Vol. 8, No. 4, August 2026

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Residual Learning-Based Hybrid ARIMA–LSTM for Digital Retail Demand Forecasting

Dwi Hartanti[1], **Aprilisa Arum Sari**[2]

[1,2]Department of Information Systems, Program Sarjana, Universitas Duta Bangsa, Surakarta, Indonesia

Received:
February 28, 2026
Revised:
June 18, 2026
Accepted:
July 29, 2026
Published:
August 22, 2026

Corresponding Author:
Author Name*:
Dwi Hartanti
Email*:
dwihartanti@udb.ac.id

DOI:
10.63158/journalisi.v8i4.1706

Abstract. The rapid growth of digital retail has increased the need for accurate demand forecasting to support inventory planning and operational decision-making. This study proposes and evaluates a residual learning-based Hybrid ARIMA–LSTM framework for daily retail demand forecasting. A univariate dataset containing 1,826 daily demand observations collected between January 2021 and December 2025 was preprocessed using data cleaning, Min–Max normalization, Augmented Dickey–Fuller (ADF) testing, first-order differencing, and sliding window transformation. ARIMA was employed to model the linear component of the demand series, while LSTM was used to learn nonlinear residual patterns. The proposed framework was evaluated against ARIMA, LSTM, Exponential Smoothing, Moving Average, and Naïve Forecast using MAE, MSE, and RMSE on the original demand scale after inverse normalization. The Hybrid ARIMA–LSTM achieved the lowest MSE (4668.71) and RMSE (68.32) among the evaluated forecasting models, whereas the standalone LSTM achieved a slightly lower MAE (49.87). Furthermore, residual analysis showed that the Hybrid ARIMA–LSTM produced stable forecasting performance, with residuals randomly distributed around the zero-reference line, indicating minimal systematic prediction bias. This study demonstrates that a residual learning-based Hybrid ARIMA–LSTM framework can effectively improve daily digital retail demand forecasting by integrating statistical and deep learning models under identical experimental settings.

Keywords: Demand Forecasting, ARIMA, LSTM, Hybrid ARIMA–LSTM, Residual Learning, Digital Retail.

1. INTRODUCTION

The rapid expansion of e-commerce ecosystems has fundamentally transformed how consumers interact with retail services, resulting in increasingly dynamic digital transaction patterns. The growth of digital retail activities requires companies to implement accurate demand forecasting strategies to maintain inventory efficiency and improve customer service quality [1], [2]. Inaccurate forecasts may lead to overstock and stockout conditions, increasing operational costs, reducing inventory management efficiency, and negatively affecting customer satisfaction due to product unavailability [3], [4], [5]. Therefore, accurate demand forecasting has become an essential component of inventory management, particularly for supporting stock planning, replenishment scheduling, and stockout prevention in digital retail environments.

Digital retail demand is characterized by highly dynamic patterns influenced by promotional campaigns, seasonal sales periods, and changes in consumer purchasing behavior [6], [7], [8]. These factors are widely recognized in the literature as potential drivers of demand fluctuations. However, they were not explicitly recorded in the dataset used in this study and are therefore discussed only as possible explanations for the observed demand spikes rather than as analyzed variables. These conditions often generate both linear and nonlinear patterns in demand data, making forecasting a challenging task. Since inventory-related decisions are commonly made on a daily basis, daily-level demand forecasting is particularly important for ensuring that inventory levels remain aligned with actual customer demand. Reliable daily forecasts can help retailers optimize inventory allocation, improve replenishment decisions, and reduce the risks associated with excess stock and product shortages.

Among various forecasting approaches, the Autoregressive Integrated Moving Average (ARIMA) model has been widely adopted because of its effectiveness in capturing linear patterns and temporal dependencies in time-series data [9], [10]. ARIMA remains one of the most widely used statistical approaches for time-series forecasting and continues to be applied in various forecasting domains [11]. However, ARIMA has limitations in modeling nonlinear fluctuations and complex demand variations frequently observed in digital retail datasets [12], [13], [14].

On the other hand, Long Short-Term Memory (LSTM), introduced by Hochreiter and Schmidhuber [15], is a specialized recurrent neural network designed to capture long-term temporal dependencies while addressing the vanishing gradient problem. Owing to its ability to learn complex nonlinear relationships in sequential data, LSTM has become one of the most widely adopted deep learning models for time-series forecasting [16], [17], [18]. Recent review studies have also highlighted its effectiveness across a broad range of forecasting applications, demonstrating its ability to model complex temporal patterns more effectively than conventional deep learning approaches in many forecasting tasks [19]. Nevertheless, LSTM models generally require higher computational resources and may be prone to overfitting when training data are limited or when model parameters are not properly optimized.

To overcome the limitations of standalone forecasting models, recent studies have investigated hybrid ARIMA–LSTM frameworks in various domains. Fardana et al. investigated a Hybrid LSTM–ARIMA model for forecasting motor vehicle CO_2 emissions. Their results showed that the hybrid model achieved promising forecasting performance; however, the standalone ARIMA model produced lower forecasting errors on the evaluated dataset, indicating that the effectiveness of hybrid forecasting depends on the characteristics and complexity of the data [20]. Collectively, these studies suggest that hybrid ARIMA–LSTM models have the potential to improve forecasting performance by exploiting the complementary strengths of statistical and deep learning approaches. Although hybrid ARIMA–LSTM models have been widely investigated, studies specifically applying a residual learning-based framework for daily digital retail demand forecasting and evaluating it against both statistical and deep learning benchmark models remain limited. Therefore, this study evaluates a residual learning-based Hybrid ARIMA–LSTM framework for daily digital retail demand forecasting and systematically compares its forecasting performance with ARIMA and LSTM under identical experimental settings and evaluation metrics.

Residual learning integrates ARIMA and LSTM by using ARIMA to model the linear component of the demand series and LSTM to learn the nonlinear patterns from the ARIMA residuals. This study evaluates whether the proposed Hybrid ARIMA–LSTM framework improves forecasting accuracy compared with standalone ARIMA, LSTM, and benchmark models under identical experimental settings. The main contributions are: (1)

I S I Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

investigating a residual learning-based Hybrid ARIMA–LSTM framework for daily digital retail demand forecasting; (2) providing a fair comparison using identical preprocessing procedures, chronological training–testing splits, model configurations, and evaluation metrics; and (3) demonstrating its potential to support inventory planning and replenishment decisions in digital retail environments.

2. METHODS

This study was conducted through six sequential stages, namely data collection, data preprocessing, ARIMA modeling, LSTM modeling, hybrid ARIMA–LSTM development, and model evaluation, as illustrated in Figure 1.

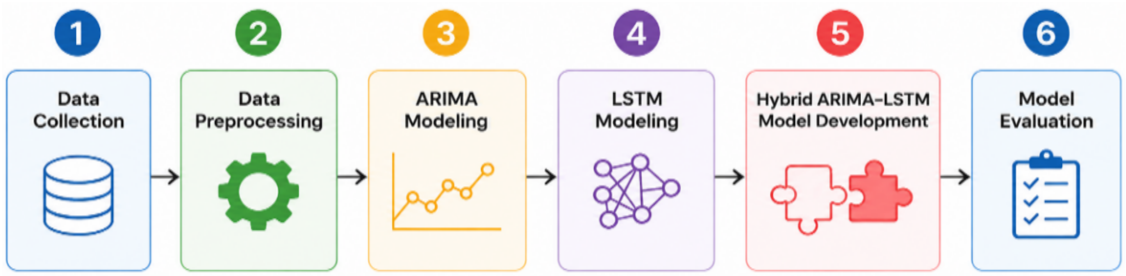

Figure 1. Research Stages

2.1. Data Collection

Daily digital retail transaction data collected between January 2021 and December 2025 were used in this study, comprising 1,826 observations. The target variable was daily product demand, defined as the total number of products sold per day. Although the dataset also contained visitor counts, conversion rates, transaction volumes, discounts, advertising costs, revenue, stock levels, and stockout status, only historical demand values were used because this study adopted a univariate forecasting approach. The overall experimental setup is summarized in Table 1.

