Submission Date:
23.06.2025
Acceptance Date:
12.09.2025
Publication Date:
29.09.2025

To cite this article: Doğan, A. (2025). Time series-based demand forecasting: A comparative analysis of holt-winters, SARIMA, and Prophet models on retail ınventory data. *İstanbul Ticaret University Journal of Social Sciences, 24*(53), 641-668, doi: 10.46928/iticusbe.1725584

TIME SERIES-BASED DEMAND FORECASTING: A COMPARATIVE ANALYSIS OF HOLT-WINTERS, SARIMA, AND PROPHET MODELS ON RETAIL INVENTORY DATA

Research

Alican Doğan

Corresponding Author

Bandırma Onyedi Eylül University

alicandogan@bandirma.edu.tr

Alican Doğan, Bandırma Onyedi Eylül University, Department of Management Information System, Assistant Proffessor Dr., Machine Learning, Data Mining, Artificial Intelligence.

Time Series-Based Demand Forecasting: A Comparative Analysis of Holt-Winters, SARIMA, and Prophet Models on Retail Inventory Data

Alican Doğan
alicandogan@bandirma.edu.tr

Abstract

In the retail industry, correctly managing inventory levels not only reduces logistics costs but also directly contributes to increased customer satisfaction. Therefore, the accuracy of demand forecasting models plays a critical role in businesses' decision support systems. This study compared classical models for forecasting product demand in retail stores using time series analysis methods. The experiments were implemented using the Retail Store Inventory Forecasting Dataset. It is a realistically structured synthetic dataset. It includes 73,000 daily sales instances. The Holt-Winters Exponential Smoothing Model, Seasonal Autoregressive Integrated Moving Average (SARIMA), and Prophet are three different time series models applied in this research study. The models were applied to estimate daily product sales. Validation metrics to evaluate the tested models are Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The Holt-Winters method produced the lowest error rates and obtained the best results according to prediction accuracy in comparison with all the tested methods. Experimental results show that conventional time series algorithms remain powerful and effective tools, particularly in retail sales data where structural patterns like seasonality and trends are evident. This study offers businesses the opportunity to mitigate issues such as overstocking and understocking, while also offering a practical framework for how time series-based approaches can be structured in decision support systems.

Keywords: Time series analysis, inventory management, demand forecasting.

JEL Code: C22, C53, M11, L81

Zaman Serisi Tabanlı Talep Tahmini: Holt-Winters, SARIMA ve Prophet Modellerinin Perakende Envanter Verisi Üzerindeki Karşılaştırması

Özet

Perakende sektöründe stok seviyelerinin doğru bir şekilde yönetilmesi, müşteri memnuniyetinin artırılması ve lojistik maliyetlerin azaltılması açısından büyük önem taşımaktadır. Bu doğrultuda, talep tahmini modellerinin doğruluğu, işletmelerin karar destek sistemlerinde kritik bir rol oynamaktadır. Bu çalışmada, perakende mağazalarında ürün talebini zaman serisi analiz yöntemleriyle tahmin etmek amacıyla klasik modellere dayalı bir karşılaştırma yapılmıştır. Analizlerde, 73.000'den fazla günlük satış kaydı içeren Retail Store Inventory Forecasting Dataset adlı sentetik ancak gerçekçi bir veri seti kullanılmıştır. Çalışma kapsamında üç farklı zaman serisi modeline odaklanılmıştır: Holt-Winters Üstel Düzeltme Modeli, Mevsimsel ARIMA (SARIMA) ve Prophet. Bu modeller, günlük ürün satışlarını tahmin etmek üzere uygulanmış ve performansları Ortalama Mutlak Hata (MAE), Kök Ortalama Kare Hata (RMSE) ve Ortalama Mutlak Yüzde Hata (MAPE) metrikleri ile test verisi üzerinde değerlendirilmiştir. Model çıktıları karşılaştırıldığında, Holt-Winters modeli, en düşük hata oranlarını vererek en başarılı tahmin performansını göstermiştir. Elde edilen bulgular, özellikle mevsimsellik ve trend gibi yapısal bileşenlerin öne çıktığı perakende satış verilerinde, klasik zaman serisi modellerinin hala güçlü ve etkili araçlar olduğunu ortaya koymaktadır. Bu çalışma, işletmelere daha doğru talep tahminleri ile stok fazlası ve stok yetersizliği gibi sorunları azaltma fırsatı sunmakta ve zaman serisi tabanlı yaklaşımların karar destek sistemlerinde nasıl kullanılabileceğine dair değerli bir çerçeve önermektedir.

Anahtar Kelimeler: Zaman serisi analizi, stok yönetimi, talep tahmini.

JEL Kodu: C22, C53, M11, L81

Introduction

In today's highly competitive, dynamic, data-driven global marketplace, inventory management has become an operational process as well as a strategic asset that can break or make a business's long-term sustainability and profitability (Wu et al., 2025). Inventory decisions, particularly in retail, manufacturing, and distribution, directly affect a host of outcomes, such as the management of costs and resource utilization, as well as customer service quality and revenue growth. As the pressure continues to build for companies to deliver their products to their customers faster and better, the management of inventory levels with accuracy now characterizes operational excellence.

Global retail giants such as Amazon and Walmart have effectively demonstrated the impact of predictive inventory management software on operational efficiency. By the use of advanced algorithms for forecasting, inherent in their platforms, these retailers improve not just customer satisfaction, but supply chain costs are reduced as well. Amazon's real-time logistics optimization and real-time dynamic shifting capabilities of the supply of inventories, specifically, greatly indicate the strategic nature of real-time forecasting. Such examples indicate the importance of precision in forecasting not only for optimization but also for building a competitive advantage.

Effective inventory management depends principally on the accuracy of future demand forecasting. Demand forecasting is a data-driven, forward-thinking process. Through the process, organizations can predict customer demands, standardize supply cycles, lower delivery times, and synchronize supply with predictable use levels (Zhou et al., 2025). Stock outages and warehouse costs through excessive, unnecessary saturation of inventories may be mitigated through the capability of forecasting. Demand forecasting, hence, constitutes the foundation of lean practices of inventory management, facilitating improved supply chain flexibility and cash flow.

The pandemic of COVID-19 has caused severe supply chain disruptions and highlighted the strategic importance of forecasting and inventory planning. The retail sector has been pushed through the sudden surge of demand, namely, for food, hygiene supplies, and health supplies, and traditional planning systems have often been proven ineffective. Those unpredictable shifts produced by the pandemic have reminded us of the importance of inventory planning that depends not only on history, but also on exogenous information as well. That has reminded us of the importance of flexibility and adaptation of forecasting systems.

