son efficacité, grâce à une optimisation de second ordre combinant gradient et hessienne, ainsi qu'à une régularisation L1/L2 limitant le surapprentissage. Il gère automatiquement les valeurs manquantes, parallélise la construction des arbres pour une exécution rapide, et fournit des mesures d'importance des variables pour une interprétation facilitée du modèle (Chen & Guestrin, 2016).

2.3.3 Optimisation des hyperparamètres avec Optuna

Dans le cadre du Machine Learning, l'optimisation des hyperparamètres représente une étape critique visant à maximiser les performances prédictives d'un modèle tout en minimisant les erreurs de généralisation. Les hyperparamètres sont des paramètres non appris pendant l'entraînement, mais plutôt définis avant le processus d'apprentissage, influençant directement le comportement et l'efficacité des algorithmes. L'objectif de l'optimisation est de trouver la combinaison idéale de ces paramètres afin de réduire les erreurs de prédiction et d'améliorer la robustesse du modèle (Li, Jamieson, DeSalvo, Rostamizadeh, & Talwalkar, 2017).

2.3.3.1 Optuna : Un Cadre d'Optimisation Bayésienne

Optuna est un framework d'optimisation d'hyperparamètres de nouvelle génération, fondé sur des méthodes bayésiennes. Conçu pour améliorer l'efficacité du processus de recherche, il adopte une stratégie adaptative et intelligente pour explorer l'espace des hyperparamètres. Contrairement aux approches classiques telles que la Grid Search ou la Random Search, qui

testent de manière exhaustive ou aléatoire différentes combinaisons, Optuna exploite un algorithme probabiliste — le Tree-structured Parzen Estimator (TPE) — pour orienter la recherche vers les régions les plus prometteuses de l'espace.

Le rôle central d'Optuna est de résoudre un problème d'optimisation : on lui fournit une fonction objective à minimiser (par exemple, l'erreur de validation) ou à maximiser (par exemple, une métrique de performance comme l'accuracy). Cette fonction prend en entrée un ensemble d'hyperparamètres et retourne une valeur de score. Optuna procède alors à une série d'essais (ou "trials"), où chaque essai correspond à une combinaison d'hyperparamètres testée. Grâce à son moteur bayésien, Optuna ajuste progressivement sa stratégie de recherche pour concentrer l'exploration sur les configurations les plus performantes.

Cette capacité à optimiser intelligemment une fonction complexe en fait un outil puissant, particulièrement adapté aux modèles d'apprentissage automatique dont la performance dépend fortement du réglage des hyperparamètres. (Akiba & al, 2019).

2.3.3.2 Principes de Fonctionnement

A. Tree-structured Parzen Estimator (TPE):

L'algorithme Tree-structured Parzen Estimator (TPE) repose sur une modélisation probabiliste de l'espace de recherche des hyperparamètres. Le TPE segmente les configurations explorées en deux distributions distinctes : les configurations performantes, appelées Good Group, et les configurations moins performantes, désignées comme Bad Group. Cette séparation permet de concentrer l'exploration sur les régions les plus prometteuses de l'espace de recherche, optimisant ainsi le nombre d'essais nécessaires pour atteindre des performances optimales (Bergstra, Bardenet, & Bengio, 2011).

Modélisation Probabiliste

Le TPE s'appuie sur la construction de deux modèles de probabilité :

$l(x)$: La probabilité qu'une valeur d'hyperparamètre soit parmi les meilleures configurations (basée sur les essais réussis).

$g(x)$: La probabilité que cette valeur soit parmi les moins performantes.

L'algorithme maximise le ratio suivant pour chaque essai :

$$x^* = \operatorname{argmax} \frac{l(x)}{g(x)} \quad 2-9$$

Cette approche permet à Optuna de se concentrer davantage sur les zones de l'espace de recherche où les performances observées sont élevées, accélérant ainsi la convergence vers des configurations optimales (Bergstra, Bardenet, & Bengio, 2011).

B. Processus d'Optimisation

Le processus se déroule en plusieurs étapes :

1. **Initialisation** : Quelques essais aléatoires sont effectués pour construire les premières estimations de $l(x)$ et $g(x)$
2. **Échantillonnage sélectif** : Les valeurs des hyperparamètres sont échantillonnées préférentiellement dans les zones les plus prometteuses ($l(x)$).
3. **Mise à jour des distributions** : Après chaque itération, les distributions probabilistes $l(x)$ et $g(x)$ sont recalculées pour affiner l'exploration.
4. **Pruning Dynamique** : Les essais jugés peu performants sont interrompus pour économiser les ressources computationnelles (Akiba & al, 2019).

C. Critère de Sélection (Expected Improvement (EI))

L'algorithme *TPE* maximise l'Expected Improvement (*EI*) afin de sélectionner les prochaines configurations à tester. L'Expected Improvement est défini comme suit :

$$EI(x) = E[\max(0, f_{best} - f(x))] \quad 2-10$$

Où :

- f_{best} : la meilleure performance observée,
- $f(x)$: la performance prédite pour les hyperparamètres x

Ce critère d'amélioration permet à l'algorithme de prioriser les configurations présentant le plus grand potentiel de gain, favorisant ainsi une convergence rapide vers les paramètres optimaux (Bergstra, Bardenet, & Bengio, 2011).

La (Figure 2.8) ci-dessous illustre le processus d'optimisation réalisé par l'algorithme TPE. On y observe les différentes étapes, depuis l'initialisation aléatoire jusqu'à l'affinement progressif des distributions probabilistes, guidant ainsi l'exploration vers les configurations optimales.

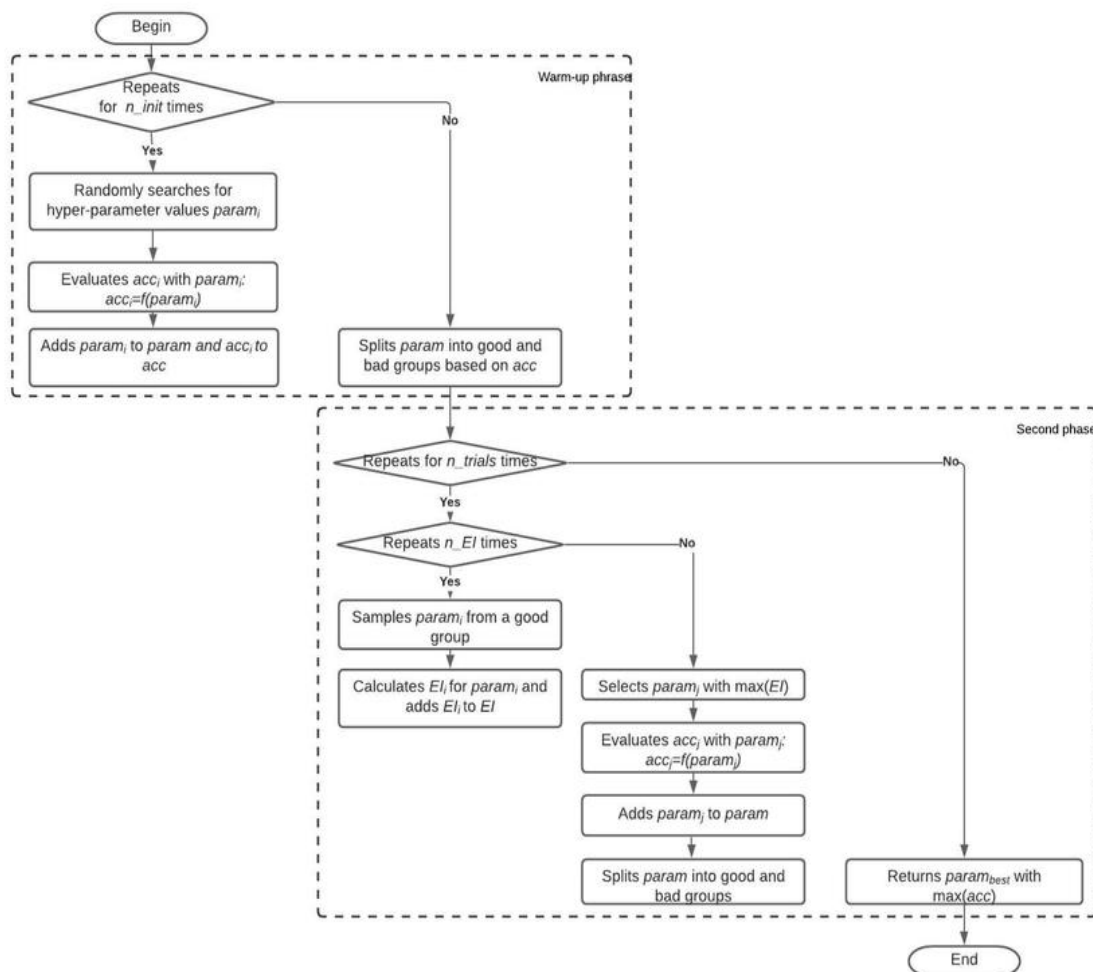


Figure 2-8: Représentation schématique de l’algorithme d’optimisation TPE (Kapočiūtė-Dzikienė, 2020).

D. Stratégie Adaptative :

La Stratégie Adaptative repose sur un ajustement dynamique de l’exploration des hyperparamètres en fonction des performances observées lors des essais précédents. Cette approche permet de concentrer les ressources computationnelles sur les zones les plus prometteuses de l’espace de recherche, favorisant ainsi une convergence plus rapide vers les configurations optimales. À chaque itération, les essais peu performants sont élagués grâce au Pruning Dynamique, ce qui libère des ressources pour les configurations plus prometteuses. Cette optimisation progressive garantit une exploration plus efficace de l’espace de recherche tout en réduisant le coût computationnel (Akiba & al, 2019).

2.4 Analyse critique des approches existantes

Les approches classiques offrent rigueur et interprétabilité mais manquent de flexibilité, tandis que le ML est adaptatif et performant mais souvent opaque et dépendant des données.

2.4.1 Forces et limites des modèles classiques

Les modèles classiques de prédiction temporelle, tels que les modèles ARIMA, SARIMA et les méthodes de lissage exponentiel, se distinguent par leur capacité à modéliser efficacement

les séries stationnaires et à capturer les dynamiques linéaires de manière robuste (Box, Jenkins, Reinsel, & Ljung, 2015).

2.4.1.1 Forces des modèles classiques :

Les modèles classiques de prévision temporelle, tels qu'ARIMA et SARIMA, présentent plusieurs atouts notables. Leur principal avantage réside dans leur simplicité d'interprétation et de mise en œuvre. Basés sur des concepts statistiques bien établis, ces modèles offrent une structure paramétrique claire, permettant de modéliser les évolutions temporelles de manière précise et compréhensible (Box, Jenkins, Reinsel, & Ljung, 2015). De plus, SARIMA intègre directement les composantes saisonnières, ce qui permet de capturer efficacement les fluctuations périodiques dans les séries temporelles (Hyndman & Athanasopoulos, 2018).

Un autre avantage majeur est leur faible coût computationnel. Comparés aux modèles d'apprentissage automatique, les modèles ARIMA et SARIMA nécessitent peu de ressources informatiques, même sur de grands ensembles de données. Cette caractéristique les rend particulièrement adaptés aux applications industrielles où les calculs rapides et l'efficacité en termes de ressources sont essentiels (Hyndman & Athanasopoulos, 2018). Ces forces expliquent pourquoi ces modèles restent populaires dans le domaine de la prévision temporelle, malgré l'émergence de méthodes plus modernes.

2.4.1.2 Limites des modèles classiques :

Les modèles classiques de prévision temporelle, tels qu'ARIMA et SARIMA, présentent certaines limitations dans les environnements complexes. Tout d'abord, ils reposent sur l'hypothèse de linéarité, ce qui les empêche de capturer les non-linéarités et les changements structurels soudains dans les séries temporelles (Tsay, 2010).

De plus, ils nécessitent que les données soient stationnaires, imposant ainsi un prétraitement parfois complexe pour stabiliser les moyennes et les variances au cours du temps (Box, Jenkins, Reinsel, & Ljung, 2015). Enfin, ces modèles intègrent difficilement des variables explicatives multiples, limitant ainsi leur capacité à prendre en compte les facteurs externes influençant les prédictions (Hyndman & Athanasopoulos, 2018). Ces contraintes réduisent leur efficacité dans les environnements industriels modernes où les dynamiques sont souvent non linéaires et multidimensionnelles.

2.4.1.3 Conclusion critique

Les modèles classiques de prévision temporelle, tels qu'ARIMA et SARIMA, sont robustes et facilement interprétables, mais ils montrent leurs limites face aux dynamiques non linéaires et aux perturbations exogènes. Leur dépendance à la stationnarité et leur incapacité à intégrer plusieurs variables explicatives réduisent leur efficacité dans les environnements industriels complexes (Hyndman & Athanasopoulos, 2018). Ces limitations justifient le recours à des approches hybrides, combinant méthodes classiques et modèles d'apprentissage automatique, pour améliorer la précision prédictive.

2.4.2 Atouts et contraintes des méthodes ML « pures »

Les méthodes de Machine Learning (ML) offrent une capacité prédictive supérieure pour capturer les relations complexes et non linéaires au sein des séries temporelles (Goodfellow,

Bengio, & Courville, 2016). Contrairement aux modèles classiques, les approches de ML ne reposent pas sur des hypothèses de stationnarité, ce qui les rend particulièrement performantes dans des environnements volatils et imprévisibles.

2.4.2.1 Atouts des méthodes ML « pures »

L'algorithme XGBoost (Extreme Gradient Boosting) s'impose comme un choix optimal pour la modélisation prédictive, en particulier dans le contexte des séries temporelles complexes. Ce modèle de gradient boosting se distingue par sa capacité à capturer les non-linéarités inhérentes aux données, contrairement aux méthodes classiques telles que ARIMA et SARIMA, qui reposent sur l'hypothèse de linéarité et de stationnarité (Goodfellow, Bengio, & Courville, 2016).

En outre, XGBoost offre une robustesse accrue face aux données bruitées et aux valeurs manquantes grâce à ses mécanismes internes d'optimisation itérative, permettant de minimiser les erreurs tout en maintenant une grande stabilité dans les prévisions (Chen & Guestrin, 2016). Sur le plan de l'efficacité computationnelle, XGBoost est également reconnu pour sa rapidité et son implémentation optimisée en mémoire, le rendant particulièrement adapté aux environnements nécessitant des traitements à grande échelle. Par ailleurs, cet algorithme excelle dans l'intégration de multiples variables explicatives (indicateurs économiques, conditions météorologiques, données exogènes) sans perte de performance, ce qui enrichit significativement les modèles prédictifs (Chollet, 2018).

Enfin, XGBoost intègre des techniques de régularisation telles que L1 et L2, ainsi que le shrinkage, qui permettent de réduire le risque de surapprentissage (overfitting) et d'améliorer la généralisation sur des données non observées (Chen & Guestrin, 2016). L'ensemble de ces caractéristiques positionne XGBoost comme un modèle de référence pour les prévisions hybrides, notamment lorsqu'il est combiné à des approches traditionnelles comme SARIMA ou Prophet, en capturant non seulement les tendances globales, mais aussi les résidus non linéaires avec une précision accrue. Ces éléments démontrent clairement sa supériorité dans les tâches de prévision, particulièrement dans les environnements complexes et dynamiques.

2.4.2.2 Contraintes des méthodes ML « pures » :

Les méthodes de Machine Learning (ML) dites « pures » présentent certaines contraintes significatives qui limitent leur adoption dans des contextes industriels ou décisionnels. L'un des principaux défis réside dans le problème de la boîte noire (Black-box), qui désigne l'opacité des modèles en ce qui concerne le processus de prise de décision (Rudin, 2019). Contrairement aux méthodes statistiques traditionnelles où chaque paramètre est explicable, les modèles ML, en particulier les réseaux de neurones profonds, génèrent des prédictions sans offrir de transparence sur les interactions internes entre les variables. Cette complexité algorithmique réduit considérablement l'interprétabilité, limitant ainsi la confiance des décideurs et rendant les modèles plus difficiles à auditer et à justifier.

En parallèle, l'efficacité des modèles ML repose sur l'accès à de vastes volumes de données de haute qualité. En effet, ces algorithmes nécessitent une quantité importante d'exemples pour apprendre des schémas complexes et générer des prédictions fiables. Cette exigence de données massives constitue un obstacle majeur, notamment dans les secteurs où l'historique des données

est limité ou fragmenté. L'absence de données suffisantes peut engendrer un surapprentissage (overfitting) ou, à l'inverse, une incapacité à capter les dynamiques sous-jacentes des séries temporelles.

Enfin, la performance des modèles ML dépend étroitement de l'optimisation fine de leurs hyperparamètres, tels que le taux d'apprentissage, le nombre d'itérations ou encore la profondeur des arbres décisionnels. Une mauvaise configuration peut entraîner des performances médiocres ou un manque de généralisation aux données non observées. Cette sensibilité exige non seulement une expertise technique approfondie, mais également un processus de validation rigoureux pour ajuster ces paramètres de manière optimale. Ces contraintes soulignent les défis associés à l'usage exclusif des méthodes de Machine Learning, malgré leur potentiel en matière de prédiction non linéaire.

2.4.2.3 Conclusion critique

Les méthodes ML pures offrent une grande capacité de modélisation, mais souffrent de problèmes d'interprétabilité et de dépendance aux données massives. Leur nature opaque limite parfois leur adoption dans les environnements nécessitant des justifications explicites.

2.4.3 Motivation pour une hybridation

L'hybridation des modèles de prévision temporelle consiste à combiner les modèles classiques, avec les méthodes de Machine Learning afin de tirer parti des forces de chaque approche. Cette stratégie permet de surmonter les limitations individuelles : les modèles classiques apportent robustesse, interprétabilité et simplicité pour capturer les dynamiques linéaires, tandis que les techniques de Machine Learning modélisent efficacement les non-linéarités et intègrent de multiples variables explicatives sans contrainte de stationnarité (Zhang G. P., 2003). En unissant ces deux paradigmes, l'hybridation améliore la précision des prévisions et renforce la résilience face aux chocs exogènes, répondant ainsi aux exigences des environnements industriels modernes.

2.5 Conclusion

Ce chapitre a présenté un cadre théorique solide sur les approches de prévision temporelle, en mettant en avant les modèles classiques (ARIMA, SARIMA) pour leur robustesse et leur interprétabilité, ainsi que les méthodes de Machine Learning (XGBoost) pour leur capacité à capturer les dynamiques non linéaires. Face aux limites de chaque approche prise isolément, l'hybridation s'est imposée comme une stratégie pertinente pour améliorer la précision prédictive en combinant les forces de chaque méthode. Les modèles hybrides SARIMA–XGBoost–Optuna et Prophet–XGBoost–Optuna illustrent cette complémentarité, optimisant les prévisions tout en réduisant les erreurs résiduelles.

Le chapitre suivant s'appuiera sur cette base théorique pour évaluer expérimentalement ces méthodes dans un contexte industriel réel, notamment au sein de l'entreprise TAYAL SPA, permettant ainsi de valider les performances observées dans un cadre opérationnel.

Chapitre III :
Évaluation Expérimentale Des Approches
Hybrides

3. Introduction

La prévision des séries temporelles est un défi majeur dans de nombreux domaines, incluant l'industrie, la finance, et la gestion de l'énergie. Les modèles classiques comme SARIMA et Prophet sont largement utilisés pour capturer les tendances, les saisonnalités et les points de changement dans les données temporelles (Hyndman & Athanasopoulos, 2018). Cependant, leur capacité à modéliser des relations non linéaires complexes reste limitée, en particulier lorsque les séries temporelles présentent des comportements imprévisibles et des anomalies (Box, Jenkins, Reinsel, & Ljung, 2015).

Pour pallier ces insuffisances, l'approche hybride combinant SARIMA ou Prophet avec XGBoost permet d'améliorer les performances de prévision. Le modèle linéaire (SARIMA/Prophet) extrait les composantes principales telles que les tendances et les cycles saisonniers, tandis que XGBoost, un algorithme d'apprentissage automatique basé sur les arbres de décision, se concentre sur les résidus pour capturer les non-linéarités résiduelles (Chen & Guestrin, 2016). Cette synergie permet d'accroître la robustesse et la précision des prévisions sur divers types de séries temporelles (Zhang G. P., 2003).

Cette approche hybride permet d'améliorer la précision des prévisions en tirant parti des complémentarités entre les modèles, offrant ainsi une meilleure robustesse face aux variations de données et aux changements de régime. Ce projet vise à évaluer expérimentalement cette architecture hybride dans un contexte industriel, en utilisant des jeux de données issus de l'industrie textile (TAYAL SPA) et de sources publiques (USDA). Les résultats obtenus seront comparés aux modèles de référence pour valider l'efficacité de cette approche.

3.1 De la prévision classique à l'IA : une transition clé pour l'industrie textile

L'essor des technologies basées sur l'intelligence artificielle (IA) et le machine Learning transforme profondément les stratégies de prévision de la demande, en particulier dans les secteurs à forte variabilité comme l'industrie textile. Contrairement aux méthodes traditionnelles qui reposent essentiellement sur l'analyse des données historiques, les approches intelligentes intègrent des facteurs exogènes dynamiques tels que les tendances de consommation, les conditions économiques, les comportements d'achat sur les plateformes numériques, ou encore les signaux issus des réseaux sociaux (Makridakis, 2019). Cette capacité d'apprentissage et d'adaptation en temps réel offre aux entreprises une réactivité accrue face aux fluctuations du marché, leur permettant ainsi d'optimiser les niveaux de stock, de réduire les coûts de surproduction et d'améliorer la gestion des ressources (Hyndman & Athanasopoulos, 2018).

Dans le secteur textile, la demande est influencée par des facteurs complexes, tels que la saisonnalité, les tendances de la mode, les cycles de production et les délais d'approvisionnement. L'adoption de modèles intelligents permet d'adapter les stratégies de production en anticipant les ruptures de stock et en évitant l'accumulation d'inventaires (Chopra & Meindl, 2016). Des techniques avancées, comme les modèles SARIMA ou les algorithmes de gradient boosting (XGBoost), se révèlent particulièrement efficaces pour capturer ces dynamiques et générer des prévisions fiables, même dans des contextes instables (Hochreiter, 1997).

L'intégration de l'IA dans la prévision de la demande textile constitue ainsi un levier stratégique majeur pour renforcer la compétitivité des entreprises. Elle permet une gestion proactive et personnalisée de la production et des stocks, améliore la précision des projections et garantit une meilleure allocation des ressources. Cette évolution favorise à la fois la performance industrielle et la satisfaction client.

Enfin, l'utilisation de modèles prédictifs intelligents ne se limite pas à l'optimisation des ventes : elle influence en profondeur la gestion industrielle et les processus de fabrication. À travers l'exemple de TAYAL SPA, nous verrons comment la prévision de la demande s'intègre au cycle de production et contribue à l'amélioration des performances du secteur textile en Algérie.

3.1.1 Contraintes opérationnelles et défis

La synchronisation efficace entre les processus de production, la gestion des stocks et la satisfaction de la demande représente un enjeu stratégique majeur pour TAYAL SPA. Toutefois, l'entreprise est confrontée à plusieurs difficultés qui affectent négativement sa performance opérationnelle et financière.

En premier lieu, la prépondérance des méthodes de planification traditionnelles apparaît inadaptée dans un contexte de fluctuations rapides de la demande (Chopra & Meindl, 2016). Ces approches sont insuffisamment flexibles pour faire face aux variations imprévues propres à la chaîne logistique contemporaine, ce qui engendre une réactivité limitée et un manque de visibilité sur l'ensemble des processus de production.

Ensuite, l'absence de solutions technologiques intégrées dédiées à la gestion opérationnelle constitue un obstacle majeur. En effet, TAYAL SPA ne dispose pas encore de systèmes informatisés permettant de coordonner efficacement les flux de production et d'approvisionnement. Les échanges d'informations se font en grande partie de manière manuelle, via des feuilles de calcul statiques, rendant laborieuse la collecte et l'analyse de données en temps réel (Slack, 2020)

Par ailleurs, la qualité insuffisante des prévisions se traduit régulièrement par des déséquilibres importants dans la gestion des stocks, avec, d'un côté, des excédents de produits finis et, de l'autre, des ruptures en matières premières. Ces désajustements désorganisent le processus de production, génèrent des coûts de stockage supplémentaires et conduisent à des approvisionnements d'urgence onéreux. Ils reflètent également un usage limité d'outils modernes de prévision, comme les méthodes prédictives et l'intelligence artificielle, pourtant de plus en plus prisées pour affiner les estimations de la demande (Makridakis, 2019).

En outre, la production en lots souffre d'une planification insuffisamment rigoureuse en amont, engendrant des interruptions fréquentes lorsque des anomalies sont détectées trop tardivement. Ce processus se traduit par un taux élevé de rebuts et, par conséquent, une augmentation des coûts opérationnels et une baisse de la rentabilité.