Table 1. Summary of Experimental Setup

Component	Parameter	Configuration
Dataset	Observation Period	January 2021 – December 2025
Dataset	Number of Observations	1,826 Daily Records
Dataset	Target Variable	Daily Product Demand
Dataset	Forecasting Approach	Univariate Time Series

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Component	Parameter	Configuration
Data Split	Training : Testing	80 : 20 (Chronological Split)
Preprocessing	Normalization Method	Min-Max Scaling
Preprocessing	Stationarity Test	Augmented Dickey-Fuller (ADF)
Preprocessing	Sliding Window Size	3
ARIMA	Model Order	ARIMA(2,1,1)
ARIMA	Parameter Selection	ACF, PACF, and AIC
LSTM	Number of Units	64
LSTM	Dropout Rate	0.2
LSTM	Optimizer	Adam
LSTM	Learning Rate	0.001
LSTM	Loss Function	Mean Squared Error (MSE)
LSTM	Batch Size	32
LSTM	Epochs	100
LSTM	Validation Split	0.2
LSTM	Early Stopping	Patience = 10
LSTM	Random Seed	42
Hybrid ARIMA–LSTM	Hybrid Strategy	Residual Learning
Hybrid ARIMA–LSTM	Residual Input	ARIMA Residuals
Hybrid ARIMA–LSTM	Final Forecast	ARIMA Forecast + LSTM Residual Forecast
Evaluation	Metrics	MAE, MSE, RMSE
Evaluation	Evaluation Scale	Original Demand Scale (after inverse normalization)

2.2. Data Preprocessing

The preprocessing stage consisted of missing value inspection, outlier identification, Min–Max normalization, stationarity testing using the Augmented Dickey–Fuller (ADF) test, first-order differencing, and sliding window transformation to generate sequential data structures [21]. The dataset was chronologically divided into 1,460 training observations (80%) and 366 testing observations (20%). Min–Max normalization was fitted using only the training data and subsequently applied to the testing data to prevent data leakage.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

A sliding window size of three was employed to transform the residual series into supervised learning sequences for LSTM training.

2.3. ARIMA Modeling

The ARIMA modeling stage was conducted to capture the linear patterns of the demand series using the Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and the Akaike Information Criterion (AIC) [22]. The candidate ARIMA models and their corresponding AIC values are presented in Table 2. Based on the lowest AIC value, ARIMA(2,1,1) was selected as the forecasting model and subsequently used to generate both forecasts and residual values for the proposed hybrid ARIMA–LSTM model.

Table 2. Candidate ARIMA Models Based on AIC

Candidate ARIMA Model	AIC
ARIMA(0,1,1)	15638.44
ARIMA(1,1,0)	15903.33
ARIMA(1,1,1)	15538.25
ARIMA(2,1,1)	**15534.49**

2.4. LSTM Modeling

The LSTM model was employed to learn the temporal dependencies and nonlinear characteristics contained in the residual series generated by the ARIMA model [23], [24]. The residual series was transformed into sequential input data using a sliding window size of three before model training. The model was trained using the Adam optimizer with Mean Squared Error (MSE) as the loss function. Early Stopping with a patience of 10 epochs was applied to reduce overfitting, while training and validation were performed chronologically without data shuffling to preserve the temporal order of the observations. The LSTM architecture and hyperparameter settings are summarized in Table 1.

2.5. Hybrid ARIMA–LSTM Development

The proposed Hybrid ARIMA–LSTM model combined the linear forecasting capability of ARIMA with the nonlinear learning capability of LSTM through a residual learning strategy. The ARIMA model first generated forecasts and residual values, after which the LSTM

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

model predicted the residual component. The final demand forecast was obtained by combining the ARIMA forecast with the predicted residual values.

2.6. Model Evaluation

Model performance was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). All evaluation metrics were calculated on the original demand scale after inverse normalization, where lower error values indicate better forecasting performance [25]. All experiments were conducted in Google Colaboratory (Google Colab) using Python 3.12.13 with TensorFlow, Statsmodels, and Scikit-learn. The benchmark models (Naïve Forecast, Moving Average, and Simple Exponential Smoothing) used the same chronological training–testing split as the proposed model. The hardware and software specifications used in this study are summarized in Table 3.

Table 3. Experimental Environment

Component	Specification
Platform	Google Colaboratory (Google Colab)
CPU	Intel Xeon Processor *(Google Colab Runtime)*
RAM	Approximately 12 GB
GPU	NVIDIA Tesla T4 *(if GPU runtime was enabled)*
Python	3.12.13
TensorFlow	2.20.0
Keras	3.13.2
Statsmodels	0.14.6
Scikit-learn	1.6.1
NumPy	2.0.2
Pandas	2.2.2
Random Seed	42

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3. RESULTS AND DISCUSSION

3.1. Dataset Description and Data Preprocessing

3.1.1. Dataset Characteristics

This study used daily digital retail demand data from e-commerce transactions collected between January 2021 and December 2025, comprising 1,826 observations. The demand series exhibited trends, fluctuations, and seasonal patterns with both linear and nonlinear characteristics, supporting the use of a Hybrid ARIMA–LSTM forecasting approach. Descriptive statistical analysis and demand trend visualization were performed to provide an initial understanding of the dataset Figure 2.

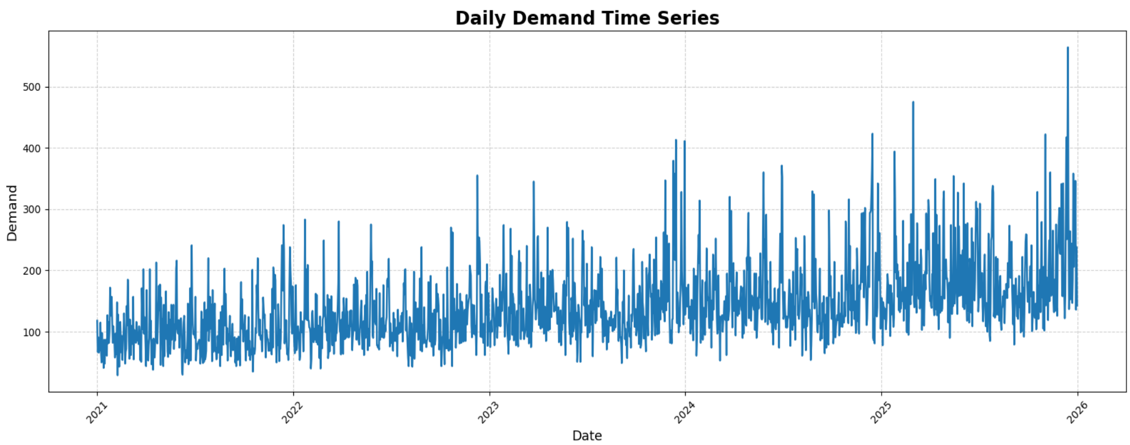

Figure 2. Product Demand

Based on Figure 2, digital retail product demand exhibits fluctuating patterns with an overall increasing trend during the 2021–2025 period. Several demand spikes can be observed at specific periods, which may be associated with promotional activities, seasonal sales events, or changes in customer purchasing behavior. These patterns indicate the presence of both linear and nonlinear characteristics, supporting the use of a hybrid forecasting approach to model complex demand dynamics more effectively.

3.1.2. Data Cleaning

The data cleaning process was conducted to ensure the quality and consistency of the dataset before model development. First, the dataset was inspected for missing values and duplicate records. The results showed that no missing values or duplicate observations were found among the 1,826 daily records. Subsequently, the Date column

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

was converted into the datetime format and sorted chronologically to preserve the temporal sequence required for time series forecasting. Finally, the Date column was set as the dataset index to facilitate subsequent preprocessing and forecasting analyses.

3.1.3. Data Normalization

To prepare the data for the LSTM model, Min-Max normalization was applied to the daily demand values. This process scales the original data into the range of 0–1, thereby reducing differences in numerical scale while preserving the temporal characteristics of the time series. As illustrated in Figure 3, the normalized data maintain the same trend and fluctuation patterns as the original series, indicating that the normalization process changes only the value scale without altering the underlying information. The normalized data were subsequently used as input for the LSTM model.

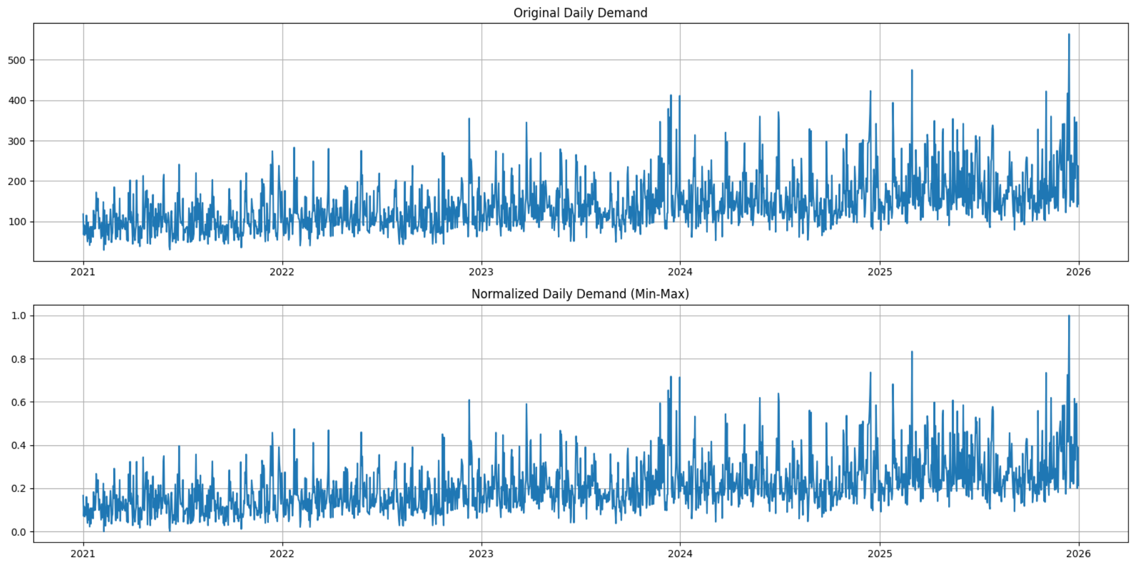

Figure 3. Original Daily Demand and Min-Max Normalized Daily Demand

3.1.4. Stationarity Testing

Stationarity was assessed using the Augmented Dickey–Fuller (ADF) test before ARIMA modeling. Although Figure 4 shows fluctuations and a gradual upward trend, the ADF results in Table 2 (ADF = −3.194379, p = 0.020312) confirmed that the original demand series satisfied the stationarity assumption.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