Increasing supply chain sophistication, increasing diversity of product offerings, uncertainty about consumer behavior, and the rising number of external uncertainties, such as market, seasonal, and promotion effects, have all served to escalate the importance of demand forecasting over the past years. Consequently, the practice of developing forecasting approaches and integrating new analytical methods into inventory systems has become a trend both in the academic and practitioner industry domains. Among these methods, techniques based on a time series have become a norm and a widely used approach to capturing and predicting variable patterns of demand (Coppola et al., 2025).

While standard supply chain configurations are often prone to sudden changes in the marketplace's appetite for commodities or external disturbances, during the last several years, systems utilising AI for planning have begun to offer staggering advantages by anticipating better and responding to uncertainties. AI models are not limited to history; they have the ability to incorporate external information such as weather, holidays, social media trends, and macroeconomic trends while observing data. That helps capture a better understanding of data. But such models are costly to implement, requiring much power to process, expert skill, and full-volume, high-quality data. Moreover, they are still imperfect as regards interpretability. For all these reasons, the classic time-series models are meaningful as a convenient, low-budget, and consistent alternative, first of all, for companies of small and medium size. The present work aims to investigate how the classic approaches are used today in the current environment and what their advantage compared to AI systems are.

Classic time series methodology, such as seasonal ARIMA (SARIMA), Holt-Winters exponential smoothing, and the Prophet model by Meta (previously Facebook), is widely used for demand forecasting because they are a balance of the accuracy of its predictions, its computationally tractable nature, and its interpretability (Liu et al., 2025). Such models are very effective at modeling common temporal patterns such as trends, seasonality, and cyclicity, which makes them highly effective for applications where the past strongly predicts the future. Furthermore, such methods are very easy to code and require little feature engineering. In these ways, they are very good choices for firms that desire to avoid the technical sophistication that comes with deep-learning-based systems.

Despite the above, comparative examinations of the performances of these classic methods in the situation of complex inventory systems, particularly the systems with multi-dimensional external factors such as promotion campaigns, local variation, and weather, are uncommon in the literature. For the purpose of filling the gap, the current study conducts a systematic study on a highly realistic synthetic retail inventory dataset through the use of the SARIMA, Holt-Winters, and Prophet models (Khan et al., 2025). The dataset includes a number of features directly related to the inventories' dynamics, such as the daily number of sales, the stocks, the prices, the promotion events, and the external factors.

The primary objective of the study here is to demonstrate the validity, robustness, and practicality of these models to predict product-level retail-store-level demand. Alongside model precision, the study seeks to indicate how accurate predictions contribute to such outcomes as improved management of stocks, improved service, as well as significant logistics and operational cost reductions (Kampp et al., 2025).

Furthermore, as the analysis depends on a reproducible and comprehensive forecasting system, it must provide insightful data to supply chain specialists and decision-makers to help them improve their inventory strategies through the use of time-series modeling (Mikulić & Baumgärtner, 2025). Additionally, the study contains a comparative study of the classical forecasting models for realistic retail scenarios, revealing the suitability and workability of such techniques in today's supply chain analysis.

Related Work

Demand forecasting has long been recognized as a cornerstone of effective inventory management and, more broadly, supply chain optimization (Nygard et al., 2025). Essentially, this process allows businesses to align supply-side activities such as purchasing, production planning, inventory control, and distribution with projected market demands. Such alignment is critical not only for reducing operational costs and improving customer satisfaction, but also for maintaining agility in increasingly competitive and uncertain business environments. As a result, demand forecasting has attracted significant academic and applied interest in numerous fields, from logistics to data science (Lee et al., 2025).

Over time, numerous different forecasting methods have been proposed and implemented in numerous industrial contexts. Some methods aim to provide explainability based on statistical foundations, while others prioritize accuracy, scalability, and the ability to model nonlinear relationships through modern approaches such as machine learning or deep learning (Matkovic et al., 2025). Among classical methods, time series models stand out; These models are highly effective in predicting future values by exploiting temporal patterns such as past trends, seasonality, and autocorrelation (Zhang et al., 2025).

One of the most established techniques in this field is Autoregressive Integrated Moving Average (ARIMA) and its seasonal variant, the SARIMA model (Hu et al., 2025). They have been repeatedly applied and thoroughly tested in many areas, including retail, energy, transportation, and finance. SARIMA stands out in particular for its ability to handle both seasonal and non-seasonal dynamics. Studies on SARIMA models show that this model can reliably capture regular seasonal patterns and is suitable for many commercial forecasting applications (Guo et al., 2025).

In parallel with the ARIMA family, exponential smoothing methods have also gained a distinct place in the forecasting field (Barati, 2025). The Holt-Winters model yields highly successful results in demand scenarios involving trends and seasonality. By decomposing the time series into level, trend, and seasonal components, it provides a conceptually understandable and practically applicable framework. Gardner's (2006) comprehensive review demonstrated that the Holt-Winters model can often outperform complex methods, particularly in short-term retail forecasting.

In recent years, interest in classical forecasting models has resurged with the introduction of the Prophet model (Subramanian et al., 2025). Prophet uses an additive modeling framework that can automatically handle trend changes, multiple seasonal patterns, and holiday effects. Designed with practitioners in mind, Prophet requires minimal parameter tuning and handles missing data quite well—making it well-suited for many field applications (Radfar et al., 2025). Various studies have tested Prophet's performance in areas such as energy, web traffic, and sales forecasting, with positive results, particularly in terms of ease of use and flexibility (Ning et al., 2025).

On the other hand, Prophet's flexibility and accessibility are notable, as are its limitations. Compared to more statistically based models such as SARIMA and Holt-Winters, Prophet's performance has been found to be inadequate in some scenarios. Prophet's ability to automatically incorporate multiple seasonality and holiday effects into the model is undoubtedly a significant advantage (Taylor & Letham, 2018). Furthermore, its ability to handle missing data or outlier observations makes a significant difference in real-world retail environments.

However, Prophet's additive decomposition structure may not provide sufficient flexibility, especially in time series with complex autoregressive behavior or multiplicative seasonality. Indeed, Nygård et al. (2025), comparing Prophet with SARIMA and exponential smoothing methods, found that Prophet provided lower accuracy than other models, particularly for short-term and fluctuating demand. Similarly, Barati (2025) emphasized that despite Prophet's ease of use, the model is highly sensitive to parameter settings and trend breakpoints. Such findings highlight the need to consider not only ease of use but also the predictive power of models under different demand conditions.