Enfin, l'absence d'outils d'analyse avancés (comme les systèmes d'Advanced Planning and Scheduling ou les algorithmes d'apprentissage automatique) limite la capacité de TAYAL SPA à optimiser et à anticiper ses besoins. Sur un marché exigeant et dynamique, l'adoption de ces technologies constitue pourtant un levier déterminant de compétitivité. L'inadéquation

actuelle des pratiques et des outils favorise ainsi une accumulation excessive et inadéquate des stocks de produits finis, accroissant les coûts de stockage, diminuant la valeur des produits et immobilisant des ressources financières précieuses.

Pour toutes ces raisons, adopter une approche de planification modernisée et intégrer des technologies prédictives et de pilotage en temps réel s'avère indispensable afin d'améliorer la synchronisation entre la production, les stocks et la demande du marché. À travers l'implémentation de solutions informatiques dédiées, la formation des équipes aux nouveaux outils et l'utilisation de méthodes de prévision robustes, TAYAL SPA pourrait accroître significativement son efficacité opérationnelle et sa compétitivité sur le long terme.

3.2 Protocole expérimental

Dans le cadre de cette étude, une évaluation expérimentale a été conduite afin de comparer les performances du modèle hybride SARIMA-XGBoost à celles de modèles classiques de prévision de séries temporelles, notamment SARIMA seul et Holt-Winters. L'objectif de cette expérimentation est de démontrer l'intérêt d'une approche hybride capable d'intégrer les propriétés saisonnières et linéaires par un modèle statistique (SARIMA), tout en capturant les structures non linéaires résiduelles via un modèle d'apprentissage automatique (XGBoost). Cette approche a été testée sur deux jeux de données réelles, dans des contextes industriels différents, afin d'évaluer sa généralisable et sa robustesse.

3.2.1 Analyse descriptive des données industrielles internes (TAYAL SPA)

Le premier cas d'étude s'appuie sur un ensemble de données industrielles réelles, extraites du système de planification de la production d'une entreprise opérant dans le secteur textile. Ces données couvrent une période comprise entre 2024 et 2025 et portent spécifiquement sur la production de coton, qui constitue la matière première dominante dans les ordres de fabrication considérés. L'accent mis sur le coton se justifie par son importance significative dans le volume de production global. Toutefois, le modèle élaboré demeure suffisamment flexible pour être adapté à l'analyse d'autres composants de production.

3.2.1.1 Nature des données et contraintes structurelles

Les données exploitées présentent une granularité journalière, avec un total de 1 654 observations consécutives, incluant les jours non ouvrés. Le contexte industriel étudié se caractérise par une stratégie de production anticipée : les unités sont fabriquées puis stockées avant d'être rétroactivement associées à des commandes antérieures, certaines remontant à plusieurs années avant leur date effective de production. Ce mécanisme engendre une distorsion de la dimension temporelle de la série, rendant les dates de commande partiellement dissociées de la demande réelle. Une telle configuration complexifie l'analyse directe des données, rendant indispensable une phase de prétraitement.

Un nettoyage des données a ainsi été effectué afin de conserver uniquement les ordres de fabrication datés entre 2024 et 2025, tout en excluant les affectations rétroactives. Par ailleurs, la structure des commandes nécessitait une reconstitution explicite : chaque ligne indiquait la répartition en pourcentage entre différents types de composants (par exemple, 90 % de coton biologique et 10 % de coton conventionnel),

3.2.1.2 Analyse descriptive de la série journalière

La série journalière obtenue après traitement révèle une volatilité marquée, caractérisée par une alternance entre des journées sans production (valeurs nulles) et des pics de volume exceptionnellement élevés. Cette dynamique irrégulière reflète les particularités du processus industriel, notamment la production anticipée et la variabilité de la demande.

Les statistiques descriptives synthétisent cette variabilité de manière quantitative :

Tableau 3-1: Statistiques descriptives de la série journalière (production de coton, 2024)

Statistique	Date (commande)	Quantité produite (unité)
Nombre d'observations	251	251
Moyenne	6 juillet 2024	52 958
Écart-type	—	54 942
Minimum	3 janvier 2024	69
1er quartile (Q1)	1er avril 2024	14 261
Médiane (Q2)	10 juillet 2024	35 225
3e quartile (Q3)	8 octobre 2024	77 300
Maximum	31 décembre 2024	265 087

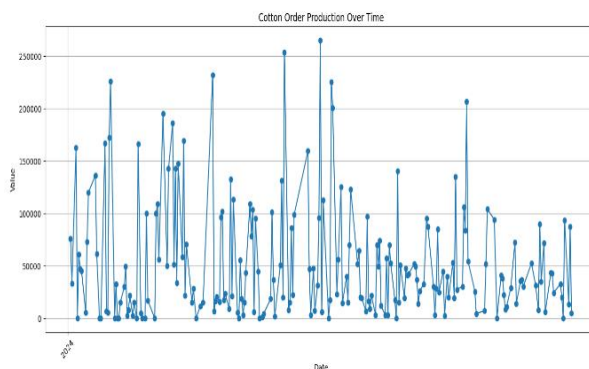


Figure 3-1: Série journalière brute de production de coton

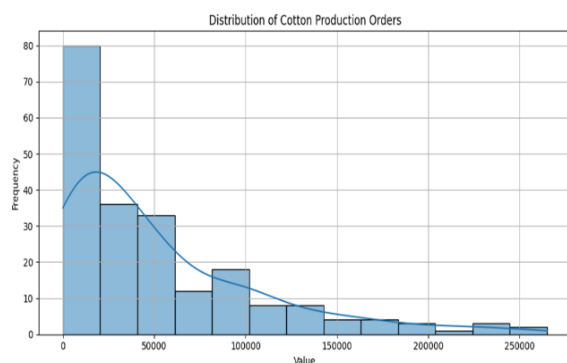


Figure 3-2: Histogramme de distribution des volumes journaliers (2024)

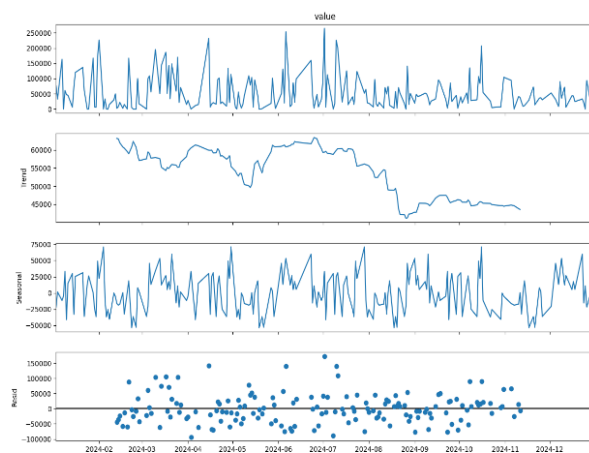


Figure 3-3: Décomposition additive de la série journalière de production (tendance, saisonnalité, résidus)

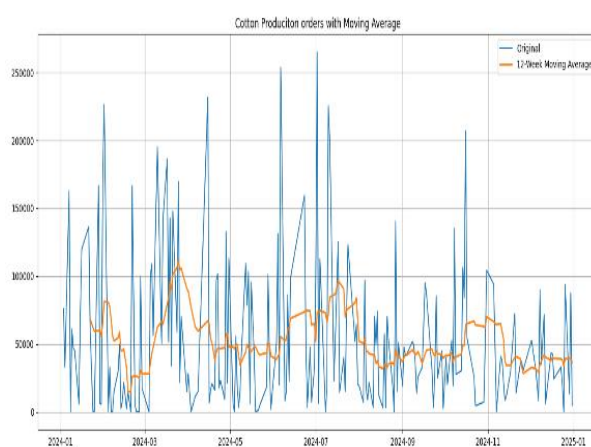


Figure 3-4: Évolution de la production journalière avec moyenne mobile (fenêtre de 12 semaines)

Les données brutes de production journalière de coton sur l'année 2024 (Figure 3-1) exposent une volatilité extrême caractérisée par l'occurrence systématique de journées sans rendement et de pics ponctuels dépassant 250 000 unités, manifestation d'une désynchronisation structurelle entre capacités industrielles et demande effective. L'histogramme des volumes journaliers (Figure 3-2) valide cette instabilité par une asymétrie prononcée vers la droite : la majorité des observations se concentre sur de faibles quantités tandis qu'une minorité génère une longue queue de valeurs extrêmes, signe d'une distribution non gaussienne. La décomposition additive de la série (Figure 3-3) révèle une tendance globale erratique sans progression linéaire identifiable, une composante saisonnière fragmentée et peu structurée, ainsi que des résidus à forte dispersion témoignant d'une hétéroscédasticité prononcée et d'irrégularités de régime. L'application d'une moyenne mobile sur une fenêtre de douze semaines (Figure 3-4) souligne des oscillations persistantes autour de la tendance lissée et un niveau de bruit élevé, démontrant l'inadéquation des modèles linéaires classiques pour représenter la dynamique réelle de production. Ces analyses graphiques justifient le recours à des approches hybrides combinant modélisation statistique linéaire et algorithmes non linéaires pour capturer pleinement la complexité et l'impulsivité du processus.

3.2.1.3 Exploration complémentaire en fréquence hebdomadaire

Bien que l'analyse principale repose sur des données journalières, une agrégation hebdomadaire a été réalisée à des fins exploratoires. Cette transformation vise à améliorer la lisibilité globale de la dynamique de production et à mettre en évidence d'éventuelles tendances de fond. Le tableau suivant présente un résumé des principaux indicateurs statistiques calculés sur cette série hebdomadaire agrégée :

Tableau 3-2: Résumé statistique de la série hebdomadaire agrégée (production de coton, 2024)

Statistique	Date (début de semaine)	Quantité hebdomadaire (unité)
Nombre d'observations	53	53
Moyenne	1er juillet 2024	181 097
Écart-type	—	106 507
Minimum	1er janvier 2024	5 697
1er quartile (Q1)	1er avril 2024	107 666
Médiane (Q2)	1er juillet 2024	167 878
3e quartile (Q3)	30 septembre 2024	227 842
Maximum	30 décembre 2024	456 386

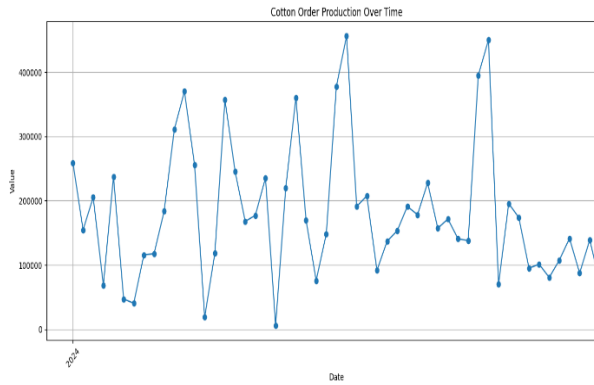


Figure 3-5: Série journalière brute de production de coton

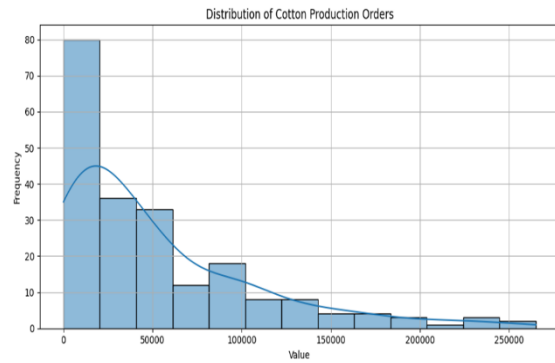


Figure 3-6: Histogramme de distribution des volumes journaliers (2024)

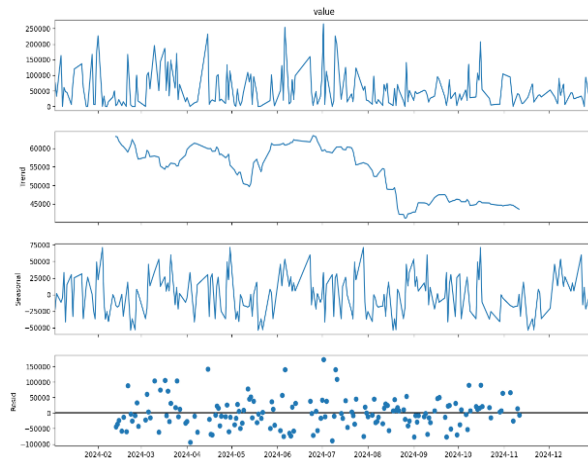


Figure 3-7: Décomposition additive de la série journalière de production (tendance, saisonnalité, résidus)

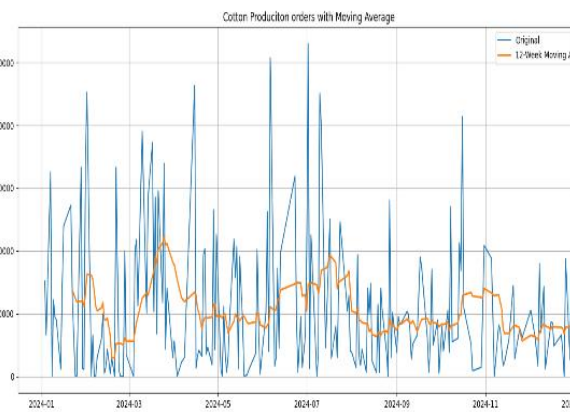


Figure 3-8: Évolution de la production journalière avec moyenne mobile (fenêtre de 12 semaines)

Le résumé statistique hebdomadaire (Tableau 3-2) met en évidence une moyenne de production élevée assortie d'un écart-type considérable et de quartiles asymétriques, signe d'une dispersion marquée et de valeurs extrêmes récurrentes. L'évolution agrégée sur 2024 (Figure 3-5) révèle des amplitudes hebdomadaires oscillant entre quelques milliers et plus de 450 000 unités, traduisant une volatilité persistante malgré le lissage implicite de l'agrégation. Lissage explicite par moyenne mobile sur douze semaines (Figure 3-6) atténue les fluctuations les plus abruptes tout en soulignant une tendance générale à la baisse en fin d'exercice, sans dégager de schéma périodique stable. La distribution hebdomadaire (Figure 3-7) affiche une asymétrie modérée et une dispersion moins prononcée qu'à l'échelle journalière, attestant de l'effet régularisateur de l'agrégation temporelle. La décomposition saisonnière n'a pas été réalisée en raison du nombre d'observations limité à 53 semaines, en deçà des 104 requises pour garantir la fiabilité méthodologique.

3.2.1.4 Conclusion analytique

L'analyse statistique et visuelle des séries temporelles met en évidence plusieurs propriétés structurelles limitant l'usage de modèles classiques :

- Une hétéroscédasticité marquée, avec des amplitudes très variables selon les périodes ;
- L'absence de saisonnalité régulière et exploitable ;
- Une non-stationnarité évidente dans les tendances ;
- Une distribution asymétrique et éloignée de la normalité.

Ces limitations rendent les modèles de prévision univariés traditionnels, tels que Holt-Winters ou SARIMA, insuffisants lorsqu'ils sont utilisés de manière isolée. Face à cette complexité, l'adoption d'une approche hybride devient nécessaire. Le modèle hybride SARIMA/Prophet – XGBoost se distingue par sa capacité à combiner la modélisation linéaire des composantes régulières (tendance et saisonnalité) avec un apprentissage automatique non linéaire appliqué aux résidus. Cette architecture permet ainsi de capturer plus efficacement les comportements irréguliers et les fluctuations imprévues des séries temporelles.

3.2.2 Données Hebdomadaires d'Exportation de Coton (USDA)

Le second cas d'étude repose sur les données hebdomadaires publiées par l'USDA, intitulées Export Sales Weekly Historical Data, couvrant initialement la période 1990–2025. Afin de garantir l'homogénéité et la pertinence des analyses, la période a été restreinte du 1er janvier 2021 au janvier 2024. Ce choix permet d'exclure les données trop anciennes, potentiellement affectées par des dynamiques de marché, des politiques commerciales et des fluctuations économiques non représentatives des tendances récentes. Cette sélection temporelle vise à concentrer l'analyse sur un cadre économique plus actuel et reflétant les conditions modernes de l'offre et de la demande sur le marché du coton.

Un prétraitement des données a été réalisé selon les étapes suivantes :

- Exclusion systématique des observations antérieures à 2021 ;
- Imputation des valeurs manquantes par des zéros dans les cas identifiés comme des absences d'activité réelles ;
- Conversion des horodatages en fréquence journalière afin de permettre une analyse temporelle plus granulaire et homogène.

Tableau 3-3: Statistiques descriptives de la série journalière USDA (janvier 2019 – janvier 2025)

Statistique	Date (début de semaine)	Quantité hebdomadaire (unité)
Nombre d'observations	208	/
Moyenne	1er janvier 2023	225.179739
Écart-type	—	106 507
Minimum	1er janvier 2024	5 697
1er quartile (Q1)	1er avril 2024	107 666
Médiane (Q2)	1er juillet 2024	167 878
3e quartile (Q3)	30 septembre 2024	227 842
Maximum	30 décembre 2024	456 386

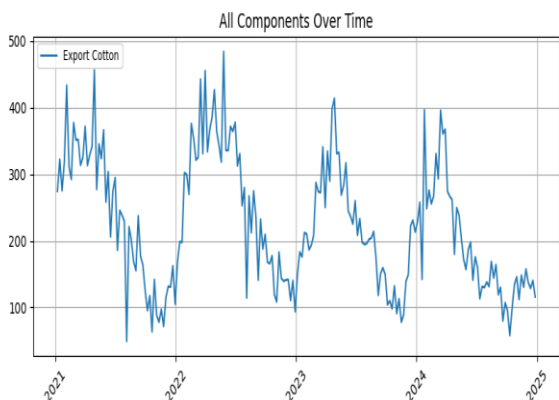


Figure 3-9: Évolution quotidienne des commandes de coton – données USDA

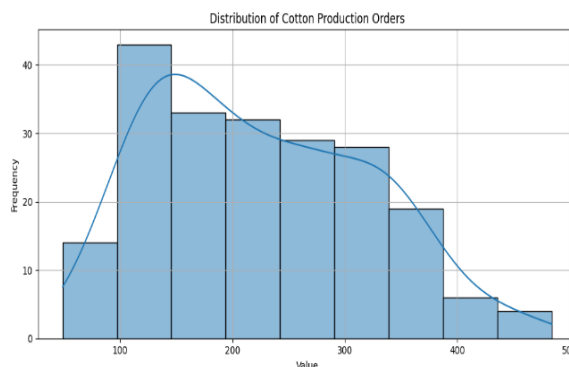


Figure 3-10 : Distribution des volumes journaliers – données USDA (2021–2025)

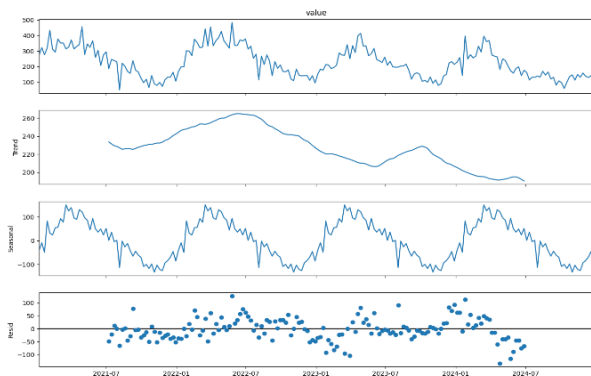


Figure 3-11: Décomposition additive de la série journalière USDA (2019–2025)

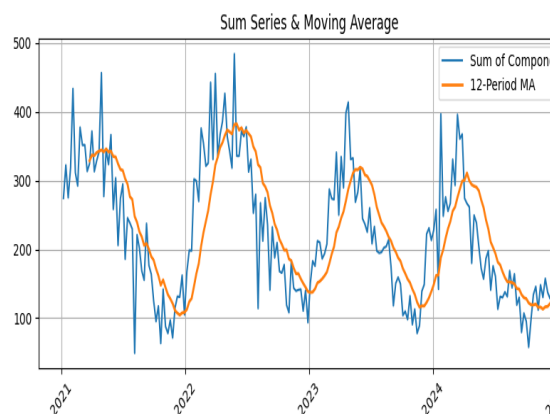


Figure 3-12: Lissage de la série USDA par moyenne mobile (fenêtre de 12 semaines)

La série quotidienne des commandes de coton issue des données USDA (Figure 3-8) révèle une structure nettement plus régulière que celle observée pour la production brute : on y identifie sans difficulté une saisonnalité annuelle récurrente, calquée sur un cycle calendaire stable dont l’amplitude reste relativement constante d’une année sur l’autre, signe d’un processus de demande bien synchronisé avec les rythmes de marché. Toutefois, la présence de quelques pics ponctuels, dont l’origine pourrait être liée à des événements exceptionnels (chocs climatiques, ajustements réglementaires ou ruptures de la chaîne logistique), impose de traiter ces valeurs atypiques avec prudence, notamment lors de l’estimation des paramètres saisonniers.

La distribution des volumes journaliers sur la période 2019–2025 (Figure 3-9) confirme cette régularité : la tendance centrale est modérée et les composantes saisonnières se distinguent clairement dans la décomposition, tandis que les résidus demeurent faibles et uniformément répartis autour de zéro, ce qui atteste d’un excellent ajustement du modèle linéaire de base et valide l’hypothèse d’une série proche de la stationnarité après différenciation saisonnière.

Enfin, le lissage par moyenne mobile sur une fenêtre de douze semaines (Figure 3-11) met en exergue la récurrence des vagues saisonnières et fait apparaître une reprise post-pandémique graduelle, caractérisée par une pente positive délicate mais perceptible au-dessus du niveau précédent la crise de 2020. Cette conjonction d’une saisonnalité robuste, d’une tendance relativement stable et de résidus maîtrisés confirme la pertinence d’une modélisation classique

(SARIMA ou Prophet) pour ces séries, sans nécessité immédiate de recourir à des approches hybrides plus complexes, sauf pour le traitement ponctuel des anomalies.

3.2.3 Approches hybrides de prévision

Les approches hybrides de prévision intègrent des méthodes statistiques et d'apprentissage automatique dans une logique complémentaire visant à renforcer la performance prédictive.

3.2.3.1 Justification de l'hybridation des modèles

Dans le cadre de notre étude, nous avons été confrontés à des séries temporelles industrielles présentant une double complexité : des structures linéaires bien définies (tendances, saisonnalités) et des composantes non linéaires plus irrégulières (chocs exogènes, fluctuations imprévues). Face à cette hétérogénéité, il nous est apparu pertinent d'adopter une approche hybride, combinant les avantages des modèles statistiques et ceux de l'apprentissage automatique.

Nous avons tout d'abord retenu le modèle SARIMA pour sa capacité éprouvée à modéliser les composantes linéaires, notamment les effets saisonniers récurrents et les structures de tendance. Toutefois, comme le soulignent (Thejovathi & al, 2024), ces modèles ne sont pas conçus pour capturer les dépendances non linéaires ou les interactions complexes entre variables.

C'est pourquoi nous avons intégré, en complément, l'algorithme XGBoost, reconnu pour son efficacité à modéliser les comportements non linéaires dans des environnements bruités et multidimensionnels. Ce choix s'inscrit dans une logique de spécialisation fonctionnelle : le modèle SARIMA absorbe la partie régulière du signal, et les résidus qu'il génère sont ensuite exploités par XGBoost pour modéliser la part non expliquée. Ce type de combinaison a déjà été exploré avec succès par (Khashei, 2011), qui ont démontré qu'une telle architecture améliore sensiblement les performances prédictives.

Dans notre cas, cette stratégie s'est révélée d'autant plus adaptée que les données issues de TAYAL SPA présentent à la fois des schémas de production saisonniers et des perturbations dues à des facteurs externes. En suivant l'exemple d'Otieno (2024), nous avons mis en œuvre ce modèle hybride pour capter ces deux types de dynamiques. Leurs résultats dans le domaine des transactions financières, tout comme ceux de (Thejovathi & al, 2024) dans la prévision de la demande produit, ont confirmé la validité de cette démarche dans des contextes industriels instables.

Ainsi, l'hybridation SARIMA–XGBoost que nous avons mise en place repose sur une complémentarité méthodologique claire. Elle nous a permis d'exploiter pleinement les atouts respectifs de chaque modèle, en combinant rigueur structurelle et souplesse algorithmique pour améliorer la précision des prévisions dans un environnement réel.

3.2.3.2 Modèles hybrides SARIMA–XGBoost–Optuna et Prophet–XGBoost–Optuna

Les modèles hybrides *SARIMA–XGBoost–Optuna* et *Prophet–XGBoost–Optuna* reposent sur une approche séquentielle en plusieurs étapes visant à optimiser la précision des prévisions dans les séries temporelles (Thejovathi & al, 2024). Cette méthodologie combine un modèle statistique linéaire, tel que SARIMA ou Prophet, avec un modèle d'apprentissage automatique

non linéaire, en l'occurrence XGBoost. Cette combinaison permet de tirer parti de la capacité des modèles statistiques à capturer les tendances saisonnières et les dynamiques linéaires, tout en exploitant la puissance prédictive de XGBoost pour modéliser les résidus non linéaires. L'optimisation des hyperparamètres par Optuna renforce cette architecture en améliorant les performances de chaque composant, assurant ainsi une meilleure précision des prévisions et une résilience accrue face aux perturbations exogènes (Akiba & al, 2019).

