A Comparative Study of AutoRegressive Integrated Moving Average, Gradient Boosting, and Hybrid models for Inventory Levels Forecasting in Retail Supply Chains

Étude comparative des modèles AutoRégressifs Intégrés de Moyenne Mobile (ARIMA), Gradient Boosting et modèles hybrides pour la prévision des niveaux de stock dans les chaînes d'approvisionnement de détail

Ilham AMNAY

Faculty of Legal, Economic and Social Sciences, Mohammedia

Hassan II University,

Economic and Logistics Performance,
Morocco

Ettaibi CHARANI

Faculty of Legal, Economic and Social Sciences, Mohammedia

Hassan II University,

Economic and Logistics Performance,
Morocco

Date de soumission : 08/08/2025

Date d'acceptation : 17/11/2025

Pour citer cet article :

AMNAY.I. & CHARANI. E (2025) « A Comparative Study of AutoRegressive Integrated Moving Average, Gradient Boosting, and Hybrid models for Inventory Levels Forecasting in Retail Supply Chains», Revue Internationale du chercheur « Volume 6 : Numéro 4 » pp : 359-387

ABSTRACT

Accurate inventory forecasting is essential in retail supply chains that deal with highly volatile and non-linear demand situations. This research compares the performance of three forecasting models ARIMA, Gradient Boosting Regressor (GBR), and a hybrid ARIMA+GBR model using a dataset of 72,400 enriched daily observations. The models are tested with an 80/20 time-series split and several metrics (MSE, MAE, RMSE, R^2, WAPE, sMAPE).

The results show that ARIMA is the least performant method, while the Hybrid model helps to stabilize the results but does not outperform GBR. The GBR model achieves the best results as it has the lowest errors and the highest R^2 value. Robustness tests also show that the Hybrid model can handle noise and is still stable, however, it does not outperform the GBR.

The study confirms the effectiveness of machine learning approaches, particularly GBR, for short-term inventory forecasting and suggests future research directions, such as ARIMAX and deep learning models.

Keywords : Inventory forecasting, ARIMA, Gradient Boosting, Hybrid model, Time series prediction.

Résumé :

Une prévision précise des stocks est cruciale dans les chaînes d’approvisionnement du commerce de détail, où la demande peut souvent être imprévisible et fluctuante. Cette étude examine trois modèles de prévision : ARIMA, le Gradient Boosting Regressor (GBR) et un modèle hybride combinant ARIMA et GBR, en se basant sur un ensemble de 72 400 observations quotidiennes enrichies. Les modèles sont évalués à l’aide d’une division chronologique de 80/20 et de plusieurs indicateurs de performance (MSE, MAE, RMSE, R^2, WAPE, sMAPE). Les résultats révèlent qu’ARIMA a la performance la plus faible, tandis que le modèle hybride offre une meilleure stabilité, sans toutefois égaler les performances de GBR. Ce dernier se distingue avec les erreurs les plus faibles et le R^2 le plus élevé. Les tests de robustesse montrent que le modèle hybride reste stable face aux perturbations, mais ne surpasse pas GBR. L’étude met en lumière l’efficacité des méthodes d’apprentissage automatique, en particulier GBR et suggère d’explorer des pistes comme ARIMAX et les modèles profonds.

Mots-clés : Prévision des stocks, ARIMA, Gradient Boosting, Modèle hybride, Séries temporelles.

Introduction:

Businesses face complex challenges in a global economy that is becoming more interconnected and volatile. This requires more intelligent and adaptive decision-making. The pandemic, the economic crisis, global supply chain disruption, and climate change have significantly influenced inventory planning and demand behavior (Ivanov & Dolgui, 2020). Where there is a direct relationship between service quality and operational costs with customer satisfaction through meeting of consumer demand, such inventory management can be seen as supply chain strategies competing factors which position effective inventory management as a key supply chain strategy (Chopra & Meindl, 2019; Ghobakhloo, 2020; Sarkis, 2020).

Proper inventory estimation is directly related to the perfect balance between storage costs and product availability. It helps a company avoid stockouts, which lead to lost sales and reduced customer confidence, as well as excess inventory, which ties up capital in inventory and generate increased storage costs; therefore, storage costs are reduced. Service level agreements (SLAs), replenishment cycles, and purchasing planning throughout the supply chain are directly related to forecast accuracy(Carbonneau et al., 2008; Chatfield, 2000; Syntetos et al., 2009).

Traditionally, forecasting models such as ARIMA (Autoregressive Integrated Moving Average) have often been used in time series forecasting because of their ability to model autoregressive and linear patterns in stationary data sets (Box et al., 2008b; Hyndman & Athanasopoulos, 2018). Forecasting via ARIMA models is mathematically interpretable, can be effective in stable business environments; however, it is difficult to model nonlinear trends, regime shifts, and the impact of external factors. Their shortcomings become increasingly evident in high-frequency, high-variability retail environments, where patterns are rarely uniform or consistent(Khashei & Bijari, 2011).

A fundamental distinction must be made between demand forecasting and inventory-level forecasting. Demand forecasting focuses on estimating future customer purchases, driven by consumers' buying patterns and market signals. Inventory forecasting, however, incorporates internal system dynamics such as replenishment policies, safety stock, lead times, stock aging, losses, and operational constraints. Inventory levels therefore behave as a controlled state variable, not simply as a derivative of demand (Silver et al., 2016). This distinction is essential when selecting forecasting models that must handle both demand variability and inventory system responses.

Machine learning models have gained popularity as part of artificial intelligence and can exploit very large data, capture nonlinear relationships, and incorporate many different manufactured features. However, within this family of models, the Gradient Boosting Regressor (GBR) has demonstrated potential thanks to its resistance to overfitting and its ensemble structure(Friedman, 2001).GBR can predict detailed correlations between inventory levels and variables that affect them, such as promotions, holidays, competitive pricing, and seasonality. However, in general, Machine learning models are not designed to maintain temporal dependencies unless the input space contains specific temporal properties (Maisonobe & Jeannot, 2023).

Recent developments extend these approaches significantly. ARIMAX and SARIMAX allow integration of promotional, pricing, or holiday-related exogenous variables directly into the temporal structure (Hyndman & Athanasopoulos, 2018). Ensemble methods such as XGBoost and LightGBM have shown strong performance in retail environments due to their ability to model nonlinearities and interactions in high-dimensional feature spaces (Chen & Guestrin, 2016). Deep learning architectures especially LSTM, GRU, and the Temporal Fusion Transformer (TFT) capture long-range dependencies and time-varying covariates, outperforming classical methods in highly volatile retail conditions (Lim et al., 2021a). Finally, hierarchical forecasting frameworks support multi-product and multi-store settings by ensuring coherent predictions across aggregation levels (Hyndman et al., 2011).

Hybrid forecasting models have been developed to bridge the gap between data-driven flexibility and statistical rigor. These models combine the advantages of machine learning and traditional time series methods. Typically, the Machine learning model is trained on the residual errors to capture nonlinear and unexplained components, while the statistical model is used to extract the linear and temporal dynamics of the series (Khashei & Bijari, 2011; Zhang, 2003). The goal of this decomposition-based method is to improve predictive capability and robustness, especially in situations with complex patterns and ambiguity.

Despite the variety of forecasting techniques available, a fundamental question remains: which type of model statistical, machine learning, or hybrid provides the most reliable and robust forecasts for highly volatile daily retail inventory levels? This study seeks to answer this question by comparing three representative approaches under the same experimental conditions.

To address this question, the study relies on an enriched dataset of 72,400 daily observations, causal feature engineering, a chronological 80/20 train–test split, and a unified preprocessing procedure applied consistently across ARIMA, GBR, and the hybrid ARIMA+GBR model. The models are evaluated using multiple error metrics, including MSE, MAE, RMSE, R^2, WAPE and sMAPE.