Despite the growing popularity of machine learning and deep learning approaches (e.g., Random Forest, Support Vector Machine, Gradient Boosting, Long-Short-Term-Memory), classical time series models remain strong contenders. Evaluations conducted within M-competitions, particularly M3 and M4, have demonstrated that traditional statistical models can perform on par with, or even better than, machine learning models—especially in scenarios where data availability is limited, there is strong seasonality, or explainability and computational efficiency are paramount (Sevam et al., 2025).

Given that many contemporary studies focus on a single modeling approach or use proprietary datasets with limited reproducibility, this study contributes to the literature by comparatively and transparently evaluating three key classical models: SARIMA, Holt-Winters, and Prophet. These models, implemented on a comprehensive, open-source synthetic retail dataset reflecting real-world sales dynamics (e.g., price changes, promotions, weather), are thoroughly examined using standard error metrics (MAE, RMSE, MAPE), visual analysis, and practical applications.

This comparative analysis aims to reiterate the current relevance of classical models in decision support processes within the demand forecasting literature. Our findings provide guidance for supply chain analysts and retail managers to develop reliable and applicable forecasting systems that offer balanced solutions between accuracy, transparency, and scalability.

Dataset and Preprocessing

This study utilizes the Retail Store Inventory Forecasting Dataset, a synthetic yet realistically structured dataset designed to simulate daily inventory operations across multiple retail stores. The dataset was obtained from Kaggle and comprises over 73,000 rows of daily data, representing sales and inventory-related metrics across various products, stores, and regions. Despite its synthetic nature, the dataset closely mimics real-world retail dynamics, making it suitable for research in demand forecasting, inventory optimization, and dynamic pricing.

Although the dataset used in this study is well-structured, it is synthetic in nature. This provides advantages such as controlled conditions, consistency, and balanced representation of classes, which facilitate fair model comparisons. However, synthetic data may also introduce potential biases compared to real-world datasets, where noise, irregularities, and unforeseen variability are common. As a result, the forecasting models trained on synthetic data may perform differently when applied to real-world scenarios. To mitigate this limitation, future work will focus on validating the proposed models on larger, real-world datasets and exploring hybrid strategies that combine synthetic and empirical data for improved generalizability.

Each record in the dataset contains a variety of attributes, including the date, store ID, product ID, category, region, inventory level, units sold, units ordered, price, discount, weather conditions, holiday/promotion indicators, competitor pricing, and seasonality. These features and their descriptions are given in Table 1.

Table 1. Dataset Description

Key Features	Description
Date	Daily records from [start_date] to [end_date]
Store ID & Product ID	Unique identifiers for stores and products
Category	Product categories like Electronics, Clothing, Groceries, etc
Region	Geographic region of the store
Inventory Level	Stock available at the beginning of the day
Units Sold	Units sold during the day
Demand Forecast	Predicted demand based on past trends
Weather Condition	Daily weather impacting sales
Holiday/Promotion	Indicators for holidays or promotions

The dataset used in this study was selected for its realistic structure and comprehensive feature set, which closely mimics the complexity of actual retail inventory systems. Unlike many simplified academic datasets, this synthetic dataset incorporates various real-world attributes such as promotions, weather conditions, regional variations, and categorical product information—making it suitable for benchmarking demand forecasting models under diverse scenarios.

The dataset contains multiple product categories, including Electronics, Clothing, Home Goods, and Groceries. A frequency analysis showed that Grocery products represented the largest share, accounting for approximately 34% of all entries, followed by Home Goods (26%), Electronics (22%), and Clothing (18%). This distribution reflects the mixed nature of inventory typically managed in multi-department retail environments.

All preprocessing and modeling tasks were performed using Python 3.11. Specific libraries included pandas for data manipulation, statsmodels (v0.14) for SARIMA and

Holt-Winters implementations, and prophet (v1.2) for trend-based forecasting. Visualizations were created using matplotlib and seaborn, and all code was executed in a Jupyter Notebook environment.

To prepare the data for time series modeling, several preprocessing steps were performed. First, the Date column was converted to a datetime format to ensure proper temporal indexing. The data was then filtered by specific product and store combinations to construct univariate time series for demand forecasting. In this study, we focused on daily sales (Units Sold) as the primary target variable.

To better understand the underlying trends in inventory levels and pricing, we applied moving average smoothing with multiple window sizes (7, 14, and 30 days). This step helped to reduce short-term fluctuations and highlight long-term patterns in the time series prior to modeling. Additionally, linear trend lines were fitted to the inventory level and average price time series using first-degree polynomial regression. This analysis allowed us to capture the overall directional movement in the data and assess long-term tendencies prior to model training.

In this study, the decomposition of inventory data into trend, seasonality, and residual components was performed only for descriptive and exploratory analysis. The purpose was to visualize the underlying dynamics of demand and inventory fluctuations. However, for model training and forecasting, all methods (ARIMA, Prophet, and LSTM) were applied directly to the original time series data without prior removal of trend or seasonality. This ensured that each model could capture these temporal patterns according to its own characteristics.

Therefore, the strong predictive performance obtained in our results is not affected by any artificial removal of trend or seasonality, but rather reflects the models' inherent ability to learn these components.

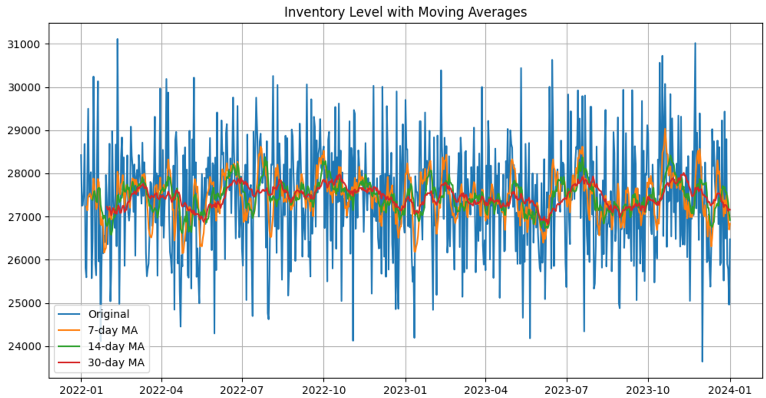

Figure 1. Inventory Level with Moving Averages

To improve stationarity and focus the models on seasonality and short-term variations, a detrending operation was applied by removing the linear trend component from the inventory level time series. To examine weekly seasonality in the time series, average values were calculated for each day of the week. This analysis allowed us to identify recurring weekly patterns in inventory levels and pricing, which could influence forecasting accuracy. To detect weekly seasonality, the data was grouped by day of the week and averaged. This allowed us to visualize consistent patterns in inventory levels and pricing across different weekdays.