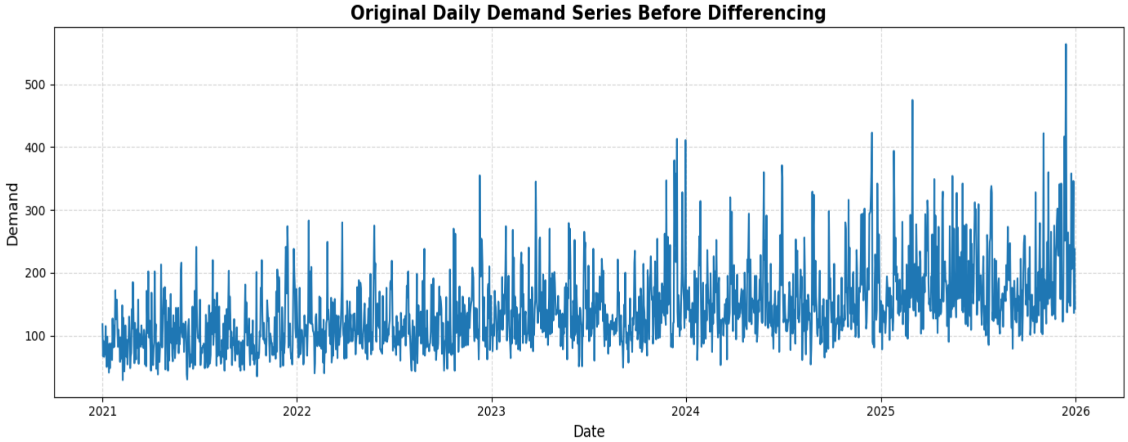

Figure 4. Original Daily Demand Series Before Differencing

3.1.5. Differencing Process

Although the original demand series satisfied the stationarity assumption based on the Augmented Dickey–Fuller (ADF) test, first-order differencing was retained during ARIMA model identification because the candidate ARIMA models with first-order differencing (d = 1) achieved lower Akaike Information Criterion (AIC) values than the alternative models evaluated. Figure 5 shows that the differenced series fluctuated around a relatively constant mean, while Table 4 summarizes the ADF results for both the original and differenced series

Table 4. ADF Test Results for the Differenced Series

Series	ADF Statistic	p-value	Lags Used	Observations	Decision
Original Series	-3.194.379	0.020312	25	1800	Stationary
First-order Differenced Series	-14.787.547	0.000000	25	1799	Stationary

Table 4 shows that the original demand series was already stationary (ADF = −3.194379, p = 0.020312). After first-order differencing, the ADF statistic decreased to −14.787547 with a near-zero p-value, indicating stronger evidence of stationarity. Although the original series satisfied the stationarity assumption, ARIMA(2,1,1) was selected because it achieved the lowest Akaike Information Criterion (AIC) among the candidate models evaluated during ARIMA model identification.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3.1.6. Sliding Window Transformation

To prepare the normalized demand series for LSTM training, a sliding window transformation with a window size of three was applied. This transformation converts sequential time-series observations into supervised learning samples by using the previous three observations as input features to predict the next demand value. As shown in Table 4, the sliding window transformation was performed sequentially across the entire dataset by shifting the window forward by one observation to generate the next training sample. The transformed dataset was subsequently used as the input for the LSTM model to capture temporal dependencies among consecutive observations.

Table 5. Sliding Window Transformation (Window Size = 3)

Sample	t-3	t-2	t-1	Target (t)
1	0.1664	0.0710	0.1159	0.0692
2	0.0710	0.1159	0.0692	0.1084
3	0.1159	0.0692	0.1084	0.0766
4	0.0692	0.1084	0.0766	0.1607
5	0.1084	0.0766	0.1607	0.0729

3.2. Benchmark Model

To evaluate the performance of the proposed forecasting framework, three benchmark forecasting methods were implemented, namely Naïve Forecast, Moving Average (MA), and Simple Exponential Smoothing (SES). These benchmark models were selected because they are widely used in time-series forecasting and provide appropriate baseline references for comparison with the proposed Hybrid ARIMA–LSTM model. All benchmark models were trained using the same training dataset and evaluated on the identical testing dataset to ensure a fair and consistent comparison.

3.2.1. Naïve Forecast

The Naïve Forecast model was implemented as the first benchmark method to evaluate the performance of the proposed forecasting framework. This method predicts the demand at the next time step by assuming that it is equal to the most recently observed demand value. Although simple, the Naïve Forecast is widely used as a baseline model because it provides a straightforward reference for comparing the forecasting

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

performance of more advanced models. The forecasting results obtained using the Naïve Forecast model are presented in Figure 5.

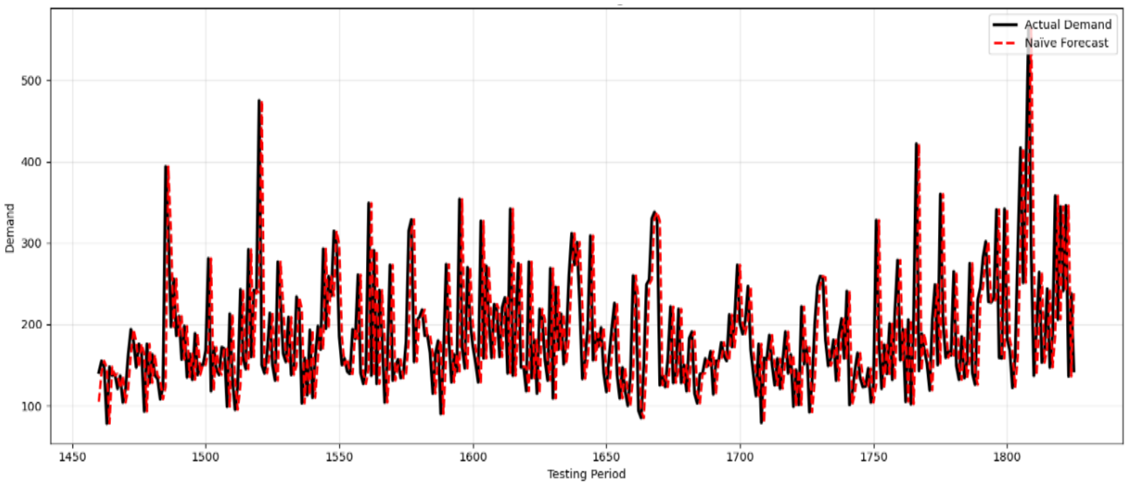

Figure 5. Actual vs. Predicted Demand Using the Naïve Forecast Mode

As shown in Figure 5, the Naïve Forecast model was able to follow the general demand trend during the testing period. However, because each prediction was based solely on the immediately preceding observation, the model exhibited a one-step lag and was less responsive to sudden demand fluctuations. Consequently, larger prediction errors occurred during periods with sharp demand spikes and high variability. These results indicate that the Naïve Forecast provides a useful baseline for performance comparison but has limited capability in capturing complex temporal patterns.

Table 6. Performance of the Naïve Forecast Model

Metric	Value
MAE	64.80
MSE	7405,13
RMSE	86,05

Table 6 summarizes the forecasting performance of the Naïve Forecast model on the testing dataset. The model achieved an MAE of 64.80, an MSE of 7405.13, and an RMSE of 86.05. These results indicate that the Naïve Forecast model produced relatively large prediction errors, particularly during periods of rapid demand fluctuations, because each

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

prediction relied solely on the most recent observation. Therefore, the Naïve Forecast model serves as a baseline for evaluating more advanced forecasting approaches.