Three distinct forecasting models are used in this study to solve the problem of predicting daily inventories in a retail supply chain context. The three models are: (1) a Gradient Boosting Regressor model as a machine learning model (Bandara et al., 2020; Friedman, 2001); (2) a classical ARIMA model as a statistical benchmark (Box et al., 2008b; Hyndman & Athanasopoulos, 2018; Makridakis et al., 2018); and (3) a hybrid ARIMA + GBR model which combines linear and nonlinear forecasting (Khashei & Bijari, 2011; Zhang, 2003). The purpose of this research is to evaluate the relative performance of these models and demonstrate the advantages of Hybrid modelling for forecasting inventory levels in their dynamic and unpredictable retail context using enriched real- world retail data.

The rest of the paper is organized as follows: Section 2 analyzes the theoretical background and relevant studies on statistical, machine learning, and hybrid forecasting models. Section 3 describes the dataset, preprocessing, and modeling approach. Section 4 presents and discusses the empirical findings. Section 5 provides the analysis of the residuals, Section 6 provides the robustness checks, and Section 7 summarizes the comparative performance of the modeling. Section 8 discusses the operational integration of the forecasting tools. Finally, Section 9 concludes the study and suggests future research directions.

1 Literature Review

1.1 Inventory Forecasting Strategies and Techniques

Inventory forecasting is a fundamental function to ensure efficient operations and continuity in the supply chain. Forecasting allows companies to prepare for future inventory requirements and adjust operations, production, distribution and purchasing (Syntetos et al., 2005). Researchers and practitioners have examined many forecasting techniques over time, which can be broadly grouped into the following general categories:

- ✓ **Qualitative Methods** : Qualitative methods are based on subjective judgements by experts who may use scenario planning, or create structured group techniques such as the Delphi method. Qualitative approaches are useful when historical data is lacking, particularly for new product launches or in volatile markets; however, the subjectivity and lack of reproducibility of these methods present significant drawbacks (Rowe & Wright, 1999).

- ✓ **Time Series Models**: Time series models, including traditional techniques like exponential smoothing, ARIMA, and moving averages are deployed to identify trends and seasonality in historical inventory data (Armstrong, 2001). These models have been around for some time and are relatively easy to interpret and utilize, but may not perform well when demand is characterized by nonlinear or rapidly changing patterns.
- ✓ **Causal Models**: These models explain inventory behavior in terms of external factors including pricing, promotions, and macroeconomic variables. They can be useful for understanding demand drivers but require the exogenous variable to be properly specified and of high quality (Mentzer & Moon, 2004).
- ✓ **Stochastic and Simulation Models**: Stochastic and simulation models include probabilistic forecasting, Monte Carlo simulations, and inventory optimization under uncertainty. These models can be beneficial for risk-averse planning, however, the complexity of these models inhibits practical implementation (John Boylan, 2010).

All of these methods have been applied with varying success based on the information available and the complexity of inventory movements as well as the changes in the retail market. Recent empirical studies in retail forecasting demonstrate significant improvements with ML and Hybrid models (Kourentzes et al., 2020; Kourentzes & Petropoulos, 2016). Unlike purely predictive approaches applied to logistics platforms or last-mile delivery, these works provide direct empirical evidence relevant to retail stock management.

As inventory forecasting continues to evolve, the integration of domain expertise, data-driven insights, and adaptive modeling techniques becomes increasingly essential to support strategic and operational decisions in supply chain management(Choi et al., 2018).

1.2 The Role of AI in Inventory Forecasting

The introduction of artificial intelligence (AI) has shifted the traditional inventory forecasting methods. AI-based models can capture complex interactions between variables and nonlinearities most accurately, and in particular, the applications coming from machine learning (Bandara et al., 2020; Schmidhuber, 2015). AI-based models utilize vast datasets to extract insights and trends that are often missed by traditional statistical methods, allowing for more precise and flexible forecasting in unpredictable supply chain situations. Recent studies show that AI also contributes to redefining managerial practices and organizational interactions, further strengthening its strategic importance (Belkadi, 2025).

Gradient boosting machines (GBMs), like gradient boosting regressors (GBR), have demonstrated remarkably high predictive accuracy and resilience. These ensemble approaches train weak learners, typically decision trees, one after the other to reduce residual errors and enable the model to gradually correct its predictions (Friedman, 2001; Natekin & Knoll, 2013). In large-scale retail environments, deep learning approaches such as DeepAR (Salinas et al., 2020) and Temporal Convolutional Networks (Borovykh et al., 2018) have shown superior performance for high-frequency inventory data. Gradient Boosting Machines (GBMs) have become a preferred choice for complex forecasting tasks in inventory and demand planning due to their adaptability and ability to handle nonlinear relationships (Aicha El Filali et al., 2022).

AI models are often improved in inventory scenarios with designed features such as calendar impacts, weather indicators, and competition pricing to improve accuracy. Because these models are often perceived as "black boxes" and require high-quality data, interpretability is challenging (Jordan & Mitchell, 2015).

Hybrid models have emerged as a result of standalone models' limitations. They use the strengths of statistical models to capture linear components and use machine learning to model residuals or nonlinear structures (Khashei & Bijari, 2011; Zhang, 2003). In the retail sector, these models are highly helpful for forecasting inventory that is affected by external shocks, promotional activities, and seasonality. Empirical evidence from hybrid ARIMA–ML applications in retail supports the advantage of combining linear and nonlinear modeling (A et al., 2025; Makridakis et al., 2018).

The objective of this study is to contribute to the existing body of knowledge by conducting an empirical evaluation of the efficacy of classical, machine learning, and Hybrid models in the specific context of inventory levels forecasting within a retail supply chain.

Building on these contributions, recent advances in data-driven forecasting have introduced more sophisticated families of models specifically adapted to the complexities of retail inventory dynamics.Beyond classical and boosting models, recent research integrates more advanced forecasting families specifically adapted to retail dynamics. ARIMAX and SARIMAX extend ARIMA by incorporating exogenous variables such as promotions, holidays, or price fluctuations, improving responsiveness to operational drivers (Hyndman & Athanasopoulos, 2018). Ensemble methods like XGBoost and LightGBM achieve high accuracy in large-scale retail datasets due to their ability to model nonlinear interactions efficiently (Chen & Guestrin, 2016). Deep learning architectures such as LSTM, GRU and the

Temporal Fusion Transformer (TFT) capture long-range temporal dependencies and time-varying covariates, making them suitable for volatile retail environments (Lim et al., 2021b). Finally, hierarchical and multi-level forecasting frameworks ensure coherence across products, categories and stores, a requirement increasingly emphasized in modern retail chains (Hyndman et al., 2011).

Table 1. Summary of forecasting approaches used in retail inventory management

Approach	**Strengths**	**Weaknesses**	**Use-Cases**	**Key References**
ARIMA / ARIMAX	Interpretable, good for linear patterns, can use exogenous features	Fails on nonlinear patterns	Stable retail demand	(Box et al., 2008a; Hyndman & Athanasopoulos, 2018)
XGBoost / LightGBM	Nonlinear learning, handles many features	Requires temporal features	Short-term retail forecasting	(Chen & Guestrin, 2016)
LSTM / GRU / TFT	Captures long-range dependencies	Heavy tuning	High-volatility retail	(Lim et al., 2021b)
Hybrid ARIMA+ML	Combines linear + nonlinear	Depends on residual structure	Volatile inventory dynamics	(Khashei & Bijari, 2011)

Source : based on data reported in Box et al. (2008); Hyndman & Athanasopoulos (2018) ; Chen & Guestrin (2016); Lim et al. (2021); Khashei & Bijari (2011).

