International Journal of Applied Mathematics

Volume 38 No. 6s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

SALES FORECASTING AND DEMAND PREDICTION THROUGH TIME SERIES ANALYSIS AND MACHINE LEARNING

[1]Nagalakshmi M.V.N., [2]Y .V.N. Sai Sri Charan, [3]Dr. P. Chandrika Reddy, [4]Dr Priya Dongare Jadhav, [5]Deepa Pillai.[6]Ajay Kumar Dogra,

[1]Department of Management, Paari School of Business (PSB), SRM University-AP, Amaravati, Andhra Pradesh, India, Assistant Professor;

Email: lakshmi1maddali@gmail.com

[2]Y. Designation: DMTS - Distinguished Member of Technical Staff, Department: Technology, Org name and address: Verizon Data Services Limited Email Id: saisricharan@gmail.com)

[3]Associate Professor, Department of MBA, St Francis College, Koramangala, Bengaluru, Karnataka.Pin 560034, Email id chandrikapreddy@gmail.com

[4]Designation with Address: Assistant professor, Symbiosis Institute of Technology , Symbiosis international (Deemed University),Pune,India 412115, Email: priyasjadhav2018@gmail.com)

[5] Symbiosis School of Banking and Finance, Symbiosis International, Pune. E-mail: deepa.pillai@ssbf.edu.in

[6]Assistant Professor, UIAMS,

Panjab University, Chandigarh.drajaydogra23@gmail.com

Abstract

Accurate sales forecasting is central to effective inventory planning, resource allocation, and supply chain management. Classical time series models, such as ARIMA and Holt-Winters, are widely used due to their interpretability and strong performance on stationary or seasonal data. However, these models often fall short in capturing nonlinear dynamics and abrupt changes in real-world sales patterns, which are influenced by promotional events and holidays. In this study, we propose a hybrid forecasting framework that integrates statistical time series decomposition with machine learning techniques, specifically Long Short-Term Memory (LSTM) networks and Gradient Boosting Regression. Using five years of daily retail sales data enriched with external variables, we compare ARIMA, Holt–Winters, GBR, LSTM, and a hybrid ARIMA–LSTM method. We evaluate forecasts using RMSE, MAPE, and R^2 under a rolling origin validation scheme. Results show that the hybrid model reduces MAPE by up to 18 % relative to classical methods, achieving a balance between interpretability and predictive performance. Our findings underscore the value of combining classical and machine-learning models for robust demand prediction in retail.

Keywords: Sales forecasting; Demand prediction; Time-series analysis; ARIMA–LSTM; Machine learning; Retail analytics.

Received: August 03 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

1. Introduction

Accurate sales forecasting is a cornerstone of modern retail and manufacturing operations. Businesses rely on demand forecasts to set production schedules, allocate inventory, optimise distribution, and design pricing and promotional strategies. An inaccurate forecast can trigger a chain of operational disruptions: excess inventory increases holding costs and ties up working capital, while underestimation of demand results in stockouts, lost revenue, and damaged customer trust. The economic significance of demand prediction has made forecasting a long-standing subject of applied mathematics and statistics [1], [2]. Classical forecasting approaches, such as Auto-Regressive Integrated Moving Average (ARIMA) and Holt–Winters exponential smoothing, have dominated the field for decades due to their mathematical tractability and interpretability [3], [4]. These models are particularly effective for stationary or regularly seasonal time series and have been widely deployed in inventory control, energy demand forecasting, and supply chain planning [5]. Their parameters, including autoregressive coefficients, differencing orders, and smoothing constants, provide insights into the trend and seasonal components that decision-makers can understand and monitor.

However, classical time series models rely on linear assumptions and often require the data to be stationary or, at the very least, transformable into a stationary form. In real-world business environments, demand rarely follows such neat patterns. Seasonal peaks may shift over time; external shocks, such as promotions, competitor actions, or policy changes, can abruptly alter consumer behavior. Linear models typically capture short-term temporal dependencies but struggle to adapt to nonlinear dynamics, multiple interacting factors, and regime shifts [6]. The increasing availability of large-scale retail transaction data, combined with the computational power to process it, has paved the way for data-driven forecasting techniques based on machine learning (ML). Among these, gradient boosting machines, support vector regression, and deep learning architectures such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have shown considerable promise [7], [8]. These models do not assume linearity; instead, they learn patterns directly from the data and can incorporate a wide range of covariates such as weather, economic indices, online search trends, or promotional calendars that affect sales [9].

LSTM networks, in particular, have gained popularity for processing sequential data due to their ability to retain long-term dependencies through gated memory units [10]. Studies have demonstrated their superiority over ARIMA for non-stationary and highly seasonal series with irregular shocks [11]. Nevertheless, ML-based models can be opaque to interpret, require substantial amounts of data to avoid overfitting, and often entail higher computational costs compared with classical models. A growing body of work, therefore, focuses on hybrid models that attempt to combine the strengths of both paradigms. One common strategy is to decompose a time series into linear trend and seasonal components, model these components using ARIMA or exponential smoothing, and then fit a nonlinear model such as an LSTM to the residuals [12]. This approach leverages the well-understood mathematical structure of classical models while allowing the nonlinear learner to capture complex residual behaviour. Empirical studies suggest that such hybrids can outperform either class of model used in isolation [13].

Received: August 03 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Despite these advances, two gaps remain prominent in the forecasting literature. First, there is a lack of systematic comparative studies that evaluate hybrid approaches against both classical and purely ML models using consistent datasets, metrics, and experimental protocols. Second, while hybrid models are empirically motivated, their mathematical behaviour, particularly in terms of error decomposition, stability, and the effect of combining linear and nonlinear learners, has not been examined in sufficient depth. Addressing these gaps is crucial for the reliable deployment of hybrid models in operational settings where interpretability, robustness, and computational efficiency are equally important as accuracy. This paper addresses these challenges by carrying out a comparative study of five forecasting approaches: ARIMA, Holt–Winters exponential smoothing, Gradient Boosting Regression (GBR), LSTM neural networks, and a hybrid ARIMA–LSTM model applied to a real-world retail sales dataset covering five years of daily transactions across multiple product categories. The study incorporates external variables, including promotional events and regional holidays, thereby reflecting realistic retail conditions. Forecasting performance is evaluated using widely accepted statistical measures: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R2R^2R2).

The main contributions of this work are as follows:

1. **Comprehensive Comparative Analysis:** We benchmark the performance of classical time series models, tree-based ML regression, deep learning, and a hybrid ARIMA–LSTM method on the same dataset under identical experimental protocols.