3.2.2. Moving Average

The Moving Average (MA) model forecasts future demand by calculating the average of a specified number of previous observations. This smoothing mechanism reduces short-term fluctuations and random noise, producing more stable forecasts than the Naïve Forecast model. The forecasting results obtained using the Moving Average model are presented in Figure 6.

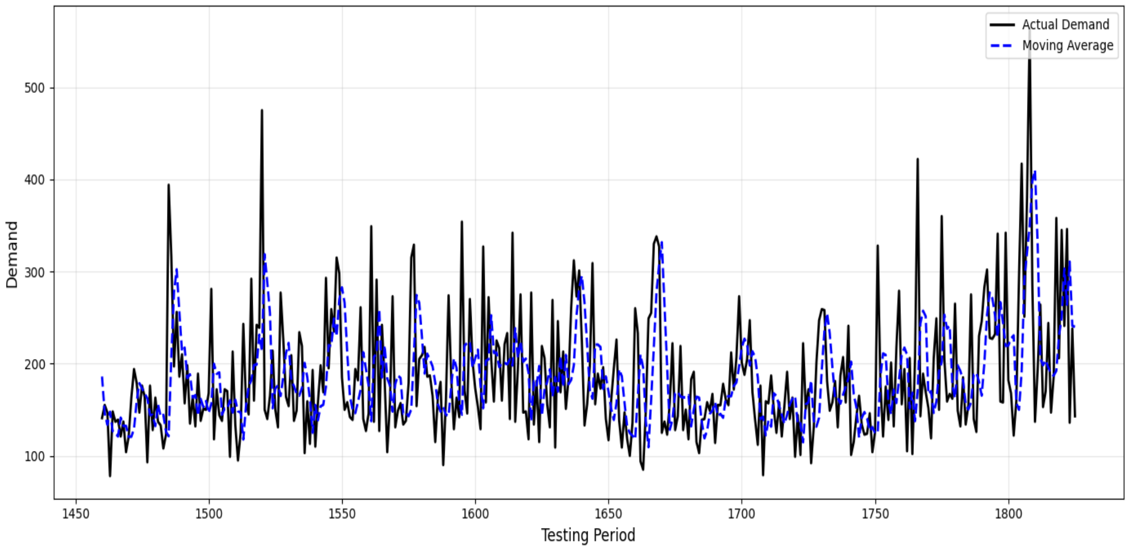

Figure 6. Actual vs. Predicted Demand Using the Moving Average Model

As shown in Figure 6, the Moving Average model successfully captured the overall demand trend during the testing period while producing smoother prediction curves than the Naïve Forecast model. The averaging process reduced short-term fluctuations and random noise, resulting in more stable forecasts. However, the model responded more slowly to abrupt demand changes, leading to the underestimation of several extreme demand peaks. These results indicate that although the Moving Average model effectively represented the general demand trend, its ability to capture highly dynamic demand fluctuations remained limited.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 7. Performance of the Moving Average Model

Metric	Value
MAE	56,11
MSE	5442,98
RMSE	73,77

Table 7 presents the forecasting performance of the Moving Average model on the testing dataset. The model achieved an MAE of 56.12, an MSE of 5442.98, and an RMSE of 73.78. These results indicate that, although the Moving Average model was able to capture the general demand trend, its forecasting accuracy remained limited due to the smoothing effect, which reduced its responsiveness to sudden demand fluctuations. Consequently, the model produced relatively larger prediction errors, particularly during periods characterized by rapid changes in demand.

3.2.3. Simple Exponential Smoothing

The Simple Exponential Smoothing (SES) model was employed as the third benchmark forecasting method to evaluate the proposed Hybrid ARIMA–LSTM model. Unlike the Moving Average model, SES assigns greater weights to more recent observations, allowing the model to adapt more effectively to recent demand patterns. The forecasting results obtained using the Rolling One-Step SES model are presented in Figure 7.

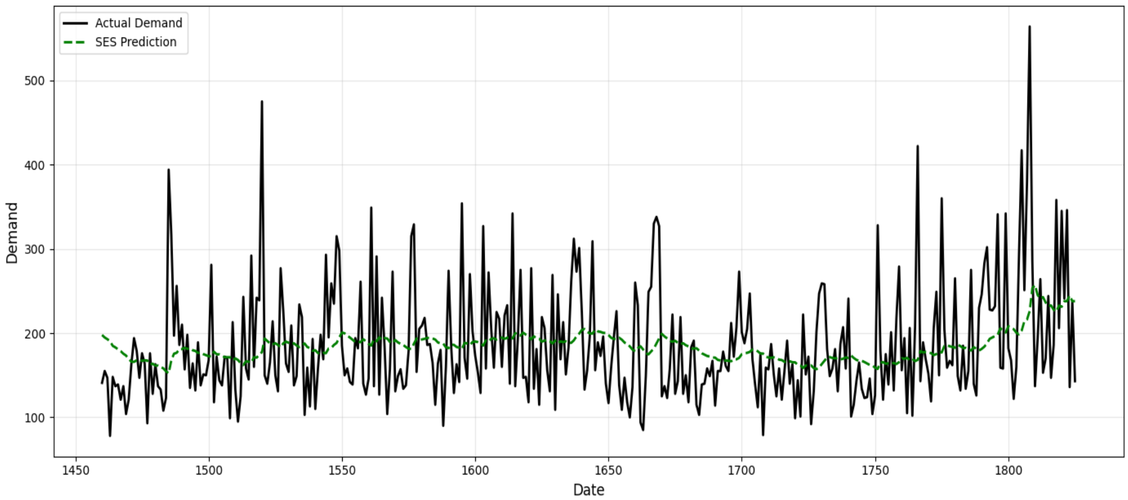

Figure 7. Actual vs. Predicted Demand Using the Rolling One-Step Simple Exponential Smoothing Model

As shown in Figure 7, the Rolling One-Step Simple Exponential Smoothing model generally followed the overall movement of the demand series and produced smoother forecasts than the actual observations. However, the model remained less responsive to sudden demand spikes and large fluctuations, resulting in higher prediction errors during periods of highly dynamic demand. These findings indicate that although the SES model effectively captured the general demand trend, its ability to represent abrupt changes in retail demand remained limited.

Table 8. Performance of the Simple Exponential Smoothing Model

Metric	Value
MAE	56,10
MSE	4910,01
RMSE	70,07

Table 8 summarizes the forecasting performance of the Rolling One-Step Simple Exponential Smoothing model. The model achieved an MAE of 56,10 an MSE of 4910,01, and an RMSE of 70,07 representing the best performance among the benchmark forecasting models. Compared with the Naïve Forecast and Moving Average models, SES produced lower prediction errors by assigning greater weights to recent observations. Nevertheless, the model still showed limitations in accurately representing abrupt demand spikes, indicating the need for more advanced forecasting models to capture the complex temporal characteristics of daily retail demand.

3.2.4. Benchmark Comparison

The forecasting performance of the benchmark models (Naïve Forecast, Moving Average, and Simple Exponential Smoothing) is summarized in Table 8. All models were trained and evaluated using the same chronological training and testing datasets to ensure a fair comparison. Although the benchmark models captured the general demand trend, their prediction errors increased during periods of rapid fluctuations and demand peaks, indicating limitations in modeling dynamic demand patterns. These results provide baseline performance for comparison with the ARIMA, LSTM, and Hybrid ARIMA–LSTM models using MAE, MSE, and RMSE.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 9. Performance Comparison of Benchmark Forecasting Models

Model	MAE	MSE	RMSE
Naïve Forecast	64,80	7405,13	86,05
Moving Average	56,11	5442,98	73,77
Simple Exponential Smoothing	56,10	4910,01	70,07

Table 9 shows that the Simple Exponential Smoothing (SES) model achieved the best forecasting performance among the benchmark methods, followed by the Moving Average and Naïve Forecast models. The lower MAE, MSE, and RMSE values obtained by SES indicate that assigning greater weights to recent observations improved forecasting accuracy compared with the other benchmark models. Nevertheless, all benchmark methods exhibited reduced accuracy during periods with rapid demand fluctuations, establishing a baseline for comparison with the ARIMA, LSTM, and Hybrid ARIMA–LSTM models presented in the following section.

3.3. ARIMA Model

This section presents the development and evaluation of the ARIMA model. The analysis includes parameter identification using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), followed by candidate model selection based on the Akaike Information Criterion (AIC).

3.3.1. ARIMA Parameter Identification

Although the original demand series satisfied the stationarity assumption based on the Augmented Dickey–Fuller (ADF) test, first-order differencing was retained during ARIMA model identification because the candidate ARIMA models with a differencing order of d = 1 achieved lower Akaike Information Criterion (AIC) values than the alternative candidate models. Consequently, first-order differencing was adopted as part of the model selection process to obtain the most appropriate ARIMA specification. Figure 5 shows that the differenced series fluctuated around a relatively constant mean, while Table 4 summarizes the ADF results for both the original and differenced series.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3.3.2. ARIMA Forecasting Results

After parameter identification, the selected ARIMA(2,1,1) model was trained using the chronological training dataset and applied to forecast daily product demand during the testing period. The forecasting results are presented in Figure 8.