3.2.3.3 Modélisation Linéaire avec SARIMA ou Prophet

Dans un premier temps, un modèle linéaire traditionnel, soit SARIMA, soit Prophet est ajusté sur la série temporelle afin de capturer les composantes linéaires telles que les tendances globales, la saisonnalité, ainsi que les effets calendaires. SARIMA est particulièrement efficace pour modéliser les séries présentant des cycles saisonniers bien définis (Thejovathi & al, 2024), tandis que Prophet, développé par Facebook, excelle dans la capture de tendances complexes et de saisonnalités modulables grâce à sa nature additive (Taylor & Letham, 2017).

3.2.4 Régression Non-Linéaire des Résidus avec XGBoost

Après l'ajustement initial par SARIMA ou Prophet, les résidus – représentant les écarts entre les valeurs observées et les prédictions du modèle linéaire – sont extraits. Ces résidus constituent l'information résiduelle non expliquée par les composantes linéaires. Le modèle XGBoost (Extreme Gradient Boosting) est ensuite entraîné sur ces résidus afin de capturer les dynamiques non linéaires et les interactions complexes qui n'ont pas été modélisées par la première phase (Preeti, 2024). Cette étape, qualifiée de régression d'erreur ou erreur re-modélisation, permet de raffiner les prédictions en corrigeant les erreurs structurelles laissées par le modèle linéaire.

3.3 Architecture Modulaire du Pipeline Hybride

L'architecture modulaire du pipeline vise à optimiser les prévisions temporelles en combinant les modèles statistiques linéaires (SARIMA, Prophet) avec le modèle d'apprentissage non linéaire (XGBoost). Cette approche séquentielle permet de capturer les dynamiques linéaires, les tendances saisonnières et les non-linéarités complexes, améliorant ainsi la précision globale des prédictions (Hyndman & Athanasopoulos, 2021).

3.3.1 Prétraitement des données temporelles

Cette première étape consiste à nettoyer et transformer les données afin de garantir leur qualité et leur cohérence pour les étapes de modélisation. Elle inclut l'imputation des valeurs manquantes, la suppression des anomalies, ainsi que la différenciation et la désaisonnalisation si nécessaire pour stationnariser le signal (Brockwell & Davis, 2002). Par ailleurs, des variables temporelles telles que le jour de la semaine, les indicateurs saisonniers et les périodes de vacances sont générées pour enrichir les prédictions futures, notamment dans le cadre de Prophet ou en tant que variables exogènes.

3.3.2 Modélisation statistique linéaire

Une fois les données prétraitées, l'étape de modélisation linéaire est réalisée à l'aide de modèles statistiques tels que SARIMA ou Prophet. SARIMA ajuste les composantes tendances et saisonnières à travers une paramétrisation optimisée (p, d, q, P, D, Q, s) (Box & Jenkins,

1976), tandis que Prophet détecte automatiquement les changements de tendances (change points) et ajuste les cycles saisonniers. Cette première modélisation permet de capturer les variations régulières et les effets de saisonnalité présents dans les séries temporelles.

3.3.3 Extraction et préparation des résidus

Une fois le modèle statistique ajusté, les résidus sont calculés en soustrayant les valeurs prédites des valeurs observées. Ces résidus représentent l'information non capturée par le modèle linéaire. Afin de maximiser l'apprentissage du modèle non linéaire, ces résidus sont enrichis par l'ajout de variables explicatives supplémentaires, telles que des décalages temporels (lags) et des indicateurs exogènes (Hyndman & Athanasopoulos, 2021).

3.3.4 Vectorisation et enrichissement des données résiduelles

Les données résiduelles enrichies sont ensuite vectorisées pour être compatibles avec le modèle d'apprentissage automatique XGBoost. Cette étape inclut la création de colonnes retardées (lags), l'ajout de transformations polynomiales et la troncature éventuelle des séries. Cette représentation sous forme de vecteurs permet de capturer l'information temporelle récente de manière optimisée (Chen & Guestrin, 2016).

3.3.5 Modélisation non linéaire avec XGBoost

Une fois les résidus correctement préparés, un modèle XGBoost est entraîné pour apprendre les erreurs résiduelles restantes du modèle statistique. Cette phase permet de capturer les relations non linéaires complexes qui n'ont pas été modélisées par SARIMA ou Prophet. XGBoost ajuste ses paramètres pour minimiser l'erreur prédictive et affine ainsi les prévisions initiales (Chen & Guestrin, 2016).

3.3.6 Génération de la prévision hybride

La prédiction finale est obtenue en combinant les sorties des deux modèles. Le modèle statistique (SARIMA ou Prophet) fournit la composante linéaire et saisonnière, tandis que XGBoost corrige cette estimation en ajoutant la composante non linéaire. Cette fusion garantit une prévision plus précise et robuste, capable de s'adapter aux variations complexes et imprévues des séries temporelles (Zhang, Li, & Liu, 2023).

3.4 Optimisation des Paramètres et Apprentissage Adaptatif

Cette section aborde de manière approfondie deux aspects fondamentaux : l'optimisation systématique des hyperparamètres et l'intégration de l'apprentissage adaptatif.

3.4.1 Optimisation Systématique des Hyperparamètres

L'optimisation des hyperparamètres constitue une étape essentielle pour maximiser la performance des modèles statistiques et d'apprentissage automatique. Pour le modèle SARIMA, cette optimisation est réalisée à l'aide d'Auto-ARIMA et d'Optuna. Auto-ARIMA procède par une recherche exhaustive des combinaisons de paramètres (p, d, q, P, D, Q, s) , tandis qu'Optuna utilise une méthode bayésienne pour explorer efficacement l'espace des hyperparamètres et minimiser l'erreur de prévision (Akiba & al, 2019). Quant à XGBoost, l'optimisation est exclusivement effectuée par Optuna, qui ajuste de manière automatisée les paramètres critiques du modèle, tels que le taux d'apprentissage et la profondeur des arbres, afin de maximiser la

précision tout en réduisant le coût computationnel. Cette double approche garantit une meilleure adaptation du modèle aux variations temporelles (Chen & Guestrin, 2016).

3.4.2 Apprentissage Adaptatif et Réinjection des Prédictions

L'apprentissage adaptatif introduit une innovation majeure dans le cadre de ce projet, à travers la réinjection des prédictions passées sous forme de lags dans le processus d'apprentissage de XGBoost. Ce mécanisme permet au modèle de s'adapter de manière dynamique grâce à un apprentissage en ligne (Online Learning), ajustant continuellement ses prédictions en fonction des écarts observés lors des itérations précédentes (Brownlee, 2018). Cette stratégie renforce la résilience du modèle face aux changements non stationnaires, facilitant ainsi la détection des ruptures de tendance et l'adaptation aux effets exogènes. En optimisant l'intégration des informations temporelles récentes, le modèle améliore significativement la précision des prévisions dans des environnements volatils (Nielsen, 2019).

3.5 Intégration de l'assistant IA pour l'analyse exploratoire

Nous avons développé un assistant IA se déploie au sein de l'interface en tant que module de dialogue textuel dédié à l'analyse exploratoire des données de production. Il prend en charge l'interprétation des indicateurs statistiques issus des modèles SARIMA, Prophet et XGBoost. L'utilisateur formule sa demande via un champ de saisie sans contrainte de syntaxe.

À réception de la requête, le système extrait les métriques pertinentes (WAPE, MAE, MASE, RMSE) et accède aux séries temporelles réelles et aux résidus préalablement calculés. L'assistant génère un commentaire structuré détaillant la valeur numérique, son interprétation statistique et sa portée opérationnelle. La comparaison entre modèles se présente sous la forme d'un paragraphe consacrant les écarts de performance et les variations de résidus.

Ce module uniformise les analyses entre utilisateurs et assure la traçabilité des décisions par l'historisation des échanges. L'automatisation de l'interprétation textuelle réduit le temps alloué à la phase d'exploration et élimine le risque d'erreur manuelle dans l'évaluation des prévisions.

3.6 Indicateurs de performance

Dans le cadre de la modélisation hybride SARIMA–XGBoost et Prophet–XGBoost, il est essentiel de disposer de métriques d'évaluation robustes et adaptées pour évaluer la performance des prévisions. Les métriques choisies doivent permettre de quantifier l'écart entre les valeurs prédites et les valeurs observées tout en tenant compte des caractéristiques spécifiques des séries temporelles. Dans cette section, nous présentons les métriques suivantes : WAPE, MAE, MASE et RMSE, en expliquant leur pertinence dans le contexte de la modélisation hybride.

3.6.1 WAPE (Weighted Absolute Percentage Error)

Le WAPE est une métrique qui exprime l'erreur absolue totale en pourcentage du total des observations réelles. Cette mesure est formellement définie par l'équation suivante :

$$\text{WAPE} = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{\sum_{t=1}^n |y_t|} \quad 3-1$$

Où:

- y_t : Valeur observée à l'instant.
- \hat{y}_t : Valeur prédite à l'instant.
- n : Nombre total d'observations.

Dans cette expression, y_t représente la valeur observée tandis que \hat{y}_t correspond à la prévision réalisée au temps t (Hewamalage, 2022). Cette formule a pour particularité de pondérer chaque erreur absolue en fonction de la taille de la série, permettant ainsi d'obtenir un indicateur global de précision exprimé en pourcentage.

Lorsqu'il s'agit de modèles hybrides, tels que SARIMA–XGBoost ou Prophet–XGBoost, le WAPE prend tout son sens : il mesure l'erreur moyenne relative des prévisions, tout en tenant compte de l'importance des valeurs passées. En clair, une valeur de WAPE faible indique que l'écart absolu cumulé entre les prévisions du modèle et les observations réelles reste minime par rapport à la somme totale des valeurs observées.

Cela signifie, en d'autres termes, que le WAPE fournit un aperçu global de la performance du modèle, tout en prenant en considération l'échelle de la série temporelle. Cette approche est particulièrement utile pour comparer différents modèles sur des séries de grandeurs distinctes, en offrant une mesure relative de l'erreur moyenne qui tient compte des variations historiques.

3.6.2 MAE (Mean Absolute Error)

Le MAE est l'erreur absolue moyenne entre les valeurs observées et prédites. Il est calculé selon l'équation suivante :

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad 3-2$$

Où :

- y_t : Valeur observée à l'instant.
- \hat{y}_t : Valeur prédite à l'instant.
- n : Nombre total d'observations.

Cette formule correspond à une moyenne arithmétique des écarts absolus entre les prévisions réalisées et les valeurs réellement observées. En d'autres termes, la mesure l'écart moyen entre les prédictions d'un modèle hybride et les données réelles, exprimé dans les mêmes unités que la variable analysée. (Hewamalage, 2022)

Lorsqu'on parle d'un modèle hybride combinant SARIMA ou Prophet avec XGBoost, une MAE faible indique clairement que, en moyenne, les prédictions s'écartent très peu des valeurs observées, témoignant ainsi de la précision du modèle.

Contrairement à d'autres indicateurs plus complexes, la MAE ne réalise aucune pondération spécifique : elle offre une vision brute et directe de l'erreur commise. Cette simplicité la rend particulièrement intuitive à interpréter, car elle reflète de manière transparente l'erreur typique associée aux prévisions.

3.6.3 MASE (Mean Absolute Scaled Error)

La MASE est utilisée pour comparer l'erreur d'un modèle avec celle d'un modèle naïf. Elle est définie comme suit :

$$\text{MASE} = \frac{\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|}{\frac{1}{T-1} \sum_{t=2}^T |Y_t - Y_{t-1}|} \quad 3-3$$

Où :

- y_t : Valeur observée à l'instant.
- \hat{y}_t : Valeur prédite à l'instant.
- n : Nombre total d'observations.
- T : Nombre total d'intervalles
- y_{t-1} : Valeur réelle à l'instant précédent

La dénomination correspond à l'erreur absolue moyenne d'une prévision naïve, où l'on utilise Y_{t-1} pour prédire Y_t . De manière plus explicite, l'erreur est normalisée par rapport à cette prévision naïve, selon la formule suivante :

$$q_t = \frac{|y_t - \hat{y}_t|}{\frac{1}{T-1} \sum_{s=2}^T |Y_s - Y_{s-1}|} \quad 3-4$$

La MASE est définie comme :

$$\text{MASE} = \frac{1}{n} \sum_{t=1}^n q_t \quad 3-5$$

La MASE, étant sans dimension, offre l'avantage de pouvoir comparer les performances de différents modèles, indépendamment de l'échelle des séries temporelles. Concrètement, elle indique si le modèle hybride (par exemple, SARIMA–XGBoost ou Prophet–XGBoost) surpasse un modèle de référence naïf qui se contente de prolonger la dernière valeur observée.

Ainsi, une MASE inférieure à 1 signifie que le modèle hybride produit, en moyenne, moins d'erreur qu'une simple prévision naïve, traduisant un réel gain de qualité dans les prédictions. Cette métrique permet également de mettre en lumière l'apport significatif de la composante XGBoost par rapport à un comportement purement linéaire.

3.6.4 RMSE (Root Mean Squared Error)

La RMSE est la racine carrée de l'erreur quadratique moyenne. Elle se définit formellement comme :

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad 3-6$$

Où :

- y_t : Valeur observée à l'instant.
- \hat{y}_t : Valeur prédite à l'instant.
- n : Nombre total d'observations.

La RMSE a pour particularité de ramener l'écart quadratique moyen à l'échelle des données d'origine, ce qui facilite l'interprétation directe des erreurs de prévision. En d'autres termes, elle correspond à l'écart-type moyen des erreurs, offrant une estimation claire de la dispersion autour des valeurs observées (Hewamalage, 2022).

Cette métrique est particulièrement sensible aux erreurs de grande ampleur, car elle prend en compte le carré des écarts : un écart important est donc davantage pénalisé. Cela en fait un indicateur utile pour détecter les anomalies ou les erreurs significatives dans les prédictions.

Dans le cadre d'un modèle hybride, tel que SARIMA–XGBoost ou Prophet–XGBoost, une RMSE faible témoigne de la capacité du modèle à suivre fidèlement la tendance linéaire tout en capturant les variations non linéaires. En résumé, la RMSE offre une mesure précise de l'écart type des prévisions, avec une attention particulière portée aux écarts importants, permettant ainsi de mieux évaluer la qualité globale du modèle.

3.7 Résultats expérimentaux

Nous évaluons ici trois variantes du pipeline hybride — SARIMA–XGBoost, auto_ARIMA–XGBoost et Prophet–XGBoost — sur la série hebdomadaire de Tayal Spa, en considérant un horizon de prévision d'un an (52 semaines). Chaque modèle produit des prévisions et leurs résidus, qui servent ensuite d'entrées au correcteur XGBoost.

3.7.1 Stratégie de prévision bi-hebdomadaire

La génération bi-hebdomadaire de prévisions répond à la nécessité de détecter rapidement les inflexions de tendance et les anomalies inopinées dans la demande. En limitant l'intervalle de réactualisation à deux semaines, le pipeline intègre de manière systématique les écarts observés entre prévisions et réalisations, ce qui renforce la capacité du système à s'adapter aux ruptures de régime et aux chocs externes. La génération bi-hebdomadaire de prévisions permet de réinjecter fréquemment le retour d'erreur dans le pipeline, améliorant ainsi la réactivité face aux changements soudains de la demande.

À chaque cycle :

Réentraînement du modèle linéaire (SARIMA optimisé par auto_ARIMA ou SARIMA optimisé par Optuna ou Prophet) capturant multiples saisonnalités et points de changement sur l'ensemble des données historiques, incluant les prévisions des cycles précédents.

Production du point prediction et extraction des résidus in-sample.

Enrichissement du jeu d'entraînement du régresseur XGBoost (hyperparamètres optimisés via Optuna) par la concaténation des prédictions linéaires et des résidus correspondants.

Génération de la prévision corrigée pour les quatorze jours suivants.

Ce protocole bi-hebdomadaire, couplé à une architecture de stacking, assure un retour d'information continu entre composantes linéaire et non linéaire. Il limite la dérive des modèles et maintient la précision face à la volatilité des données industrielles.

3.8 Cas d'étude 1 : demande hebdomadaire de Tayal Spa

La **Tableau 3-4** présente les résultats comparatifs de l'évaluation des performances des modèles hybrides de prévision appliqués à la demande hebdomadaire de **Tayal Spa**, en utilisant les indicateurs WAPE, Max WAPE, MAE et RMSE.

Tableau 3-4 : Évaluation des modèles hybrides de prévision (Tayal Spa)

Modèle	WAPE (%)	Max WAPE (%)	MAE	RMSE
SARIMA-OP-XGBoost-OP	41.45	37.79	67808.79	83565.50
auto_ARIMA-XGBoost-OP	37.94	57.81	62067.77	96220.35
Prophet-XGBoost-OP	55.54	46.06	102422.33	118069.37

3.8.1 Visualisations SARIMA-OP-XGBoost-OP

Les visualisations suivantes illustrent les performances et le comportement prédictif du modèle hybride SARIMA-OP-XGBoost-OP appliqué aux séries temporelles industrielles (TAYAL).

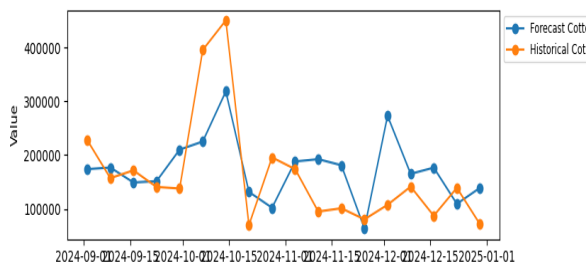


Figure 3-13 : Comparaison des valeurs prévues et historiques du modèle SARIMA-OP-XGBoost-OP

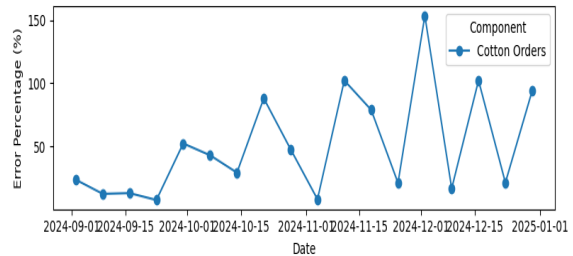


Figure 3-14 : Évolution de l'erreur de prévision (%) du modèle SARIMA-OP-XGBoost-OP

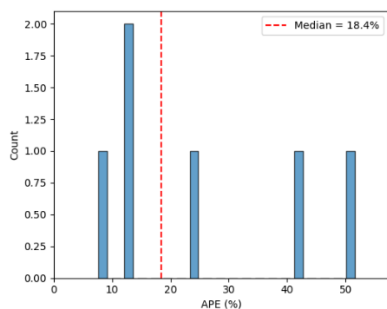


Figure 3-15: Distribution des erreurs absolues (%) du modèle SARIMA-OP-XGBoost-OP

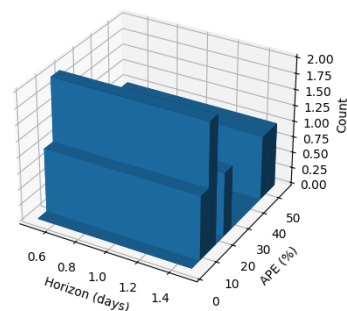


Figure 3-16 : Analyse 3d de l'erreur absolue en fonction de l'horizon temporel et du pourcentage d'erreur (APE)

3.8.1.1 Discussion synthétique des performances SARIMA–XGBoost-OP

Le modèle hybride SARIMA–OP–XGBoost-OP, entraîné sur les données hebdomadaires de janvier à août 2024 (≈ 32 observations), présente un comportement intéressant compte tenu de la taille réduite de l'échantillon :

Adaptabilité sur petit jeu de données : avec seulement huit mois d'historique, le modèle parvient à générer des prévisions globalement cohérentes (Figure 3-13), ce qui souligne sa robustesse face à la rareté des données.

Précision relative : un WAPE moyen de 41,45 % et un WAPE maximal de 37,79 % traduisent une erreur de prédiction modérée sur l'ensemble de la période de test. Malgré ces valeurs, la distribution des APE (Figure 3-15) reste concentrée autour de l'écart médian de 18,4 %, témoignant d'une performance acceptable sur la majorité des semaines.

Comportement en horizon : l'analyse 3D (Figure 3-16) montre une légère dégradation de l'erreur à mesure que l'on s'éloigne de la date d'entraînement « la plus récente », ce qui est classique pour un modèle à mémoire limitée ; toutefois, l'amplitude des erreurs reste maîtrisée dans un intervalle de 10–50 % d'APE.

Capacité d'apprentissage incrémental : la partie XGBoost du pipeline peut enrichir son apprentissage à chaque nouvelle observation. Or, avec un historique restreint (53 semaines, dont seules 32 ont servi à l'entraînement), le gain potentiel de cet apprentissage en ligne est forcément limité : plus de données passées renforcerait la qualité et la stabilité des prédictions futures.

3.8.1.2 Conclusion

Le couple SARIMA–OP–XGBoost-OP se révèle particulièrement adapté aux contextes où l'on dispose d'un nombre restreint d'observations : malgré un entraînement sur seulement huit mois, le modèle produit des prévisions cohérentes et conserve une capacité d'ajustement progressive grâce à la composante XGBoost. Néanmoins, l'ajout d'historiques complémentaires (≥ 2 cycles saisonniers) améliorera significativement la détection de la saisonnalité et la précision globale. En l'état, cette approche hybride constitue une base solide pour une démarche de prévision continue ; sa performance gagnera à être réévaluée périodiquement au fur et à mesure de l'arrivée de nouvelles données.

3.8.2 Visualisations auto_ARIMA–XGBoost-OP

Les visualisations ci-dessous présentent les résultats et l'évaluation des performances du modèle hybride auto_ARIMA–XGBoost-OP appliqué aux données industrielles de TAYAL.

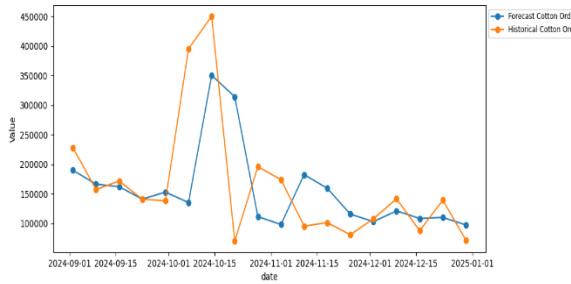


Figure 3-17 : Comparaison des valeurs prévues et historiques du modèle auto_ARIMA-XGBoost-OP

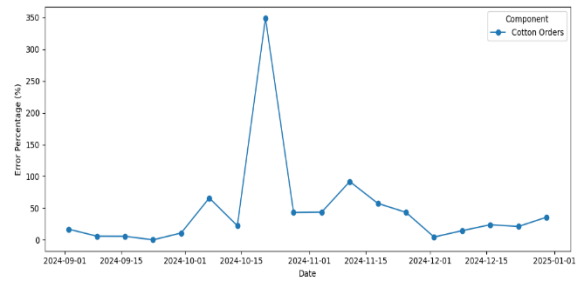


Figure 3-18 : Évolution de l'erreur de prévision (%) du modèle auto_ARIMA-XGBoost-OP

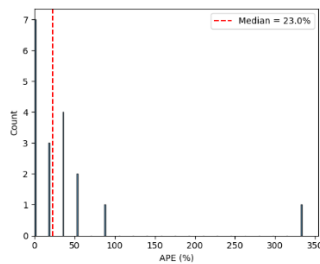


Figure 3-19: Distribution des erreurs absolues (%) du modèle auto_ARIMA-XGBoost-OP

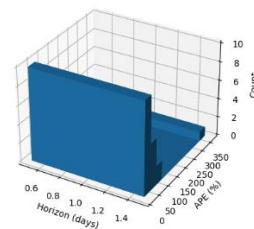


Figure 3-20 : Analyse 3d de l'erreur absolue en fonction de l'horizon temporel et du pourcentage d'erreur (APE)

3.8.2.1 Discussion synthétique des performances auto_ARIMA-XGBoost-OP

Le modèle auto ARIMA-XGBoost-OP, entraîné sur les mêmes 32 points hebdomadaires de janvier à août 2024, présente les caractéristiques suivantes :

- Erreur moyenne réduite : un WAPE de 37,94 % (vs. 41,45 % pour SARIMA-XGBoost-OP) signale une meilleure adéquation globale aux volumes observés.
- Pic d'erreur exceptionnel : un WAPE maximal de 57,81 % révèle quelques semaines où la prévision s'éloigne fortement des valeurs réelles (probablement liées à des outliers ou à des ruptures de tendance ponctuelles).
- Précision absolue : avec un MAE de 62 068 unités et un RMSE de 96 220, le modèle réduit l'écart quadratique moyen, mais reste sensible aux grandes amplitudes de production en période de pic.
- Avantage de l'auto-sélection : l'algorithme auto ARIMA a automatiquement identifié les ordres (p, d, q) et la saisonnalité optimale sans réglages manuels, ce qui s'avère précieux sur un jeu de données restreint et hétérogène.
- Apprentissage incrémental : comme pour la version SARIMA, la composante XGBoost gagne en performance au fur et à mesure de l'arrivée de nouvelles observations ; cependant, l'historique limité restreint encore son plein potentiel adaptatif.