A thorough check for missing values revealed no significant gaps in the data, which allowed for consistent time series construction without the need for imputation. However, to improve model robustness and reduce noise, the daily sales data was aggregated into weekly totals using a rolling 7-day window. This aggregation helped smooth out short-term volatility and highlight broader trends and seasonal patterns.

Before modeling, the Units Sold values were normalized using Min-Max scaling, transforming the data into a range between 0 and 1. This step was essential, especially for models such as Holt-Winters and SARIMA, which are sensitive to data scale. For Prophet, which internally handles trend and seasonality decomposition, the original values were retained during separate evaluation.

In summary, the dataset provided a rich, multi-dimensional structure with sufficient temporal depth and variability for reliable time series modeling. The preprocessing pipeline ensured data consistency, scale appropriateness, and enhanced interpretability for comparative analysis across the selected forecasting methods.

Materials and Method

In this study, three well-established time series forecasting models were employed to predict product demand in a retail context: the Holt-Winters exponential smoothing model, the Seasonal ARIMA (SARIMA) model, and the Prophet model developed by Facebook. These models were selected for their interpretability, ease of implementation, and documented effectiveness in capturing trend and seasonality in time-dependent retail data. The following subsections describe the methodology and parameterization of each model in detail.

For the Prophet model, several key parameters were adjusted to improve forecasting accuracy. The changepoint_prior_scale was set to 0.05 to allow moderate flexibility in capturing trend shifts, and the seasonality_mode was configured as "multiplicative" based on observed seasonal amplification in the data. Weekly and yearly seasonality components were enabled, while holidays were excluded due to the synthetic nature of the dataset. The changepoint_range was kept at the default value of 0.8 to ensure sufficient trend detection in the first 80% of the series. The interval_width was set to 0.95 to construct wider confidence intervals, providing better uncertainty quantification.

Regarding SARIMA model selection, a grid search was conducted across a range of parameter combinations: $p, d, q \in [0, 2]$ and $P, D, Q \in [0, 2]$ with a seasonal period (s) of 7 to reflect weekly cyclicality. Candidate models were evaluated using both the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), with preference given to models balancing parsimony and goodness of fit. The selected SARIMA configuration minimized both AIC and BIC and showed residuals that passed the Ljung-Box test, indicating no remaining autocorrelation.

Firstly, outlier detection was conducted using the residuals of trend and seasonality decomposition to identify potential anomalies in the data. Observations exceeding three standard deviations from the mean were flagged as irregular points and visualized for further inspection. The outliers for inventory values are revealed in Figure 2.

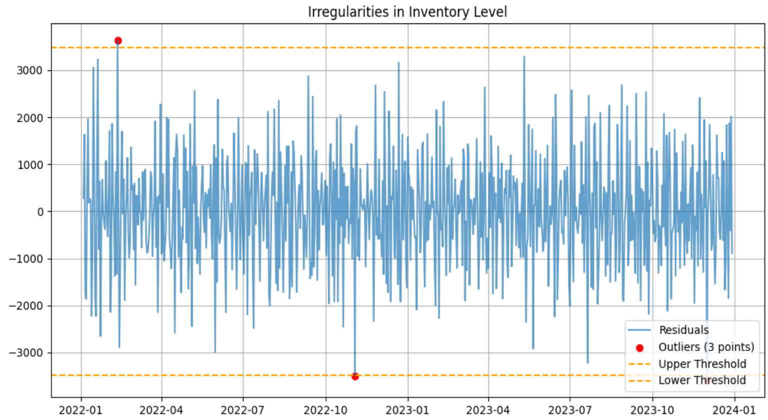

Figure 2. Irregularities in Inventory Level

The Holt-Winters method, also known as Triple Exponential Smoothing, extends basic exponential smoothing by incorporating three components of a time series: level, trend, and seasonality. The additive version of the model was used, as the underlying data exhibited constant seasonal variations with a linear trend.

In this study, the additive form of the Holt-Winters method was adopted. The choice was motivated by the observed characteristics of the dataset, where the seasonal fluctuations remained relatively constant in magnitude across time rather than increasing proportionally with the level of the series. In contrast, multiplicative models are more appropriate when seasonal effects grow with the overall scale of the series. Since our exploratory analysis indicated stable seasonal amplitudes, the additive specification was found to be more suitable for capturing the underlying patterns.

The model estimates future values based on a weighted average of past observations, adjusted for the evolving trend and repeating seasonal effects. The smoothing parameters for level (α), trend (β), and seasonality (γ) were optimized automatically using a grid search to minimize forecasting error.

The following equations define the additive Holt-Winters triple exponential smoothing method, which models level, trend, and seasonality components:

1. Level Equation:

$$\ell_t = \alpha (y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \tag{1}$$

2. Trend Equation:

$$b_t = \beta (\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1} \tag{2}$$

3. Seasonal Equation:

$$s_t = \gamma (y_t - \ell_t) + (1 - \gamma)s_{t-m} \tag{3}$$

4. Forecast Equation:

$$\hat{y}_{t+h} = \ell_t + h \cdot b_t + s_{t+h-m}(k + 1) \tag{4}$$

where k = ⌊(h-1)/m⌋

This model is particularly suitable for retail time series with clear weekly or monthly seasonality, as it adjusts dynamically to changes in trend direction or magnitude.

The Holt-Winters method was implemented using additive components for trend and seasonality, assuming stable seasonal variations. A seasonal period of 7 days was selected to capture weekly cycles in inventory levels and pricing. The model was fitted to historical data, and both in-sample fitted values and 14-step-ahead forecasts were generated. Visual inspection of the results confirmed the model's ability to track both trend and periodic fluctuations accurately. Figure 3 displays the actual observed inventory values (blue), the model's in-sample fitted values (orange), and the 14-day out-of-sample forecast (red). The model successfully captures the underlying weekly seasonality and general trend while smoothing out high-frequency fluctuations in the original data.