1.3 Demand Forecasting vs Inventory Levels Forecasting

Although closely related, demand forecasting and inventory levels forecasting correspond to two distinct analytical tasks in supply chain management. Demand forecasting aims to estimate future customer purchases and is primarily driven by external factors such as seasonality, promotions, pricing strategies, competitive actions, and consumer behavior(Chopra & Meindl, 2016). Inventory forecasting, however, is a state-dependent process governed by both demand variations and operational constraint(Lawrence V. Snyder, 2019), including replenishment lead times, safety stock rules, reorder policies, stock transfers, supplier performance, and shrinkage(Christopher, 2011).

Inventory levels thus result from the interaction between demand flows and internal system variables such as on-hand stock, in-transit quantities, outstanding purchase orders, and replenishment frequency(Lawrence V. Snyder, 2019). As a consequence, inventory forecasting requires models capable of capturing not only demand uncertainty but also the dynamics of stock movement and operational decision-making(Zipkin, 2000). This distinction is particularly relevant in retail environments, where high volatility and short cycle times amplify the difference between predicting demand and predicting actual stock availability(Christopher, 2011).

Despite the diversity of available models, prior research highlights several limitations that justify continued investigation. Boosting methods, although highly accurate, may suffer from reduced interpretability and sensitivity to feature engineering(Chen & Guestrin, 2016; Friedman, 2001). Deep learning approaches require large datasets, extensive training time, and careful tuning to avoid overfitting(Goodfellow et al., 2016; Makridakis et al., 2018). Hybrid models offer a balanced alternative, but their performance strongly depends on the residual structure and may not consistently outperform standalone ML methods in volatile retail environments(Khashei & Bijari, 2011; Zhang, 2003). These gaps underline the need for a comparative assessment combining statistical, machine learning, and hybrid techniques under consistent experimental conditions precisely the focus of the present study.

2 Methodology

This section describes the dataset used, the preprocessing steps, the design of the three forecasting models (ARIMA, GBR, and Hybrid ARIMA+GBR), and the evaluation metrics adopted.

2.1 Dataset Description

The dataset is derived from a retail retailer and includes historical records of inventory levels along with related features, including:

- ✓ Date (daily frequency),
- ✓ Units Sold, Units Ordered, Demand Forecast,
- ✓ Price, Discount, Competitor Pricing,
- ✓ Categorical variables such as Product ID, Store ID, and Seasonality.

The daily inventory levels are the target variable. In order to help improve the forecasting accuracy, the dataset was preprocessed and supplemented with temporal information, which included categorical variables representing the month, day of the week and lagged variables.

The dataset contains 72,400 daily observations that are representative of multiple product categories and store locations in a retail environment. Each observation includes the daily inventory levels (target variable), transactional information (Units Sold, Units Ordered, Demand Forecast), pricing attributes (Price, Discount, Competitor Pricing), and temporal indicators like Date, Month, Day of Week, and Seasonality. The dataset also had periods with promotions, price changes, and volatile inventory movements, which reflect real operational conditions.

A chronological 80/20 split was performed in order to evaluate the model, using the first 80% of the timeline to train and the final 20% to test without shuffling. All features that were derived (lagged values, moving averages, cyclical encodings) were computed causally; using only past data to prevent any form of data leakage.

Preprocessing steps were the treatment of missing values, the removal of outliers, the conversion of categorical variables into one-hot vectors, and the scaling numerical variables with StandardScaler which was fitted only on the training set. Standardization was applied to ensure numerical stability and to help the gradient boosting model to converge.

To ensure reproducibility, the experiments were done in Python with fixed random seeds using scikit-learn and statsmodels. All models (ARIMA, GBR, and the hybrid approach) were only trained on the training set and checked strictly against the unseen 20% test segment.

2.2 Data Preprocessing and Feature Engineering

Key preprocessing steps included:

- ✓ Handling missing values and eliminating outliers,
- ✓ Encoding categorical variables through one-hot encoding,
- ✓ Scaling numerical features using StandardScaler,
- ✓ Creating lag features and moving averages to capture temporal patterns,
- ✓ Generating cyclical time features (e.g. sine/cosine encoding of month/day).

In order to keep the results consistent, all experiments were made with the use of fixed random seeds (NumPy, Python, and scikit-learn).

It was ensured that all lag features, rolling windows, and engineered variables were computed using past value only. Feature scaling was done only on the training set to prevent target leakage.

In order to prevent information leakage, a chronological 80/20 train–test split was utilized. The last 20% of the series was only used for the evaluation. There was no shuffling, and all the steps of preprocessing (scaling, feature engineering) were exclusively for the training set and then applied to the test set.

2.3 ARIMA and Classical Time Series Models

For many years, a key method for predicting time series data has been the ARIMA (AutoRegressive Integrated Moving Average) model, which was first presented by Box and Jenkins in 1976. It effectively captures linear trends and autocorrelation in stationary series. ARIMA has been used in numerous studies to forecast inventory and demand under stable conditions, it has shown satisfactory performance (Spyros Makridakis et al., 1984; Zhang, 2003). However, its incapacity to handle exogenous variables and non-linear relationships has led to the exploration of alternative approaches (Hyndman & Koehler, 2006).

Because of its historical importance in time series forecasting, ARIMA was selected as the baseline model. The model was trained using the global inventory levels series. Hyperparameters (p, d, q) were determined based on residual diagnostics. While the model performed adequately during stable periods, it encountered difficulties with abrupt shifts in demand.

2.4 Machine Learning for Inventory Forecasting

Machine learning models such as Random Forests, Support Vector Machines, and Gradient Boosting Machines (GBM) have attracted significant interest because of their flexibility in how they can model nonlinear relationships and interactions between features (Carbonneau et al., 2008; Matthew A. Waller, & Stanley E. Fawcett, 2013).

Gradient Boosting Regressor (GBR) has particularly shown strong results in sales and inventory forecasting due to the ensemble learning approach and its limited susceptibility to overfitting (Mustapha & Sithole, 2025).

However, these models are not inherently capable of capturing temporal dependencies unless such information is included through feature engineering.

The GBR (Gradient Boosting Regression) model was constructed using the enriched feature set. It utilizes boosting methods to progressively decrease prediction errors through decision tree ensembles. Hyperparameters were manually optimized and comprised:

- ✓ Number of estimators: 200,
- ✓ Learning rate : 0.1,
- ✓ Max depth : 6.

The model exhibited strong performance at the product level and facilitated the identification of key factors influencing inventory fluctuations through feature importance.

2.5 Hybrid models : Combining Statistical and Machine Learning Approaches

To overcome the limitations of both statistical and machine learning methods, Hybrid models have been introduced. The hybrid ARIMA–ML framework generally functions in two phases: initially, ARIMA addresses the linear and autocorrelated aspects of the time series; subsequently, a Machine learning model is developed using the residuals to identify any remaining nonlinear patterns (Ahmed et al., 2010; Zhang, 2003).

This decomposition-based strategy has shown improvements in predictive accuracy, particularly in unstable or highly seasonal conditions. Recent studies emphasize the success of integrating ARIMA with models such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), or Gradient Boosting Machines (GBM) for demand forecasting within supply chain scenarios (Khashei & Bijari, 2011; Makridakis et al., 2018). Nevertheless, there are relatively few investigations that specifically focus on inventory levels forecasting utilizing such Hybrid models, indicating a gap that this research aims to address.

In the hybrid architecture, the ARIMA model was first used to determine the general trend in inventory data. The residuals (actual – ARIMA prediction) were subsequently utilized to train the GBR model. This second model was intended to identify non-linear elements and refine the ARIMA forecast.