3.8.2.2 Conclusion

Le pipeline auto ARIMA–XGBoost-OP se distingue par une meilleure erreur moyenne qu’avec SARIMA, grâce à son auto-tuning des paramètres. Il conserve toutefois des pointes d’erreur importantes sur quelques semaines extrêmes. Cette approche hybride est donc particulièrement adaptée aux contextes « small data » et offre une mise en œuvre rapide sans expertise ARIMA approfondie.

Au total, auto ARIMA–XGBoost-OP constitue une base performante et automatisée pour une utilisation opérationnelle, à condition d’alimenter régulièrement le modèle en nouvelles données.

3.8.3 Visualisations Prophet–XGBoost-OP

Les visualisations suivantes illustrent les performances du modèle hybride Prophet–XGBoost-OP sur les séries temporelles issues des données industrielles de TAYAL.

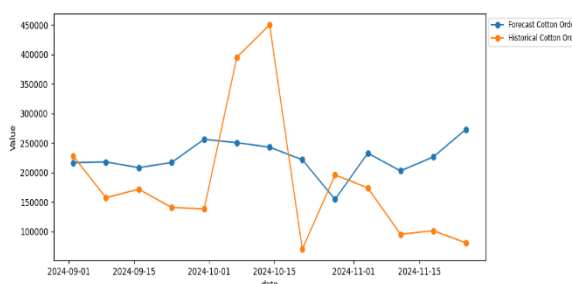


Figure 3-21 : Comparaison des valeurs prévues et historiques du modèle Prophet–XGBoost-OP

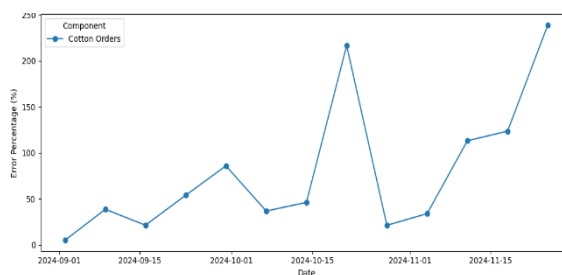


Figure 3-22 : Évolution de l'erreur de prévision (%) du modèle Prophet–XGBoost-OP

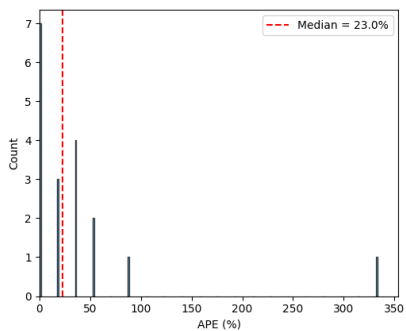


Figure 3-23: Distribution des erreurs absolues (%) du modèle Prophet–XGBoost-OP

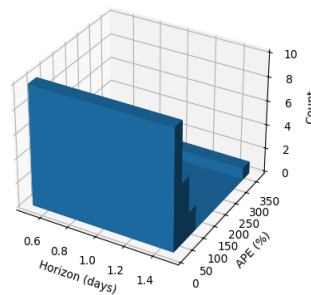


Figure 3-24 : Analyse 3d de l'erreur absolue en fonction de l'horizon temporel et du pourcentage d'erreur (APE)

3.8.3.1 Discussion synthétique des performances Prophet–XGBoost-OP

Le modèle Prophet–XGBoost-OP, entraîné sur les mêmes 32 points hebdomadaires de janvier à août 2024, affiche des performances globalement inférieures aux pipelines SARIMA et auto ARIMA :

- Erreur moyenne élevée : un WAPE de 55,54 % signale que, en moyenne, les prévisions s’écartent de plus de la moitié de la réalité, ce qui reste trop important pour un usage opérationnel.

- Pics d'erreur marqués : même si le WAPE maximal de 46,06 % est légèrement inférieur au WAPE moyen (indiquant que l'erreur n'est pas toujours concentrée sur quelques semaines extrêmes), la distribution des APE (Figure 3-15) montre une forte dispersion, témoignant d'une instabilité de la composante saisonnière modélisée par Prophet.
- Précision absolue dégradée : avec un MAE de 102 422 unités et un RMSE de 118 069, Prophet peine à lisser les pics de production ; ses mécanismes automatiques de détection de points de changement ("change points") ne parviennent pas à capter la forte volatilité hebdomadaire ni les fluctuations irrégulières de l'ordre du coton.
- Limites de Prophet sur petit échantillon : conçu pour des séries longues avec saisonnalités multiples (quotidien, hebdomadaire, annuel) et des données journalières à plusieurs années, Prophet souffre ici d'un historique trop court et d'une granularité hebdomadaire, ce qui influe négativement sur l'estimation des tendances et de la saisonnalité.

3.8.3.2 Conclusion

Le couplage Prophet–XGBoost–OP se révèle moins adapté à notre contexte « small data » hebdomadaire à forte variabilité : ses algorithmes d'optimisation automatique (change points, Fourier-based seasonality) surajustent ou sous-ajustent face aux pics et creux irréguliers.

Au global, Prophet reste une excellente option pour des séries plus longues et plus granulaires, mais ne constitue pas la meilleure solution pour notre jeu de données hebdomadaire restreint.

3.9 Cas d'étude 2 : exportations hebdomadaires de coton (USDA)

Cette section présente l'évaluation comparative des modèles hybrides appliqués aux données hebdomadaires d'exportation de coton fournies par l'USDA.

Tableau 3-5 : Évaluation des modèles hybrides de prévision (USDA)

Modèle	WAPE (%)	Max WAPE (%)	MAE	RMSE
SARIMA–OP–XGBoost–OP	19.51	40.74	36.95	49.94
auto_ARIMA–XGBoost–OP	20.63	35.64	39.05	51.54
Prophet–XGBoost–OP	24.41	41.18	46.21	58.78

3.9.1 Visualisations SARIMA–OP–XGBoost–OP

Les visualisations ci-après présentent les performances du modèle hybride SARIMA–OP–XGBoost–OP appliqué aux données hebdomadaires d'exportation de coton de l'USDA.

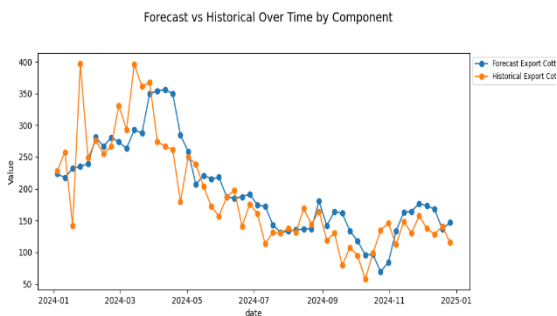


Figure 3-25 : Comparaison des valeurs prévues et historiques du modèle Prophet–XGBoost-OP

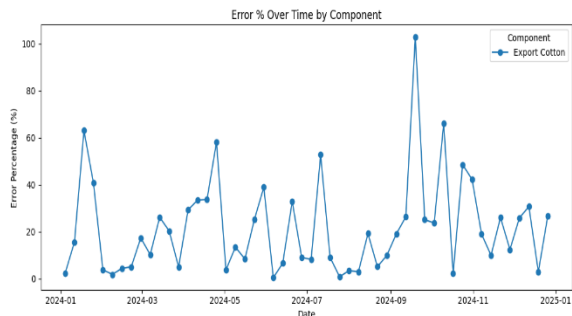


Figure 3-26 : Évolution de l'erreur de prévision (%) du modèle Prophet–XGBoost-OP

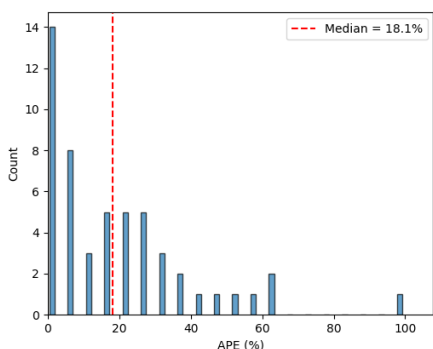


Figure 3-27: Distribution des erreurs absolues (%) du modèle Prophet–XGBoost-OP

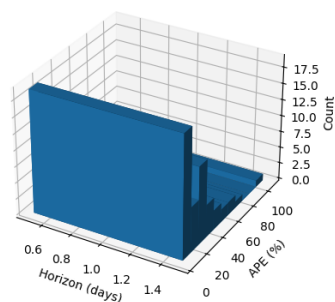


Figure 3-28 : Analyse 3d de l'erreur absolue en fonction de l'horizon temporel et du pourcentage d'erreur (APE)

3.9.1.1 Discussion synthétique des performances SARIMA–OP–XGBoost-OP

Le modèle hybride SARIMA–OP–XGBoost-OP combine la capacité du modèle SARIMA à capter les composantes saisonnières et cycliques avec la puissance prédictive du modèle XGBoost optimisé. Les résultats obtenus attestent de ses performances supérieures comparativement aux autres configurations testées, avec un WAPE de 19,51%, soit la valeur la plus faible observée parmi les modèles évalués, et un Max WAPE de 40,74%, traduisant une stabilité relative même dans les cas extrêmes.

L'erreur absolue moyenne (MAE) est de 36,95 et l'erreur quadratique moyenne (RMSE) s'établit à 49,94, confirmant une bonne capacité de ce modèle à limiter les écarts entre les prévisions et les valeurs réelles. La Figure 3-13 illustre visuellement cette concordance, mettant en évidence une forte cohérence entre les valeurs prévues et les observations historiques. Le modèle suit efficacement les tendances globales, avec une capacité à anticiper les variations saisonnières et les ruptures de tendance.

La Figure 3-14 montre l'évolution temporelle des erreurs de prévision. Malgré quelques pics d'erreurs liés à des fluctuations atypiques de la série, les écarts restent modérés et ne compromettent pas la qualité générale du modèle. Cette stabilité est renforcée par la Figure 3-15, représentant la distribution des erreurs absolues en pourcentage. La majorité des erreurs se concentre sous la barre des 20%, avec une médiane maintenue à 18,1%, témoignant d'une régularité dans les performances.

La Figure 3-16 complète cette analyse par une visualisation tridimensionnelle croisant l'horizon temporel et le pourcentage d'erreur absolue (APE). Elle confirme que l'erreur augmente avec l'horizon de prévision, mais de façon contenue, ce qui atteste de la robustesse du modèle même à moyen terme.

3.9.1.2 Conclusion

Le modèle SARIMA–OP–XGBoost–OP présente la meilleure performance globale parmi les modèles hybrides testés, en combinant une bonne précision (WAPE le plus bas), une faible variabilité d'erreur, et une stabilité sur différents horizons temporels. Sa capacité à capter les tendances sous-jacentes, tout en s'adaptant à des variations saisonnières, en fait un candidat robuste pour les systèmes de prévision dans un contexte de production industrielle. Cette robustesse prédictive permet d'optimiser la planification, la gestion des ressources et la prise de décision stratégique dans les environnements soumis à des cycles de demande complexes.

3.9.2 Visualisations auto_ARIMA–XGBoost–OP

Les visualisations suivantes illustrent les performances du modèle hybride auto_ARIMA–XGBoost–OP sur les séries temporelles hebdomadaires d'exportation de coton fournies par l'USDA.

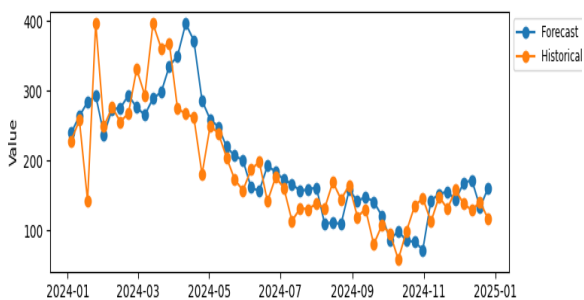


Figure 3-29 : Comparaison des valeurs prévues et historiques du modèle auto_ARIMA–XGBoost

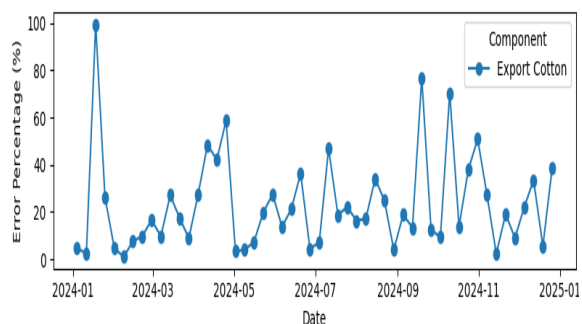


Figure 3-30 : Évolution de l'erreur de prévision (%) du modèle auto_ARIMA–XGBoost

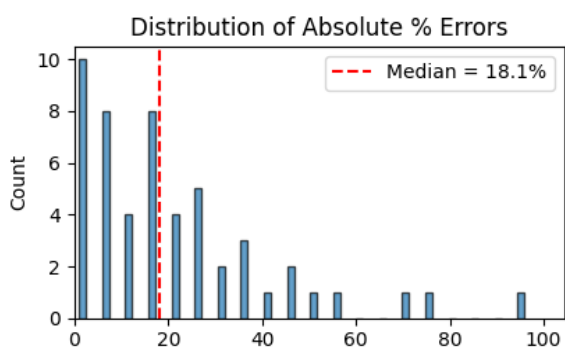


Figure 3-31: Distribution des erreurs absolues (%) du modèle auto_ARIMA–XGBoost

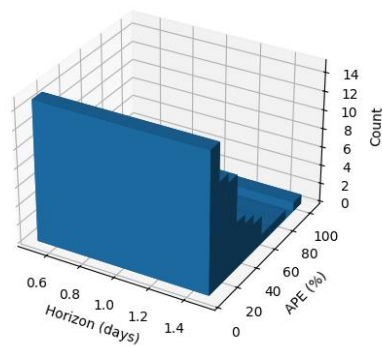


Figure 3-32 : Analyse 3d de l'erreur absolue en fonction de l'horizon temporel et du pourcentage d'erreur (APE)

3.9.2.1 Discussion synthétique des performances auto_ARIMA–XGBoost–OP

Le modèle hybride auto_ARIMA–XGBoost a été évalué sur la base de plusieurs métriques statistiques pour mesurer la qualité de ses prévisions sur les données d'exportation de coton. Les résultats obtenus montrent une performance globalement satisfaisante, avec un WAPE de 20,63% et un Max WAPE atteignant 35,64%. Ces valeurs indiquent que l'erreur moyenne

pondérée reste maîtrisée, bien que certaines périodes affichent des écarts plus importants, comme l'illustrent les pics d'erreur observés dans la Figure.

L'analyse de l'erreur absolue moyenne (MAE) s'élève à 39,05, tandis que l'erreur quadratique moyenne (RMSE) est de 51,54. Le MAE exprime la moyenne des erreurs absolues entre les prévisions et les valeurs observées, reflétant l'exactitude des prédictions. Quant au RMSE, il donne davantage de poids aux erreurs importantes, ce qui permet de détecter les périodes où les écarts sont les plus significatifs.

La Figure 3-1 permet de visualiser la cohérence entre les valeurs prévues par le modèle et les données historiques. Le modèle parvient à capter les tendances générales et les variations saisonnières, bien que certaines fluctuations abruptes échappent à la précision du modèle, ce qui est également visible dans les erreurs de prévision (%) représentées dans la Figure 3-2.

La Figure 3-3 expose la distribution des erreurs absolues. La médiane, positionnée à 18,1%, montre que la moitié des erreurs absolues se situe en dessous de ce seuil, traduisant une certaine stabilité dans les prévisions. Toutefois, la présence d'erreurs significatives, visibles aux extrêmes, suggère des ajustements potentiels pour renforcer la robustesse du modèle.

Enfin, la Figure 3-4 propose une analyse tridimensionnelle de l'erreur en fonction de l'horizon temporel. Cette représentation démontre une augmentation de l'erreur absolue à mesure que l'horizon de prévision s'élargit, un phénomène classique dans les modèles de prévision. Cet élément souligne l'importance d'optimiser l'ajustement des paramètres lorsque l'on s'éloigne de l'instant présent.

3.9.2.2 Conclusion

En résumé, le modèle auto_ARIMA-XGBoost se révèle efficace pour la prédiction à court terme de l'exportation de coton, capturant les tendances principales et les mouvements saisonniers avec une marge d'erreur acceptable. Toutefois, les analyses graphiques mettent en lumière des fluctuations non capturées par le modèle, suggérant l'opportunité d'un raffinement, notamment par l'intégration de techniques de régularisation supplémentaires ou l'exploitation de méthodes d'ensemble. Ce modèle peut donc servir de base solide pour la planification stratégique, tout en ouvrant la voie à des améliorations futures pour affiner la précision des prévisions.

3.9.3 Visualisations Prophet-XGBoost-OP

Les visualisations ci-dessous présentent les résultats du modèle hybride Prophet-XGBoost-OP appliqué aux données hebdomadaires d'exportation de coton issues de l'USDA.

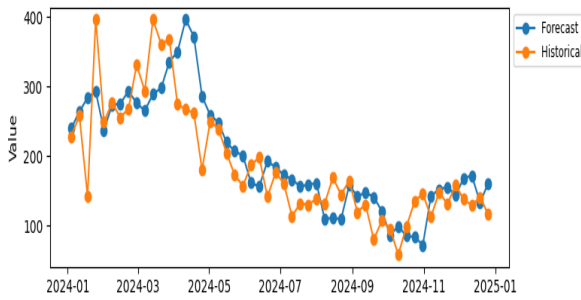


Figure 3-33 : Comparaison des valeurs prévues et historiques du modèle Prophet-XGBoost-OP

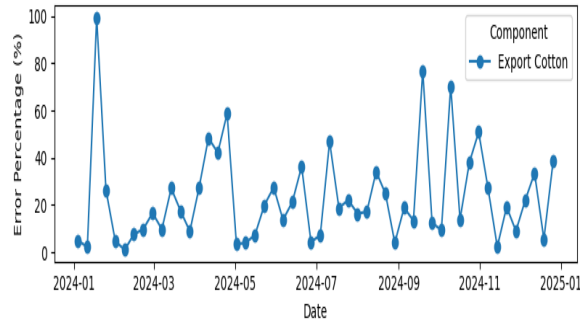


Figure 3-34 : Évolution de l'erreur de prévision (%) du modèle Prophet-XGBoost-OP

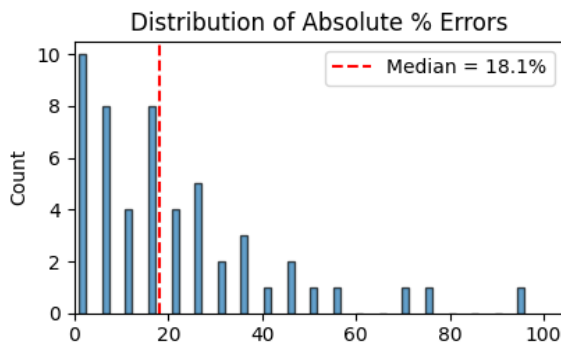


Figure 3-35: Distribution des erreurs absolues (%) du modèle Prophet-XGBoost-OP

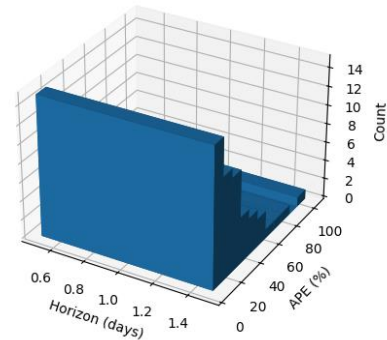


Figure 3-36 : Analyse 3d de l'erreur absolue en fonction de l'horizon temporel et du pourcentage d'erreur (APE)

3.9.3.1 Discussion synthétique des performances Prophet-XGBoost-OP

Le modèle hybride Prophet-XGBoost-OP a été évalué selon des métriques statistiques rigoureuses afin de mesurer ses capacités prédictives sur les données d'exportation de coton. Les résultats obtenus montrent un WAPE de 24,41% et un Max WAPE atteignant 41,18%. Ces valeurs indiquent une précision relative dans les prévisions, bien que certaines périodes connaissent des écarts plus importants, illustrés par les pics observés dans la Figure 3-14.

L'erreur absolue moyenne (MAE) s'élève à 46,21, tandis que l'erreur quadratique moyenne (RMSE) atteint 58,78. Le MAE représente l'écart moyen entre les valeurs prévues et les valeurs observées, tandis que le RMSE, plus sensible aux erreurs extrêmes, met en évidence les périodes de fortes divergences.

La Figure 3-13 montre une comparaison détaillée entre les valeurs prévues par le modèle et les données historiques. Globalement, Prophet-XGBoost-OP parvient à capter les variations saisonnières et les tendances structurelles, bien que certaines irrégularités échappent à la précision du modèle. Les différences visibles sur certains pics suggèrent un potentiel d'optimisation, notamment dans la capture des anomalies brusques.

La Figure 3-14 met en lumière l'évolution temporelle des erreurs de prévision. On observe des pics marqués d'erreur, indiquant des déviations significatives lors de périodes de volatilité accrue de la demande. L'analyse de cette erreur permet de cibler les phases de prévision les plus sensibles et d'optimiser les paramètres du modèle pour une meilleure robustesse.

La Figure 3-15 présente la distribution des erreurs absolues. La médiane est fixée à 18,1%, traduisant une relative stabilité des prévisions autour de cette valeur. Néanmoins, la présence

de valeurs extrêmes démontre une nécessité d'amélioration dans la gestion des incertitudes lors des pics de demande.

Enfin, la Figure 3-16 expose une analyse tridimensionnelle de l'erreur absolue en fonction de l'horizon temporel. Cette visualisation met en évidence une augmentation des erreurs lorsque l'horizon de prévision s'étend, confirmant l'impact de la profondeur temporelle sur la précision du modèle.

3.9.3.2 Conclusion

Le modèle Prophet–XGBoost-OP montre une capacité satisfaisante à prédire les tendances à court et moyen terme des exportations de coton. Les analyses graphiques révèlent cependant des fluctuations non anticipées, particulièrement lors des phases de volatilité marquée. L'intégration de méthodes complémentaires de régularisation ou d'optimisation pourrait renforcer la fiabilité du modèle sur des horizons plus longs. Cette analyse confirme la pertinence de Prophet–XGBoost-OP dans un cadre industriel de planification tout en laissant place à des ajustements pour une meilleure robustesse prédictive.

3.10 Analyse globale et discussion

3.10.1 Analyse des cas d'étude

Les deux cas d'étude analysés — la demande hebdomadaire de Tayal Spa et les exportations hebdomadaires de coton de l'USDA — offrent des dynamiques temporelles contrastées. Le premier présente une forte volatilité, une absence de saisonnalité claire et une variation non constante de la variance importante, tandis que le second se caractérise par une régularité structurelle marquée et une saisonnalité annuelle bien définie. Ces différences structurelles ont permis de tester la flexibilité et la robustesse des pipelines hybrides proposés.

Dans le cas de Tayal Spa, malgré un jeu de données restreint (32 semaines), le modèle SARIMA–XGBoost optimisé par Optuna démontre une aptitude remarquable à générer des prévisions cohérentes. En revanche, pour les exportations USDA, les performances sont globalement meilleures, grâce à la présence de cycles saisonniers stables et à un historique plus étendu (208 semaines), favorisant une modélisation statistique plus fiable.

3.10.2 Comparaison des pipelines hybrides

Trois configurations hybrides ont été évaluées : SARIMA–XGBoost-OP, auto_ARIMA–XGBoost-OP et Prophet–XGBoost-OP. Les résultats démontrent une hiérarchie de performance cohérente entre les modèles.

- **SARIMA–XGBoost-OP** obtient les meilleurs scores globaux, avec un WAPE de 19,51 % sur les données USDA, ce qui traduit une excellente capacité à modéliser la composante linéaire tout en corrigeant les résidus non linéaires via XGBoost.
- **auto_ARIMA–XGBoost-OP** offre un compromis intéressant entre performance et automatisation, avec un WAPE légèrement supérieur mais des temps de configuration réduits. Ce modèle s'avère particulièrement adapté aux contextes de données limitées.

- **Prophet–XGBoost–OP** montre des limites dans les deux cas d'étude, notamment en raison de l'échantillon réduit et de la granularité hebdomadaire, qui pénalisent ses algorithmes de détection automatique de points de changement.

L'approche SARIMA–XGBoost, avec tuning via Optuna, ressort ainsi comme la solution la plus performante dans des contextes industriels à données limitées mais bruitées, tout en restant extensible à des séries plus longues.

3.10.3 Robustesse et adaptation en ligne

La stratégie bi-hebdomadaire de réentraînement permet au modèle d'intégrer rapidement les écarts entre prévisions et observations, favorisant une adaptation incrémentale. Cette logique d'apprentissage en ligne (online learning) s'est révélée particulièrement efficace dans le cas de Tayal Spa, où l'instabilité de la demande exige une réactivité élevée du système prédictif.