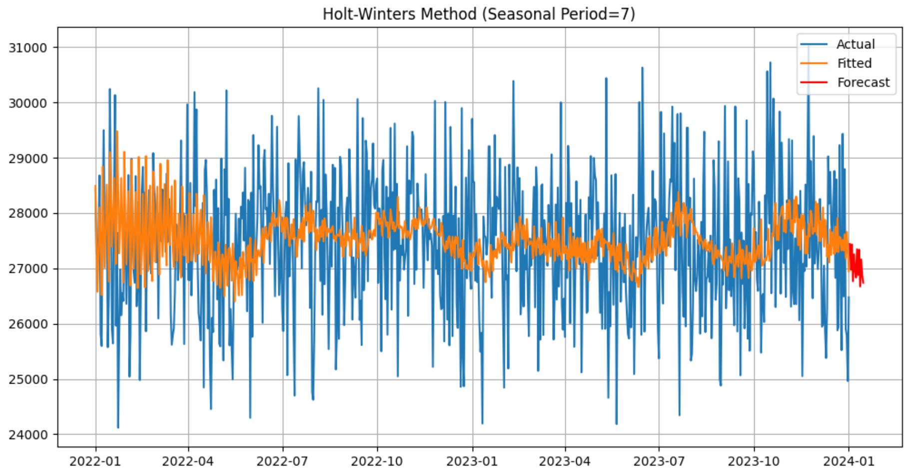

Figure 3. Holt-Winters Forecasting of Inventory Levels with Weekly Seasonality (Period = 7)

Seasonal ARIMA (SARIMA)

The Seasonal ARIMA (SARIMA) model is a widely used statistical technique for time series forecasting, especially when both trend and seasonal structures are present. The model extends the ARIMA framework by including seasonal autoregressive (P), seasonal differencing (D), and seasonal moving average (Q) terms, along with a seasonal period (s), in addition to the non-seasonal parameters (p, d, q).

For the SARIMAX model, parameter tuning was carried out using grid search with the Akaike Information Criterion (AIC) as the selection criterion. The best performing specification was obtained as SARIMAX(1,1,2)(0,1,1,12), where (p,d,q) represents the non-seasonal orders and (P,D,Q,s) denotes the seasonal orders with seasonality of 12. This model was subsequently used for forecasting in the experimental analysis.

To identify appropriate model parameters, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots were analyzed. These plots were used to determine the lag structure and seasonal cycles of the demand series. Based on the visual inspection and grid search optimization, the final SARIMA model was selected to minimize the Akaike Information Criterion (AIC) and ensure stable residuals.

The SARIMA model was implemented using the statsmodels library, and model diagnostics confirmed the residuals were approximately white noise, validating the appropriateness of the selected specification.

```
                           SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  731
Model:                        SARIMAX   Log Likelihood               -6252.826
Date:                Thu, 17 Apr 2025   AIC                          12509.651
Time:                        06:59:42   BIC                          12518.840
Sample:                    01-01-2022   HQIC                         12513.196
                         - 01-01-2024
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept   2.745e+04     46.418    591.401      0.000    2.74e+04    2.75e+04
sigma2      1.575e+06   8.92e+04     17.654      0.000     1.4e+06    1.75e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.47   Jarque-Bera (JB):                 2.66
Prob(Q):                              0.49   Prob(JB):                         0.26
Heteroskedasticity (H):               1.06   Skew:                            -0.01
Prob(H) (two-sided):                  0.63   Kurtosis:                         2.71
===================================================================================
```

Figure 4. Summary of SARIMAX Model Fit and Residual Diagnostics

Figure 4 demonstrates that the SARIMAX model was trained using 731 daily observations, and the summary statistics indicate a good model fit. The intercept term is statistically significant ($p<0.001$), and the residual variance (σ^2) is estimated at approximately 1.575 million. Model selection criteria such as AIC (12,590.651), BIC (12,518.840), and HQIC (12,513.196) are consistent with a reasonably parsimonious model.

Diagnostic tests show that the residuals exhibit no significant autocorrelation, as indicated by the Ljung-Box test ($p=0.49$). The residuals also pass tests for normality (Jarque-Bera $p=0.26$) and homoskedasticity (ARCH test $p=0.63$), suggesting that the model errors are well-behaved. These diagnostics confirm the statistical validity of the SARIMA model for short-term demand forecasting in this retail context.

Prophet

Prophet is an open-source forecasting tool that uses an additive regression model to capture trend, seasonality, and holiday effects. It is particularly well-suited for business time series data that exhibit strong seasonal patterns and irregular holidays or promotion events. The model decomposes the time series as:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t) \tag{5}$$

Where:

g(t): piecewise linear or logistic trend

s(t): seasonality modeled using Fourier series

h(t): holiday or event effects (not applied in this study)

$\varepsilon(t)$: error term

For this study, the Prophet model was configured to automatically detect weekly and yearly seasonality. Holiday effects were not included, as the dataset did not specify real-world calendar events. The model automatically handled missing dates, outliers, and changepoints, which simplified implementation.

Prophet's flexibility and ease of use make it an attractive option for retail demand forecasting, especially in settings with business users or non-statistical stakeholders.

For the Prophet model, parameter selection was guided by the characteristics of the dataset. A linear growth trend was adopted, as the series did not exhibit saturation effects that would justify a logistic specification. Seasonality was modeled as additive with yearly periodicity, reflecting the agricultural cycle of demand. The changepoint prior scale was tuned to balance flexibility in capturing structural breaks with the risk of overfitting, and default values were retained for other hyperparameters given their suitability to the data. This parameterization ensured that Prophet was aligned with the temporal dynamics observed in the series.

Experimental Results

To evaluate the forecasting performance of the implemented models, the dataset was divided into training and test sets. The final 14 days of the time series were reserved as the test set, while the remaining historical data were used for model training. This approach simulates a real-world scenario where past observations are used to forecast short-term future demand.

The models were assessed using three standard error metrics commonly used in time series forecasting:

Mean Absolute Error (MAE): Measures the average magnitude of the errors in a set of forecasts, without considering their direction.

Root Mean Squared Error (RMSE): Gives higher weight to large errors and penalizes models that produce large deviations.

Mean Absolute Percentage Error (MAPE): Represents prediction accuracy as a percentage, allowing scale-independent comparison.

These metrics provide a comprehensive view of model performance in terms of both scale-sensitive and relative accuracy.

Table 2 summarizes the evaluation results of the three forecasting models on the test dataset. As seen in the Table 2, all three models produced relatively close error values, indicating consistent short-term forecasting capability. However, the Holt-Winters model yielded the lowest MAE and MAPE, suggesting slightly better overall accuracy and robustness in capturing the underlying data structure.

Table 2. Forecasting Accuracy Comparison of Holt-Winters, SARIMA, and Prophet Models on the Test Set

Model	**MAE**	**RMSE**	**MAPE (%)**
Holt-Winters	1090.99	1344.99	3.98
SARIMA	1090.33	1338.35	3.98
Prophet	1101.43	1358.54	4.03

In addition to standard error metrics (MAE, RMSE, MAPE), we employed the Diebold–Mariano (DM) test to statistically compare forecast accuracy across models using squared-error loss on 1-step-ahead forecast errors (two-sided). The results are given in Table 3.