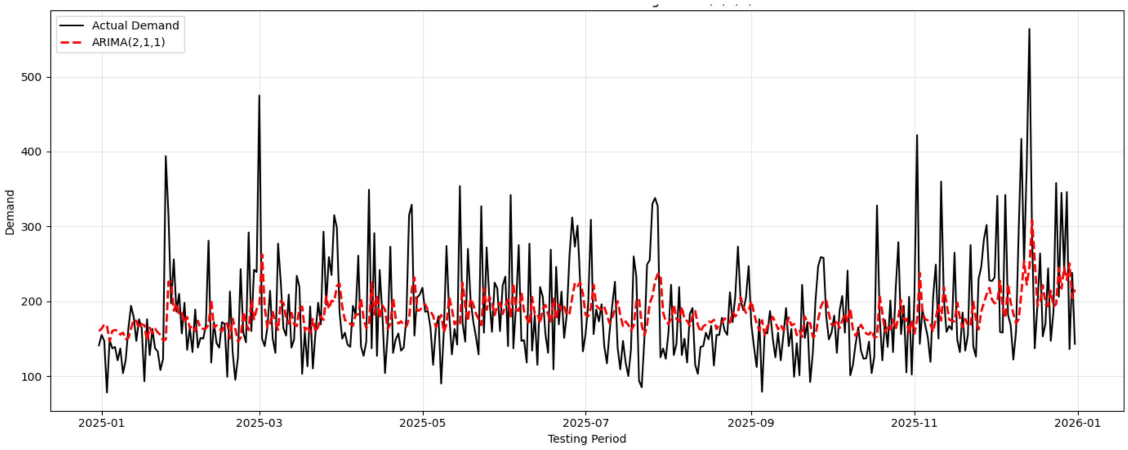

Figure 8. Forecasting Results Using the ARIMA(2,1,1) Model

As shown in Figure 8, the ARIMA(2,1,1) model successfully captured the overall trend of the daily demand series throughout the testing period. The predicted values generally followed the direction of the actual observations during relatively stable periods, indicating that the model effectively represented the linear temporal patterns in the demand data. However, noticeable deviations were observed during periods with abrupt demand changes and extreme demand spikes. The predicted values appeared smoother than the actual observations, resulting in the underestimation of several extreme demand peaks. This behavior reflects the limitation of the ARIMA model in representing complex nonlinear variations in the demand series.

Table 10. Performance of the ARIMA(2,1,1) Model

Metric	Value
MAE	51.76
MSE	4781,93
RMSE	69,15

Table 10 shows that the ARIMA(2,1,1) model achieved lower forecasting errors than the benchmark models evaluated in the previous section, indicating an improvement in

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

forecasting accuracy. Nevertheless, the remaining prediction errors suggest that the linear ARIMA model was not sufficient to fully capture the highly dynamic and nonlinear characteristics of the daily demand series. Therefore, the following section evaluates the LSTM model to investigate its capability in modeling nonlinear temporal patterns.

3.4. LSTM Model

3.4.1. LSTM Model Training

Following the data preprocessing, normalization, and sliding window transformation stages, the dataset was chronologically divided into 80% training data and 20% testing data to preserve the temporal order of the daily demand series. The training dataset was used to develop the LSTM model, while the testing dataset was reserved for evaluating the forecasting performance. The transformed input sequences were generated using a sliding window with a window size of three, where the demand values from the previous three time steps were used to predict the demand value at the subsequent time step. The proposed LSTM architecture consisted of 64 LSTM units, a dropout rate of 0.2, and a Dense output layer. Model training was performed using the Adam optimizer with a learning rate of 0.001, Mean Squared Error (MSE) as the loss function, a batch size of 32, 100 training epochs, a validation split of 0.2, and early stopping to prevent overfitting. The complete training configuration is summarized in Table 12.

Table 12. LSTM Model Training Parameters

Parameter	Value
Window Size	3
LSTM Units	64
Dropout Rate	0.2
Epoch	100
Batch Size	32
Optimizer	Adam
Learning Rate	0.001
Loss Function	Mean Squared Error (MSE)
Validation Split	0.2
Early Stopping	Applied
Random Seed	42

During the training process, the LSTM model optimized its parameters by minimizing the training loss over successive epochs. The training and validation loss curves are presented in Figure 9.

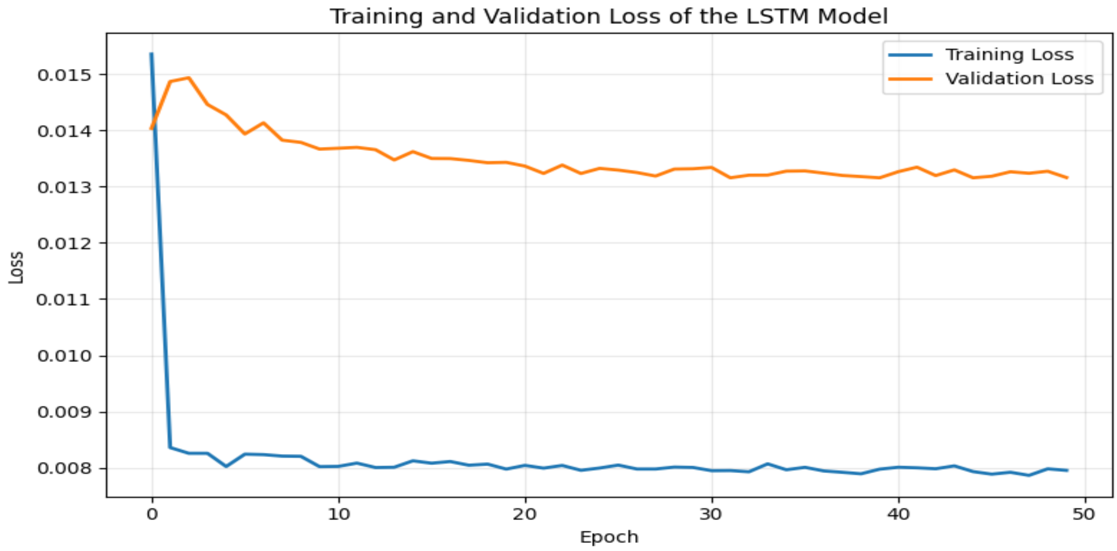

Figure 9. Training and Validation Loss Curves of the LSTM Model

Figure 9 illustrates the training and validation loss curves of the proposed LSTM model. The training loss decreased rapidly during the initial epochs before converging to a stable value, indicating that the model effectively learned the temporal patterns from the training data. Meanwhile, the validation loss remained relatively stable throughout the training process without exhibiting a substantial increase. Although the validation loss was consistently higher than the training loss, the gap between the two curves remained relatively small and stable, suggesting that the model achieved good generalization performance without significant overfitting. Consequently, the trained LSTM model was considered suitable for forecasting nonlinear temporal patterns in the daily retail demand series.

3.4.2. LSTM Forecasting Performance

After completing the training process, the trained LSTM model was used to forecast daily retail demand on the testing dataset. The forecasting process utilized input sequences generated through the sliding window transformation, and the predicted values were compared with the actual demand observations. The forecasting results are presented in Figure 10.

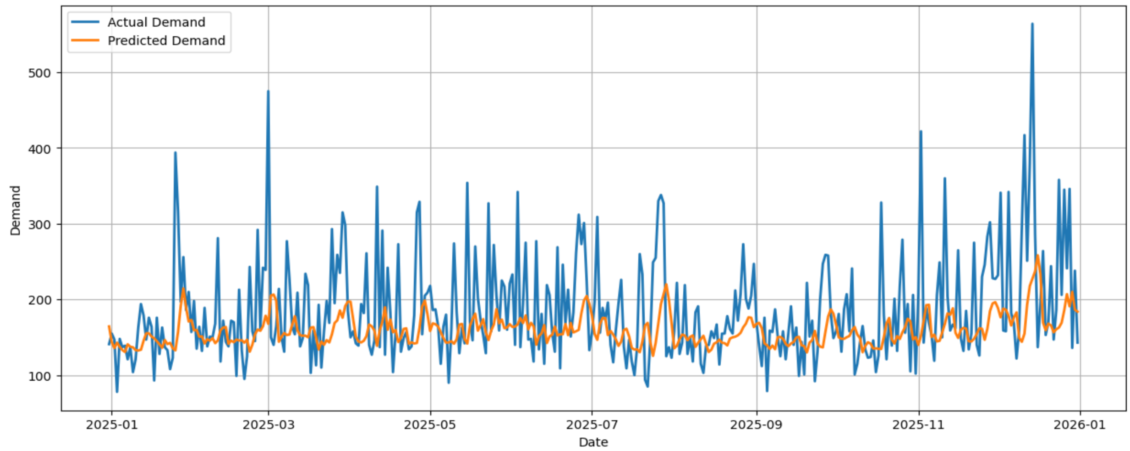

Figure 10. Actual and Predicted Daily Demand Using the LSTM Model

Figure 10 shows that the LSTM model successfully captured the overall temporal pattern of daily retail demand, with predictions generally following the actual demand trend. However, the predicted series was smoother than the actual observations and underestimated several extreme demand peaks. The model achieved an MAE of 49.87, an MSE of 4893.06, and an RMSE of 69.95, indicating good performance in learning nonlinear temporal relationships but limited accuracy during abrupt demand fluctuations. These limitations motivated the evaluation of the proposed Residual Learning-Based Hybrid ARIMA–LSTM model, which combines ARIMA's linear forecasting capability with LSTM's nonlinear learning capability to improve forecasting accuracy.