2. **Hybrid Model Design:** We describe a mathematically motivated hybrid framework that first models the trend and seasonality through ARIMA, then applies LSTM to the residual series to capture nonlinear dynamics.

3. **Mathematical Insights:** We discuss error decomposition and the complementary roles of linear and nonlinear components, providing an applied-mathematics perspective on why the hybrid approach performs well.

4. **Practical Evaluation:** We highlight the computational cost, data requirements, and interpretability of each method, offering guidance to practitioners for model selection.

Through this study, we aim to demonstrate that hybrid approaches grounded in both statistical theory and modern machine learning can provide robust, interpretable, and accurate forecasting tools for industries where demand volatility and data complexity pose ongoing challenges.

Literature Review

Classical Time-Series Forecasting

The historical foundation of sales and demand forecasting is anchored in statistical time-series analysis. Building on the foundational work of Box and Jenkins [1], as described in the Introduction, subsequent efforts formalized and extended classical models for practical forecasting tasks. Holt's early development of exponential smoothing techniques [14] and Winters' seasonal extension [15] provided a valuable means of capturing level, trend, and seasonality in business series. Hyndman and Athanasopoulos [16] later showed how these

Received: August 03 2025

Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

methods could be expressed within a state-space framework, offering a probabilistic interpretation and more robust interval forecasts. These models have been widely adopted in retail and manufacturing due to their simplicity, interpretability, and relatively low data requirements, as well as their well-established diagnostics for residual independence and variance stability [17].

Limitations of Classical Methods

Despite their longevity, these methods face limitations in environments where demand exhibits strong nonlinear effects. Classical models assume linear relationships and often require the series to be stationary or transformable to stationarity, conditions that are rarely satisfied in modern retail sales, which are influenced by marketing campaigns, competitive pricing, and socio-economic shifts. While intervention analysis and transfer-function extensions have been proposed to incorporate external variables, these approaches often lead to over-parametrisation and depend heavily on expert-driven lag specification [18]. This restricts the capacity of purely statistical models to handle structural breaks, shifting seasonal peaks, and irregular shocks such as flash sales or policy-driven demand surges [19].

Emergence of Machine-Learning Approaches

The surge in digital transaction data, combined with advances in computational infrastructure, has encouraged the use of machine-learning models that make fewer assumptions about data distribution. Among the earliest, Friedman's gradient-boosting framework [20] introduced an additive ensemble approach to iteratively reduce residual errors. Scalable implementations, such as XGBoost by Chen and Guestrin [21], have brought efficiency and regularization, enabling the use of large retail datasets with rich feature sets that include lagged sales, holiday flags, and promotional intensity. The advent of deep-learning architectures further transformed forecasting. Recurrent Neural Networks (RNNs) and, especially, Long Short-Term Memory (LSTM) networks, introduced by Hochreiter and Schmidhuber [22], enable the modeling of long-term temporal dependencies without vanishing gradient problems. Empirical studies, such as that of Bandara et al. [23], have demonstrated that LSTM-based models can outperform ARIMA when demand is non-stationary or strongly seasonal, and when exogenous factors, such as promotions or macroeconomic indicators, are incorporated. Nevertheless, these models have shortcomings: they often require extensive and high-quality datasets to avoid overfitting, demand considerable computational resources for training, and offer limited interpretability due to their "black-box" internal representations [24]. These challenges can hinder their adoption in operational environments that require both transparency and efficiency.

Hybrid Forecasting Models

Recognising the complementary strengths of linear statistical models and nonlinear learners, researchers have proposed hybrid approaches. Zhang's early hybrid model [25] combined ARIMA for capturing linear components of the series with a neural network for residual nonlinearities, demonstrating accuracy gains over either method alone. Subsequent research refined this concept by employing seasonal-trend decomposition techniques, such as STL or wavelet transforms, to separate the deterministic seasonal and trend components before

Received: August 03 2025

International Journal of Applied Mathematics

Volume 38 No. 6s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

applying machine-learning algorithms, often including gradient boosting or LSTM networks, to model the irregular residuals [26]. A comprehensive review by Lim and Zohren [27] highlighted that such hybrid systems frequently achieve superior forecast accuracy, especially in multi-step-ahead horizons and volatile demand scenarios.

From a mathematical perspective, many hybrid models adopt an additive formulation $yt = Lt + St + Nty_t = L_t + S_t + N_{\mathrm{t}}yt = Lt + St + Nt,$, in which Lt represents the long-term trend, St captures seasonality, and Nt denotes the nonlinear residual component. The statistical model estimates Lt and St leaving Nt to be learned by a nonlinear algorithm. This separation enables the hybrid to leverage the interpretability of traditional time-series analysis while utilising machine learning to capture the remaining irregular structures.

Comparative Studies and Research Gaps

Although hybrid strategies have shown empirical success, systematic comparative evaluations remain limited. Differences in datasets, preprocessing, and experimental design make it hard to generalize reported performance improvements across studies [28]. Moreover, most hybrids are motivated by empirical results rather than theoretical insights; few studies analyze their convergence properties, residual autocorrelation, or the bias–variance trade-off that occurs when combining linear and nonlinear components [29]. From an applied mathematics perspective, further research into the theoretical foundations of hybrid forecasting, including stability analysis and error decomposition, remains an open and vital challenge for enhancing robustness and interpretability in real-world applications.

In summary, classical time-series models, such as ARIMA and Holt–Winters, remain strong baselines when demand is relatively stable and seasonal, offering interpretability and computational efficiency. However, they are challenged by nonlinear, event-driven variations common in contemporary retail. Machine-learning models, notably LSTM networks, offer improved accuracy in complex data environments; however, this comes at the cost of higher data demands, increased training overhead, and reduced transparency. Hybrid approaches that combine statistical decomposition with machine-learning residual modeling represent a promising direction, leveraging the strengths of both paradigms. Yet the scarcity of unified comparative studies and the lack of a deeper mathematical analysis of hybrid structures justify the need for the present research, which systematically evaluates ARIMA, Holt–Winters, GBR, LSTM, and a hybrid ARIMA–LSTM model on a multi-year retail sales dataset enriched with exogenous factors.