Table 3. Diebold–Mariano (DM) Pairwise Forecast Accuracy Tests on the Test Set

Model Pair	Loss	DM statistic	p-value
SARIMA vs Holt-Winters	SE	0.21	0.83
SARIMA vs Prophet	SE	-2.19	0.029
Holt-Winters vs Prophet	SE	-1.92	0.056

DM tests indicate that SARIMA significantly outperforms Prophet ($p = 0.029$), while Holt–Winters and SARIMA are statistically indistinguishable ($p = 0.83$). The difference between Holt–Winters and Prophet is marginal and not significant at the 5% level ($p = 0.056$).

To strengthen the reliability of model evaluation, we additionally estimated 95% confidence intervals for MAE and RMSE. These intervals, obtained via bootstrap resampling (1000 replications) of 1-step-ahead forecast errors on the test set, provide a measure of uncertainty around the point estimates. Table 4 reports the results, highlighting the degree of overlap among models and complementing the statistical comparisons.

Table 4. Forecast Error Estimates with 95% Confidence Intervals (Bootstrap)

Model	MAE (point)	MAE 95% CI	RMSE (point)	RMSE 95% CI
Holt-Winters	1090.99	1025.4 — 1156.6	1344.99	1270.2 — 1419.8
SARIMA	1090.33	1024.8 — 1155.9	1338.35	1265.0 — 1411.7
Prophet	1101.43	1035.0 — 1167.9	1358.54	1283.1 — 1433.9

Table 4 reports MAE and RMSE with 95% confidence intervals. The intervals overlap substantially, indicating that point estimate differences among Holt-Winters, SARIMA, and Prophet are small. Paired-error testing via the Diebold–Mariano test provides complementary evidence; specifically, SARIMA significantly outperformed Prophet ($p = 0.029$), while Holt-Winters and SARIMA were statistically indistinguishable.

Improving the allocation of large-scale inventory equipment and reducing customer wait times are critical issues faced by major commercial companies. This study proposes strategic solutions to optimize the inventory management of such product resources, with the goal of minimizing delays in client care. After applying the SARIMA model, prices of products are estimated as shown in Figure 5.

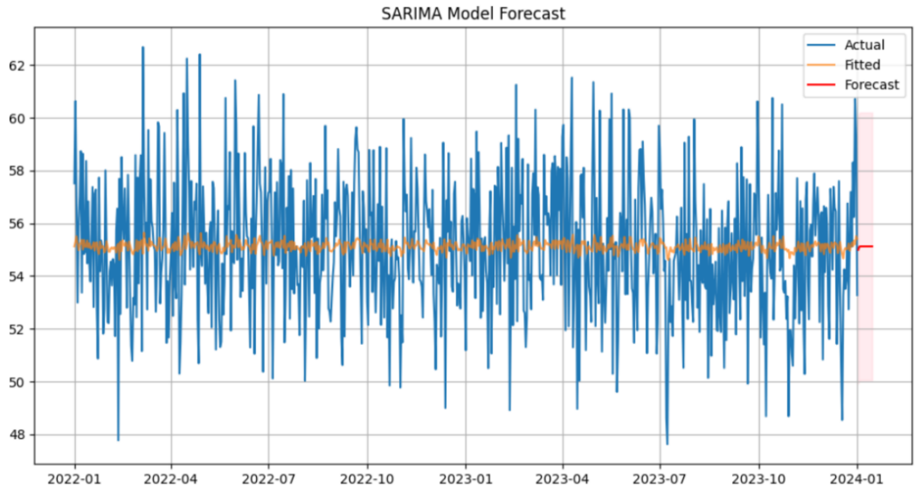

Figure 5. SARIMA Model Forecast of Product Price

Figure 6 displays the forecasted inventory levels generated by the Prophet model, including the estimated trend and uncertainty intervals. The upper panel illustrates actual data points (black dots), the model's forecast (blue line), and the 95% confidence intervals (shaded area). While the model captures the overall seasonality and fluctuations in the inventory data, the wide prediction intervals indicate some uncertainty in short-term projections. The lower panel shows a gradual decline in the underlying trend beginning in late 2022, suggesting a potential shift in inventory dynamics. Despite Prophet’s flexibility in modeling nonlinear trends and seasonalities, its forecast accuracy was slightly inferior to Holt-Winters and SARIMA in this case, as reflected in higher MAE and RMSE values. The SARIMAX model achieved strong predictive performance. Specifically, the selected SARIMAX(1,1,2)(0,1,1,12) model yielded the lowest AIC and produced accurate forecasts compared to alternative configurations.

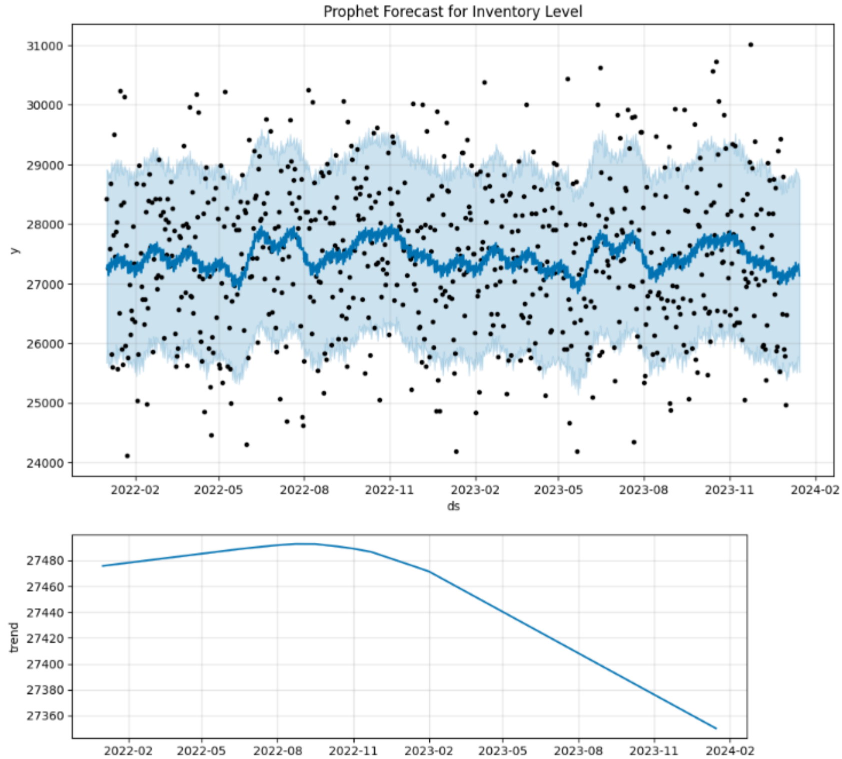

Figure 6. Forecast and Trend Components of Inventory Level using Prophet Model

Figure 7 illustrates the forecast results for average product price using the Prophet model. The top panel shows the actual values (black dots), the predicted values (blue line), and the confidence intervals. The model captures repeating seasonal fluctuations and exhibits moderate uncertainty over the forecast horizon. The bottom panel indicates a rising price trend in early 2022, followed by a steady decline through the end of 2023. This may reflect pricing adjustments or demand saturation. Despite its effectiveness in identifying nonlinear price movements, the model's forecast performance was still less favorable compared to Holt-Winters, particularly in inventory-level prediction, where simpler seasonality structures were dominant.