3.3.4. LSTM Model Evaluation

The predictive performance of the LSTM model was evaluated by comparing the forecasting results with the actual daily retail demand on the testing dataset. Model performance was assessed using three standard error metrics, namely Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). All evaluation metrics were calculated on the original demand scale after inverse normalization to ensure meaningful interpretation of forecasting performance. The evaluation results are presented in Table 13.

Table 13. LSTM Model Evaluation

Model	MAE	MSE	RMSE
LSTM	49,87	4893,06	69.95

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

The LSTM model achieved an MAE of 49.87, an MSE of 4893.06, and an RMSE of 69.95, indicating its ability to learn the general temporal pattern of daily retail demand. As illustrated in Figure 14, the predicted values generally followed the overall demand trend but tended to underestimate several extreme demand peaks, resulting in higher prediction errors during periods of substantial demand variability. These findings suggest that, although LSTM effectively captured nonlinear temporal relationships, its forecasting performance remained limited in handling abrupt demand fluctuations.

3.5. Hybrid ARIMA–LSTM

3.5.1. Hybrid Model Architecture

To improve the modeling of both linear and nonlinear patterns in daily retail demand, this study proposes a Residual Learning-Based Hybrid ARIMA–LSTM framework. The hybrid approach combines the statistical capability of the ARIMA(2,1,1) model in capturing linear temporal patterns with the learning capability of the LSTM model in modeling nonlinear information contained in the residual series. Instead of directly forecasting the demand series, the LSTM model is trained using the residuals generated by the ARIMA model. The final forecast is obtained by combining the ARIMA forecast with the residual prediction produced by the LSTM model. The overall architecture of the proposed hybrid forecasting framework is illustrated in Figure 15.

$$\hat{Y}_t^{Hybrid} = \hat{Y}_t^{ARIMA} + \hat{R}_t^{LSTM}$$

where $\hat{Y}_t^{Hybrid}$ denotes the final hybrid forecast, $\hat{Y}_t^{ARIMA}$ represents the forecast generated by the ARIMA(2,1,1) model, and $\hat{R}_t^{LSTM}$ denotes the residual prediction produced by the LSTM model.

Figure 11 presents the architecture of the proposed Residual Learning-Based Hybrid ARIMA–LSTM model. Historical demand data are first preprocessed and modeled using ARIMA(2,1,1) to capture the linear component. The resulting residuals are transformed using a sliding window (window size = 3) and used as input to the LSTM model to learn the nonlinear component. The final hybrid forecast is obtained by combining the ARIMA forecasts with the residual predictions generated by the LSTM model. Model performance is subsequently evaluated using MAE, MSE, and RMSE.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

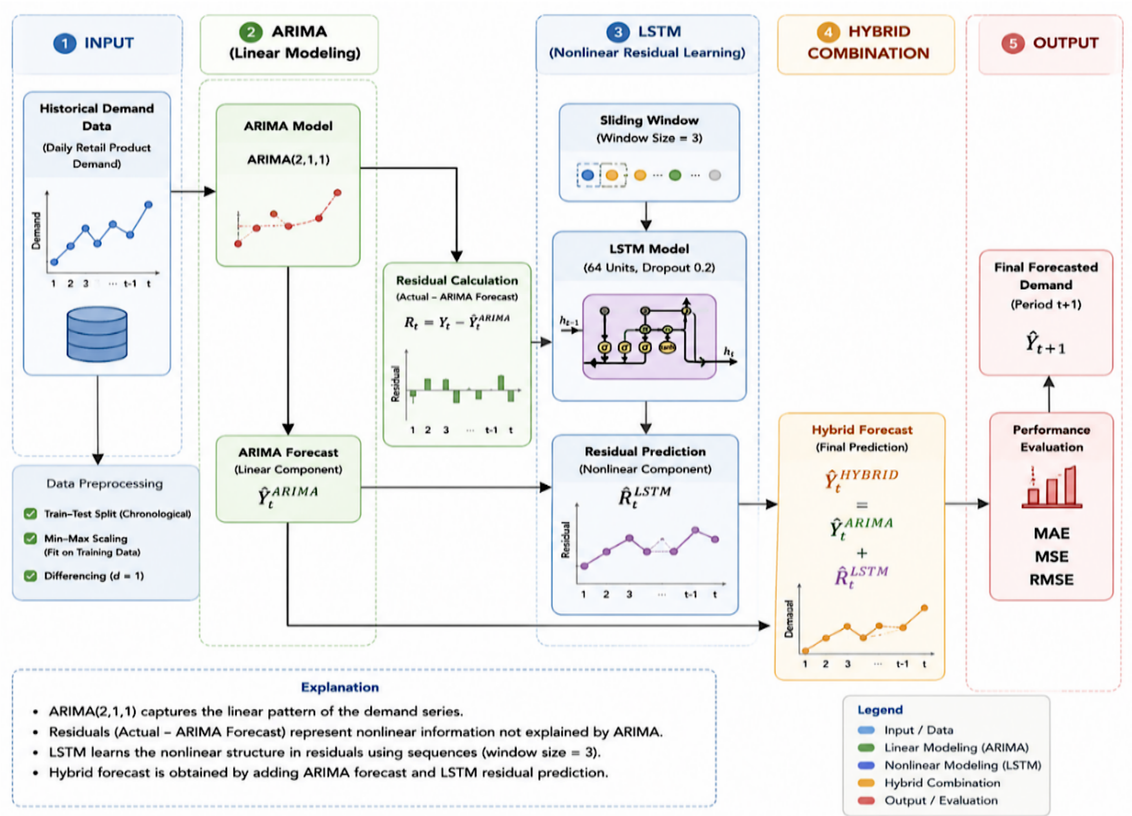

Figure 11. Hybrid Model Architecture

3.5.2. Hybrid Forecasting

The proposed Hybrid ARIMA–LSTM model was applied to forecast daily retail demand on the testing dataset using the residual learning strategy. The forecasting results show that the model successfully captured the overall demand trend while effectively modeling the nonlinear information contained in the ARIMA residual series. The final forecasts closely followed the actual demand observations, indicating that the integration of ARIMA and LSTM improved the representation of both linear and nonlinear demand patterns. Overall, the proposed Hybrid ARIMA–LSTM framework demonstrated accurate and stable forecasting performance under the experimental settings adopted in this study.

3.5.3. Hybrid Model Evaluation

To evaluate the forecasting performance of the proposed Hybrid ARIMA–LSTM model, three forecasting accuracy metrics were employed: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Lower values of these metrics indicate better forecasting performance by reflecting smaller prediction errors.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 14. Forecasting Performance of the Hybrid ARIMA–LSTM Model

Metric	Value
MAE	51.94
MSE	4668.71
RMSE	68.32

As presented in Table 14, the proposed Hybrid ARIMA–LSTM model achieved an MAE of 51.94, an MSE of 4668.71, and an RMSE of 68.32. These results indicate that the proposed model provided accurate and stable forecasting performance under the adopted experimental settings. A comparative evaluation with the benchmark forecasting models is presented in the following section.

3.6. Model Performance Comparison

To comprehensively evaluate the forecasting performance of the proposed Hybrid ARIMA–LSTM model, the prediction results were compared with those of the benchmark forecasting models, including Naïve Forecast, Moving Average (MA), Simple Exponential Smoothing (SES), ARIMA, and LSTM. The comparison was performed using three widely adopted forecasting evaluation metrics, namely Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), where lower values indicate better forecasting accuracy. The quantitative comparison of all forecasting models is presented in Table 15.