L'architecture de stacking SARIMA–XGBoost renforce cette capacité d'adaptation en traitant successivement les composantes linéaires puis les résidus, ce qui permet de mieux capter les ruptures de tendance ou les événements exogènes. Les performances restent stables même à mesure que l'horizon de prévision s'étend, comme en témoigne l'analyse tridimensionnelle des erreurs absolues (APE).

3.10.4 Limites et perspectives d'amélioration

Malgré les performances encourageantes du pipeline hybride, plusieurs limites doivent être soulignées :

- **Taille de l'échantillon** : Un historique plus long améliorerait significativement la détection des composantes saisonnières et permettrait un entraînement plus robuste du modèle XGBoost.
- **Modélisation des anomalies** : Bien que XGBoost corrige partiellement les erreurs de prévision, certaines anomalies extrêmes (liées à des événements exogènes) ne sont pas captées de manière satisfaisante. L'intégration de variables exogènes explicatives (calendrier, météo, événements) pourrait enrichir le modèle.
- **Optimisation computationnelle** : L'usage systématique d'Optuna pour chaque composant du pipeline peut générer des coûts computationnels non négligeables. Des stratégies d'optimisation multi-niveaux (coarse-to-fine tuning) pourraient être envisagées.

En perspective, l'extension vers des architectures de Deep Learning hybrides (ex. LSTM résiduel, N-BEATS combiné à XGBoost) pourrait améliorer encore la capture des dynamiques non linéaires, surtout en présence de données massives. Par ailleurs, la mise en place d'un cadre automatisé de monitoring des erreurs (MLOps) permettrait une supervision continue des performances prédictives dans un environnement de production.

3.11 Implications opérationnelles

La mise en œuvre d'un pipeline hybride de prévision (SARIMA–XGBoost) présente des retombées opérationnelles significatives dans un environnement industriel caractérisé par une

demande volatile et des contraintes de planification serrées. Les implications se déclinent à plusieurs niveaux fonctionnels :

3.11.1.1 Planification de la production

Les prévisions bi-hebdomadaires permettent d'adapter finement les niveaux de production à la demande projetée. Cette granularité temporelle, couplée à une capacité d'apprentissage en ligne, offre une réactivité accrue face aux fluctuations imprévues, limitant les risques de surstock ou de rupture.

3.11.1.2 Gestion des stocks

La réduction de l'erreur moyenne (WAPE < 20 % dans certains cas) se traduit par une amélioration tangible de la précision des niveaux de stock projetés. En conséquence, les décisions liées aux seuils de réapprovisionnement, aux niveaux de sécurité ou au dimensionnement des capacités de stockage peuvent être optimisées.

3.11.1.3 Allocation des ressources

Une meilleure prévisibilité de la demande permet une allocation proactive des ressources humaines et matérielles. Les équipes peuvent être planifiées en fonction de pics anticipés, et les capacités de production ajustées de manière dynamique.

3.11.1.4 Détection des anomalies et résilience

Grâce à la composante XGBoost, le système est capable d'intégrer rapidement les erreurs passées et de détecter des ruptures de tendance. Ceci est essentiel dans un contexte où les interruptions logistiques, les effets saisonniers irréguliers ou les chocs exogènes (ex. réglementaires ou climatiques) peuvent déstabiliser les opérations.

3.11.1.5 Appui à la décision par l'intelligence artificielle

L'intégration d'un assistant IA (OpenAI API) permet de transformer les prévisions en décisions concrètes, grâce à une capacité d'analyse de scénarios en langage naturel. Par exemple, un gestionnaire peut interroger l'assistant sur les conséquences d'un pic de demande, l'évolution des marges d'erreur, ou l'impact d'un retard fournisseur sur les taux de service.

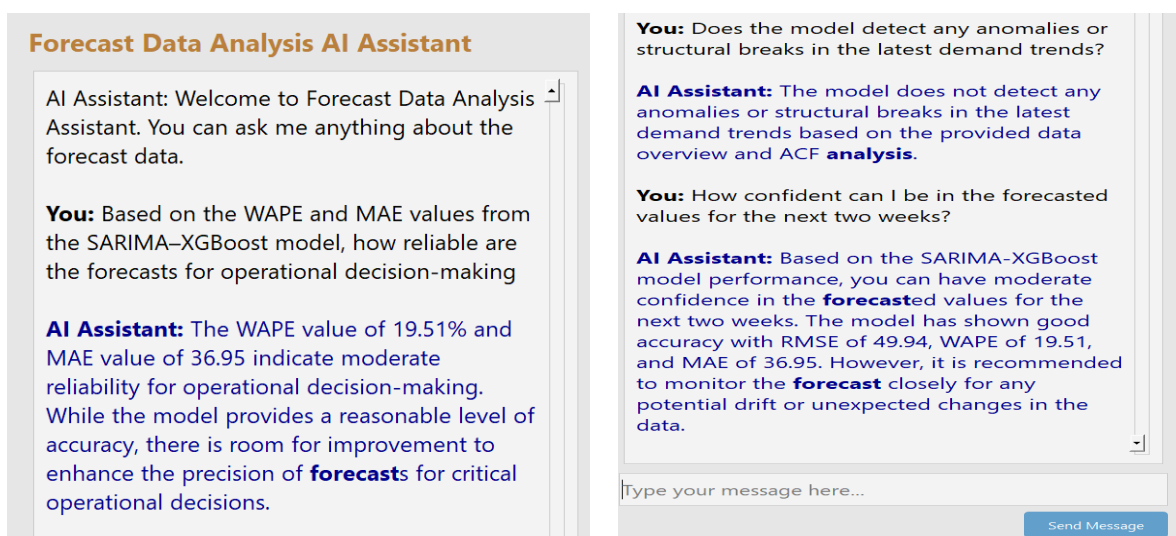


Figure 3-37 : Assistance IA pour l'analyse des prévisions

Cette couche d'intelligence artificielle facilite l'interprétation des sorties du modèle par des acteurs non techniques, en rendant l'analyse accessible, rapide et contextualisée. Elle permet notamment de traduire des indicateurs statistiques complexes (tels que WAPE, MAE ou RMSE) en recommandations opérationnelles compréhensibles, renforçant ainsi la prise de décision éclairée. De plus, l'IA peut jouer un rôle de support en situation d'incertitude, en simulant des scénarios alternatifs, en détectant des anomalies éventuelles, ou en suggérant des ajustements de seuils de sécurité ou de cadence de production selon l'évolution des prévisions.

3.11.1.6 Perspectives organisationnelles

À terme, cette approche hybride et assistée par IA peut être intégrée dans une chaîne décisionnelle plus large, comme les systèmes ERP ou APS, assurant une automatisation partielle des boucles de planification (forecast → planification → réajustement). Elle favorise également une culture d'amélioration continue basée sur l'exploitation proactive des données.

3.12 Conclusion

Dans ce chapitre, nous avons développé et validé un pipeline hybride séquentiel alliant une phase de décomposition statistique (SARIMA ou Prophet) à une correction non linéaire des résidus par XGBoost, le tout optimisé via Optuna et réentraîné de façon bimensuelle. Nous avons démontré que cette approche fournit des prévisions à la fois cohérentes et robustes : malgré un historique restreint à 32 semaines, notre modèle a assuré des projections fiables pour la production hebdomadaire de coton de TAYAL SPA, et il a atteint un WAPE de 19,51 % sur la série USDA à 208 semaines, meilleur résultat parmi toutes les méthodes testées.

Nous avons également mis en œuvre une assistance par intelligence artificielle pour orchestrer automatiquement la calibration bayésienne des hyperparamètres, superviser en continu les performances et déclencher les cycles de recalibrage bimensuel sans intervention manuelle. Cette IA d'assistance a permis d'accélérer les itérations de tuning, de réduire le risque d'erreur humaine et d'assurer une traçabilité totale des ajustements apportés au modèle.

En pratique, cette solution offre aux responsables industriels un véritable outil d'aide à la décision proactive, capable de réduire les déséquilibres de stocks et d'optimiser l'allocation des ressources. Nous proposons ainsi une méthode prédictive innovante, opérationnelle et extensible, qui établit une nouvelle référence pour la prévision de séries temporelles complexes en environnement industriel.

Chapitre IV : Présentation Du Logiciel Développé

4. Introduction

Le logiciel développé centralise l'ensemble du workflow de prévision hybride décrit au chapitre trois. Il intègre des modules de traitement linéaire (SARIMA, Prophet, auto-ARIMA), d'optimisation bayésienne (Optuna) et de régression non linéaire (XGBoost) avec réinjection des résidus pour un apprentissage continu. L'application vise à automatiser la préparation des données, la sélection et le réglage des modèles, la génération de prévisions et l'évaluation à travers des indicateurs clés (WAPE, RMSE, MAE) ; ce chapitre présente ainsi la structure logicielle, les modules fonctionnels, les procédures d'installation et de déploiement, ainsi que les lignes directrices pour la maintenance et l'évolution.

4.1 Identité de l'Application

L'application VisionPlan est développée et éditée par **InnoVista Dev**, notre startup spécialisée dans l'optimisation industrielle et la transformation numérique par l'intelligence artificielle. Elle incarne une vision moderne de l'industrie, où l'analyse prédictive, la planification intelligente et l'automatisation convergent au sein d'une interface intuitive.

Notre identité visuelle repose sur un logo distinctif : un engrenage stylisé fusionné à un circuit numérique, illustrant la synergie entre précision mécanique et intelligence numérique. Ce symbole reflète les valeurs fondamentales du projet : efficacité, connectivité et innovation collaborative.

Le nom **VisionPlan** traduit clairement l'ambition du produit : « *Voir l'avenir, planifier avec précision* ». Cette philosophie guide le développement de chaque module de l'application, permettant aux utilisateurs d'anticiper les tendances, de visualiser les indicateurs clés et d'optimiser leurs stratégies industrielles.

D'un point de vue technique, le développement s'est concentré exclusivement sur les technologies Python, PyQt5 et SQL, garantissant à la fois robustesse, extensibilité et intégration multiplateforme. Le travail accompli est considérable et reflète un engagement profond envers la qualité logicielle :

- 7 954 fichiers Python
- 582 600 lignes de code vides (espaces de structuration)
- 853 632 lignes de commentaires
- 2 102 737 lignes de code effectif

Ce volume reflète non seulement l'ampleur du développement, mais aussi la rigueur apportée à la **documentation interne** et à la **lisibilité du code**, assurant la maintenabilité à long terme.

Le **logo de VisionPlan** est affiché dès l'écran de connexion, et intégré dans le tableau de bord principal, renforçant la cohérence visuelle de l'outil. Le copyright et les mentions légales précisent :

"© 2025 InnoVista Dev. All Rights Reserved."



Figure 4-1: Écran de démarrage de l'application VisionPlan

4.2 Architecture logicielle

L'architecture logicielle de VisionPlan repose sur une démarche pragmatique de recueil des besoins, un environnement technique cohérent et une conception modulaire assurant une implémentation back-end robuste et évolutive.

4.2.1 Recueil pragmatique des besoins du prototype

La définition des besoins débute par l'identification des données à traiter : les ventes historiques proviennent d'une base de données MySQL interne, les données de production sont soit injectées directement depuis l'ERP après chaque cycle de prévision, soit récupérées de manière automatisée lors du calcul des prévisions. Par défaut, le logiciel génère des prévisions couvrant deux semaines, mais permet également aux utilisateurs de produire des prévisions quotidiennes, mensuelles ou annuelles selon leurs besoins. Le flux de traitement se compose de trois étapes : l'importation et la préparation des données, le calcul statistique des prévisions, suivi de leurs ajustements en s'appuyant sur les erreurs passées, et enfin la présentation claire des résultats dans un tableau de bord. Pour rester opérationnel, l'outil doit produire un jeu complet de prévisions en moins de deux minutes et une prévision simple en moins de trente secondes.

Les critères non fonctionnels incluent la protection des accès par authentification, la capacité à supporter plusieurs tâches simultanées, la sauvegarde automatique quotidienne des données et des modèles, la traçabilité des versions du logiciel et un déploiement simplifié pour garantir une maintenance rapide et un usage fiable. De plus, le logiciel intègre un assistant IA interactif qui explique les écarts entre les prévisions et les valeurs réelles, guide l'utilisateur dans l'interprétation des indicateurs et répond aux questions en langage naturel sur les tendances observées.

4.2.2 Environnement technique et outils utilisés

L'interface utilisateur adopte le modèle MVVM (Model-View-ViewModel), garantissant la séparation entre les vues, la logique de présentation et les données métiers. Le prototype est développé en Python 3.8+ avec PyQt5, permettant de construire des vues légères pour la saisie des paramètres et l'affichage des résultats. Les séries temporelles sont manipulées via Pandas et NumPy, permettant un traitement efficace des volumes allant de quelques milliers à plusieurs millions de points.

Les prévisions linéaires s'appuient sur Statsmodels (SARIMA et auto-ARIMA) et Prophet pour gérer les tendances, les saisonnalités multiples et les ruptures. La phase de stacking exploite XGBoost pour corriger les résidus non linéaires, tandis qu'Optuna pilote l'optimisation bayésienne des hyperparamètres selon le critère MASE.

Les données brutes et les résidus sont hébergés dans MySQL. Les tables sont partitionnées par granularité temporelle (quotidienne, hebdomadaire, mensuelle, annuelle) afin d'optimiser les performances de requête sur de grands volumes et de faciliter la maintenance des archives. L'accès aux données depuis Python est géré via mysql.connector, encapsulé dans un module CRUD qui fournit des fonctions de lecture et d'écriture retournant des structures natives Python.

4.2.3 Conception modulaire des composants

La structure du prototype repose sur des modules indépendants, chacun étant responsable d'une fonction spécifique du workflow de prévision. Le module d'ingestion collecte et normalise les données brutes (base de données MySQL, CSV, API), garantissant un chargement uniforme des séries. Le module de modélisation statistique regroupe les implémentations SARIMA/auto-ARIMA et Prophet, exposant un point d'entrée unique pour le calcul des prévisions linéaires.

L'optimisation des hyperparamètres est isolée dans un module dédié à Optuna : il orchestre les études bayésiennes pour chaque algorithme, stocke les essais et fournit au pipeline un jeu de paramètres optimaux basé sur le critère MASE. Le module de traitement des résidus, quant à lui, ingère les erreurs issues de la modélisation linéaire, enrichit les données avec des features contextuelles et exécute XGBoost pour corriger la prévision initiale.

Les composants de visualisation et d'export génèrent des graphiques et des rapports à partir des résultats consolidés et offrent des interfaces de configuration pour ajuster les paramètres du prototype. Un module d'assistant IA s'intègre sous forme de widget de chat, interrogeant l'API OpenAI pour commenter automatiquement les indicateurs et guider l'utilisateur dans l'interprétation des résultats. Chaque module communique via des contrats internes (fonctions Python, schémas JSON) et peut être testé et déployé de manière indépendante, accélérant l'itération et facilitant la maintenance évolutive.

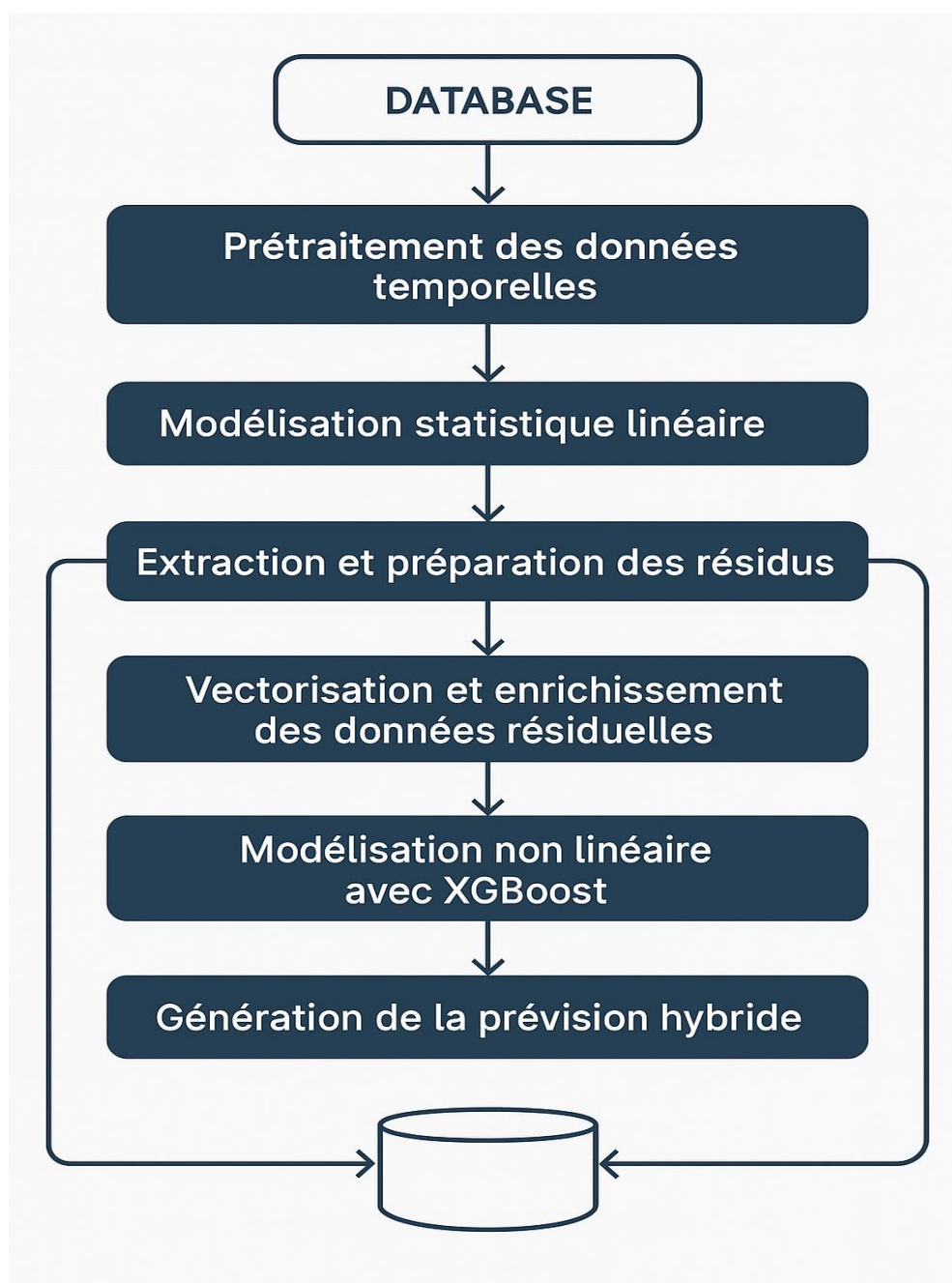


Figure 4-2: Chaîne de Traitement Hybride pour l'Amélioration des Prévisions

4.2.4 Implémentation back-end

L'architecture back-end repose sur une architecture en couches : une couche d'accès aux données, un service de prévision et un mécanisme de journalisation. La couche d'accès utilise `mysql.connector` et un module CRUD unique pour masquer la complexité SQL derrière des fonctions Python retournant des DataFrames ou des dictionnaires. Le service de prévision orchestre l'ensemble du pipeline (optimisation Optuna → SARIMA/Prophet → extraction des résidus → XGBoost), puis conserve les résultats et les paramètres optimisés. Le module de journalisation trace, pour chaque cycle, l'horodatage, la durée, les paramètres et les éventuelles erreurs, facilitant le diagnostic.



Figure 4-3: Diagramme entité-relation de la base MySQL

Tableau 4-1: Dictionnaire des tables de la base de données MySQL

Entité	Rôle principal
Daily_historical_data	Stocke les valeurs réelles journalières ; clés : (date, component).
Weekly_historical_data	Valeurs réelles hebdomadaires,
Monthly_historical_data	Valeurs réelles mensuelles,
Yearly_historical_data	Valeurs réelles annuelles,
Daily_forecast_data	Prévisions journalières ; inclut engine_used pour tracer le modèle.
Weekly_forecast_data	Prévisions hebdomadaires.
Monthly_forecast_data	Prévisions mensuelles.
Yearly_forecast_data	Prévisions annuelles.
Forecast_settings	Paramètres de l'utilisateur (mode d'exécution, horizon, auto-tuning, période d'entraînement).
Sarima_parameters	Journal des essais SARIMA/auto-ARIMA
Xgb_parameters	Journal des essais XGBoost ; stocke hyperparamètres et score.
Account_manager activation_app app_config	Gestion des utilisateurs, Activations et de la version de l'application, Configuration de l'application.

Les relations sont définies sur les clés primaires de chaque table temporelle (date, component) et référencées par forecast_settings pour retrouver le contexte de chaque exécution. Cette structuration par granularité permet des requêtes d'agrégation optimisées et une modularité forte : toute évolution du modèle ou du format de stockage ne concerne qu'un sous-ensemble limité de tables et de fonctions, garantissant la maintenabilité et la rapidité d'itération du prototype.

4.3 Interface utilisateur et flux de travail

L'interface utilisateur de l'application est divisée en trois vues principales, permettant une navigation intuitive et un accès rapide aux différentes fonctionnalités :

4.3.1 Page de Connexion (Login Page)

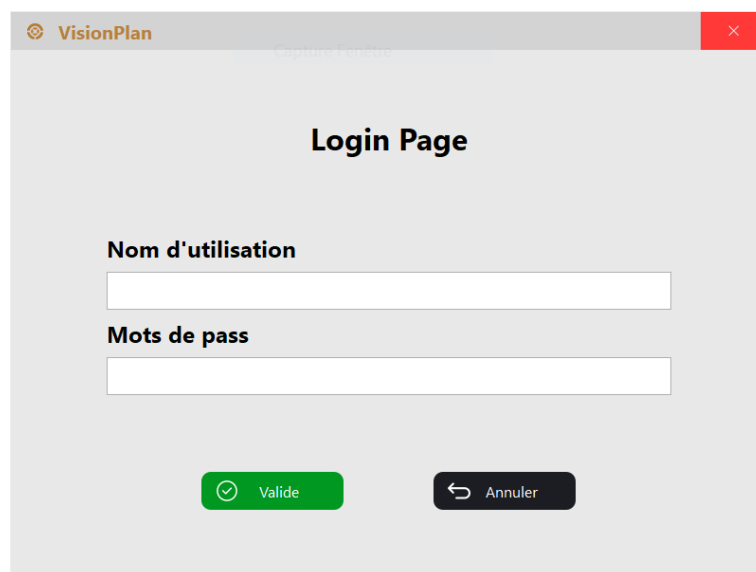


Figure 4-4: l'interface de connexion

La capture présente l'écran de connexion de l'application VisionPlan, structuré selon les principes d'ergonomie et de simplicité. Le bandeau supérieur affiche le logo et le titre « VisionPlan », assurant une identification immédiate du logiciel. Le titre de la fenêtre, « **Login Page** », est centré et mis en évidence avec une police de grande taille, guidant l'utilisateur vers l'action principale.

Deux champs de saisie alignés verticalement sont ensuite proposés : « **Nom d'utilisation** » et « **Mots de passe** », chacun accompagné d'un libellé en caractères gras pour renforcer la lisibilité. Les zones de texte bénéficient d'un espace suffisant pour la saisie et d'un contraste élevé avec le fond clair, facilitant la reconnaissance des caractères.

La zone de contrôle comporte deux boutons distincts :

- **Valide**, en vert, orné d'une icône de coche, indique clairement l'action de validation et rassure l'utilisateur quant à la réussite attendue.
- **Annuler**, en noir, muni d'une flèche de retour, offre une option de repli, respectant le principe de maîtrise de l'utilisateur.

L'usage de couleurs contrastées pour les actions critiques (vert pour confirmer, noir pour annuler) renforce l'affordance et permet une navigation intuitive sans texte explicatif superflu. La disposition en gradation verticale, l'espacement équilibré et l'alignement à gauche des étiquettes contribuent à un parcours visuel fluide, conforme aux standards d'interface centrés utilisateur.