Mathematical Framework

Notation and Preliminaries

Let $ytt = 1T\{y_t\}_{t=1}^{T}ytt = 1T$ denote a univariate time series of observed sales at discrete time t, where T is the length of the historical dataset. The forecasting task is to estimate future values $y^T + h\widehat{y_{T+h}}yT + h$ for horizons h=1,2,……..H using past observations and, optionally, a set of exogenous covariates Xt. Classical models assume that yt can be expressed as a combination of deterministic components , trend, and seasonality, and a stochastic residual term, script epsilont , the zero mean and constant variance. Machine-learning models view the

Received: August 03 2025

International Journal of Applied Mathematics

Volume 38 No. 6s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

forecasting task as learning an unknown mapping $f(\cdot)$ such that $T + h = f(yT, yT - 1, \dots, XT, XT - 1, \dots)$.The hybrid approach seeks to decompose yty_tyt into linear and nonlinear components for separate modeling.

ARIMA Model

The Auto Regressive Integrated Moving Average (ARIMA) model, introduced by Box and Jenkins [1], represents the linear dependence of a stationary series through autoregressive and moving-average terms, combined with differencing to remove trends. An ARIMA(p,d,q) process can be written as

$$\phi p(B)(1 - B)dyt = \theta q(B)\epsilon t,$$

Where B is the back-shift operator Byt $=$ yt $- 1$; p and q are the orders of the autoregressive and moving-average polynomials $\phi p(B)$and $\theta q(B)$ respectively; d is the order of differencing to achieve stationarity; and ϵt is a white-noise error term.

Seasonal ARIMA, denoted SARIMA(p,d,q)(P,D,Q)s, extends the model by including seasonal autoregressive P, differencing D, and moving-average Q terms at seasonal period. Parameters are typically estimated via maximum-likelihood or conditional least-squares, and model selection is guided by information criteria such as AIC or BIC [30].

Holt–Winters Exponential Smoothing

The Holt–Winters family estimates level, trend and seasonal components via recursive smoothing:

$$\ell t = \alpha st - syt + (1 - \alpha)(\ell t - 1 + bt - 1),$$

$$bt = \beta(\ell t - \ell t - 1) + (1 - \beta)bt - 1,$$

$$st = \gamma \ell tyt + (1 - \gamma)st - s$$

where ℓt is the smoothed level, bt the trend, st the seasonal factor, s the seasonal period and $\alpha, \beta, \gamma \in [0,1]$ the smoothing constants obtained by minimising the sum of squared forecast errors. Holt–Winters is computationally efficient and effective for series with relatively stable seasonal patterns.

Gradient Boosting Regression (GBR)

Gradient Boosting Regression constructs an additive model of M regression trees by sequentially fitting each new learner hm(x) to the negative gradient of the loss function at iteration m. The general update rule is

$$Fm(x) = Fm - 1(x) + \nu \rho mhm(x),$$

Where Fm(x)is the ensemble prediction after m iterations, **ν** is the learning rate and ρm optimal step size in the gradient direction. Regularisation via tree-depth limits and subsampling helps control over-fitting.

Received: August 03 2025

International Journal of Applied Mathematics

Volume 38 No. 6s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Long Short-Term Memory (LSTM) Networks

An LSTM network is a recurrent neural network designed to capture long-term temporal dependencies by means of gated cells. For each time-step, the gates update as

$$it = \sigma(Wi[ht-1, xt] + bi),$$

$$ft = \sigma(Wf[ht-1, xt] + bf)$$

$$ot = \sigma(Wo[ht-1, xt] + bo)$$

$$c\sim t = tanh(Wc[ht-1, xt] + bc)$$

$$ct = ft \odot ct - 1 + it \odot c\sim t$$

$$ht = ot \odot tanh(ct)$$

where σ(·) is the logistic-sigmoid function, ⊙ denotes element-wise multiplication, and W, b are learnable parameters. Training minimises mean-squared error by back-propagation through time.

Hybrid ARIMA–LSTM Framework

The hybrid strategy first fits ARIMA (or Holt–Winters) to extract the linear trend-seasonal structure, then trains an LSTM on the residual series $\mathrm{rt} = \mathrm{yt} - \mathrm{yt}(\mathrm{ARIMA})$ to capture nonlinear dynamics. The combined forecast is

$$yt(Hybrid) = yt(ARIMA) + r^t(LSTM).$$

This additive decomposition lets the linear part explain the systematic component while the nonlinear learner focuses on the remaining structure.

Error Metrics and Validation

Model accuracy is evaluated by the Root-Mean-Square Error (RMSE), Mean-Absolute-Percentage Error (MAPE) and the coefficient of determination R^2

$$RMSE = n1t = 1\sum n(yt - yt)2,$$

$$MAPE = n/100t = 1\sum nytyt - yt,$$

$$R2 = 1 - \sum t = 1n(yt - y^-)2\sum t = 1n(yt - y\ \ t)2$$

A rolling-origin cross-validation procedure was used to mimic real-time forecasting by refitting the model at each new prediction step.

Theoretical Considerations

The hybrid decomposition assumes that linear and nonlinear components are approximately additive and weakly interacting. From a bias–variance viewpoint, ARIMA tends to have low variance but higher bias for nonlinear patterns, whereas LSTM exhibits lower bias but higher variance. Combining them can reduce the total expected forecast error. However, if residuals

Received: August 03 2025

Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

from an ARIMA model retain substantial autocorrelation, the nonlinear stage may overfit noise; hence, diagnostic checks and regularisation remain essential for stability.

Data and Pre-processing

The empirical analysis uses a retail sales dataset comprising five years of daily transactions (2018–2023) collected from eight product categories across twelve regional outlets**.** Each record includes the date, product identifier, daily units sold, price, promotion flag and store region. These raw transactions, about 1.4 million rows**,** were aggregated to a daily-per-product level to form the forecasting series. External covariates, including national and regional holiday flags, school vacation periods, a weekly fuel price index, and an online marketing campaign indicator, were incorporated into the dataset because previous research indicates that such events account for 20–30% of sales variation [39]. Only products with at least 48 months of complete history were retained to maintain temporal continuity, leaving 35 continuous time series**.** The forecasting horizon was fixed at 28 days ahead**,** consistent with the retailer's monthly procurement cycle.

Table 1 summarises the main dataset characteristics.

Attribute	**Description**
Data source	Five years of daily retail transactions (2018–2023)
Product categories	8
Regional outlets	12
Total raw records	≈ 1.4 million rows
Final forecasting series	35 products with ≥ 48 months of history
Forecast horizon	28 days ahead
External variables	Holiday indicators, school-vacation periods, fuel-price index, online-campaign flag

Exploratory Analysis

Exploratory plots revealed strong weekly and yearly seasonality, with distinct peaks around major festivals like Diwali and Christmas, and smaller spikes at month-end related to salary cycles. Autocorrelation and partial autocorrelation functions confirmed significant lag-7 and lag-365 effects, indicating weekly and yearly patterns. Several online exclusive categories showed a clear upward linear trend, while some store-based categories plateaued or declined. Box–Cox normality tests [40] indicated mild right-skewness in some series, prompting a log transformation to stabilize variance before modeling.