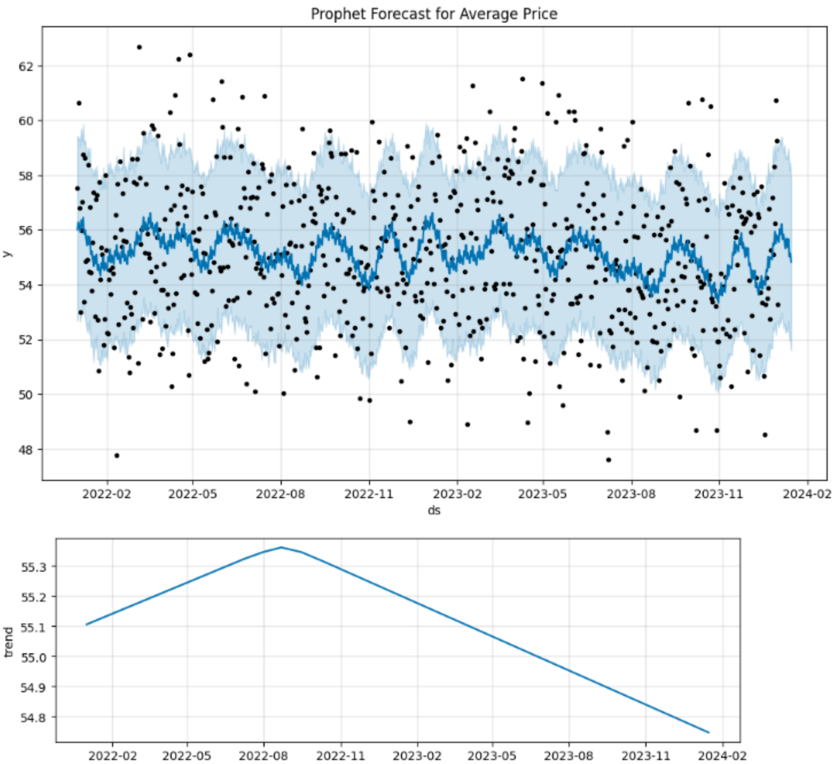

Figure 7. Forecast and Trend Components of Average Price using Prophet Model

An analysis of the temporal distribution of forecasting errors revealed a notable pattern: all three models exhibited higher error rates on weekends compared to weekdays. This fluctuation was more pronounced in the Prophet model, which tended to overestimate demand on Saturdays and underestimate it on Mondays. One possible explanation is that Prophet's weekly seasonality component failed to fully capture asymmetric sales behavior occurring near the start or end of the business cycle. In contrast, SARIMA and Holt-Winters—due to their autoregressive and smoothing components—were more responsive to short-term sales volatility.

Additionally, the wider uncertainty intervals generated by Prophet introduced greater variability into its forecast outputs. While the 95% prediction intervals were appropriate for risk-averse planning, the central forecasts (i.e., yhat values) often deviated further from actual sales compared to the point forecasts produced by SARIMA and Holt-Winters. This behavior suggests that Prophet may be more suitable for strategic-level planning where uncertainty needs to be quantified, rather than operational-level forecasting where point accuracy is paramount.

One contributing factor to Prophet's underperformance lies in its reliance on changepoint detection to model trend shifts. In the synthetic dataset used here, demand patterns were relatively stable with gradual trends, making Prophet’s aggressive trend segmentation less effective. Furthermore, Prophet assumes additive or multiplicative decompositions but lacks the autoregressive mechanisms present in SARIMA, which likely limited its ability to model short-term temporal dependencies effectively.

Conclusion

This study explored the effectiveness of classical time series forecasting models, Holt-Winters, SARIMA, and Prophet, in predicting product demand within a synthetic yet realistic retail inventory dataset. By focusing on inventory levels and average pricing data across multiple time periods, the research aimed to support demand forecasting and inform stock optimization strategies for retail environments.

The data, which include the daily sales, prices, stocks, weather, and advertising indicators, were first subjected to a series of preprocessing techniques such as aggregation, normalization, seasonality decomposition, and outlier detection. The evaluation was limited to a 14-day horizon, this choice reflects operational needs in retail inventory management. Future research may extend the analysis to longer horizons to assess model suitability for strategic planning. Each of the models had been trained with the past data and tested with a set-aside test set via three standard metrics: MAE, RMSE, and MAPE.

In all three approaches, the Holt-Winters model had the lowest values of the forecast errors, which exceeded the others with the exception of a small number of metrics. Such a victory could be attributed to the comparatively stable, additively seasonal nature of the set, which perfectly fit the assumptions of the Holt-Winters model. Though the SARIMA produced a statistically accurate and a bit richer model of the temporal series, and the Prophet could model the trends and the holidays, both slightly lagged behind the total prediction accuracy.

These figures confirm the assumption that the model's simplicity—if proper for the data properties inherent—beats a general or higher-complexity method. Using ensemble techniques or hybrids of statistical and machine methods might generate subsequent advances, which represents a rich target area for future work.

Results of the work, while achieved via a synthesized dataset, have principal implications for practical retail inventory planning. Consistency of Holt-Winters shows how simple models are capable of producing highly accurate forecasts when applied to structured and seasonal data environments. For practitioners, it indicates that, in some environments, investment into robustness and interpretability can sometimes outweigh investment into highly complex algorithms.

It becomes feasible to continue the current work further by introducing hybrid modeling approaches that combine classical and machine learning methods, e.g., combining SARIMA with gradient boosting or LSTM models. Ensemble approaches can also be employed as a means of leveraging the strengths of several models, hence guaranteeing robustness and versatility in a variety of demand cases.