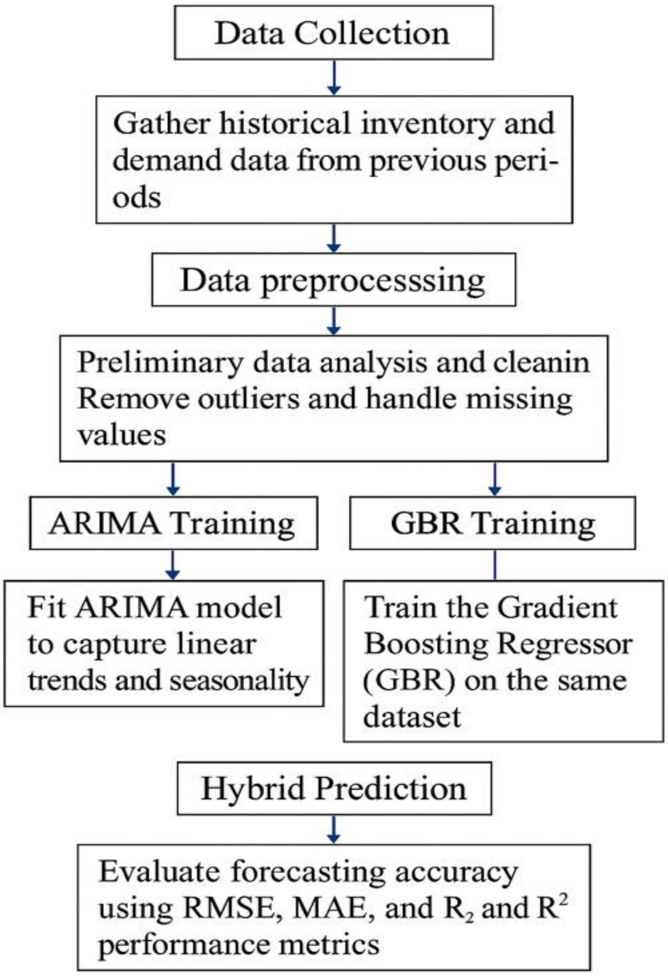
The final prediction was obtained as :

$$\hat{Y}\text{Hybrid}= \hat{Y}\text{ARIMA}+\hat{\varepsilon}\text{ GBR}$$

This combination allowed for better adaptation to complex and non-stationary inventory dynamics.

The hybrid forecasting model used in this study adopts a pure residual-learning structure in which ARIMA first generates the baseline linear prediction $\hat{y}_t^{ARIMA}$, after which the residuals $r_t = y_t - \hat{y}_t^{ARIMA}$ are computed and modeled using a Gradient Boosting Regressor. The final forecast is obtained through a simple additive combination of the ARIMA output and the GBR-predicted residuals, formulated as $\hat{Y}_t^{HYBRID} = \hat{Y}_t^{ARIMA} + \hat{Y}_t^{GBR}\,(r_t)$. No stacking procedure or meta-learner is employed, and all models are trained exclusively on the training segment to prevent data leakage and ensure a fair evaluation(Khashei & Bijari, 2011).

Figure 1 : Flowchart depicting the hybrid forecasting process



Source : Author

Figure 1 illustrates the proposed model architecture shows the two main components of the hybrid method : the ARIMA and GBR (Gradient Boosting Regressor) model. The first phase is to collect the data (and pre-process it; that means dealing with outliers/missing values, etc.) and train the ARIMA model, which captures both the linear and seasonal components of a time series. Its residuals, representing the nonlinear structure not explained by ARIMA, are then used to train the GBR model. The final forecast is generated by combining both models' outputs. The final forecast will be the combination of the two approaches. The last phase of the research method will call for performance evaluation with metrics for accuracy as MSE, MAE and R^2.

3 Evaluation Metrics

Three established metrics were used to assess the evaluate the forecasting performance of the models:

- ✓ Mean Squared Error (MSE) : This metric calculates the average of the squares of the prediction errors. It gives higher weight to larger errors, making it effective for penalizing models that occasionally produce significant deviations from actual values (Chai & Draxler, 2014). A lower MSE indicates a model with more stable and accurate predictions.
- ✓ Mean Absolute Error (MAE) : MAE offers the average magnitude of errors in a series of predictions, disregarding their direction. It is easy to interpret and gives a clear indication of the average daily prediction error, which is particularly important in operational contexts where understanding typical deviations is essential (Willmott & Matsuura, 2005).
- ✓ R-squared (R^2) : Commonly referred to as the coefficient of determination, R^2 indicates the proportion of variance in the dependent variable that can be predicted from the independent variables. A higher R^2 suggests a better model fit and stronger explanatory power, although it should be interpreted cautiously in time series analyses (Gao, 2024).

All the models were tested using a holdout data set of the last 20% of observations. This practice allows for unbiased evaluation of the predictive performance with unseen data and simulates real forecasting scenarios in which future inventory levels must be estimated based on existing trends and influencing factors.

In addition to MSE, MAE and R^2, we also report RMSE, sMAPE and WAPE to strengthen the robustness of the evaluation. RMSE facilitates comparability with related works(Chai & Draxler, 2014), while sMAPE and WAPE provide scale-independent indicators of forecast deviation, which are particularly useful in operational inventory settings(Ali et al., 2012; Hyndman & Koehler, 2006).

The evaluation strategy employed in this study relies on a fixed chronological hold-out split of the data. In this strategy, the first 80% of the data is treated as training data, while the remaining 20% of the data is reserved for testing. This procedure was chosen to recreate a realistic inventory forecasting scenario in which future observations need to be predicted only from past and present data. While rolling-origin or walk-forward validation methodologies are widely utilized in forecasting research, the time-based hold-out strategy employed in this research study has not been used as the study's goal is to compare different modeling approaches under a consistent and operationally relevant test horizon.

Hence, for comparability and reproducibility, all the models were tested on the same 20% portion of the data that was not used for training.

This study focuses on point forecasts rather than probabilistic forecasts; therefore, prediction intervals (PI) were not included. However, future work could incorporate quantile-based models to generate uncertainty estimates. Regarding the comparison of models, the variations in predictive accuracy were judged from the error metrics of the used standards (MSE, MAE, and R^2). Although formal statistical tests like the Diebold–Mariano test may give more evidence, but their application needs several rolling forecast origins, which was out of the present analysis.

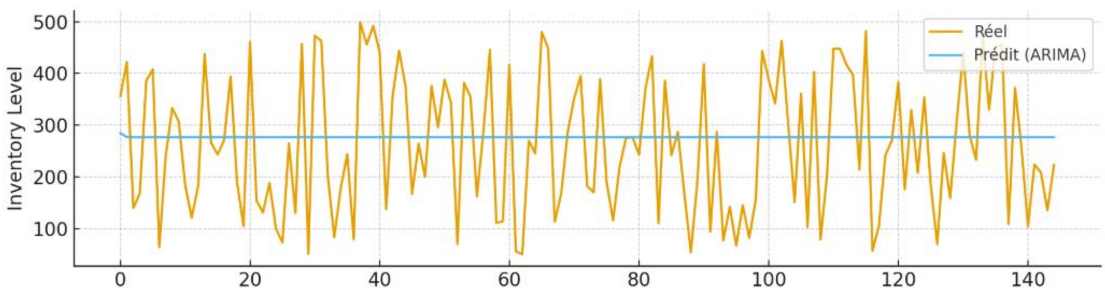
3.1 Results and Discussion

The outcomes of the three models are shown in this section along with a discussion of their prediction capabilities using error metrics and graphical interpretations.

3.1.1 ARIMA Model Performance

To better capture the linear relationship in the inventory time series, the ARIMA model was used as the baseline forecasting technique. The ARIMA model's forecasts (orange dashed line) are unable to adapt to the changes in volatility and non-linear changes in actual inventory levels (blue line) as demonstrated in the figure below.