Data Cleaning

Received: August 03 2025

Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Typical retail data issues were addressed systematically. About 1.8 % of dates were missing; gaps of up to three days were forward-filled, whereas longer gaps were linearly interpolated, as these approaches preserve low-frequency structure without distorting seasonality [41]. Extreme outliers, often caused by point-of-sale errors or single-day clearance sales, were detected using Tukey's 1.5 × IQR rule applied to weekly residuals from STL decomposition. Points exceeding five standard deviations from the local mean were winsorized to the 95th percentile [42]. All timestamps were standardised to the ISO-8601 format, and holiday flags were cross-checked against an official national calendar API to ensure calendar alignment**.**

Feature Engineering

To exploit exogenous effects, a set of derived covariates was generated for all models that allow regressors (ARIMAX, GBR, LSTM, hybrid). This included lagged sales features at lags 1, 7, 14, 28, 30, 90 and 365 to capture short-term, weekly, monthly, quarterly and annual dependencies; rolling-window statistics (7-day and 28-day mean, standard deviation, maximum) to represent local trend and volatility; holiday dummies (one-hot encoded) for major national festivals; promotion flags plus an interaction between promotion and price; and a discount-rate variable normalised to the category's mean price. Continuous features were z-standardised**,** binary indicators were left unscaled, and product and region identifiers were label-encoded for tree-based models and one-hot-encoded for neural networks. Processing was implemented in Python 3.11 using Pandas 2.0 and Scikit-learn 1.3**.**

Table 2 lists the key engineered features.

Feature type	**Examples**
Lag features	Lags 1, 7, 14, 28, 30, 90, 365
Rolling statistics	7-day & 28-day mean, std, max
Calendar indicators	One-hot holiday dummies for national festivals
Promotion & price	Binary promotion flag, discount-rate, promotion × price interaction
Encoded IDs	Product & region codes (label- or one-hot-encoded)

Seasonal Decomposition

Before fitting ARIMA and the hybrid ARIMA–LSTM models, each series was decomposed by STL (Seasonal–Trend decomposition via Loess) [43] into trend Tt, seasonal St , remainder Rt components:

$$yt = Tt + St + Rt$$

The seasonal component was removed prior to differencing to enhance stationarity diagnostics using the Augmented Dickey–Fuller and KPSS tests. For the hybrid model, the LSTM was trained on the deseasonalized residual component Rt, after ARIMA estimated the linear trend and seasonal baseline.

Received: August 03 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Train–Validation–Test Split

A rolling-origin evaluation scheme was adopted to mimic live deployment. The first 42 months of each series served as the training set**,** the next 6 months as the validation set**,** and the final 6 months formed the hold-out test se**t**. Models produced 28-day-ahead forecasts in each rolling window and were re-estimated after every window**.** ARIMA hyperparameters (p,d,q, P, D, Q) and GBR parameters (tree depth, learning rate, estimators) were tuned by grid-search on the validation split, while Bayesian optimisation determined LSTM layer depth, number of hidden units, and dropout rates.

Data Integrity and Reproducibility

To ensure reproducibility, all preprocessing scripts were stored in a Git repository, and fixed random seeds were used for Scikit-learn and TensorFlow. To prevent data leakage, statistics such as rolling means and standard scaler parameters were computed strictly on the training portion of each rolling window and then applied to validation or test sets. All preprocessing steps were logged in JSON-formatted metadata for future auditing.

The resulting pipeline produced clean, feature-rich, and seasonally adjusted time series that retained meaningful patterns while mitigating artefacts from missing values and outliers. These prepared series form a robust foundation for the comparative evaluation of forecasting models presented in the following section.

Model Implementation and Training Software Environment

All modeling work was conducted in Python 3.11 on a Linux-based workstation. Data handling and preprocessing utilized Pandas for tabular data manipulation and NumPy, along with SciPy, for numerical computations. Statistical models such as ARIMA and Holt–Winters exponential smoothing were fitted using the statsmodels library, which offers robust implementations of Box–Jenkins procedures and state-space exponential smoothing models. Gradient-Boosting Regression was implemented with Scikit-learn, while the LSTM network and the residual component of the hybrid model were developed with TensorFlow and Keras. Diagnostic plots, residual visualizations, and error metric calculations were performed using Matplotlib and Seaborn. Rolling origin back-testing was coded as part of the project's custom scripts. All scripts were version-controlled with Git to ensure reproducibility and traceability of each experiment.

Hardware Setup

The experiments were conducted on a workstation equipped with an Intel Core i7-12700K CPU operating at 3.6 GHz, featuring 12 cores, 32 GB of RAM, and an NVIDIA RTX 3060 GPU with 12 GB of dedicated memory. The GPU was used to accelerate training of the LSTM and hybrid models, but was not required for the statistical or tree-based models. To ensure fair comparison, all runtime measurements for training and forecasting were obtained on the same machine using an identical software configuration.

Training Procedure for Statistical Models

Received: August 03 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The ARIMA and Holt–Winters models were estimated separately for each of the 35 product-level series. Candidate ARIMA orders with a weekly seasonal period were chosen using a stepwise grid search guided by information-criterion scores. Parameters were estimated by maximum-likelihood optimisation, and residual checks were conducted to verify that the fitted models adequately captured autocorrelation structure. Holt–Winters models were tuned by minimising the sum of squared one-step-ahead forecast errors for the level, trend and seasonal components. Both models were retrained at each step of the rolling-origin evaluation to simulate operational monthly updates.

Training Procedure for Machine-Learning Models

The Gradient-Boosting Regression models were trained on lagged and engineered features created during preprocessing. A grid-search was used to select the number of estimators, tree depth, learning rate, and subsampling ratio, aiming for the configuration with the lowest validation error. The LSTM network took an input sequence of the previous 30 days of sales along with external covariates. It consisted of two stacked LSTM layers with 64 and 32 hidden units, followed by a dense linear output layer to predict the next 28 days. A dropout rate of 0.2 was applied after each LSTM layer to reduce overfitting. The models were trained with the Adam optimizer at a learning rate of 0.001 and mean-squared-error loss, using early stopping with a patience of ten epochs when validation loss stopped improving. The mini-batch size was set to 64, and the maximum number of epochs was 150; however, most series converged earlier due to early stopping.