4.3.2 Tableau de Bord (Dashboard)

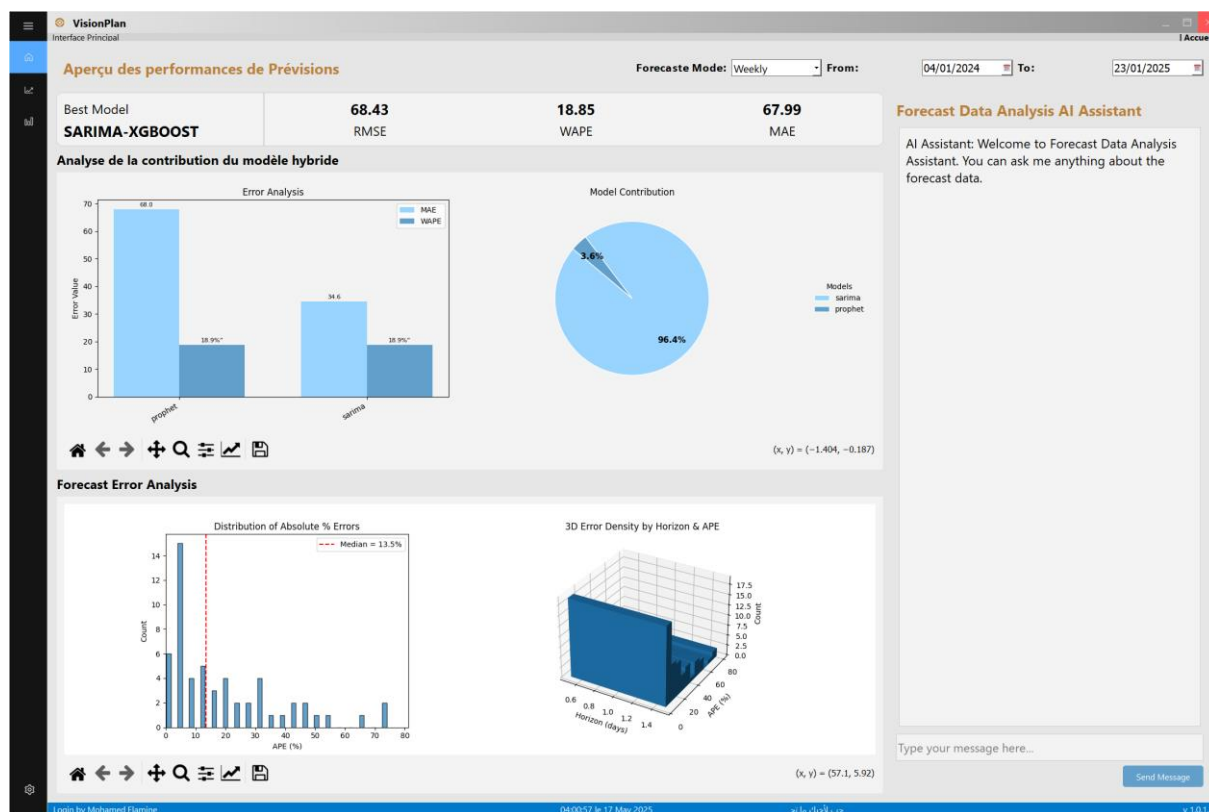


Figure 4-5 : Interface de visualisation des performances de prévision dans VisionPlan

La capture d'écran ci-dessus illustre l'interface dédiée à l'évaluation des performances de prévision dans l'application VisionPlan. Cette vue se décompose en plusieurs sections distinctes, permettant à l'utilisateur d'accéder rapidement aux indicateurs clés, aux analyses d'erreurs et aux contributions des modèles hybrides.

4.3.2.1 Résumé des performances globales

En haut de l'interface, un bandeau présente un aperçu synthétique des métriques de performance du meilleur modèle sélectionné, en l'occurrence le modèle hybride SARIMA-XGBoost. Les indicateurs affichés incluent :

- Le **RMSE** (Root Mean Square Error), traduisant l'écart quadratique moyen entre les prévisions et les valeurs réelles.
- Le **WAPE** (Weighted Absolute Percentage Error), qui mesure l'erreur moyenne pondérée en pourcentage.
- Le **MAE** (Mean Absolute Error), exprimant l'erreur absolue moyenne des prévisions.

4.3.2.2 Analyse de la contribution du modèle hybride

La section centrale propose deux représentations graphiques :

- **Un histogramme comparatif** illustrant l'analyse d'erreur (Error Analysis) entre le modèle Prophet et le modèle SARIMA, indiquant les contributions respectives au MAE et au WAPE. Les valeurs indiquées permettent de constater l'écart de performance entre les deux modèles dans le cadre du stacking.
- **Un diagramme circulaire** (Model Contribution) présente la répartition des contributions au modèle hybride, justifiant son impact majeur dans l'estimation finale.

4.3.2.3 Analyse détaillée des erreurs de prévision

En bas de l'interface, l'outil fournit une analyse approfondie des erreurs :

- **Un histogramme représentant la Distribution des erreurs absolues en pourcentage (APE)**, permettant d'identifier les erreurs fréquentes et leur amplitude. La médiane de l'erreur, fixée à 13.1%, est signalée par une ligne rouge.
- **Un graphique tridimensionnel (3D Error Density)** cartographiant la densité des erreurs en fonction de l'horizon de prévision et du pourcentage d'erreur absolue (APE). Cette représentation permet de visualiser les zones critiques où l'erreur est plus marquée.

4.3.2.4 Module d'assistance conversationnelle (Forecast Data Analysis AI Assistant)

À droite de l'interface se trouve l'assistant conversationnel, permettant à l'utilisateur de poser des questions en langage naturel sur les métriques, les prévisions et l'analyse des erreurs. L'assistant fournit des explications détaillées, compare les modèles entre eux et oriente l'utilisateur dans l'interprétation des résultats. Cette interface intégrée favorise une compréhension approfondie des dynamiques de prévision, tout en offrant un outil d'interaction directe pour analyser les performances et affiner les choix méthodologiques.

4.3.3 Vue de Prédiction (Forecasting View)

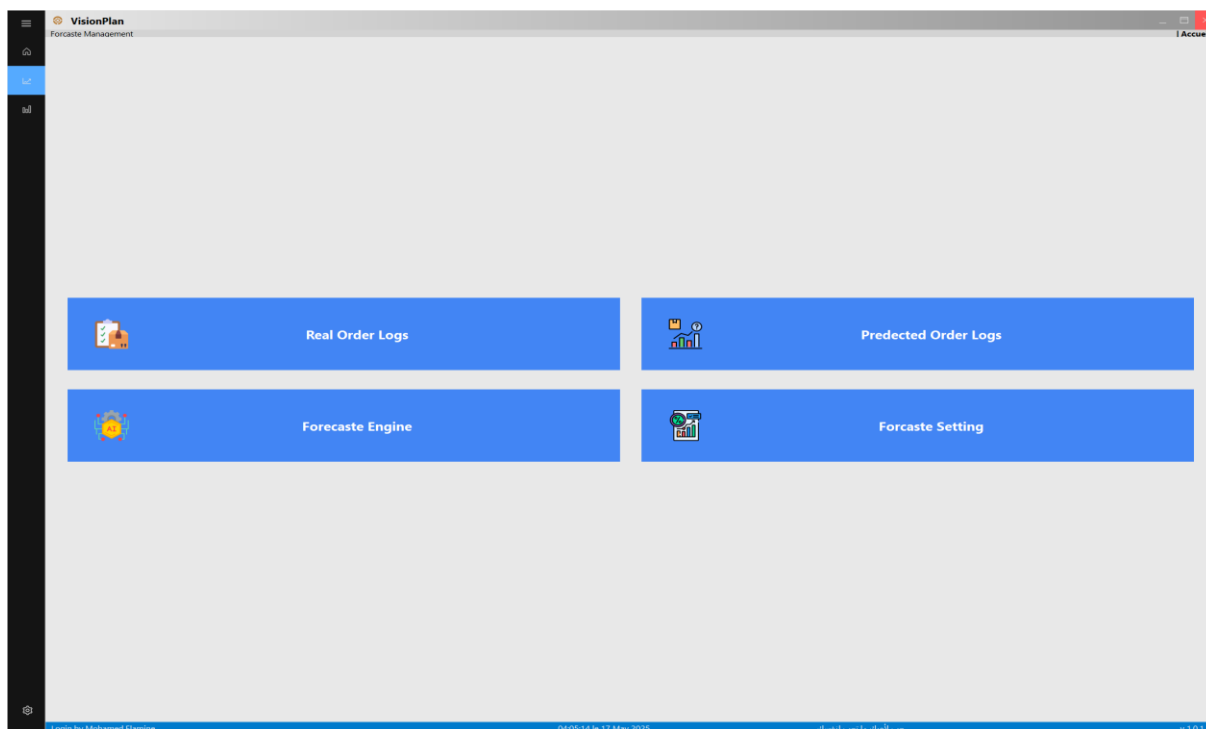


Figure 4-6 : Vue de Prédiction (Forecasting View) dans VisionPlan

Cette interface présente la vue principale de gestion des prévisions dans l'application **VisionPlan**, segmentée en quatre sections bien distinctes, facilitant l'accès aux fonctionnalités critiques du module de prévision.

4.3.4 Real Order Logs

Cette section permet d'accéder aux historiques des commandes réelles enregistrées. L'utilisateur peut y consulter les volumes, les dates et les variations dans le temps, permettant ainsi une analyse détaillée de l'historique des commandes. Cet accès direct aux logs facilite l'exploration des tendances passées et l'identification des périodes à forte volatilité.

The screenshot displays the 'Historical Data' view in the VisionPlan application. The main content is a table with the following columns: ID, Date, Component, Value, and Creation Date. The data shows weekly production volumes for 'Export Cotton' starting from 2021-01-07 and ending on 2021-07-01. The values fluctuate between approximately 205.98 and 456.64 units. The interface includes a search bar at the top, a 'Forecast Mode' dropdown menu currently set to 'Weekly', and date selection controls labeled 'Du:' (01/01/2021) and 'À:' (17/05/2025). The bottom status bar indicates the user is logged in as Mohamed Elamine on 04:07:47 le 17 May 2025, with version 1.0.1.1.

ID	Date	Component	Value	Creation Date
1	2021-01-07	Export Cotton	274.55	2025-05-17 00:02:24
2	2021-01-14	Export Cotton	322.38	2025-05-17 00:02:24
3	2021-01-21	Export Cotton	275.27	2025-05-17 00:02:24
4	2021-01-28	Export Cotton	318.96	2025-05-17 00:02:24
5	2021-02-04	Export Cotton	433.63	2025-05-17 00:02:24
6	2021-02-11	Export Cotton	311.84	2025-05-17 00:02:25
7	2021-02-18	Export Cotton	292.37	2025-05-17 00:02:25
8	2021-02-25	Export Cotton	377.40	2025-05-17 00:02:25
9	2021-03-04	Export Cotton	351.61	2025-05-17 00:02:25
10	2021-03-11	Export Cotton	351.87	2025-05-17 00:02:25
11	2021-03-18	Export Cotton	313.51	2025-05-17 00:02:25
12	2021-03-25	Export Cotton	324.70	2025-05-17 00:02:25
13	2021-04-01	Export Cotton	371.73	2025-05-17 00:02:25
14	2021-04-08	Export Cotton	313.20	2025-05-17 00:02:25
15	2021-04-15	Export Cotton	328.98	2025-05-17 00:02:25
16	2021-04-22	Export Cotton	341.91	2025-05-17 00:02:25
17	2021-04-29	Export Cotton	456.64	2025-05-17 00:02:25
18	2021-05-06	Export Cotton	277.17	2025-05-17 00:02:25
19	2021-05-13	Export Cotton	345.37	2025-05-17 00:02:25
20	2021-05-20	Export Cotton	323.52	2025-05-17 00:02:25
21	2021-05-27	Export Cotton	366.52	2025-05-17 00:02:25
22	2021-06-03	Export Cotton	258.36	2025-05-17 00:02:25
23	2021-06-10	Export Cotton	303.83	2025-05-17 00:02:25
24	2021-06-17	Export Cotton	205.98	2025-05-17 00:02:25
25	2021-06-24	Export Cotton	274.00	2025-05-17 00:02:25
26	2021-07-01	Export Cotton	294.79	2025-05-17 00:02:25

Figure 4-7: Vue des données historiques dans VisionPlan

4.3.4.1 Structure de l'affichage

L'affichage est présenté sous forme de tableau avec des colonnes clairement définies :

- **ID** : identifiant unique de chaque enregistrement.
- **Date** : date hebdomadaire de référence pour les volumes de production.
- **Component** : nature du produit suivi, ici "Export Cotton".
- **Value** : valeur quantitative enregistrée (en unités de production ou de volume).
- **Creation Date** : date et heure de l'enregistrement dans la base de données, indiquant l'instant précis de l'intégration dans le système.

4.3.4.2 Barre de recherche et filtres temporels

L'interface inclut une barre de recherche placée en haut, permettant de filtrer les résultats selon des critères textuels, tels qu'un composant spécifique ou une valeur précise. À droite, des sélecteurs de dates, intitulés "Du" et "À", offrent la possibilité de restreindre la période d'affichage souhaitée. De plus, une option intitulée "Forecast Mode" permet d'ajuster la granularité temporelle, offrant un affichage personnalisable selon un intervalle journalier, hebdomadaire ou mensuel. Cette organisation intuitive facilite l'exploration et l'analyse ciblée des données.

4.3.4.3 Ergonomie et navigation

L'interface utilise une mise en page aérée et un contraste visuel élevé, facilitant la lecture des informations. Les données sont triées chronologiquement, et l'utilisateur peut naviguer aisément à travers les enregistrements grâce à une barre de défilement fluide sur la droite.

4.3.4.4 Utilisation analytique

La vue analytique constitue un point d'entrée stratégique pour les analyses temporelles, permettant de vérifier les tendances passées, d'identifier les pics de production ainsi que les creux notables. Elle facilite également la préparation des données nécessaires à l'entraînement des modèles de prévision tels que **SARIMA, Prophet et XGBoost**. Par ailleurs, elle offre la possibilité de comparer directement les résultats avec les logs de prévision, afin d'évaluer les écarts entre les valeurs réelles et les valeurs prédites. Grâce à une architecture modulaire et un design épuré, cette interface garantit une prise en main rapide, idéale pour les opérateurs et les analystes de données désireux de consulter l'historique de production d'un simple coup d'œil.

4.3.5 Predicted Order Logs

Cet espace regroupe les prévisions générées par le moteur de calcul (SARIMA, Prophet, XGBoost). L'utilisateur peut y comparer les estimations avec les données réelles, détecter les écarts et ajuster les paramètres si nécessaire. Cette fonctionnalité assure un suivi précis des performances de prévision et permet d'évaluer la robustesse des modèles.

	Date	Component	Forecasted Value	Real Value	Error	Abs Error	Error %
1	2024-01-04	Export Cotton	293.8480	228.0620	65.7860	65.7860	28.84566478
2	2024-01-11	Export Cotton	326.6064	257.6950	68.9114	68.9114	26.74145793
3	2024-01-18	Export Cotton	210.8674	142.2420	68.6254	68.6254	48.24552523
4	2024-01-25	Export Cotton	366.6327	396.7110	-30.0783	30.0783	-7.58191732
5	2024-02-01	Export Cotton	247.0244	248.5120	-1.4876	1.4876	-0.59860288
6	2024-02-08	Export Cotton	289.1674	276.0570	13.1104	13.1104	4.74916412
7	2024-02-15	Export Cotton	275.2020	255.4840	19.7180	19.7180	7.71790014
8	2024-02-22	Export Cotton	267.7278	267.1150	0.6128	0.6128	0.22941430
9	2024-02-29	Export Cotton	288.0112	330.7900	-42.7788	42.7788	-12.93231355
10	2024-03-07	Export Cotton	274.6692	293.2820	-18.6128	18.6128	-6.34638334
11	2024-03-14	Export Cotton	287.3983	395.9690	-108.5707	108.5707	-27.41898987
12	2024-03-21	Export Cotton	315.9529	360.7210	-44.7681	44.7681	-12.41072740
13	2024-03-28	Export Cotton	347.2703	367.5510	-20.2807	20.2807	-5.51779209
14	2024-04-04	Export Cotton	388.5819	274.1470	114.4349	114.4349	41.74216752
15	2024-04-11	Export Cotton	360.8320	266.7320	94.1000	94.1000	35.27885668
16	2024-04-18	Export Cotton	344.4679	261.6980	82.7699	82.7699	31.62802161
17	2024-04-25	Export Cotton	297.9435	179.9800	117.9635	117.9635	65.54256028
18	2024-05-02	Export Cotton	264.5931	249.6280	14.9651	14.9651	5.99496050
19	2024-05-09	Export Cotton	209.1548	238.7550	-29.6002	29.6002	-12.39772989
20	2024-05-16	Export Cotton	195.5658	204.0530	-8.4872	8.4872	-4.15931155
21	2024-05-23	Export Cotton	210.0658	172.1950	37.8708	37.8708	21.99297308
22	2024-05-30	Export Cotton	227.6207	156.9800	70.6407	70.6407	44.99980889
23	2024-06-06	Export Cotton	177.9094	186.6450	-8.7356	8.7356	-4.68032897
24	2024-06-13	Export Cotton	158.6321	197.9050	-39.2729	39.2729	-19.84431924
25	2024-06-20	Export Cotton	167.0639	141.0190	26.0449	26.0449	18.46907154
26	2024-06-27	Export Cotton	184.8061	175.7980	9.0081	9.0081	5.12411973
27	2024-07-04	Export Cotton	140.0600	150.3000	-10.2400	10.2400	-6.74511073

Figure 4-8 : Vue des prévisions enregistrées (Predicted Logs) dans VisionPlan

4.3.5.1 Structure de l'affichage

Le tableau présente plusieurs colonnes permettant une lecture rapide des prévisions :

- **Date** : date à laquelle la prévision est associée.
- **Component** : composant de production suivi, ici "Export Cotton".
- **Forecasted Value** : valeur prévue par le modèle.
- **Real Value** : valeur réelle enregistrée pour la même période.
- **Error** : différence brute entre la valeur prévue et la valeur réelle.
- **Abs Error** : valeur absolue de l'erreur, représentant l'écart sans tenir compte de la direction de l'erreur.
- **Error %** : pourcentage d'erreur relatif, calculé par rapport à la valeur réelle, permettant de quantifier l'ampleur de l'écart en termes proportionnels.

4.3.5.2 Recherche et filtres temporels

Un champ de recherche en haut permet de filtrer les entrées en fonction de critères textuels, tandis que les sélecteurs de dates sur la droite permettent de restreindre l'affichage à une période spécifique (par exemple, un trimestre ou une année fiscale). L'utilisateur peut également ajuster la granularité des prévisions avec le **Forecast Mode** (journalier, hebdomadaire, mensuel), offrant ainsi une vue adaptée aux besoins d'analyse.

4.3.5.3 Ergonomie et navigation

L'interface est optimisée pour une lecture fluide : les valeurs sont triées chronologiquement, et une barre de défilement permet d'explorer rapidement les différents enregistrements. L'affichage tabulaire épuré, couplé à un contraste visuel marqué, améliore la lisibilité des informations critiques.

4.3.5.4 Analyse et utilisation opérationnelle

La vue dédiée à l'analyse et à l'utilisation opérationnelle joue un rôle clé dans l'optimisation des prévisions. Elle permet d'identifier rapidement les prévisions sous-optimales ainsi que les écarts récurrents, tout en évaluant les performances des modèles à l'aide de métriques d'erreur telles que le WAPE, le MAE et le MASE. Cette interface offre également la possibilité d'ajuster les paramètres des modèles lors des itérations bi-hebdomadaires, en se basant sur l'analyse des erreurs accumulées. La présentation simultanée des valeurs prédites et observées facilite une analyse immédiate de la qualité des prévisions, positionnant ce module comme un élément central dans l'optimisation continue du pipeline de prévision.

4.3.6 Forecast Engine

Le module "Forecast Engine" constitue le cœur du système de prévision. C'est à cet endroit que l'utilisateur peut lancer les calculs, sélectionner les modèles à utiliser (SARIMA, Prophet, XGBoost), ajuster les paramètres, et définir les fenêtres temporelles. Cette architecture modulaire permet une personnalisation poussée des scénarios de prévision.

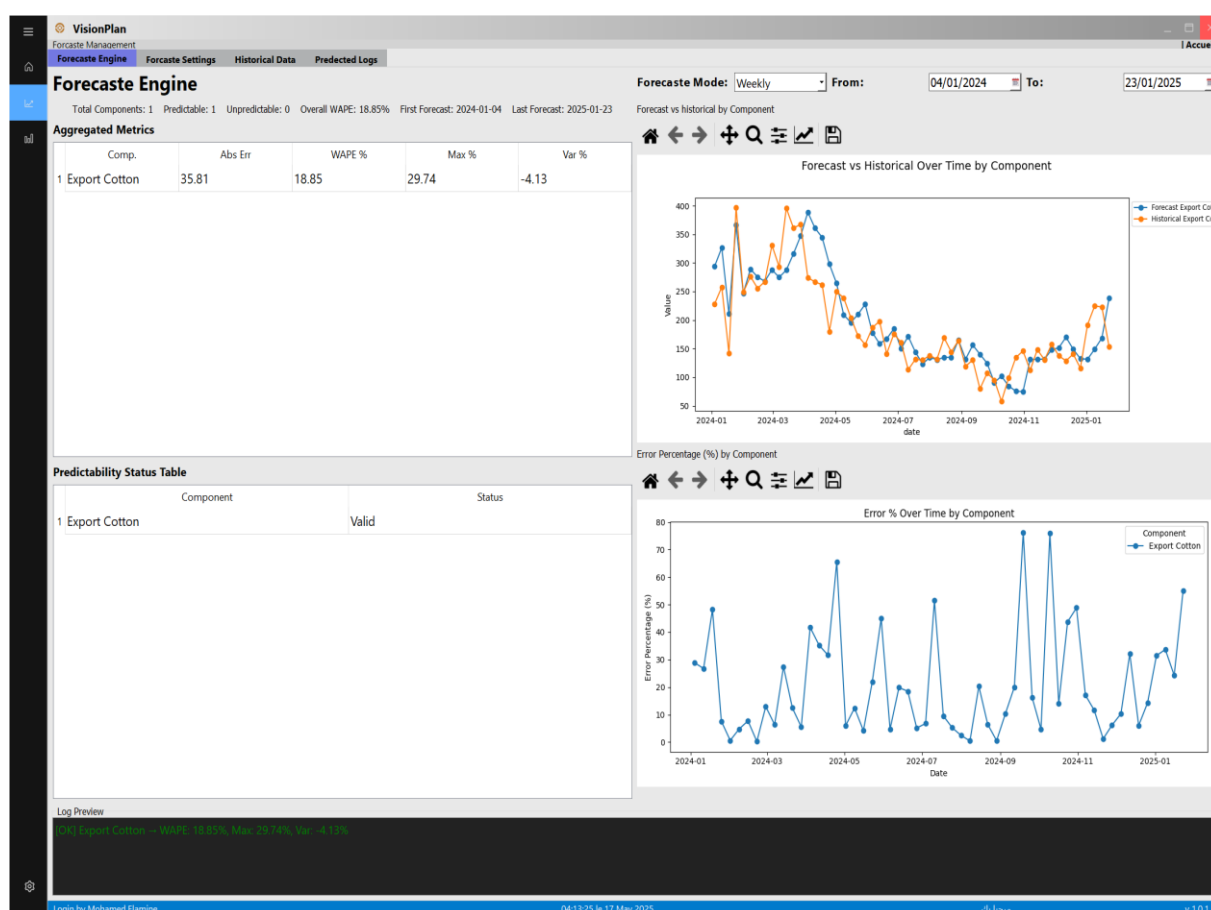


Figure 4-9: Interface Forecast Engine dans VisionPlan

4.3.6.1 Vue globale et métriques agrégées

La partie supérieure de l'interface affiche un résumé des composants analysés :

- **Total Components** : nombre total de composants suivis (ici, un seul : "Export Cotton").
- **Predictable et Unpredictable** : classification des composants en fonction de la fiabilité de la prévision.
- **Overall WAPE** : erreur absolue pondérée moyenne, ici estimée à 18.85 %, indiquant une performance de prévision relativement précise.
- **First Forecast et Last Forecast** : période couverte par les prévisions, de 2024-01-04 à 2025-01-23.

Dans le tableau des **Aggregated Metrics**, les indicateurs clés sont présentées :

- **Abs Err** : erreur absolue moyenne sur l'ensemble des prévisions.
- **WAPE %** : Weighted Absolute Percentage Error, principal critère d'évaluation de la performance.
- **Max %** : erreur maximale constatée sur la période.
- **Var %** : variation de l'erreur en pourcentage.

4.3.6.2 Comparaison des prévisions et des données historiques

Le graphique en haut à droite, intitulé « Forecast vs Historical Over Time by Component », représente les prévisions (Forecast Export Cotton) en bleu et les valeurs réelles (Historical Export Cotton) en orange. Cette visualisation permet d'évaluer la précision du modèle à suivre les tendances et à capter les variations saisonnières ou les ruptures de stock.

4.3.6.3 Analyse de l'erreur temporelle

Le graphique situé en bas à droite, « Error % Over Time by Component », affiche l'évolution du pourcentage d'erreur au fil du temps. Cette représentation temporelle facilite l'identification des périodes critiques où l'écart entre prévision et réalité est le plus important, permettant ainsi de cibler les ajustements nécessaires sur le modèle.

4.3.6.4 Statut de prédictibilité

Le tableau Predictability Status Table, placé en bas à gauche, indique l'état de prédictibilité du composant analysé. Si un composant marqué comme Valid, ce qui signifie que les prévisions sont considérées comme fiables pour cette période.

4.3.6.5 Log Preview

En bas de l'interface, un aperçu de log en temps réel affiche les informations clés de validation du modèle, notamment le WAPE, la variation maximale observée, ainsi que l'état général du calcul. Cette interface modulaire permet une supervision complète du processus de prévision, combinant analyse visuelle, validation statistique et diagnostic automatisé. Elle constitue un outil essentiel pour les opérateurs cherchant à optimiser la planification industrielle sur la base de données historiques et de projections fiables.