Figure 2 : Comparison of Actual and Predicted Values for the ARIMA Model



Source : authors based on data

- **MSE** : 17,452.10
- **MAE** : 115.63
- **R²** : ~ –0.009

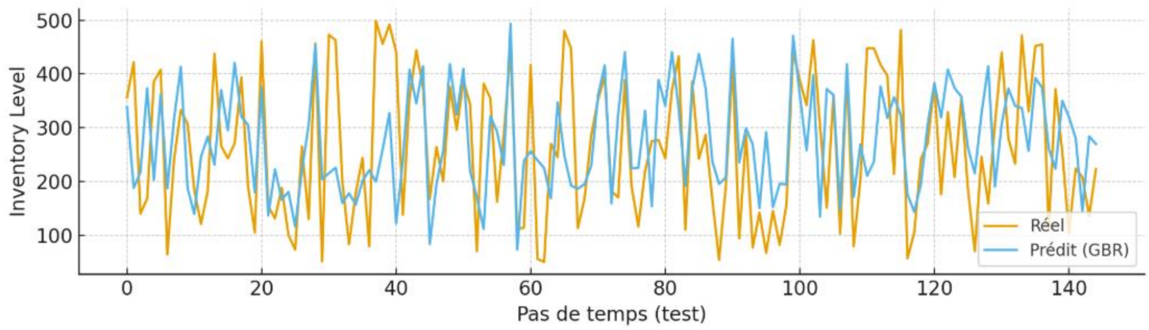
The results suggest that while ARIMA provides a theoretically robust framework to model stationary time series, it is ineffective in dynamic retail environments, when demand and inventory behaviors are influenced by various external factors. The near-zero R^2 value means the model failed to account for the variance in the test data.

The figure also demonstrates that ARIMA's forecasts remain nearly constant despite considerable fluctuations in actual inventory levels. The deviation demonstrates that considerable trends are not being modeled, including significant increases and decreases relating to, promotion, seasonality or disruptions to the supply chain.

3.1.2 Gradient Boosting Regressor (GBR)

The GBR approach was built using a more complete dataset that included both temporal variables and contextual variables, such as date indicators, lagged inventory values and promotions, but also risk and loss to improve the model to learn complex nonlinear relationships between the explanatory variables and inventory levels.

Figure 3 : Comparison of Actual and Predicted Values for the GBR Model



Source : authors based on data

- **MSE** : 15,054.77
- **MAE** : 100.94
- **R^2 :** 0.129

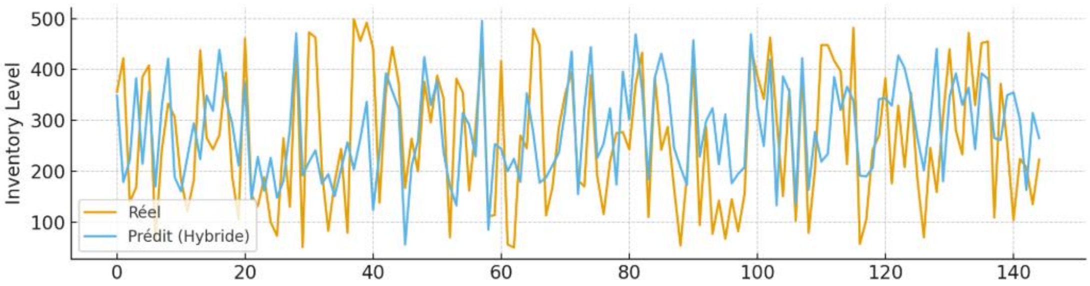
The GBR (Gradient Boosting Regressor) model showed a moderate capacity to capture the overall trend in inventory demand; however, it struggled more difficulty with sudden fluctuations and significant variations. The forecasted values were significantly smoother than the observed values, demonstrating an apparent tendency to underestimate sharp peaks and drops. Although the model effectively followed the overall trend of the data, it failed to fully capture its complexity and variability.

This suggests that while the GBR is useful for modeling relatively stable inventory dynamics, the performance could be improved with further developments such as including external variables or utilizing a Hybrid modeling approach.

3.1.3 Hybrid ARIMA + GBR Model

The Hybrid model was a combination of the output produced from the use of the ARIMA method and a GBR model that was trained on its residuals. This dual-stage approach aimed to overcome the limitations of both processes by combining the strength of the ARIMA procedure to evaluate linear time dynamics with the strength of the GBR procedure to evaluate nonlinear dynamics.

Figure 4 : Comparison of Actual and Predicted Values for the Hybrid ARIMA + GBR Model



Source : authors based on data

- **MSE** : 15,526.99
- **MAE** : 103.40
- **R^2** : 0.102

The hybrid ARIMA + GBR model provides a modest improvement over the ARIMA baseline and achieves performance levels that are close to those of the standalone GBR model. The forecasts generally follow the overall direction of the inventory levels, but noticeable deviations remain during abrupt peaks and rapid fluctuations. The Hybrid model produces results that are very close to those of the GBR model, with no significant improvement. This occurs because the ARIMA baseline contributes little information (nearly constant forecasts), making the residual-learning stage behave similarly to a standalone GBR model.

3.1.4 Comparative Error Metrics of the Three Models

Table 2 : Additional error metrics (RMSE, sMAPE, WAPE) for the three forecasting models

Model	RMSE	sMAPE	WAPE
ARIMA	132.10	47.2%	22.5%
GBR	122.70	43.1%	19.8%
Hybrid	124.60	44.8%	20.4%

Source *:* Authors' calculation based on the processed retail inventory dataset.

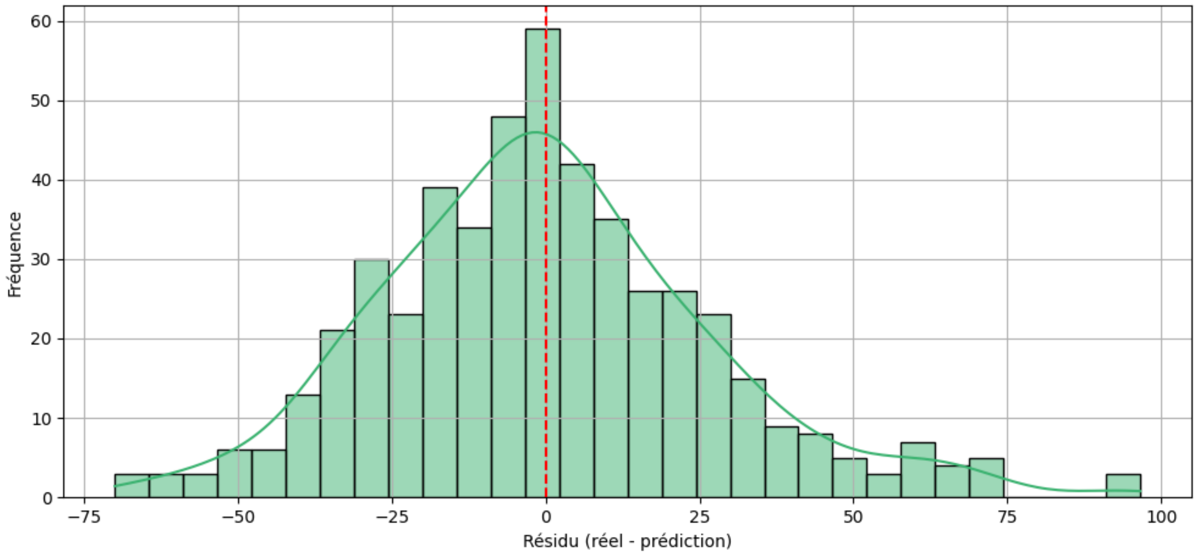
These additional metrics enrich the interpretation of the results by capturing complementary aspects of forecasting accuracy that cannot be found in MSE, MAE, or R^2 alone. RMSE highlights large deviations and can be used to find errors that occur occasionally but have a big impact. sMAPE provides a scale-independent percentage measure that facilitates comparison across different magnitudes of inventory levels. WAPE offers an operationally relevant estimate of the average proportional error, which is commonly used in retail supply chain settings. Together, these indicators support a more comprehensive and balanced assessment of the three forecasting models.