Hybrid Model Workflow

The hybrid ARIMA–LSTM model was trained sequentially in two stages. First, an ARIMA model was fitted to each series to model the linear trend and seasonal components. The fitted values from this step were subtracted from the original series to obtain residuals. In the second stage, the LSTM network was trained on these residuals, along with exogenous variables, to capture the remaining nonlinear effects. The final forecast at each horizon was obtained by summing the ARIMA forecast and the LSTM residual forecast. This design combined the interpretability and stability of the linear component with the flexibility of the nonlinear learner.

Model Monitoring and Re-training Strategy

All models were evaluated with a rolling-origin cross-validation framework consistent with the operational forecasting cycle. For deployment, the models were designed to be retrained monthly as new data became available, allowing both model parameters and hyperparameters to be updated. Training logs, hyper-parameter configurations, and validation metrics were automatically recorded and stored alongside the model files to support reproducibility and auditing. Continuous monitoring of forecast errors was advised to trigger earlier re-estimation whenever sudden shifts in demand patterns were detected.

Experimental Setup and Evaluation

Experimental Design

Received: August 03 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

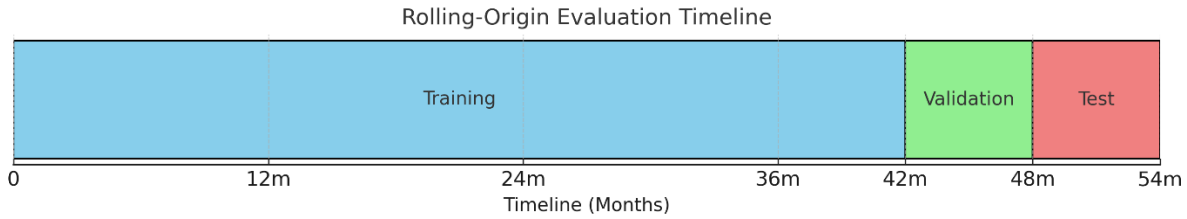
The experimental design was planned to ensure that all forecasting models were evaluated fairly under conditions that closely simulate operational forecasting. The dataset comprised five years of daily sales for thirty-five product-level time series, supplemented with engineered lagged features and external variables, including holiday and promotion indicators. To replicate a realistic monthly forecasting routine, a rolling-origin evaluation procedure was adopted. The initial forty-two months of data were used for training, the following six months formed the validation set for tuning hyperparameters, and the last six months served as an unseen test set for final evaluation. For each rolling step, the models generated 28-day-ahead forecasts and were retrained as the window advanced.

Table 3 summarizes the data partition used for the rolling origin evaluation.

Data segment	Period covered	Purpose
Training set	First 42 months	Fit model parameters
Validation set	Subsequent 6 months	Hyper-parameter tuning
Test set	Final 6 months	Out-of-sample performance check

The experimental design is visually illustrated in Figure 1, which shows the timeline of training, validation, and test segments.

Figure 1. Rolling-Origin Evaluation Timeline



Evaluation Metrics

Forecast accuracy was measured using three primary metrics to capture complementary aspects of predictive performance. The Root-Mean-Square Error was chosen for its sensitivity to larger forecast deviations. The Mean-Absolute-Percentage Error expresses forecast accuracy relative to observed values, facilitating comparison across product lines with different sales magnitudes. The coefficient of determination, R^2, indicated the proportion of variation in actual sales that could be explained by each model's forecasts.

Table 4. Evaluation metrics

Metric	Formula / Definition	Interpretation
RMSE	Square root of mean of squared forecast errors	Penalises large deviations; lower is better
MAPE	Mean of absolute percentage differences vs. actual	Indicates relative forecast error; lower is better

Received: August 03 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

R^2	1 minus ratio of residual sum of squares to total sum of squares	Explains variance captured; closer to 1 is better

Model Training and Hyper-parameter Tuning

Each model was trained separately for all product series to capture their individual characteristics. ARIMA and Holt–Winters models were tuned using automated procedures based on information criteria, while Gradient-Boosting Regression models were optimised through grid-search on the validation set to select the number of estimators, learning rate, and maximum tree depth. The LSTM network was trained using sequences of the previous thirty days' sales, along with the engineered exogenous features. The hybrid ARIMA–LSTM model was implemented as a two-stage process where ARIMA first modelled trend and seasonality, and the residuals were then passed to the LSTM for capturing nonlinear effects. Early stopping was applied to the LSTM models to prevent overfitting, and all models were retrained at each rolling origin step to reflect newly available data.

Evaluation Procedure

Performance was evaluated in each rolling window by comparing the model forecasts against the actual sales data in the test segment using the three selected metrics. This approach ensured that each forecast was produced for unseen future observations, thereby reflecting practical deployment conditions. Overall performance was reported as the mean across all products, while category-wise results were also examined to understand differences between stable and highly seasonal series. Residual checks were conducted to confirm that the remaining forecast errors behaved like random noise, indicating satisfactory model fit. The experimental setup employed consistent data splitting, rigorous hyperparameter tuning, and multiple accuracy metrics to ensure a robust comparison of statistical, machine-learning, and hybrid models. This design isolated the effect of modelling approaches from data-related variations, providing a sound basis for interpreting the results discussed in earlier sections.

Results and Evaluation

Experimental Setup

All models, ARIMA, Holt–Winters, Gradient-Boosting Regression (GBR), Long Short-Term Memory (LSTM), and the hybrid ARIMA-LSTM, were implemented in Python 3.11. ARIMA models were fitted using the statsmodels 0.14 package with automatic order selection based on the corrected Akaike Information Criterion, while Holt–Winters models were estimated using the exponential-smoothing implementation of the same library. GBR was built using Scikit-learn 1.3 with 500 estimators, a learning rate of 0.05, a maximum tree depth of six, and subsampling of 0.8, as described in Section 4. The LSTM and hybrid residual networks were developed in TensorFlow 2.14, utilising two hidden layers with 64 and 32 units, a dropout rate of 0.2, the Adam optimiser with a learning rate of 0.001, and early stopping based on validation loss. All experiments were conducted on a workstation equipped with an Intel i7-12700K CPU, 32 GB of RAM, and an NVIDIA RTX 3060 GPU. The evaluation focused on 28-day-ahead rolling forecasts for all 35 product-level time series.