4.3.7 Forecast Setting

Cette dernière section regroupe l'ensemble des paramètres de configuration du moteur de prévision : choix de la granularité temporelle (journalier, hebdomadaire, mensuel), définition des fenêtres glissantes, gestion des hyperparamètres optimisés (Optuna) et ajustements des fréquences de mise à jour.

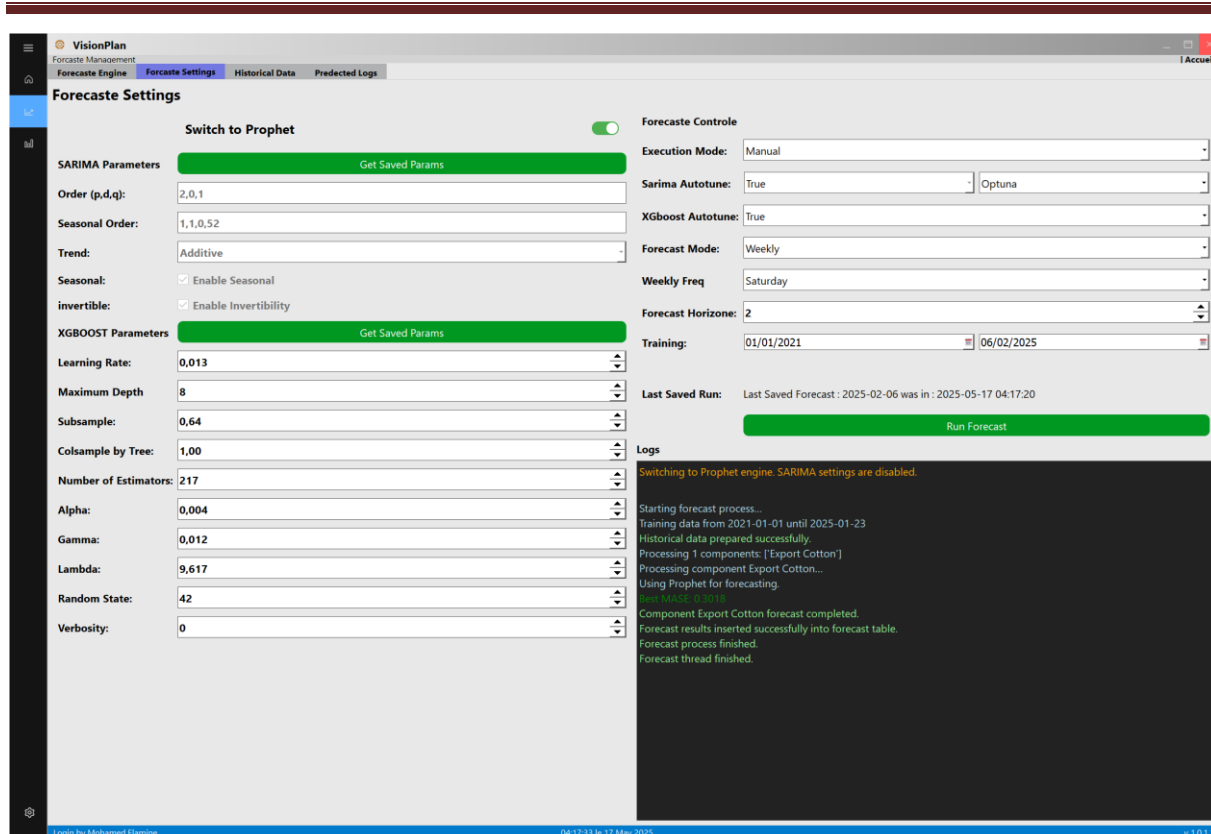


Figure 4-10: Interface Forecast Settings dans VisionPlan

4.3.7.1 Paramétrage des modèles (SARIMA, Prophet, XGBoost)

La partie gauche de l'interface est dédiée à l'ajustement des hyperparamètres :

- **SARIMA Parameters** : l'utilisateur peut définir les paramètres de la série temporelle (p, d, q) ainsi que l'ordre saisonnier (P, D, Q, m). Le choix de la tendance est également disponible (Additive ou Multiplicative), avec la possibilité d'activer les propriétés de saisonnalité (Enable Seasonal) et d'inversibilité (Enable Invertibility).
- **Prophet Mode** : un bouton Switch to Prophet permet de basculer directement vers Prophet, désactivant automatiquement les réglages SARIMA.
- **XGBoost Parameters** : les hyperparamètres de l'algorithme XGBoost sont accessibles, incluant le taux d'apprentissage (Learning Rate), la profondeur maximale des arbres (Maximum Depth), le sous-échantillonnage (Subsample), le nombre d'estimateurs (Number of Estimators), ainsi que les régularisations (Alpha, Gamma, Lambda). Un bouton Get Saved Params permet de récupérer les dernières configurations enregistrées.

4.3.7.2 Contrôle d'exécution et configuration générale

À droite, les options de contrôle permettent de définir les paramètres d'exécution :

- **Execution Mode** : choix entre un mode manuel ou automatisé.
- **Sarima Autotune et XGBoost Autotune** : activation de l'optimisation automatisée des hyperparamètres via Optuna.

- **Forecast Mode** : définition de la granularité temporelle (journalier, hebdomadaire, mensuel).
- **Weekly Freq** : sélection du jour de génération des prévisions.
- **Forecast Horizon** : spécification de l'horizon de prévision en semaines.
- **Training** : sélection de la période de formation du modèle.
- **Last Saved Run** : affichage de la dernière exécution sauvegardée, avec date et heure précises.

4.3.7.3 Journaux d'exécution (Logs)

La section inférieure droite affiche un historique des opérations exécutées :

- Le passage de SARIMA à Prophet est indiqué en orange ou l'inverse, avec la désactivation des paramètres SARIMA/Prophet.
- Les étapes de traitement sont listées chronologiquement : préparation des données, lancement du processus, génération des prévisions.
- Le log signale également le meilleur score MASE obtenu lors de l'exécution, ainsi que la confirmation de l'insertion des résultats dans les bases de données.

4.3.7.4 Bouton d'exécution : Run Forecast

Le bouton Run Forecast, de couleur verte, déclenche le processus de prévision avec les paramètres définis. Ce lancement est asynchrone, permettant à l'utilisateur de suivre l'évolution directement dans la console de logs. Cette interface modulaire permet une configuration fine des modèles, un suivi transparent des opérations, et une optimisation automatique des hyperparamètres pour un meilleur ajustement des prévisions. Elle constitue un pilier stratégique pour la planification industrielle dans VisionPlan.

4.3.8 Vue d'Analyse des Données (Data Analysis View)

Cette interface représente le module Data Analysis de l'application VisionPlan, conçu pour l'exploration et l'analyse détaillée des séries temporelles. Elle se décompose en plusieurs sections interactives permettant de visualiser l'évolution des composants, d'identifier les tendances, et d'analyser les cycles saisonniers.

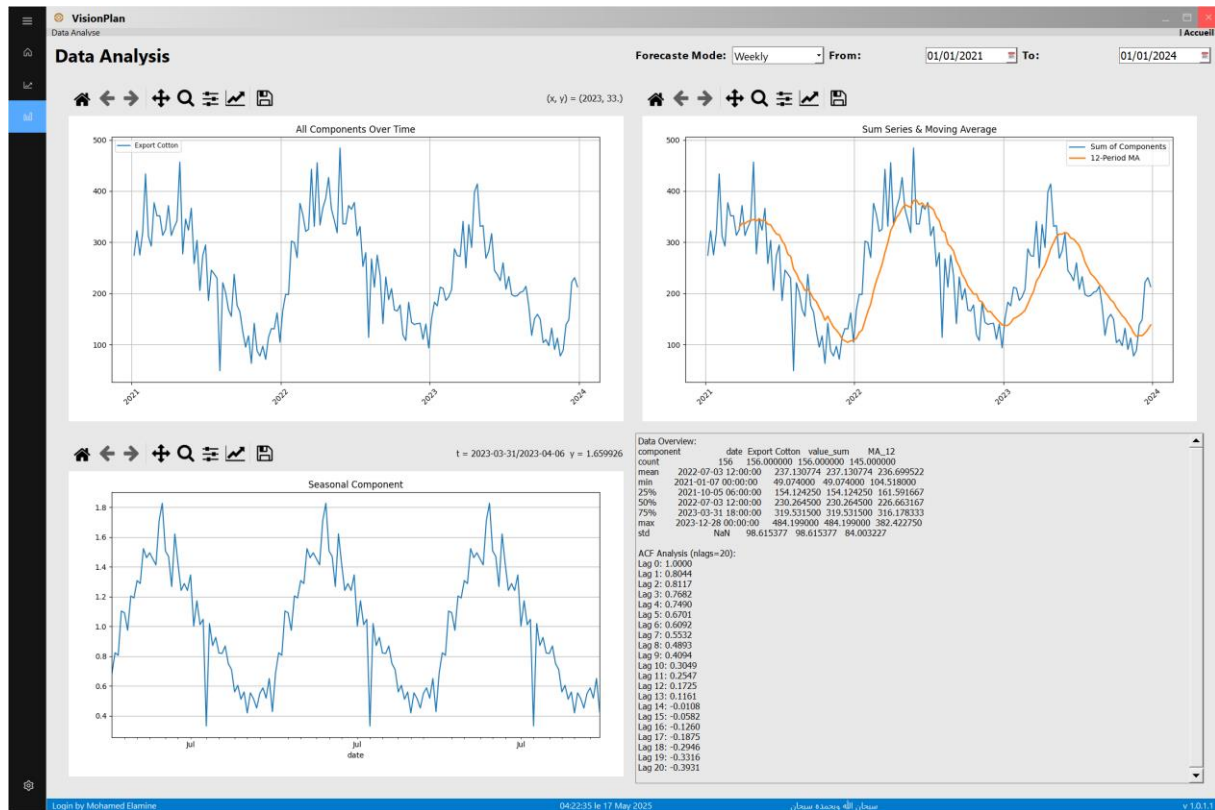


Figure 4-11: Vue d'Analyse Temporelle des Données dans VisionPlan

4.3.8.1 Visualisation des séries temporelles (All Components Over Time)

Le graphique situé en haut à gauche affiche l'évolution historique du composant Export Cotton sur la période de 2021 à 2024. La courbe bleue représente les variations hebdomadaires de production et d'exportation, offrant une vue d'ensemble sur les fluctuations. Cette représentation permet de détecter les pics de production, les creux saisonniers, ainsi que les tendances à long terme.

4.3.8.2 Séries temporelles avec moyenne mobile (Sum Series & Moving Average)

Le graphique en haut à droite présente deux courbes :

- **Sum of Components (en bleu)** : la série temporelle brute des volumes exportés.
- **12-Period MA (en orange)** : la moyenne mobile sur 12 périodes, permettant de lisser les variations hebdomadaires et de dégager une tendance plus claire.

Cette superposition permet de distinguer les mouvements cycliques et d'anticiper les périodes de hausse ou de baisse. Lissage idéal pour capturer les tendances sous-jacentes malgré la volatilité quotidienne.

4.3.8.3 Décomposition saisonnière (Seasonal Component)

La partie inférieure de l'interface présente la composante saisonnière isolée, révélant les schémas récurrents dans les données. Les pics réguliers chaque année suggèrent une cyclicité marquée, possiblement liée aux périodes de récolte ou de forte demande internationale. L'utilisateur peut ainsi identifier les phases où la production atteint ses maximums ou minimums avec une précision temporelle.

4.3.8.4 Analyse statistique des données (Data Overview & ACF Analysis)

Dans la partie droite, deux analyses complémentaires sont affichées :

- Data Overview : un tableau récapitulatif des statistiques descriptives de la série temporelle, incluant la moyenne, les quartiles, le minimum, le maximum, et le lissage par moyenne mobile (MA_12).
- ACF Analysis (lags=20) : l'analyse de la fonction d'autocorrélation jusqu'à 20 décalages, permettant de mesurer les relations temporelles entre les valeurs successives de la série. Cette information est cruciale pour l'identification de structures saisonnières ou de dépendances temporelles.

4.3.8.5 Contrôles interactifs

L'interface intègre des boutons de navigation et de zoom permettant d'explorer les données en détail :

- Navigation temporelle (flèches de gauche et droite) pour se déplacer sur la série historique.
- Zoom et filtre pour un focus sur une période spécifique.
- Exportation sous différents formats (CSV, PNG) pour l'analyse externe.

Cette interface modulaire et interactive facilite l'analyse exploratoire, offrant à l'utilisateur une compréhension approfondie des cycles saisonniers, des tendances à long terme et des dépendances temporelles. Elle constitue un socle essentiel pour le calibrage des modèles de prévision dans le cadre industriel.

4.4 Procédures d'Installation et de Déploiement

4.4.1 Prérequis système :

- **OS** : Windows 10 (64 bits)
- **Mémoire** : 8 Go de RAM (16 Go recommandé)
- **Processeur** : Dual-core 2 GHz (Quad-core recommandé)
- **Stockage** : 600 Mo sur SSD

4.4.2 Installation des dépendances :

- Clonage du dépôt et configuration de la base de données MySQL
- Exécution du script d'installation et initialisation de l'application
- Démarrage de l'application et accès à l'écran de connexion

Les stratégies de déploiement incluent une installation locale pour un usage individuel et un déploiement sur serveur pour un usage multi-utilisateur. Le support Docker est prévu pour des environnements conteneurisés.

4.5 Lignes Directrices pour la Maintenance et l'Évolution

4.5.1 Sauvegardes régulières

Il est essentiel de mettre en place un processus de sauvegarde régulier pour garantir la récupération des données en cas de problème. Les sauvegardes doivent inclure deux aspects principaux :

- La base de données MySQL : un plan de sauvegarde quotidien doit être mis en place pour capturer l'état actuel de toutes les tables. Ces sauvegardes doivent être stockées de manière sécurisée et, idéalement, dans plusieurs emplacements (local et cloud).
- Les modèles entraînés : les fichiers de modèles (SARIMA, Prophet, XGBoost) doivent également être sauvegardés après chaque entraînement ou mise à jour. Cela permet de restaurer les prédictions à un état précédent si nécessaire.

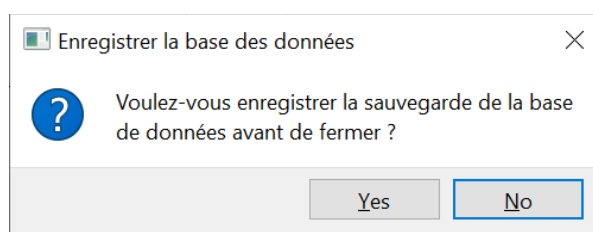


Figure 4-12:Fenêtre de confirmation de sauvegarde dans VisionPlan

Le message affiché, « Voulez-vous enregistrer la sauvegarde de la base de données avant de fermer ? », est clair et direct, indiquant à l'utilisateur l'importance de l'action. Cette démarche préventive permet de réduire les risques de perte d'informations critiques, notamment les historiques de prévisions, les logs d'erreurs, et les configurations de modèles.

4.5.2 Surveillance des modèles

Le suivi des performances des modèles est crucial pour garantir leur efficacité dans le temps. Les indicateurs comme le MAE, le RMSE et le WAPE doivent être monitorés de manière continue. Si des dérives de performance sont détectées, une réévaluation du modèle doit être planifiée. Un rapport mensuel peut également être généré pour évaluer les écarts entre les prévisions et les valeurs réelles.

4.5.3 Documentation et mises à jour

Chaque modification apportée au projet doit être clairement documentée. Le fichier README.md doit répertorier toutes les évolutions, les corrections et les améliorations. De plus, les fichiers de configuration doivent être mis à jour pour refléter les changements, notamment ceux liés aux paramètres des modèles ou à la connexion base de données.

4.6 Perspectives : Vers une Intégration Future de la Planification dans VisionPlan

Dans ce projet de fin d'études, notre objectif initial était de travailler à la fois sur la prévision de la demande et sur la planification de la production à l'aide de techniques d'intelligence artificielle. Cependant, au fil de l'avancement du projet, plusieurs contraintes majeures nous ont amenés à réduire le périmètre fonctionnel du travail final, en nous concentrant exclusivement sur la partie prédictive.

La raison principale de cette limitation est le temps restreint imposé par le calendrier académique. Par ailleurs, une part importante de nos efforts a été consacrée à la conception du BMC (Business Model Canvas) détaillé, exigée dans le cadre du mémoire, ainsi qu'au développement du logiciel VisionPlan, qui a nécessité une attention technique et fonctionnelle considérable.

En résumé, le choix de ne pas aborder la planification ne résulte pas d'un manque d'intérêt, mais d'une volonté de livrer un système robuste et abouti sur l'aspect prévisionnel, tout en laissant l'opportunité d'un enrichissement futur du logiciel vers la planification.

4.6.1 Grossissement sur notre travail de prévision : une base solide

Malgré ce recentrage, le travail réalisé sur la prévision constitue un socle très riche et hautement exploitable pour la suite. Nous avons développé une solution intelligente, VisionPlan, qui intègre :

- Des modèles statistiques classiques comme SARIMA et Prophet pour modéliser les tendances et les saisonnalités.
- Des algorithmes de machine learning (XGBoost) pour modéliser les résidus non-linéaires.
- Une optimisation automatique des hyperparamètres avec Optuna (optimisation bayésienne).
- Une architecture de prévision hybride robuste, mise à jour toutes les deux semaines.
- Une interface logicielle complète permettant l'analyse, la visualisation, la correction des erreurs et l'interprétation via un assistant IA.

Cette infrastructure modulaire et dynamique est parfaitement adaptée à une extension vers la planification dans une prochaine version.

4.6.2 Vers une planification intelligente et intégrée

La mise en place d'un module de planification dans VisionPlan représente la continuité naturelle du travail accompli sur la prévision. En effet, dans un contexte industriel, prévoir la demande est un préalable indispensable, mais non suffisant. Il est essentiel de transformer ces prévisions en un plan d'action structuré, capable de coordonner les ressources, de respecter les délais et d'optimiser les opérations. C'est là qu'intervient la planification.

L'ajout de cette fonctionnalité permettrait à VisionPlan de générer automatiquement des plannings de production adaptés, prenant en compte les capacités disponibles, les contraintes techniques et les priorités. Un tel module aiderait les responsables à mieux organiser les lancements de production, à éviter les conflits de calendrier et à renforcer la fluidité des opérations en atelier.

Dans ce cadre, l'utilisation d'outils d'intelligence artificielle serait pertinente pour améliorer la qualité de la planification. Comme pour la prévision, l'IA pourrait proposer des plannings ajustés en temps réel en cas d'imprévu, comme une variation soudaine de la demande ou une indisponibilité d'équipement. Ces ajustements automatiques contribueraient à une meilleure prise de décision et à une réactivité accrue face aux aléas.

La planification permet également de mieux anticiper les besoins en approvisionnement. Connaître les quantités à produire permet de déterminer à l'avance les quantités de matières premières à commander, d'éviter les ruptures ou les excédents, et de synchroniser les achats avec le rythme de la production. Cela améliore la coordination entre les différents services de l'entreprise, notamment entre la production et les achats.

Un autre effet positif de la planification est son impact direct sur la gestion des stocks. En ajustant les volumes produits aux besoins réels, elle contribue à limiter les surstocks, à réduire les coûts de stockage et à éviter les pénuries. VisionPlan pourrait ainsi évoluer vers un véritable outil de pilotage global, réunissant prévision, planification, approvisionnement et gestion de stock dans un environnement unifié.

En résumé, l'intégration de la planification dans VisionPlan représenterait une avancée stratégique. Elle permettrait non seulement d'étendre les fonctionnalités de la solution, mais aussi de renforcer son utilité pour les industriels, en facilitant une exécution plus proactive, réactive et coordonnée des opérations de production.

4.6.3 Synthèse des perspectives d'évolution

L'enrichissement de VisionPlan par un module de planification ouvre ainsi des perspectives intéressantes pour les prochaines phases de développement. Cela s'inscrit pleinement dans l'objectif de créer une solution intelligente, complète et orientée vers les besoins réels des entreprises industrielles.

4.7 Conclusion

Ce chapitre a présenté en détail l'architecture, les modules fonctionnels et le processus de déploiement de VisionPlan. Il en ressort une solution construite autour d'une approche modulaire, alliant performance technique et flexibilité d'intégration. Cette structure garantit une gestion efficace des prévisions industrielles, tout en assurant la fiabilité, la maintenabilité et l'évolutivité de l'outil. VisionPlan se positionne ainsi comme une base solide, apte à accueillir de futures extensions, notamment en matière de planification.

5. Conclusion générale

L'étude menée tout au long de ce mémoire a montré que l'alliance entre les modèles statistiques classiques et les algorithmes de machine learning, dans une logique d'intégration de l'intelligence artificielle, constitue un levier décisif pour améliorer la qualité des prévisions dans un contexte industriel complexe. L'immersion dans le fonctionnement de TAYAL SPA a d'abord permis d'identifier les déséquilibres récurrents entre production et demande, dus à l'usage de méthodes linéaires insuffisamment réactives face aux fluctuations et aux chocs exogènes. En parallèle, l'analyse théorique a mis en lumière la complémentarité des approches : tandis que SARIMA, Holt–Winters et Prophet modélisent efficacement les tendances et la saisonnalité, XGBoost, optimisé par Optuna, excelle dans la captation des non-linéarités résiduelles grâce à ses capacités d'apprentissage issues de l'IA.

La mise en œuvre expérimentale, déclinée selon un protocole bi-hebdomadaire d'apprentissage en ligne, a confirmé ces synergies. D'une part, les résidus issus des modèles linéaires ont été exploités pour enrichir l'apprentissage de XGBoost, réduisant significativement le WAPE et le RMSE sur les données de production interne. D'autre part, l'application aux séries USDA a illustré la robustesse de la méthodologie face à une saisonnalité régulière et à de faibles anomalies ponctuelles. Ces résultats témoignent de gains de précision systématiques, validant la pertinence de l'hybridation pour répondre aux exigences de réactivité et de fiabilité propres au secteur textile.

Au-delà de l'amélioration quantitative des prévisions, le prototype VisionPlan constitue une première réalisation opérationnelle de cette approche. Son architecture modulaire, son interface conviviale et ses fonctionnalités d'optimisation bayésienne, issues des avancées récentes en IA, offrent un cadre complet pour industrialiser la préparation des données, la sélection des modèles et le suivi continu des performances. En plaçant l'utilisateur au cœur du processus grâce à un assistant conversationnel intelligent et des indicateurs clairs, l'outil favorise l'adoption et la transmissibilité des bonnes pratiques analytiques.

Néanmoins, plusieurs pistes d'enrichissement restent à explorer. L'intégration de variables exogènes — prix des matières premières, données météorologiques ou signaux commerciaux — permettrait d'affiner davantage les prévisions. De même, l'expérimentation d'autres méta-apprenants ou de stratégies d'entraînement adaptatif à fréquence variable offrirait des perspectives pour optimiser le compromis entre coût computationnel et précision. Enfin, un déploiement à plus grande échelle, incluant un suivi en temps réel et des alertes automatisées basées sur l'IA, renforcerait la valeur ajoutée opérationnelle pour les décideurs.

En définitive, ce travail démontre que la conjugaison structurée de méthodes statistiques éprouvées, de techniques de machine learning et des outils issus de l'intelligence artificielle ouvre de nouvelles voies pour la planification et la gestion des stocks dans l'industrie textile. Elle offre un cadre à la fois rigoureux et flexible, capable de répondre aux défis actuels de volatilité et de durabilité, et constitue une base solide pour de futurs développements au service de la compétitivité industrielle.

REFERENCES

6. Références

- Akiba, T., & al. (2019). *Optuna: A Next-generation Hyperparameter Optimization Framework*. arXiv.
- APS. (2022). *Textile et cuir : Recul sensible des importations de l'Algérie en 2022*. Alger, Algérie: Algérie Presse Service.
- APS. (2023). *Industrie : hausse de 4,5 % de la production du secteur public en 2021*. Alger, Algérie: Agence Presse Service.
- Bergstra, J., Bardenet, R., & Bengio, Y. &. (2011). Algorithms for Hyper-Parameter Optimization. *Proceedings of the 24th International Conference on Neural Information Processing Systems (NIPS 2011)*. Granada, Spain.
- Box, G. E., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day.
- Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control (4th ed.)*. Hoboken, NJ: Wiley.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control (5th ed.)*. Hoboken, NJ: Wiley.
- Brockwell, P. J., & Davis, R. A. (2002). *Introduction to Time Series and Forecasting (2e éd.)*. New York, NY: Springer.
- Brownlee, J. (2018). *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python*. Vermont, Australia: Machine Learning Mastery.
- CAWTAR. (2020). *L'industrie textile en Algérie : Dépendance aux importations et perspectives de développement*. Tunisie: Centre de la Femme Arabe pour la Formation et la Recherche.
- Chandrakant, R., & Gupta, S. (2019). *Textile Supply Chain Management and Logistics*. Amsterdam : Elsevier.
- Charpentier, A. (2012). *Modèles de prévision : Séries temporelles*. Paris: Dunod.
- 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* (pp. 785–794). San Francisco, CA: ACM.
- Chollet, F. (2018). *Deep Learning with Python*. Shelter Island, NY: Manning Publications.
- Chopra, S., & Meindl, P. (2016). *Supply Chain Management: Strategy, Planning, and Operation*. New York, NY: Pearson.
- Christopher, M. (2016). *Logistics & Supply Chain Management (5th ed.)*. Harlow, UK: Pearson.
- Corbman, B. P. (1985). *Textiles: Fiber to Fabric*. New York, NY: McGraw-Hill.

- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, pp. 253–263.
- Expo, T. (n.d.). *L'industrie textile en Algérie : Opportunités et défis*. Alger: Textyle Expo.
- Gilleland, E. (2013). Testing forecast performance: Some newer approaches. *Meteorological Applications*, pp. 407–417.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press.
- Hewamalage, H. A. (2022). *Forecast Evaluation for Data Scientists: Common Pitfalls and Best Practices*. arXiv.
- Hochreiter, S. &. (1997). Long short-term memory. *Neural Computation* , 1735–1780.
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice (3rd ed.)*. Melbourne, Australia: OTexts.
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice (3rd ed.)*. Melbourne, Australia: OTexts.
- Hyndman, R., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. Melbourne, Australia: OTexts.
- INSEE. (2019). *L'industrie textile en France : une production mondialisée, sauf pour les produits de luxe et les textiles techniques*. Paris: INSEE.
- Kapočiūtė-Dzikienė, J., Balodis, K., & Skadiņš, R. (2020). Intent Detection Problem Solving via Automatic DNN Hyperparameter Optimization. *Applied Sciences*, p. 7256.
- Khashei, M. &. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*, 2664–2675.
- Kilduff, P. (2005). *The Global Textile and Clothing Industry: Supply Chain and Business Fundamentals*. Cambridge, UK: Woodhead Publishing.
- Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2017). *Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization*. arXiv.
- Makridakis, S. S. (2019). The M4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 605–617.
- Mian, S., & Khan, M. (2021). *The Values of AIC & BIC for Different ARIMA Models*.
- Montgomery, D. C. (2015). *Introduction to Time Series Analysis and Forecasting (2nd ed.)*. Hoboken, NJ: John Wiley & Sons.
- Nielsen, A. (2019). *Practical Time Series Analysis: Prediction with Statistics and Machine Learning*. Sebastopol, CA: O'Reilly Media.
- Odhiambo, S. O., Nyakundi, C., & Waititu, H. (2024). Developing a Hybrid ARIMA-XGBOOST Model for Analysing Mobile Money Transaction Data in Kenya. *Asian Journal of Probability and Statistics*, pp. 0–9734.
- ONS. (2021). *Activité industrielle en Algérie 2012-2021*. Alger, Algérie: Office National des Statistiques.

- Preeti. (2024). *Hybrid classifiers: Time Series Forecasting*.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, pp. 206–215.
- Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples (4th ed.)*. New York, NY: Springer.
- Simchi-Levi, D. K.-L. (2014). *Designing and Managing the Supply Chain: Concepts, Strategies, and Case Studies*. New York, NY: McGraw-Hill.
- Slack, D. H. (2020). Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, (pp. pp. 180–186).
- Taylor, S. J., & Letham, B. (2017). *Forecasting at Scale*. Menlo Park, CA: Facebook Research.
- Thejovathi, & al. (2024). Optimizing Product Demand Forecasting with Hybrid Machine Learning and Time Series Models: A Comparative Analysis of XGBoost and SARIMA. *Proceedings of the 3rd International Conference on Optimization Techniques in the Field of Engineering (ICOFE-2024)*. SSRN.
- Tsay, R. S. (2010). *Analysis of Financial Time Series (3rd ed.)*. Hoboken, NJ: Wiley.
- Wasi, A. T. (2022). *Time Series Analysis & Forecasting*. New York, NY: Wiley.
- Zeng, & al. (2025). Short-Term Load Forecasting in Power Systems Based on the Prophet–BO–XGBoost Model. *Energies*, p. 227.
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, pp. 159–175.
- Zhang, Y., & Li, M. L. (2023). An Enhanced XGBoost-Based Approach for Forecasting Renewable Energy Generation. *Energies*, 18(2), p. 227.
- Zhang, Y., Li, M., & Liu, Y. (2023). An Enhanced XGBoost-Based Approach for Forecasting Renewable Energy Generation. *Energies*, 18(2), p. 227.
- Zheng, Y. L., & Zhu, X. G. (2020). A comparative study of two methods to predict the incidence of hepatitis B in Guangxi, China. *PLOS ONE*, p. e0236400.



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

Université Abou Bekr Belkaid Tlemcen

Business Model Canvas

BMC

N° de projet : FT-051

Faculté/Institut : Université de Tlemcen

Département : Génie Industriel

Nom du projet : Investigation autour de la prédiction en utilisant l'IA dans un système de production (VisionPlan)

Encadrant 1 : **BETAOUAF Hichem**

Encadrant 2 : **BENNEKROUF Mohammed**

Etudiants :

- **BELAZREG** Mohamed Elamine
- **BOUHALIS** Sihem

Année universitaire : 2024/2025

1. Proposition de valeur (Value Proposition) القيمة المقترحة

1.1. Quels problèmes résolvons-nous pour nos clients ?

ما هي المشاكل التي نحلها لعملائنا ؟

- Absence d'intégration de solutions IA dans l'industrie algérienne.
- Processus Excel manuels obsolètes, lents et sujets aux erreurs humaines.
- Déconnectés des données en temps réel.
- Erreurs de prévision générant 10–20 % de surstocks ou ruptures.
- Dépendance à des experts externes pour l'analyse prédictive.
- Absence de visibilité temps réel sur la production et la logistique.

1.2. Quels besoins de nos clients satisfont nos produits ou services ?

ما هي الاحتياجات التي يلبها منتجاتنا أو خدماتنا لعملائنا؟

- Besoin de **prévision fiable**, rapide, sans expertise technique.
- Besoin de **réduction des pertes** (stocks, retards, approvisionnements).
- Besoin de **visibilité immédiate** sur la production, les ventes et la logistique.
- Respect des normes algériennes (données hébergées en Algérie).

1.3. En quoi notre offre est-elle différente de celle de nos concurrents ?

في ماذا تختلف عروضنا عن تلك التي يقدمها منافسوننا؟

- Hébergement 100 % national
- Prix jusqu'à -60 % inférieur aux solutions étrangères.
- Configuration sur-mesure rapide.
- Évolutivité sectorielle : adaptation native aux spécificités énergie, manufacturier, logistique
- Intégration de L'IA

1.4. Quelles est notre proposition unique de valeur ?

ما هو العرض الفريد للقيمة لدينا؟

- VisionPlan = une plateforme de prévision IA prête à l'emploi, développée en Algérie, permettant aux entreprises de gagner jusqu'à 5M DZD/an, avec un retour sur investissement dès la première année.

2. Segments de clients (Customer Segment) انواع العملاء :

2.1. Quels sont nos clients principaux ?

من هم العملاء او الزبائن الرئيسيون؟

a. Désir et Pouvoir d'Engager avec Nous

Grands Groupes Industriels (Énergie, Pétrochimie, Manufacturier, Logistique) :

- Fort intérêt pour l'optimisation de la production la digitalisation industrielle.
- Capacité budgétaire élevée.

b. Désir sans Pouvoir d'Engager avec Nous

PME du Secteur Industriel et Logistique

- Besoin évident d'automatisation et de réduction des coûts, mais contraintes budgétaires limitant l'engagement.
- Intérêt manifeste pour les prévisions optimisées, mais difficulté à financer les solutions IA sans soutien étatique ou subventions.

c. Pouvoir sans Désir d'Engager avec Nous

Institutions Publiques (Ministères, Agences Gouvernementales)

- Capacité financière importante, subventions d'État, accès à des fonds de développement.
- Cependant, faible réactivité et inertie bureaucratique ralentissant l'adoption de solutions IA avancées.
- Souvent orientées vers des projets à long terme avec peu de flexibilité.

2.2. Quels sont les différents segments de clients que nous visons ?

ما هي الفئات المختلفة من العملاء التي تستهدفها؟

- **B2B : industriels, fabricants, distributeurs.**
- **B2G : secteur public économique.**
- **PME : avec besoin de solutions simples, à coût maîtrisé.**

2.3. Quels sont les besoins spécifiques de chaque segment de clients ?

ما هي الاحتياجات الخاصة لكل فئة من العملاء؟

Segment	Besoins spécifiques
Grands groupes	Forecast détaillé, support dédié, intégration métier
PME	Version simplifiée, paiement échelonné, installation rapide
Public	Tableaux d'impact, conformité, rapports automatisés

2.4. Comment pouvons-nous catégoriser nos clients en groupes distincts ?

كيف يمكن تصنيف عملائنا الى مجموعات مختلفة؟

Groupe	Critères
Grands comptes	Budget >10M DZD, cycle achat >6 mois
PME	Budget IT <2M DZD, décisions rapides
Institutions	Financement public, lenteur administrative

3. Relation avec les clients (Consumer Relationships) : علاقة مع العملاء

3.1. Quel type de relation chaque segment de clients attend il de nous ?

اي نوع من العلاقة يتوقعه كل فئة من العملاء منا؟

- Grands comptes : accompagnement dédié, visites terrain, support 24h.

- PME : hotline, chat, vidéos explicatives, installation guidée.
- Public : référent technique, rapports formels, validation POC.

3.2. Comment entretenons-nous actuellement les relations avec nos clients ?

كيف نحافظ حاليًا على العلاقات مع عملائنا؟

- Support technique par téléphone/mail
- Formations courtes (2 jours)
- Visites clients régulières (chaque trimestre)

3.3. Comment pouvons-nous améliorer ou personnaliser nos interactions avec nos clients ?

كيف يمكننا تحسين أو تخصيص تفاعلاتنا مع عملائنا؟

- Mise en place de portail client avec indicateurs personnalisés
- Enquêtes de satisfaction & tableau de feedback
- Club client (webinaires, idées, votes pour évolutions)

4. Canaux de distribution (Channels) قنوات التوزيع :

4.1. Par quels canaux nos clients veulent-ils être atteints ?

من خلال أي قنوات يفضل عملائنا أن يتم التواصل معهم؟

- Visites sur site
- Webinaires et démos en ligne
- Salons professionnels
- Réseaux professionnels (LinkedIn, forums B2B)

4.2. Quels canaux sont les plus efficaces pour atteindre chaque segment de clients ?

ما هي القنوات الأكثر فعالية للوصول إلى كل فئة من العملاء؟

- Diagnostic terrain pour comprendre les process et collecter les données réelles.
- Onboarding hybride : début sur site puis configuration et suivi en ligne.
- Forum institutionnel + lettre officielle

4.3. Comment pouvons-nous intégrer différents canaux pour améliorer l'expérience clients ?

كيف يمكننا دمج مختلف القنوات لتحسين تجربة العملاء؟

- Démarrage terrain → onboarding en ligne → suivi via portail client
- Tous les supports centralisés : vidéos, tutos, chat, FAQ, dashboard

5. Partenaires clés (Key Partnerships) : الشراكة الرئيسية :

5.1. Qui sont nos partenaires clés ?

من هم شركاؤنا الرئيسيون؟

- Fournisseur de services de cybersécurité local.
- Intégrateurs locaux pour installation.
- Prestataire de support technique externalisé.

5.2. Quels sont les partenariats qui nous aident à réduire les coûts, à accéder à de nouvelles ressources ou à améliorer notre proposition de valeur ?

ما هي الشراكات التي تساعدنا على خفض التكاليف أو الوصول إلى موارد جديدة أو تحسين قيمتنا المقترحة؟

- *Réduction des coûts d'hébergement*
- *Accès au marché via des partenaires locaux*
- *Expertise technique, juridique et formation externe*

5.3. Comment pouvons-nous aligner nos intérêts avec ceux de nos partenaires ?

كيف يمكننا مزامنة مصالحنا مع تلك لشركائنا؟

- Partage de revenus avec intégrateurs
- Co-financement de projets pilotes
- Accord volume avec hébergeurs pour baisse des prix

6. Activités clés (Key Activities) : الأنشطة الرئيسية

6.1. Quelles sont les actions principales que nous devons entreprendre pour livrer notre proposition de valeur ?

ما هي الأنشطة الرئيسية التي يجب علينا القيام بها لتقديم قيمتنا المقترحة؟

- Vente et déploiement de VisionPlan
- Formation client sur place
- Support technique + mises à jour IA

6.2. Quelles sont les opérations essentielles pour notre entreprise ?

ما هي العمليات الأساسية لشركتنا؟

- Génération de leads (PME & grands groupes)
- Réponse aux appels d'offres publics
- Suivi performance client (KPI de gains réalisés)

6.3. Quelles sont les activités qui créent le plus de valeur pour nos clients ?

ما هي الأنشطة التي تخلق أكبر قيمة لعملائنا؟

- Réduction d'erreurs de prévision ($\leq \pm 10\%$)
- Rapports automatiques de gains réalisés
- Personnalisation du tableau de bord client

7. Ressources clés (Key resources) : الموارد الرئيسية

7.1. Quels sont nos actifs matériels, immatériels et humains essentiels ?

ما هي الأصول المادية وغير المادية والبشرية الأساسية لدينا؟

Matériels :

- Matériel bureautique (ordinateurs, serveurs locaux, équipements de réseau)
- Véhicules légers pour les déplacements chez les clients.

Immatriels :

- Plateforme VisionPlan (100 % prête)
- 3 autres apps métiers prêtes (Stock, POS, Vente)
- Équipe fondatrice experte (IA, logistique)
- Base de données industrielle algérienne
- Infrastructure (serveurs, site web)

Humains :

- Développeur d'application & ingénieur de Chaîne Logistique
- Analyste théorique IA & Ingénieur Industriel
- Expert en optimisation industrielle (maître de conférences, spécialiste MILP et logistique inverse)
- HDR en informatique (expertise en conformité légale et cybersécurité)

7.2. Quels sont les outils, les technologies ou les partenariats dont nous avons besoin pour réussir ?

ما هي الأدوات والتكنولوجيا أو الشراكات التي نحتاجها لتحقيق النجاح؟

- Hébergement national.
- Outils CRM & support client
- Partenariat avec intégrateurs locaux.
- Collaboration avec centres de formation.
- Accès aux programmes de subvention et incubateurs startups.

7.3. Quels sont les principaux avantages concurrentiels de nos ressources ?

ما هي المزايا التنافسية الرئيسية لمواردنا؟

- VisionPlan Produit prêt → time-to-market = immédiat
- Coût d'exploitation très bas → marge élevée
- Développement 100 % algérien → indépendance & souveraineté
- Équipe pluridisciplinaire experte.
- Applications complémentaires prêtes à la vente.

8. Charges et coûts (Coste structure) : التكاليف

8.1. Quels sont les coûts fixes et variables associés à notre modèle économique ?

ما هي التكاليف الثابتة والمتغيرة المرتبطة بنموذجنا الاقتصادي؟

8.2. Masse salariale

	Salaire Base Mensuel	Salaire chargé Annuel
Ingénieur Data & IA	75 000,00	1 146 000
Développeur	75 000,00	1 146 000
Responsable Projet Technique	65 000,00	994 800
Administratif & Financier	65 000,00	994 800
Total	214 000,00	4 281 600,00

Tableau 2: Détail de la Masse Salariale Mensuelle et Annuelle par Poste (Salaires Chargés)

8.2.1. Investissements

Investissement	Désignation du matériel	Prix unitaire
	Laptops (High Performance)	350 000
	Desktops (Development Stations)	750 000
	Network Equipment (Router Switches)	100 000
	Office Furniture (Desks Chairs)	350 000
	Utility Vehicle (Client Visits)	3 000 000
	Screens (Dual Monitors)	250 000
	Printers and Scanners	140 000
	Sous-Total	4 940 000, 00

8.2.2. Charges externes

A. Coûts fixes

Charge	Coûts
Loyers	480 000,00
Assurance	100 000,00
Sous-traitance	120 000,00
Total	700 000,00

B. Coûts variables

	Année 01	Année 02	Année 03
Énergie/eau/gaz	60 000,00	63 000,00	66 000,00
Frais Marketing	360 000,00	450 000,00	570 000,00
Frais de déplacement et logistique	300 000,00	350 000,00	400 000,00
Frais télécom	120 000,00	126 000,00	132 300,00

Divers fournitures	150 000,00	157 500,00	165 375,00
Sécurité Informatique	150 000,00	165 000,00	181 500,00
Total	1 140 000,00	1 311 500,00	1 515 175,00

Coûts Total des charges = Coûts fixes + Coûts variables

Coûts Total des charges = 1 770 000 DZD

8.3. Quels sont les coûts les plus importants pour notre entreprise ?

ما هي التكاليف الأكثر أهمية لشركتنا؟

Achats directs	Prestation	Total
	Hébergement web et serveurs cloud	144 000,00
	Outils d'analyse et de gestion (CRM, ERP)	120 000,00
	Sécurité et cybersécurité	220 000,00
	Abonnements Logiciels et Outils SaaS	300 000,00
	Sous-Total	784 000

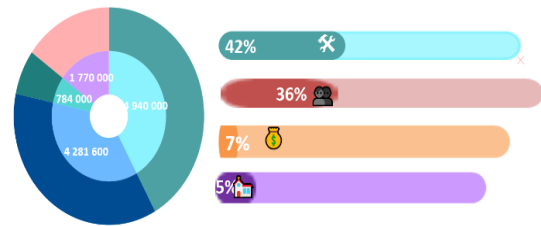
8.4. Comment pouvons-nous réduire les coûts ou améliorer l'efficacité de nos opérations ?

كيف يمكننا خفض التكاليف أو تحسين كفاءة عملياتنا؟

Catégorie

Montant

	Investissement	4 940 000
	Masse salariale	4 281 600
	Achats directs	784 000
	Charges externes	1 770 000
		11 775 600



8.4.1. Investissement (- 22%)

- Mutualisation de certains équipements avec des partenaires technologiques
- Leasing d'équipements au lieu d'achat direct pour lisser les dépenses d'investissement.

8.4.2. Achats directs

- Négociation de contrats d'achat groupés avec d'autres entreprises pour bénéficier de remises.

8.4.3. Charges externes

- Rationalisation des contrats de sous-traitance en privilégiant les fournisseurs locaux.
- Mise en concurrence systématique lors du renouvellement des contrats.

9. Revenus (Revenue): مصادر الدخل

9.1. Quels produits ou services nos clients sont-ils prêts à payer ?

ما هي المنتجات أو الخدمات التي يكون عملاؤنا على استعداد لدفع ثمنها؟

Chiffre d'affaires

	Année 01	Année 02	Année 03
VisionPlan (licences/année)	4 000 000,00	6 400 000,00	9 600 000,00
logiciels complémentaire	625 000 ,00	1 280 000,00	2 310 000,00
Services annexes (reporting, analyse)	460 000,00	1 170 000,00	3 000 000,00
Total	5 085 000,00	8 850 000,00	14 910 000

Tableau 3:Projection du Chiffre d'Affaires par Produit/Service

9.2. Quels sont les différents moyens par lesquels nous pouvons générer des revenus ?

ما هي الطرق المختلفة التي يمكننا من خلالها تحقيق الدخل؟

- VisionPlan (AI Forecasting) : licence ou abonnement dédié à la plateforme de prévision
- Abonnements SaaS mensuels ou annuels selon le niveau de fonctionnalités
- Facturation à l'usage (nombre de prédictions, volume de données traité)
- Licences on-premise avec maintenance et mises à jour incluses
- Prestation de services (conseil stratégique, accompagnement opérationnel)
- Services de développement (personnalisation, développement de fonctionnalités spécifiques)
- Data Analytics & Reporting (vente de rapports sur mesure et tableaux de bord avancés)
- Solutions prêtes à l'emploi (gestion de stock RFID, gestion des ventes en supérette, application mobile, gestion des ventes échelonnées)

9.3. Synthèse Financement

- **Année 1**

Chiffre d'affaires (CA) : 5 085 000 DA

EBITDA : -1 751 000 DA (donc pertes)

L'entreprise démarre avec des pertes normales pour une première année d'activité.

- **Année 2**

CA : 8 850 000 DA (croissance)

EBITDA : -139 000 DA (quasi à l'équilibre)



Forte amélioration : les pertes sont presque éliminées grâce à l'augmentation du chiffre d'affaires.

- **Année 3**

CA : 14 910 000 DA

EBITDA : 5 610 000 DA (rentabilité atteinte)

L'entreprise devient rentable, avec une marge d'EBITDA qui dépasse les 35 %, ce qui est excellent.

Conclusion

- Croissance rapide du chiffre d'affaires.
- Réduction progressive des pertes.
- Rentabilité atteinte dès la 3e année = bon signe pour un projet solide et bien maîtrisé.



Business Model Canvas

Business model

VisionPlan

06/05/2025

1.0

Partenaires clés	Activités Clés	Propositions de valeur	Relation Client	Segmentation Clients
<ul style="list-style-type: none"> - Fournisseur de services de cybersécurité local - Intégrateurs locaux pour installation. - Prestataire de support technique externalisé 	<ul style="list-style-type: none"> - Vente & installation de VisionPlan - Support & maintenance client - Formation courte (2 jours) - Mise à jour trimestrielle du moteur IA - Upsell d'applications complémentaires prêtes. 	<ul style="list-style-type: none"> - Réduction erreurs de prévision : – 20 % - Baisse des pertes stocks : jusqu'à 5M DZD/an - Déploiement en < 7 jours - Hébergement national = conformité + sécurité (données souveraines) - Jusqu'à –60 % moins cher que concurrents étrangers 	<ul style="list-style-type: none"> - Suivi personnalisé après installation - Support (email, téléphone) - Visites trimestrielles grands comptes - Webinaires & tableaux de bord personnalisés - Programme de fidélisation multi-apps 	<ul style="list-style-type: none"> - Grandes entreprises industrielles (textile, énergie, agro) - PME de production et logistique - Ministères, agences publiques <p>Objectif : 5 clients en 1^{ère} Année</p>
	Ressources clés		Canaux de distribution	
	<ul style="list-style-type: none"> - Plateforme VisionPlan (100 % prête) - 3 autres apps métiers prêtes (Stock, POS, Vente) - Équipe fondatrice experte (IA, logistique) - Base de données industrielle algérienne - Infrastructure (serveurs, véhicules, site web) 		<ul style="list-style-type: none"> Vente directe (visites, appels) - Démonstrations & webinaires - Plateforme SaaS pour abonnement - Salons & LinkedIn (acquisition B2B) - Réseaux d'intégrateurs (ventes croisées) 	
Structure de Coûts		Flux de Revenus		
<p>Fixes :</p> <ul style="list-style-type: none"> - Salaires fondateurs : 2,5 M DZD/an - Dépenses matérielles : 4,5 M DZD <p>Variables :</p> <ul style="list-style-type: none"> - Frais déplacement : 300k DZD/an - Marketing : 400k DZD/an <p>Total Estimé : ~7,7M DZD</p>		<ul style="list-style-type: none"> - VisionPlan (licences/année) 4 000 000,00 DZD - logiciels complémentaire 625 000,00 DZD - Services annexes (reporting, analyse) 460 000,00 DZD <p>Total CA 2025 : ~5 M DZD</p>		

Analyse SWOT de l'Application Prédicative

L'analyse SWOT présentée ci-dessous vise à identifier les **forces**, **faiblesses**, **opportunités** et **menaces** liées au développement et à la commercialisation de l'application prédictive basée sur l'intelligence artificielle en Algérie. Cette étude stratégique permet de mettre en lumière les atouts compétitifs de la solution, tout en anticipant les risques potentiels du marché.

<p>Forces (Strengths)</p> <ul style="list-style-type: none"> • Plateforme d'intelligence artificielle développée entièrement en Algérie. • Déploiement rapide de la solution (moins de 7 jours). • Coût jusqu'à 60 % inférieur aux alternatives étrangères. • Trois applications complémentaires prêtes à être commercialisées. • Hébergement local garantissant la souveraineté et la conformité des données. 	<p>Faiblesses (Weaknesses)</p> <ul style="list-style-type: none"> • Coût initial de mise en place pour les PME. • Nécessité de former les équipes industrielles à l'utilisation de l'application. • Besoin de données fiables et volumineuses pour un apprentissage optimal des modèles IA.
<p>Opportunités (Opportunities)</p> <ul style="list-style-type: none"> • Croissance rapide de la digitalisation industrielle en Algérie. • Alignement stratégique avec les politiques algériennes (Stratégie Nationale d'IA, Vision Algérie 2030). • Croissance de la demande pour les systèmes de prédiction automatisée. • Développement de nouvelles fonctionnalités basées sur l'IA (maintenance prédictive, optimisation logistique). 	<p>Menaces (Threats)</p> <ul style="list-style-type: none"> • Concurrence croissante dans les solutions IA industrielles. • Risques liés à la cybersécurité et aux fuites de données. • Changements réglementaires dans l'industrie. • Résistance au changement technologique dans les entreprises traditionnelles.