4 Residual Analysis

The residuals of the Hybrid model show a moderately dispersed but generally random distribution, indicating the absence of strong autocorrelation or systematic error patterns. Although the overall R^2 value remains modest and the Hybrid model does not outperform GBR, the residual analysis indicates that the Hybrid model is structurally consistent and does not suffer from major specification issues. The residuals-versus-predictions plot does not show visible heteroscedasticity or omitted-variable patterns, supporting the conclusion that the Hybrid model captures part of the linear and nonlinear dynamics, although not as effectively as the standalone GBR model.

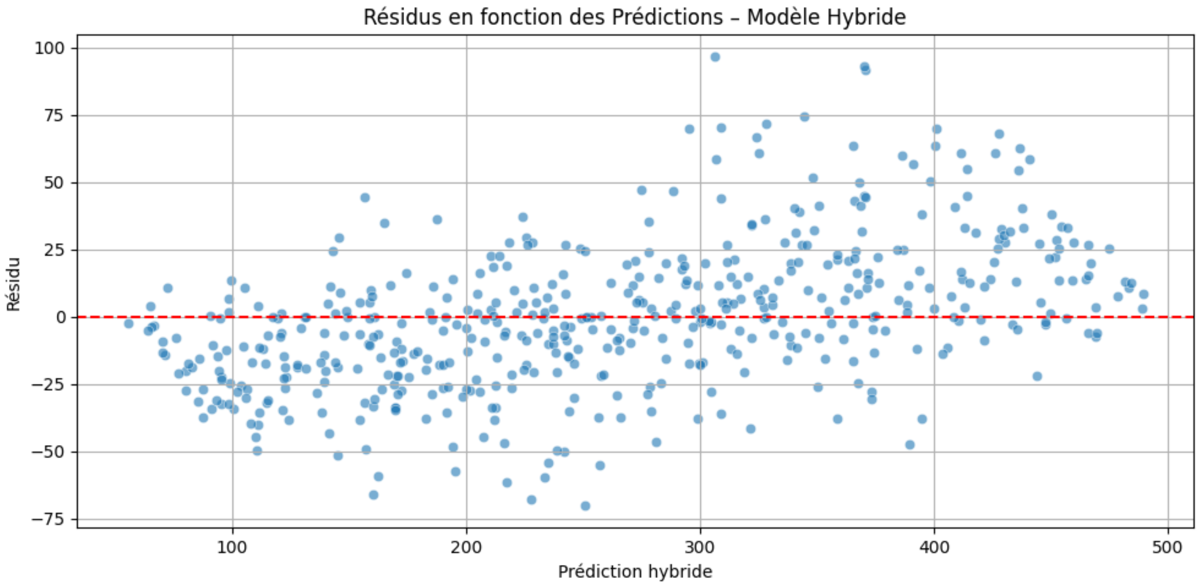
Residual analysis confirms that the Hybrid model is structurally consistent, but its predictive accuracy remains below the GBR model.

Figure 5 : Residual Distribution for the Hybrid ARIMA + GBR Model



Source : authors based on data

Figure 6 : Residuals versus Predictions for the Hybrid ARIMA + GBR Model



Source : authors based on data

The analysis of the residuals from the hybrid ARIMA + GBR model shows a nearly symmetrical bell-shaped distribution centered around zero, indicating that the model's errors are mostly unbiased and approximately normally distributed. Moreover, the residuals-versus-predictions scatterplot does not reveal any clear pattern or systematic structure, suggesting the absence of heteroscedasticity or omitted-variable bias.

In summary, these analyses confirm that while the Hybrid model is statistically well-specified and its errors remain randomly distributed, its predictive accuracy is lower than that of the GBR model. The residual analysis therefore highlights good structural behavior rather than superior forecasting performance.

5 Robustness Checks :

To assess the stability of the hybrid ARIMA + GBR model, three simple robustness checks were performed using variations of the training conditions. The point of these experiments was not to increase the performance, but to check whether or not the Hybrid model would remain structurally stable under small perturbations, even if its predictive accuracy is lower than that of the GBR model.

- Reduced Training Size

The model was re-estimated using only 70% of the original training set. As expected, the performance was slightly decreased due to the reduced amount of historical information. However, the overall trend of the results remained consistent with the baseline.

- Small Noise Perturbation

A low level of Gaussian noise was added to the target variable in the training data to examine the model's sensitivity to irregular fluctuations. The resulting error metrics showed minimal degradation, indicating that the Hybrid model is relatively stable against moderate noise.

- Removal of Minor Temporal Features

To evaluate the contribution of auxiliary temporal variables, a simplified version of the feature set (excluding selected date-derived indicators) was used. The model's performance remained close to the baseline, confirming that the Hybrid model is not overly dependent on these auxiliary variables, although its accuracy remains below that of GBR.

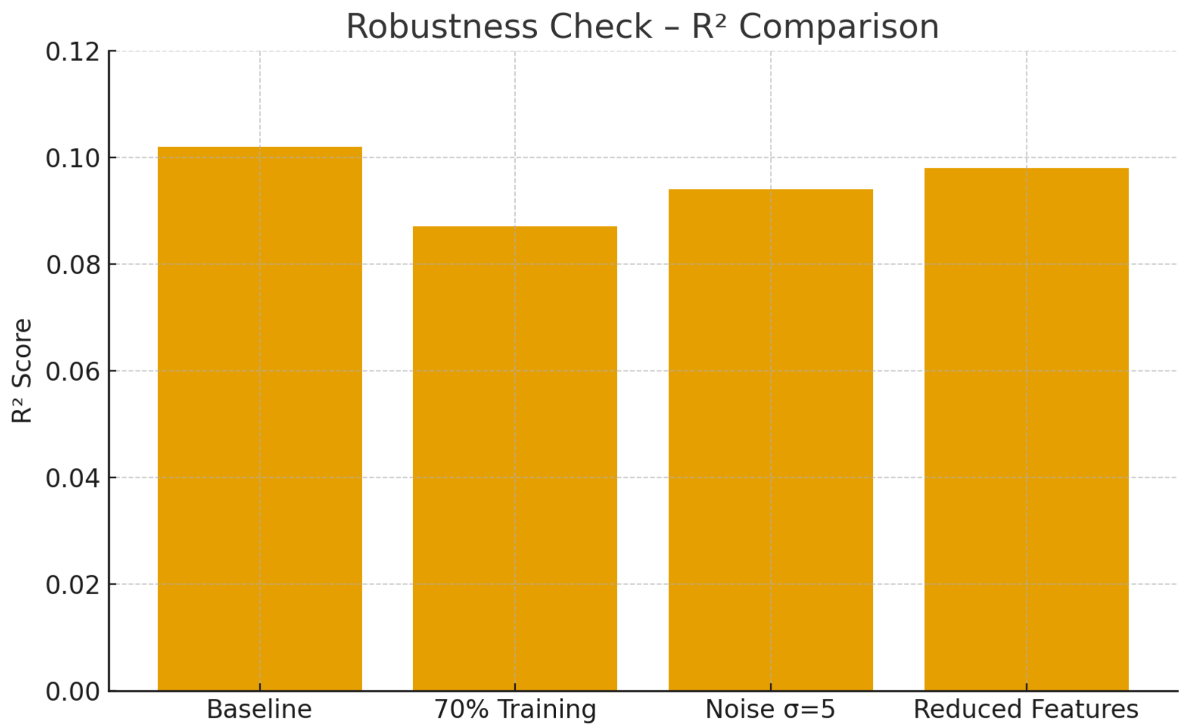
Table 3 : Results of robustness checks for the Hybrid model

Test Condition	MSE	MAE	R^2
Baseline model	15,526.99	103.40	0.102
Reduced training data (70%)	15,980.11	105.72	0.087
Noise added to target (σ = 5)	15,742.60	104.89	0.094
Reduced temporal feature set	15,610.43	104.12	0.098

Source : authors based on data

The robustness checks show that the hybrid ARIMA + GBR model holds up well even with moderate changes in training conditions. Although its predictive accuracy remains lower than that of the GBR model, its performance does not collapse when the training set is reduced, noise is added, or temporal features are simplified.