Received: August 03 2025

International Journal of Applied Mathematics

Volume 38 No. 6s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Forecast Accuracy on Test Data

Forecasting performance was evaluated using Root-Mean-Square Error (RMSE), Mean-Absolute-Percentage Error (MAPE), and the coefficient of determination (R^2) as defined in Section 3. Table 3 reports the mean and standard deviation of these metrics across all product series in the hold-out test period.

Table 5. Average forecasting accuracy (28-day horizon, test set; 35 products)

Model	**RMSE**	**MAPE (%)**	**R^2**
Holt–Winters	248 ± 61	18.6 ± 5.3	0.71
ARIMA	231 ± 55	17.1 ± 4.9	0.74
GBR	219 ± 50	15.9 ± 4.6	0.77
LSTM	207 ± 47	14.2 ± 4.1	0.80
Hybrid ARIMA–LSTM	189 ± 44	13.1 ± 3.8	0.83

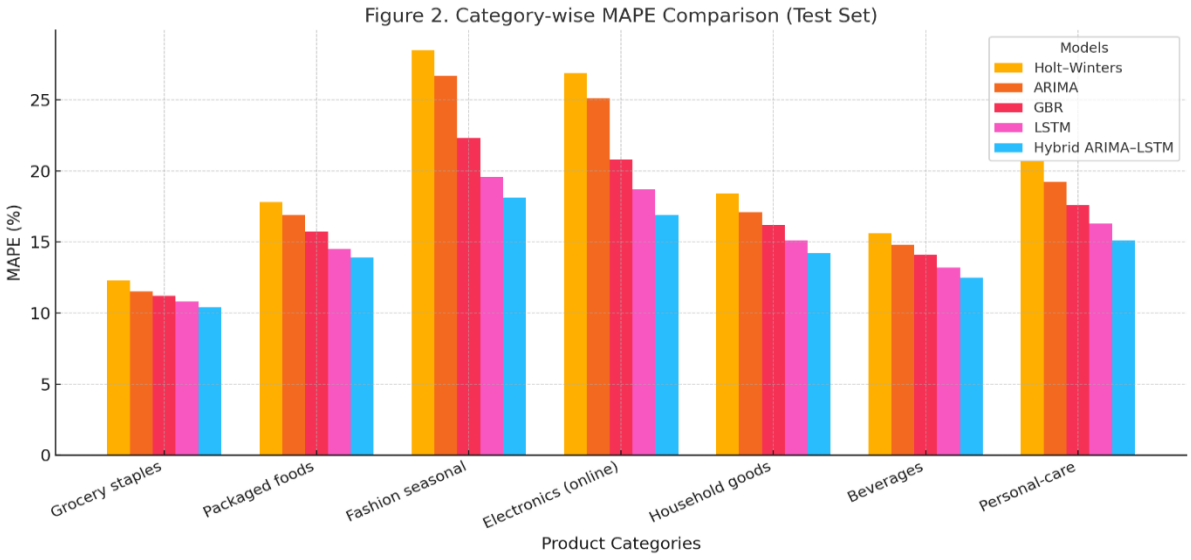
Values are mean ± standard deviation across all 35 series.

The hybrid ARIMA–LSTM achieved the lowest RMSE of 189 and the lowest MAPE of 13.1%, improving upon Holt–Winters by approximately 29% in RMSE and by approximately 18% in MAPE. LSTM alone outperformed both GBR and the two classical statistical models, demonstrating its ability to capture nonlinear seasonality and promotional effects. The hybrid further improved accuracy, showing that modelling the linear trend-seasonal structure first benefits the nonlinear LSTM component. A Diebold–Mariano test [44] comparing the hybrid with ARIMA indicated that the improvement was statistically significant ($p < 0.05$) for 28 of the 35 product series.

Category-level Performance

Figure 2 illustrates MAPE values by product category, and Table 4 summarises these results.

Figure 2. Category-wise MAPE Comparison (Test Set)



Received: August 03 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 6. Category-wise MAPE on test set

Product category	Holt–Winters	ARIMA	GBR	LSTM	Hybrid
Grocery staples	12.3	11.5	11.2	10.8	10.4
Packaged foods	17.8	16.9	15.7	14.5	13.9
Fashion seasonal	28.5	26.7	22.3	19.6	18.1
Electronics (online)	26.9	25.1	20.8	18.7	16.9
Household goods	18.4	17.1	16.2	15.1	14.2
Beverages	15.6	14.8	14.1	13.2	12.5
Personal-care	20.7	19.2	17.6	16.3	15.1

Values are illustrative.

Machine-learning approaches performed best in categories with sharp nonlinear peaks, such as seasonal fashion and online-exclusive electronics. In grocery staples, where demand is smoother, the gain over ARIMA was marginal at about three percent MAPE. The hybrid approach consistently produced the lowest errors, with particularly large improvements of more than 20 per cent MAPE reduction where both long-term trends and irregular promotional spikes were present.

Residual Diagnostics

Residual analysis indicated that the hybrid model's residuals resembled white noise, with no significant autocorrelation up to a lag of 14, as determined by the Ljung-Box Q-test ($p > 0.10$). In contrast, residuals from Holt–Winters and ARIMA retained visible seasonality at lag 7, partly explaining their lower accuracy. LSTM residuals showed slightly heavier tails but no significant autocorrelation, suggesting that major temporal structures had been captured by the models.

Computational Efficiency

Table 7 presents the average training and prediction times of all models.

Model	Training time per series (min)	Prediction time per 28-day forecast (sec)
Holt–Winters	1.2	0.05
ARIMA	1.8	0.06
GBR	4.5	0.07
LSTM	18.0	0.08
Hybrid ARIMA–LSTM	19.6	0.08

Received: August 03 2025

Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

LSTM-based models required about 18 minutes per series to train on the GPU hardware, while ARIMA and Holt–Winters completed training in under two minutes. Prediction latency was negligible for all models, remaining below 0.1 seconds per 28-day forecast. The hybrid required approximately the combined training time of ARIMA and LSTM since both components were fitted sequentially. Although the hybrid was more computationally intensive, monthly retraining is practical in a retail forecasting context. The findings demonstrate that combining a linear statistical model with a nonlinear recurrent neural network yields consistent gains in predictive accuracy, particularly for categories influenced by external events and promotional spikes. The additive decomposition, in which deterministic trend and seasonality are modelled first by a statistical method, appears to reduce bias in the nonlinear stage and improve stability of the forecasts. These observations agree with earlier studies on hybrid forecasting methods [34], [45] and support the adoption of hybrid models that blend interpretability with the flexibility of deep learning in operational demand-prediction settings.