Table 15. Performance Comparison of Forecasting Models

Model	MAE	MSE	RMSE
Hybrid ARIMA–LSTM	51,94	4668,71	68.32
ARIMA	51,76	4781,93	69,15
Exponential Smoothing	56,10	4910,01	70,07
LSTM	49,87	4893,06	69.95
Moving Average	56,11	5442,98	73,77
Naïve Forecast	64,80	7405,13	86,05

As presented in Table 15, the Hybrid ARIMA–LSTM model achieved the lowest MSE (4668.71) and RMSE (68.32), whereas the standalone LSTM model achieved the lowest MAE

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

(49.87). These results indicate that the proposed hybrid model was more effective in reducing larger forecasting errors while maintaining competitive prediction accuracy across the evaluated forecasting models. Figure 13 further illustrates that the hybrid model produced prediction errors that were generally smaller and more closely concentrated around the zero reference line, supporting the residual analysis that indicated stable forecasting performance with minimal systematic prediction bias.

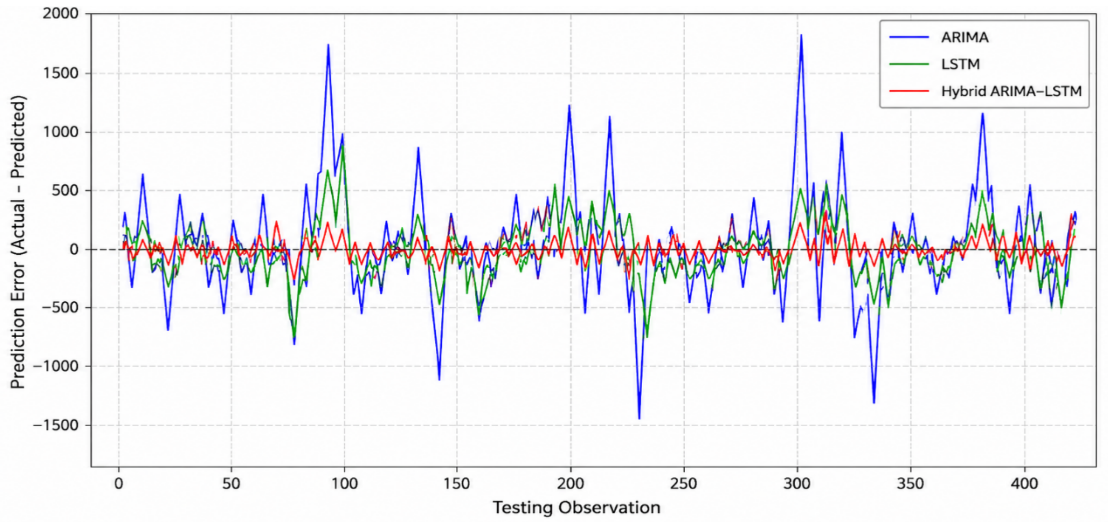

Figure 12. Prediction Error Comparison of ARIMA, LSTM, and Hybrid ARIMA–LSTM Models

As illustrated in Figure 12, the Hybrid ARIMA–LSTM model produced smaller prediction errors that were more closely concentrated around the zero reference line than the benchmark models, confirming its superior forecasting accuracy. To further evaluate the model, residual analysis was performed to examine whether the prediction errors were randomly distributed around zero without systematic patterns.

Figure 13 shows that the residuals of the Hybrid ARIMA–LSTM model are generally distributed around the zero reference line without a clear systematic pattern, indicating stable forecasting performance with minimal prediction bias. To further assess residual stability, a boxplot was used to summarize the distribution, variability, and potential outliers of the residual values.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

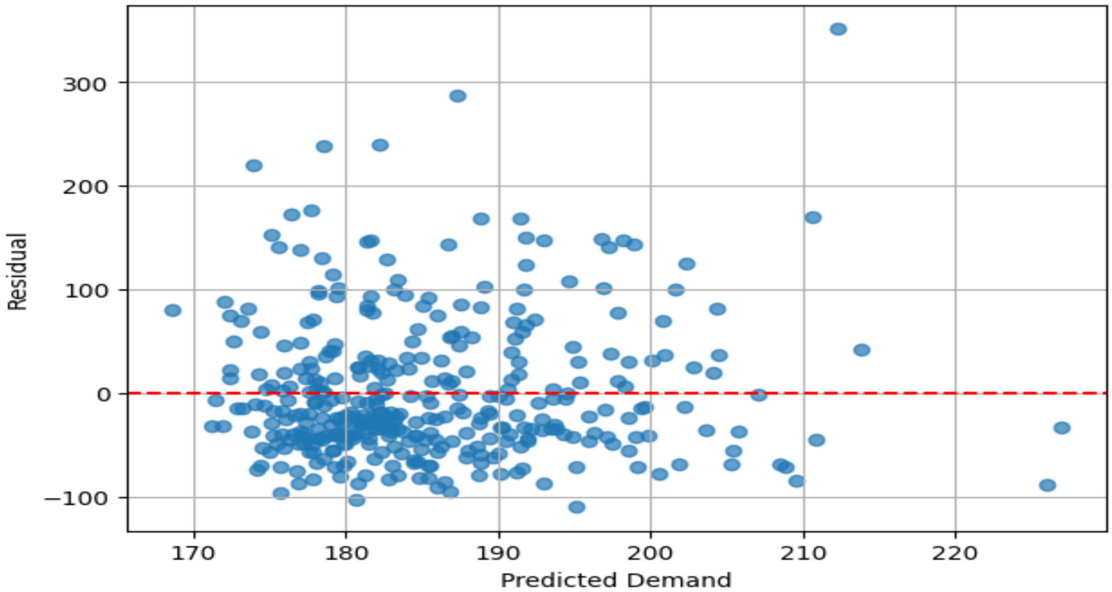

Figure 13. Residual Plot of the Hybrid ARIMA–LSTM Model during the Testing Period

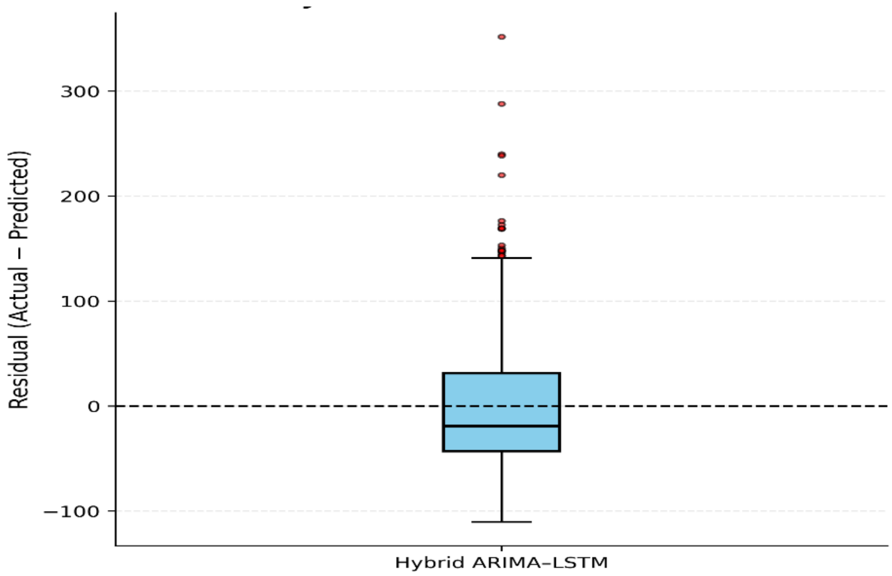

Figure 14. Residual Error Distribution of the Hybrid ARIMA–LSTM Model

Figure 14 shows that the residual errors of the Hybrid ARIMA–LSTM model are mostly concentrated around the zero reference line, indicating stable forecasting performance with minimal systematic bias. The relatively narrow interquartile range further reflects low residual variability, while the few positive outliers represent isolated demand fluctuations rather than consistent forecasting errors. Overall, these results support the quantitative findings in Table 14 and confirm the robustness of the proposed Hybrid ARIMA–LSTM model.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3.7. Discussion

The results indicate that the proposed Residual Learning-Based Hybrid ARIMA–LSTM model achieved lower forecasting errors than the benchmark forecasting methods, including Naïve Forecast, Moving Average, Simple Exponential Smoothing (SES), standalone ARIMA, and standalone LSTM. This improvement can be attributed to the complementary strengths of the hybrid framework, where ARIMA captures the linear temporal patterns of the demand series, while LSTM learns the nonlinear information contained in the ARIMA residuals. As a result, the hybrid model produced forecasts that more closely followed the observed demand patterns.

The residual analysis further supports these findings. The residual plots indicate that the prediction errors were generally distributed around the zero reference line without obvious systematic patterns, suggesting that the proposed model achieved stable forecasting performance with limited prediction bias. These results demonstrate that the residual learning strategy effectively improved forecasting accuracy by modeling the nonlinear information that remained after ARIMA forecasting.