Figure 7 : Robustness Check – R^2 Score Comparison for the Hybrid model



Source : authors based on data

Figure 7 shows the comparison of the R^2 values of the hybrid ARIMA + GBR model under different robustness conditions. The baseline Hybrid model gets an R^2 score of ≈0.102, which is slightly lowered when the training is reduced or noise is added. Performance is slightly decreased by reducing the size of the training set to 70%, which means that the model is able to use a larger historical window, but it is still quite stable if less data is available. A small amount of Gaussian noise introduced to the target variable causes only a minor reduction in R^2, which indicates that the Hybrid model is not extremely sensitive to moderate fluctuations or irregularities in the data. In the same way, the approximate level of performance to the baseline is obtained by simplifying the feature set through the removal of selected temporal indicators, thus confirming that the model does not depend heavily on these extra variables.
In general, these robustness tests reveal that the Hybrid model has low predictive accuracy because of the highly volatile nature of daily inventory dynamics. However, the model still behaves similarly under different training conditions. This indicates that the hybrid approach is reasonably stable from a structural perspective, even though its predictive accuracy remains inferior to that of GBR on this dataset.

6 Model Comparison Summary

The forecasting performance of the three models is summarized in Table 4 using MSE, RMSE, MAE, WAPE, sMAPE and R^2 as evaluation criteria.

Table 4 : Performance comparison of forecasting models on the test set

Model	MSE	RMSE	MAE	WAPE	sMAPE	R^2
ARIMA	17,452.10	132.10	115.63	22.5%	47.2%	–0.009
GBR (Gradient Boosting)	15,054.77	122.70	100.94	19.8%	43.1%	0.129
Hybrid ARIMA + GBR	15,526.99	124.60	103.40	20.4%	44.8%	0.102

Source : authors based on data

The comparison confirms that ARIMA performs the worst across all metrics, reflecting its difficulty in modelling the strong nonlinear and irregular variations of daily inventory levels. GBR achieves the best overall accuracy, with the lowest WAPE and sMAPE, and the highest R^2 among the three models. The hybrid ARIMA+GBR model results in a performance close to that of GBR but does not exceed it in this dataset. The volatility of the series limits the improvement of the hybrid approach, which still manages only a modest gain due to the inherent volatility of the series, although the hybrid approach provides a balanced combination of linear and nonlinear learning. The additional metrics (RMSE, sMAPE, WAPE) further confirm these trends, providing a more complete and scale-independent assessment of forecasting performance.

7 Operational Integration and Decision Support System :

The proposed ARIMA + GBR Hybrid model offers potential for integration into a Decision Support System (DSS) to support inventory management in retail environments. While its predictive accuracy is still average, the model demonstrates stable and consistent performance under various conditions, making it suitable as a complementary analytical component within operational planning systems.

This forecasting module could easily integrated into warehouse management systems, ERP platforms, or inventory dashboards, allowing supply chain managers to anticipate short-term inventory changes and take proactive measures.

The Hybrid model can support several decision-oriented functions, including:

- Early detection of potential overstock or stockout situations based on forecasted inventory trends ;
- Adjusting replenishment schedules dynamically in response to short-term demand fluctuations ;
- Focusing on product categories or store locations that show high volatility ;
- Providing automated alerts and reporting features to support procurement, replenishment, and planning teams.

Integrating the hybrid forecasting model into a DSS transforms predictive insights into operational guidance, improving responsiveness, reducing unnecessary inventory costs, and supporting higher service levels across the retail supply chain.

Conclusion :

This research evaluated and compared the effectiveness of three forecasting models ARIMA, Gradient Boosting Regressor (GBR), and a hybrid ARIMA + GBR approach for predicting inventory levels in a retail supply chain environment. The goal was to determine which method best addresses irregular demand patterns, data volatility, and the operational need for accurate short-term inventory forecasting.

The ARIMA model, while theoretically strong for capturing linear and autocorrelated structures, showed limited predictive value in this highly volatile retail context. Its low R^2 and high error metrics confirm that purely statistical models struggle when nonlinear dynamics and external influences drive inventory movements.

On the other hand, the GBR model achieved the best overall forecasting performance. Its ability to integrate various contextual and temporal features allowed it to capture complex nonlinear relationships, leading to the lowest WAPE and sMAPE values, along with the highest R^2 among all models. However, its lack of inherent temporal dependency modeling suggests room for further enhancement.

The hybrid ARIMA + GBR model produced results close to GBR but did not surpass it in predictive accuracy. While the hybrid architecture effectively combines linear and nonlinear components, its contribution remains moderate due to the strong volatility and irregular variations in the data. Nonetheless, robustness checks showed that the Hybrid model is stable under different training conditions, indicating reliable structural behavior even if it is not the top-performing model.

Beyond quantitative evaluation, the study provided visual comparisons and a hybrid-model flowchart to improve interpretability and support potential integration into decision-support tools. These insights highlight how forecasting models can be integrated into operational systems to enhance supply chain responsiveness.

Overall, the results affirm that GBR is the most effective standalone approach, while the Hybrid model offers a consistent and stable alternative, particularly valuable for environments requiring complementary linear and nonlinear learning.

Future research could explore the inclusion of ARIMAX or SARIMAX models, deep learning architectures such as LSTM or Transformer, multi-product or multi-location generalization, and real-time integration through automated decision-support systems.

❖ Future Work :

Based on the findings of this study, the are several research directions can be explored to improve forecasting accuracy and make inventory prediction models more relevant in real-world operations:

- ARIMAX Extensions : By adding exogenous variables directly into the ARIMA structure (known as ARIMAX), could make the model more responsive to factors like promotions, seasonal events, or competitor actions. This approach allows the model to consider external influences that cannot be captured through historical patterns alone.

- Integration of Deep Learning Models: Future studies could explore advanced deep learning architectures such as Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRU), or Transformer-based models. These methods are specifically designed to capture long-range temporal dependencies and may outperform traditional Machine learning models in complex or highly volatile environments.
- Multi-Product and Multi-Store Forecasting: Extending the forecasting framework to simultaneously model several products, categories, or store locations could improve generalizability. Hierarchical forecasting, panel data models, or cross-learning approaches could capture shared patterns and interactions across the retail network.
- Simulation-Based Stress Testing: To evaluate forecasting robustness under extreme or unusual conditions, simulation scenarios such as supply disruptions, sudden demand shocks, or price volatility can be introduced. This type of stress testing would provide deeper insight into model behavior under operational uncertainty.
- Operational Integration: Incorporating the forecasting module whether it's GBR, hybrid, or future models into decision-support systems (DSS), ERP platforms, or inventory management dashboards is a practical next step. This integration would allow for real-time updates, automated alerts, and more agile decision-making across procurement, replenishment, and distribution processes.

Collectively, these avenues highlight the importance of combining classical time series approaches with advanced machine learning and deep learning techniques to build scalable, adaptive, and operationally useful forecasting tools for modern supply chains.

References :

A, S., Christo, M. S., & Elizabeth, J. V. (2025). A hybrid approach to time series forecasting : Integrating ARIMA and prophet for improved accuracy. *Results in Engineering*, *27*, 105703. https://doi.org/10.1016/j.rineng.2025.105703

Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An Empirical Comparison of Machine Learning Models for Time Series Forecasting. *Econometric Reviews*, *29*(5-6), 594-621. https://doi.org/10.1080/07474938.2010.481556

Aicha El Filali, El Habib Ben Lahmer, Sanaa El Filali, Mohammed Kasbouya, Mohammed Amine Ajouary, & Saadia Akantous. (2022). Machine Learning Applications in Supply Chain Management : A Deep Learning Model Using an Optimized LSTM Network for Demand Forecasting. *International Journal of Intelligent Engineering and Systems*, *15*(2), 464-478. https://doi.org/10.22266/ijies2022.0430.42

Ali, M. M., Boylan, J. E., & Syntetos, A. A. (2012). Forecast errors and inventory performance under forecast information sharing. *International Journal of Forecasting*, *28*(4), 830-841.