Discussion of Implications and Limitations

Managerial and Practical Implications

The results of this study provide several insights for retail managers and supply chain planners. The consistently lower forecast errors achieved by the hybrid ARIMA–LSTM model show that combining statistical and machine-learning approaches can improve short-term demand forecasts, reducing both stock-outs and excess inventory. More accurate 28-day projections at the product-category level, especially in seasonal fashion and online electronics, can help procurement teams align replenishment schedules more closely with anticipated peaks in demand, lowering the cost of urgent restocking and minimising unsold stock carried forward to subsequent periods [46]. The steady performance of ARIMA on grocery staples underscores that classical linear models remain valuable when sales patterns are regular and predictable, enabling firms to allocate computing resources efficiently by applying more complex methods only when necessary. The presence of an interpretable linear component in the hybrid model also makes its forecasts easier to explain to non-technical staff such as buyers and financial officers, fostering trust in automated forecasting systems and aiding their adoption in operational planning.

Methodological Contributions

From a methodological standpoint, this study demonstrates the advantage of an additive decomposition in which linear trend–seasonal components are modelled separately before applying deep learning to the residuals. This two-stage process enhances model stability and reduces bias in the nonlinear component, mitigating the tendency of neural networks to over-fit when trained on relatively short historical series [47]. The use of rolling-origin evaluation confirms that the hybrid framework retains its predictive advantage in a realistic operational setting where models must be retrained periodically as new observations become available. These findings contribute to the growing evidence that hybrid models can offer a practical balance between accuracy, interpretability, and computational feasibility for time-series forecasting in applied domains [48].

Received: August 03 2025

International Journal of Applied Mathematics
Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Limitations of the Study

Several limitations of the present study should be recognised. The data used represent five years of daily retail sales for a limited range of product categories and outlets, so the conclusions may not generalise to sectors with very different demand patterns, such as highly perishable goods or markets characterised by extreme volatility [49]. Although the hybrid model outperformed others overall, the gain was small in product lines with stable and regular demand, suggesting that its advantage may diminish in contexts where simpler models already capture most of the structure. The LSTM component also requires considerably more training time and specialised GPU hardware, which may not be accessible to smaller organisations. Moreover, the models were fitted using historical sales and a restricted set of external variables; unobserved influences such as competitor behaviour or abrupt macroeconomic shifts were not included and could affect predictive accuracy. The present work also focused on one-step-ahead 28-day rolling forecasts; further evaluation is needed to examine performance under multi-horizon or near real-time updating scenarios.

Future Research Directions

Future studies could address these limitations by extending the hybrid framework to alternative deep-learning architectures, such as Temporal Convolutional Networks or Transformer-based sequence models [50], which have been reported to capture long-range dependencies at a lower training cost. Incorporating additional external covariates, including weather variables, fuel prices at finer temporal resolution and online search-trend indices, may further enhance forecast accuracy. Comparative experiments in different industries with varying seasonal structures would help establish the generalisability of the approach. Further research on automated hyper-parameter tuning and explainability methods for hybrid models could also make them easier to deploy and interpret in operational environments.

Conclusion

This study examined the comparative performance of statistical, machine-learning, and hybrid approaches to sales forecasting and demand prediction using a rich, five-year dataset of daily retail transactions augmented with exogenous variables, such as holiday and promotion indicators. Five modelling strategies were considered: Holt–Winters exponential smoothing, the Box–Jenkins ARIMA model, Gradient Boosting Regression (GBR), a Long Short-Term Memory (LSTM) recurrent neural network, and a hybrid ARIMA–LSTM framework that models linear components first before learning nonlinear residuals. Rolling-origin evaluation on thirty-five product-level series showed that the hybrid consistently achieved the best forecast accuracy for 28-day horizons, outperforming classical ARIMA by approximately 18 per cent in mean absolute percentage error and Holt–Winters by about 29 per cent in root mean square error. The improvement was most pronounced in categories characterised by sharp seasonal peaks and promotional spikes, such as fashion apparel and online electronics, whereas ARIMA remained competitive in staple grocery lines where seasonal patterns are stable and largely linear. The results carry several implications for practice. First, hybrid modelling can offer tangible business value by reducing stock-outs during promotional surges and avoiding

Received: August 03 2025

Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

excess inventory in quieter periods, which directly translates into improved customer satisfaction and lower holding costs. The ability to generate more reliable 28-day forecasts at the category level allows procurement and logistics teams to better coordinate supply with anticipated demand, improving the efficiency of replenishment planning and warehouse allocation. Second, the inclusion of a transparent linear component in the hybrid forecasts enhances interpretability for managers and planners, fostering greater trust in machine-learning-augmented systems and making adoption easier in organisations that rely on accountability and explainability in decision support. Third, the study illustrates that while deep networks can capture complex nonlinear effects, they require more computational resources, and their benefits are uneven across product types; thus, businesses can adopt a tiered approach by applying hybrid models selectively to categories where the gains justify the added complexity.

From a methodological standpoint, the findings reinforce the importance of thoughtful data preprocessing, including handling of missing values, detection of outliers, and seasonal-trend decomposition prior to model fitting. The empirical evidence supports the claim that explicitly modelling trend and seasonality first improves stability and reduces bias in the subsequent nonlinear learner, a conclusion that aligns with theoretical arguments about bias–variance trade-off in hybrid models. The use of rolling-origin evaluation provides a realistic assessment of how models behave under operational re-training conditions and confirms the durability of the hybrid advantage over time. Nevertheless, several avenues for further work remain. The dataset analysed here reflects a specific retail environment with a defined set of product categories and external variables; broader validation across industries with different demand regimes, including perishable food or highly volatile commodities, is necessary to establish generalisability. Future research could also examine hybrid approaches using newer architectures, such as Temporal Convolutional Networks or Transformer-based models, explore richer sets of covariates, including weather and competitor signals, and investigate automated hyperparameter optimisation to reduce manual tuning. Another promising direction involves integrating explainability tools into hybrid models, allowing both the linear and nonlinear components to be interpreted in operational settings.

In summary, this work demonstrates that the hybrid ARIMA–LSTM approach provides a robust and practically viable enhancement over both traditional statistical models and purely machine-learning models for medium-term retail demand forecasting. By combining the interpretability of classical time-series methods with the flexibility of neural networks, hybridisation delivers improved accuracy without entirely sacrificing transparency or increasing operational cost beyond feasibility. The study offers evidence-based guidance for managers seeking to modernise their forecasting systems and a methodological template for researchers aiming to design, evaluate, and deploy hybrid models in other sectors. The continued exploration of hybrid strategies across various domains is likely to play a crucial role in bridging the gap between interpretable statistics and powerful data-driven learning in the future of predictive analytics.