From a practical perspective, the proposed forecasting framework can support inventory planning by providing more reliable estimates of future product demand. However, because the model was developed using a univariate forecasting approach, future studies should investigate the incorporation of external variables, such as promotional activities, seasonal effects, and stock availability, to further improve forecasting performance.

This study has several limitations. First, the forecasting models were evaluated using a single chronological 80:20 holdout split instead of rolling-origin or walk-forward validation, limiting the assessment of forecasting robustness across multiple forecasting horizons. Second, the proposed framework used only univariate demand data without incorporating external factors such as promotions, advertising costs, and stock availability. Future research should evaluate the Hybrid ARIMA–LSTM model using rolling-origin or walk-forward validation, statistical significance testing, seasonal benchmark models, and multivariate forecasting approaches.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

4. CONCLUSION

This study evaluated the forecasting performance of ARIMA, LSTM, and a residual learning-based Hybrid ARIMA–LSTM model for daily retail demand forecasting. The experimental results showed that the proposed Hybrid ARIMA–LSTM model achieved the lowest MSE (4668.71) and RMSE (68.32) among the evaluated forecasting models, whereas the standalone LSTM achieved a slightly lower MAE (49.87). These findings indicate that the residual learning strategy was effective in reducing larger forecasting errors while maintaining competitive prediction accuracy. Furthermore, prediction error analysis, residual plots, and residual error distribution showed that the proposed model produced stable forecasting performance, with residuals randomly distributed around the zero reference line and minimal systematic prediction bias. This study was conducted using a univariate daily demand dataset without incorporating external factors such as promotions, pricing, or seasonal events. Future research may extend the proposed framework by incorporating multivariate features, comparing its performance with more advanced deep learning architectures, and evaluating its applicability in real-world inventory management and decision support system.

REFERENCES

[1] N. S. Dissurul, L. F. Wally, R. A. Zuleika, S. J. A. P. Antoni, R. A. J. F. Maulidina, and Rusdi Hidayat Nugroho, "Transformasi Pola Belanja Konsumen dari Ritel Konvensional Menuju Quick Commerce Melalui Layanan Pengiriman Instan," *J. Bisnis Kreat. dan Inov.*, vol. 2, no. 4, pp. 241–256, 2025, doi: 10.61132/jubikin.v2i4.1138.

[2] R. Sari Handayani, Gizka Alysah Diva, Eko Nur Hermansyah, and Ari Siswati, "Transformasi Digital E-Commerce di Era Post-Pandemic: Analisis Tren dan Dampak pada Perilaku Konsumen Indonesia," *J. Inform. dan Kesehat.*, vol. 2, no. 2, pp. 108–117, 2025, doi: 10.35473/ikn.v2i2.3704.

[3] X. Wen, J. Liao, Q. Niu, N. Shen, and Y. Bao, "Deep learning-driven hybrid model for short-term load forecasting and smart grid information management," *Sci. Rep.*, vol. 14, no. 1, pp. 1–16, 2024, doi: 10.1038/s41598-024-63262-x.

[4] G. Theodoridis and A. Tsadiras, "Retail Demand Forecasting: A Comparative Analysis of Deep Neural Networks and the Proposal of LSTMixer, a Linear Model Extension," *Inf.*, vol. 16, no. 7, 2025, doi: 10.3390/info16070596.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[5] Y. Liu, D. Kalaitzi, M. Wang, and C. Papanagnou, "A machine learning approach to inventory stockout prediction," *J. Digit. Econ.*, vol. 4, pp. 144–155, 2025, doi: 10.1016/j.jdec.2025.06.002.

[6] R. Fildes, S. Ma, and S. Kolassa, "Retail forecasting: Research and practice," *Int. J. Forecast.*, vol. 38, no. 4, pp. 1283–1318, 2022, doi: 10.1016/j.ijforecast.2019.06.004.

[7] C. Giri and Y. Chen, "Deep Learning for Demand Forecasting in the Fashion and Apparel Retail Industry," *Forecasting*, vol. 4, no. 2, pp. 565–581, 2022, doi: 10.3390/forecast4020031.

[8] K. Douaioui, R. Oucheikh, O. Benmoussa, and C. Mabrouki, "Machine Learning and Deep Learning Models for Demand Forecasting in Supply Chain Management: A Critical Review," *Appl. Syst. Innov.*, vol. 7, no. 5, 2024, doi: 10.3390/asi7050093.

[9] C. Li, X. Fang, Z. Yan, Y. Huang, and M. Liang, "Research on Gas Concentration Prediction Based on the ARIMA-LSTM Combination Model," *Processes*, vol. 11, no. 1, 2023, doi: 10.3390/pr11010174.

[10] V. Kramar and V. Alchakov, "Time-Series Forecasting of Seasonal Data Using Machine Learning Methods," *Algorithms*, vol. 16, no. 5, 2023, doi: 10.3390/a16050248.

[11] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks," *Futur. Internet*, vol. 15, no. 8, p. 255, Jul. 2023, doi: 10.3390/fi15080255.

[12] W. M. S. Bandara and W. A. R. De Mel, "Forecasting Nonlinear Time Series with ARIMA, ANN, and Hybrid Models: A Case Study on Inflation Rate in Sri Lanka," *Indones. J. Stat. Its Appl.*, vol. 9, no. 1, pp. 145–156, 2025, doi: 10.29244/ijsa.v9i1p145-156.

[13] Z. Liu, Z. Zhang, and W. Zhang, "A Hybrid Framework Integrating Traditional Models and Deep Learning for Multi-Scale Time Series Forecasting," *Entropy*, vol. 27, no. 7, 2025, doi: 10.3390/e27070695.

[14] H. Dhake, Y. Kashyap, and P. Kosmopoulos, "Algorithms for Hyperparameter Tuning of LSTMs for Time Series Forecasting," *Remote Sens.*, vol. 15, no. 8, pp. 1–17, 2023, doi: 10.3390/rs15082076.

[15] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[16] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting," *Int. J. Forecast.*, vol. 37, no. 4, pp. 1748–1764, 2021, doi: 10.1016/j.ijforecast.2021.03.012.

[17] W. Hu, Y. Yu, J. Yan, Z. Zhao, W. Sun, and X. Shen, "Water Flow Prediction Based on

Improved Spatiotemporal Attention Mechanism of Long Short-Term Memory Network," *Water (Switzerland)*, vol. 16, no. 11, 2024, doi: 10.3390/w16111600.

[18] M. Alharthi and A. Mahmood, "xLSTMTime: Long-Term Time Series Forecasting with xLSTM," *AI*, vol. 5, no. 3, pp. 1482–1495, 2024, doi: 10.3390/ai5030071.

[19] X. Song, L. Deng, H. Wang, Y. Zhang, and Y. He, "Deep learning-based time series forecasting," *Artif. Intell. Rev.*, vol. 58, no. 23, pp. 1–67, 2025, doi: 10.1007/s10462-024-10989-8.

[20] M. H. P. S. Muhammad Hakam Fardana, Wahyu Syaifullah Jauharis Saputra, "Analisis Prediksi Emisi Karbon Dioksida Kendaraan Bermotor Menggunakan Algoritma Hibrida LSTM Dan Arima," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 3, 2025, doi: 10.23960/jitet.v13i3.6782.

[21] M. S. Haque, M. S. Amin, J. Miah, D. M. Cao, M. A. Sayed, and S. Ahmed, "Retail Demand Forecasting: A Comparative Study for Multivariate Time Series," *J. Math. Stat. Stud.*, vol. 4, no. 4, pp. 40–46, 2023, doi: 10.32996/jmss.2023.4.4.4.

[22] J. Zulu, G. Mwansa, and K. Changwe, "Forecasting Inflation Rate Using the ARIMA Model: Zambia's Perspective from 2023 to 2043," *Int. J. Res. Sci. Innov.*, vol. XI, no. XII, pp. 698–713, 2025, doi: 10.51244/ijrsi.2024.11120063.

[23] Z. Zhan and S. K. Kim, "Versatile time-window sliding machine learning techniques for stock market forecasting," *Artif. Intell. Rev.*, vol. 57, no. 8, pp. 1–25, 2024, doi: 10.1007/s10462-024-10851-x.

[24] X. Kong *et al.*, *Deep learning for time series forecasting: a survey*, vol. 16, no. 7–8. Springer Berlin Heidelberg, 2025. doi: 10.1007/s13042-025-02560-w.

[25] R. A. Nampira, J. M. Sambas, I. Nur, L. Fitriana, L. Adhi, and K. Sulung, "ARIMA and LSTM Comparison for Forecasting Healthcare Service Costs in Bogor," vol. 6, no. 4, pp. 352–360, 2025, doi: 10.47065/bit.v5i2.2278.