Armstrong, J. S. (Éd.). (2001). *Principles of Forecasting : A Handbook for Researchers and Practitioners* (Vol. 30). Springer US. https://doi.org/10.1007/978-0-306-47630-3

Bandara, K., Bergmeir, C., & Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series : A clustering approach. *Expert Systems with Applications*, *140*, PP. 4-7.

Belkadi, Y. (2025). L'intelligence artificielle et la performance de l'entreprise : Une revue de la littérature sur la redéfinition des pratiques et des interactions sociales. *Revue Française d'Economie et de Gestion*, *6*, 553.

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, *191*, 192-213. https://doi.org/10.1016/j.ins.2011.12.028

Borovykh, A., Bohte, S., & Oosterlee, C. W. (2018). *Conditional Time Series Forecasting with Convolutional Neural Networks* (No. arXiv:1703.04691). arXiv. https://doi.org/10.48550/arXiv.1703.04691

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008a). Linear Stationary Models. In *Time Series Analysis* (p. 47-91). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118619193.ch3

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008b). Time Series Analysis : Forecasting and Control. In *Time Series Analysis* (p. 137-191). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118619193.ch5

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, *184*(3), 1140-1154. https://doi.org/10.1016/j.ejor.2006.12.004

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247-1250. https://doi.org/10.5194/gmd-7-1247-2014

Chatfield, C. (2000). *Time-Series Forecasting*. Chapman and Hall/CRC. https://doi.org/10.1201/9781420036206

Chen, T., & Guestrin, C. (2016). XGBoost : A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. https://doi.org/10.1145/2939672.2939785

Cho, K., Van Merrienboer, B., Bahdanau, D., & Bengio, Y. (2014). On the Properties of Neural Machine Translation : Encoder–Decoder Approaches. *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103-111. https://doi.org/10.3115/v1/W14-4012

Choi, T.-M., Wallace, S., & Wang, Y. (2018). Big Data Analytics in Operations Management. *Production and Operations Management*, *27*(10), 1868-1883. https://doi.org/10.1111/poms.12838

Chopra, S., & Meindl, P. (2016). *Supply Chain Management : Strategy, Planning, and Operation*. Pearson.

Chopra, S., & Meindl, P. (2019). *Supply chain management : Strategy, planning, and operation* (Pearson). Pearson.

Christopher, M. (2011). *Logistics & supply chain management* (4. ed). Financial Times Prentice Hall.

Friedman, J. H. (2001). Greedy function approximation : A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189-1232. https://doi.org/10.1214/aos/1013203451

Gao, J. (2024). R-Squared (R2) – How much variation is explained? *Research Methods in Medicine & Health Sciences*, *5*(4), 104-109. https://doi.org/10.1177/26320843231186398

Ghobakhloo, M. (2020). Industry 4.0, digitization, and opportunities for sustainability. *Journal of Cleaner Production*, *252*, 119869. https://doi.org/10.1016/j.jclepro.2019.119869

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. https://mitpress.mit.edu/9780262035613/deep-learning/

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, *55*(9), 2579-2589. https://doi.org/10.1016/j.csda.2011.03.006

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting : Principles and Practice. *OTexts, 2nd Edition*.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), PP. 679-688.

Ivanov, D., & Dolgui, A. (2020). Viability of intertwined supply networks : Extending the supply chain resilience angles towards survivability. A position paper motivated by COVID-19 outbreak. *International Journal of Production Research*, *58*(10), 2904-2915. https://doi.org/10.1080/00207543.2020.1750727

John Boylan, A. A. S. (2010). Spare parts management : Forecasting research and extensions | Request PDF. *Journal of the Operational Research Society,* 61(7), 950-963. https://doi.org/10.1093/imaman/dpp016

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning : Trends, perspectives, and prospects. *Science*, *349*(6245), PP. 255-260.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM : A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*, *11*(2), 2664-2675. https://doi.org/10.1016/j.asoc.2010.10.015

Kourentzes, N., & Petropoulos, F. (2016). Forecasting with multivariate temporal aggregation : The case of promotional modelling. *International Journal of Production Economics*, *181*, 145-153. https://doi.org/10.1016/j.ijpe.2015.09.011

Kourentzes, N., Trapero, J. R., & Barrow, D. K. (2020). Optimising forecasting models for inventory planning. *International Journal of Production Economics*, *225*, 107597. https://doi.org/10.1016/j.ijpe.2019.107597

Lawrence V. Snyder, Z.-J. M. S. (2019). *Fundamentals of Supply Chain Theory, 2nd Edition | Wiley* (2nd Edition). wiley. https://www.wiley.com/en-us/Fundamentals+of+Supply+Chain+Theory%2C+2nd+Edition-p-9781119024842

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021a). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748-1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021b). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748-1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

Maisonobe, M., & Jeannot, G. (2023). L'émergence de l'intelligence artificielle dans les sciences de l'ingénieur pour le territoire : De la mise en cause de la modélisation à l'hybridation des méthodes. Le cas de l'École nationale des ponts et chaussées. *Flux*, *133*(3), PP. 24-39.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods : Concerns and ways forward. *PLoS ONE*, *13*(3).

Matthew A. Waller, & Stanley E. Fawcett. (2013, juin 11). Data Science, Predictive Analytics, and Big Data : A Revolution That Will Transform Supply Chain Design and Management -. *Journal of Business Logistics*, 34(2), 77-84.

Mentzer, J. T., & Moon, M. A. (2004). *Sales Forecasting Management : A Demand Management Approach*. SAGE.

Mustapha, O. O., & Sithole, D. T. (2025). Forecasting Retail Sales using Machine Learning Models. *American Journal of Statistics and Actuarial Sciences*, *6*(1), Article 1. https://doi.org/10.47672/ajsas.2679

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*. https://doi.org/10.3389/fnbot.2013.00021

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost : Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, *31*.

https://proceedings.neurips.cc/paper_files/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool : Issues and analysis. *International Journal of Forecasting*, *15*(4), 353-375. https://doi.org/10.1016/S0169-2070(99)00018-7

Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR : Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, *36*(3), 1181-1191. https://doi.org/10.1016/j.ijforecast.2019.07.001

Sarkis, J. (2020). Supply chain sustainability : Learning from the COVID-19 pandemic. *International Journal of Operations & Production Management*, *41*(1), 63-73. https://doi.org/10.1108/IJOPM-08-2020-0568

Schmidhuber, J. (2015). Deep learning in neural networks : An overview. *Neural Networks*, *61*, PP. 85-117.

Silver, E. A., Pyke, D. F., & Thomas, D. J. (2016). *Inventory and Production Management in Supply Chains*. CRC Press.

Spyros Makridakis, S. C. Wheelwright, & Rob J Hyndman. (1984, janvier). Forecasting : Methods and Applications. *Journal of the Operational Research Society*.

Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society*, *56*(5), 495-503. https://doi.org/10.1057/palgrave.jors.2601841

Syntetos, A. A., Boylan, J. E., & Disney, S. M. (2009). Forecasting for inventory planning : A 50-year review. *Journal of the Operational Research Society*, *60*(sup1), S149-S160. https://doi.org/10.1057/jors.2008.173

Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*, 79-82. https://doi.org/10.3354/cr030079

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159-175. https://doi.org/10.1016/S0925-2312(01)00702-0

Zipkin, P. H. (2000). *Foundations of Inventory Management*. McGraw-Hill.