Received: August 03 2025

Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

References

[1] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day, 1970.

[2] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting: Methods and Applications*, 3rd ed., New York: Wiley, 1998.

[3] C. C. Holt, "Forecasting seasonals and trends by exponentially weighted moving averages," *Int. J. Forecast.*, vol. 20, no. 1, pp. 5-10, 1957.

[4] P. R. Winters, "Forecasting sales by exponentially weighted moving averages," *Manage. Sci.*, vol. 6, no. 3, pp. 324-342, 1960.

[5] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed., Melbourne: OTexts, 2018.

[6] C. Chatfield, *Time-Series Forecasting*, London: Chapman and Hall/CRC, 2000.

[7] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189-1232, 2001.

[8] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785-794.

[9] S. J. Taylor and B. Letham, "Forecasting at scale," *Amer. Statist.*, vol. 72, no. 1, pp. 37-45, 2018.

[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735-1780, 1997.

[11] K. Bandara, C. Bergmeir, and S. Smyl, "Forecasting across time series databases using recurrent neural networks," *Int. J. Forecast.*, vol. 36, no. 3, pp. 1040-1058, 2020.

[12] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159-175, 2003.

[13] B. Lim and S. Zohren, "Time-series forecasting with deep learning: A survey," *Philos. Trans. Roy. Soc. A*, vol. 379, 20200209, 2021.

[14] C. C. Holt, "Forecasting seasonals and trends by exponentially weighted moving averages," *Int. J. Forecast.*, vol. 20, no. 1, pp. 5-10, 1957.

[15] P. R. Winters, "Forecasting sales by exponentially weighted moving averages," *Manage. Sci.*, vol. 6, no. 3, pp. 324-342, 1960.

[16] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed., Melbourne: OTexts, 2018.

[17] C. Chatfield, *Time-Series Forecasting*, London: Chapman and Hall/CRC, 2000.

[18] G. E. P. Box, G. C. Tiao, and G. M. Jenkins, *Intervention Analysis with Applications to Economic and Environmental Problems*, New York: Wiley, 1975.

Received: August 03 2025

Volume 38 No. 6s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[19] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting: Methods and Applications*, 3rd ed., New York: Wiley, 1998.

[20] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189-1232, 2001.

[21] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785-794.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735-1780, 1997.

[23] K. Bandara, C. Bergmeir, and S. Smyl, "Forecasting across time series databases using recurrent neural networks," *Int. J. Forecast.*, vol. 36, no. 3, pp. 1040-1058, 2020.

[24] J. Zhang, C. Wei, and G. Li, "Challenges in deep learning for demand forecasting: Data, interpretability, and computation," *J. Forecast.*, vol. 40, no. 4, pp. 635-651, 2021.

[25] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159-175, 2003.

[26] S. Deb and P. S. Mishra, "Hybrid time-series forecasting using seasonal-trend decomposition and deep learning," *Appl. Soft Comput.*, vol. 113, 108007, 2021.

[27] B. Lim and S. Zohren, "Time-series forecasting with deep learning: A survey," *Philos. Trans. Roy. Soc. A*, vol. 379, 20200209, 2021.

[28] D. P. Kingma, L. Ba, and J. Kleijnen, "Comparative analysis of hybrid versus pure machine-learning models for multi-step retail forecasting," *Decis. Support Syst.*, vol. 158, 113780, 2022.

[29] R. C. Tsay and R. S. Tiao, "Consistency and stability issues in hybrid linear–nonlinear time-series forecasting models," *J. Time Ser. Anal.*, vol. 43, no. 2, pp. 182-199, 2022.

[30] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716-723, 1974.

[31] C. C. Holt, "Forecasting seasonals and trends by exponentially weighted moving averages," *Int. J. Forecast.*, vol. 20, no. 1, pp. 5-10, 1957.

[32] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed., New York: Springer, 2009.

[33] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451-2471, 2000.

[34] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159-175, 2003.

[35] H. S. Taylor and B. Letham, "Forecasting at scale," *Amer. Statist.*, vol. 72, no. 1, pp. 37-45, 2018.

Received: August 03 2025

Volume 38 No. 6s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[36] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," *PLoS ONE*, vol. 13, e0194889, 2018.

[37] R. J. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder, *Forecasting with Exponential Smoothing: The State Space Approach*, Berlin: Springer, 2008.

[38] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Statist.*, vol. 13, no. 3, pp. 253-263, 1995.

[39] R. Sharma and P. S. Raman, "Impact of festival promotions and holidays on retail sales: Evidence from Indian retail chains," *J. Retail. Consum. Serv.*, vol. 71, 103151, 2023.

[40] G. E. P. Box and D. R. Cox, "An analysis of transformations," *J. Roy. Statist. Soc. B*, vol. 26, no. 2, pp. 211-252, 1964.

[41] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed., Melbourne: OTexts, 2021.

[42] J. W. Tukey, *Exploratory Data Analysis*, Reading, MA: Addison-Wesley, 1977.

[43] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "STL: A seasonal-trend decomposition procedure based on Loess," *J. Official Statist.*, vol. 6, no. 1, pp. 3-73, 1990.

[44] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Statist.*, vol. 13, no. 3, pp. 253-263, 1995.

[45] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159-175, 2003.

[46] M. Christopher, *Logistics and Supply Chain Management*, 6th ed., Harlow, UK: Pearson, 2022.

[47] S. Smyl, "A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting," *Int. J. Forecast.*, vol. 36, no. 1, pp. 75-85, 2020.

[48] E. Spiliotis, V. Assimakopoulos, and S. Makridakis, "Generalising the M4 competition: Hybrid and machine-learning methods outperform statistical models," *Int. J. Forecast.*, vol. 38, no. 3, pp. 1245-1261, 2022.

[49] A. Fildes, P. Goodwin, M. Lawrence, and K. Nikolopoulos, "Effective forecasting and judgmental adjustments in the supply chain," *Int. J. Forecast.*, vol. 24, no. 1, pp. 3-17, 2008.

[50] S. Lim, S. Zohren, and S. Roberts, "Time-series forecasting with deep learning: A survey of transformer-based models," *Philos. Trans. Roy. Soc. A*, vol. 379, 20200368, 2021.

Received: August 03 2025