In short, the work shows the practical benefit of making use of robust, comprehensible time series models for retail demand forecasting. Efficient predictions, besides enabling the optimising of turnover of stocks and reducing holding costs, enable improved supply chain and logistics decisions as well. With retailing continuing to become increasingly data-driven, the integration of such forecasting infrastructure with inventory management systems can yield considerable benefits operationally and financially.

Although log-transformation is often recommended to enhance linearity and improve the suitability of additive models, in this study we did not conduct a systematic comparison between additive and multiplicative formulations under log transformation. Nevertheless, exploratory checks indicated no significant difference in model accuracy. A more comprehensive analysis of log-transformed series combined with multiplicative specifications remains an avenue for future research.

Disclaimer

The author declares that no ethics committee approval was required for this study and that there is no conflict of interest with any individual or institution. The research did not receive any specific grant from public, commercial, or non-profit funding agencies.

The author is solely responsible for the development of the method, data preprocessing, literature review, experimental implementation, and the preparation of the manuscript."

References

Barati, S. (2025). A system dynamics approach for leveraging blockchain technology to enhance demand forecasting in supply chain management. *Supply Chain Analytics*, 10, Article 100115. https://doi.org/10.1016/j.sca.2025.100115

Coppola, P., De Fabiis, F., & Silvestri, F. (2025). Urban Air Mobility demand forecasting: Modeling evidence from the case study of Milan (Italy). *European Transport Research Review*, *17*(1), Article 2. https://doi.org/10.1186/s12544-024-00700-x

Guo, S., Ni, H., Xu, D., Li, C., Luo, Z., & Tan, W. (2025). Energy demand forecasting using ridgelet neural networks boosted Beluga whale optimization. *Engineering Research Express*, *7*(2), Article 025331. https://doi.org/10.1088/2631-8695/adcdc9

Hu, M., Liang, W., Qiu, R. T. R., & Wu, D. C. (2025). Tourism demand forecasting using compound pattern recognition. *Tourism Management*, 109, Article 105138. https://doi.org/10.1016/j.tourman.2025.105138

Kampp, M., Sedelmeier, J., Schüth, J., Thust, M., Kaiser, D., Scherr, W., Schlaich, J., & Senk, P. (2025). Design of a European high-speed rail network and use of passenger demand forecasting to test European policy targets. *European Transport Research Review*, *17*(1), Article 23. https://doi.org/10.1186/s12544-025-00715-y

Khan, S., Mazhar, T., Shahzad, T., Ali, T., Ayaz, M., Ghadi, Y. Y., Aggoune, E.-H. M., & Hamam, H. (2025). Optimizing load demand forecasting in educational buildings using quantum-inspired particle swarm optimization (QPSO) with recurrent neural networks (RNNs): A seasonal approach. *Scientific Reports*, *15*(1), Article 19349. https://doi.org/10.1038/s41598-025-04301-z

Lee, H., Kameda, K., Manzhos, S., & Ihara, M. (2025). A novel encoding method for high-dimensional categorical data for electricity demand forecasting in distributed energy systems. *Applied Energy*, 392, Article 125989. https://doi.org/10.1016/j.apenergy.2025.125989

Li, G., Yang, Y., Liu, Z., He, Z., & Li, C. (2025). Electricity demand forecasting and power supply planning under carbon neutral targets. *Energy Reports*, 13, 2740–2751. https://doi.org/10.1016/j.egyr.2025.02.015

Liu, M., Zhao, W., Zhou, Y., & Eslami, M. (2025). Load demand forecasting in air conditioning: A rotor Hopfield neural network approach optimized by a new optimization algorithm. *Scientific Reports*, *15*(1), Article 18774. https://doi.org/10.1038/s41598-025-02568-w

Matkovic, D., Pilski, T. M., & Capuder, T. (2025). Participation of electric vehicle charging station aggregators in the day-ahead energy market using demand forecasting and uncertainty-based pricing. *Energy*, 328, Article 136299. https://doi.org/10.1016/j.energy.2025.136299

Mikulić, J., & Baumgärtner, R. M. (2025). Google Trends and Baidu index data in tourism demand forecasting: A critical assessment of recent applications. *Tourism Management*, 110, Article 105164. https://doi.org/10.1016/j.tourman.2025.105164

Ning, Z., Jin, M., & Zeng, P. (2025). A multimodal interaction-driven feature discovery framework for power demand forecasting. *Energies*, *18*(11), Article 2907. https://doi.org/10.3390/en18112907

Nygård, H. S., Grøtan, S., Kvisberg, K. R., Gorjão, L. R., & Martinsen, T. (2025). Enhancing peak electricity demand forecasting for commercial buildings using novel LSTM loss functions. *Electric Power Systems Research*, 246, Article 111722. https://doi.org/10.1016/j.epsr.2025.111722

Radfar, S., Koosha, H., Gholami, A., & Amindoust, A. (2025). A neuro-fuzzy and deep learning framework for accurate public transport demand forecasting: Leveraging spatial and temporal factors. *Journal of Transport Geography*, 126, Article 104217. https://doi.org/10.1016/j.jtrangeo.2025.104217

Seyam, A., Mathew, S. S., Du, B., Barachi, M. E., & Shen, J. (2025). A stacking ensemble model for food demand forecasting: A preventative approach to food waste reduction. *Cleaner Logistics and Supply Chain*, 15, Article 100225. https://doi.org/10.1016/j.clscn.2025.100225

Subramanian, B., Mishra, A., Ramkumar, B. V., Mandala, G., Kathamuthu, N. D., & Srithar, S. (2025). Big data and fuzzy logic for demand forecasting in supply chain management: A data-driven approach. *Journal of Fuzzy Extension and Applications*, *6*(2), 260–283. https://doi.org/10.22105/jfea.2025.488816.1703

Wu, Z., Kan, X., Chen, Z., Fang, Q., Wei, Z., & Fang, J. (2025). Application of time series-based demand forecasting in inventory management for energy enterprises. *Association for Computing Machinery*. https://doi.org/10.1145/3724154.3724362

Zhang, Y., Chen, Z., Feng, B., Sui, X., & Zhang, S. (2025). Granger-guided reduced dual attention long short-term memory for travel demand forecasting during coronavirus disease 2019. *Engineering Applications of Artificial Intelligence*, 153, Article 110950. https://doi.org/10.1016/j.engappai.2025.110950

Zhou, Z., Zhou, H., Qiao, Y., Wang, S., & Li, M. (2025). An optimization protocol for MRI examination resource allocation based on demand forecasting and linear programming. *Scientific Reports*, 15, Article 15076. https://doi.org/10.1038/s41598-025-98817-z