



Forecasting continuous and intermittent demand: A segmented empirical benchmarking framework for mixed- demand portfolios

Michelle Caroline, Siana Halim*

Department of Industrial Engineering, Faculty of Industrial Technology, Petra Christian University, Indonesia

Abstract

Accurate demand forecasting is a significant challenge in retail, particularly with products exhibiting both continuous and intermittent demand patterns. Most existing studies analyze these demand types separately, limiting insights into model performance. This study presents a segmented empirical benchmarking framework that evaluates various forecasting methods—statistical, intermittent-demand, and deep learning—within a mixed-demand context, using consistent validation strategies. The framework classifies 577 products based on demand characteristics and utilizes tailored models, including Facebook Prophet (additive and multiplicative), intermittent-demand methods (ADIDA, IMAPA, TSB, optimized SES), and deep learning models (RNN and LSTM). It is validated with transaction data from an Indonesian retailer in the Bicycle Parts, Accessories, and Apparel (PAA) sector, covering 2019-2025 through sliding and expanding window techniques. Results based on Mean Absolute Error (MAE) show that aggregation-based methods such as ADIDA and IMAPA achieve lower error levels for intermittent demand than the company's current forecasting approach. For continuous demand, differences across models are relatively small, with no single method consistently dominating. Overall, performance differences across models remain moderate, and Diebold-Mariano test results suggest that these differences are not consistently statistically significant. The findings indicate that forecasting performance depends on aligning models with demand characteristics rather than relying on a single approach. The proposed framework provides practical guidance for model selection in mixed-demand environments.

This is an open-access article under the [CC BY-SA](#) license



Keywords:

Facebook Prophet;
Intermittent demand forecasting;
Long Short-Term Memory;
Recurrent Neural Network;
Retail forecasting.

Article History:

Received: November 14, 2025
Revised: April 27, 2026
Accepted: July 20, 2026
Published: October 10, 2026

Corresponding Author:

Siana Halim
Department of Industrial
Engineering, Faculty of
Industrial Technology, Petra
Christian University, Indonesia
Email: halim@petra.ac.id

INTRODUCTION

Accurate demand forecasting is critical for effective retail and supply chain management, yet it remains a persistent challenge when product portfolios exhibit heterogeneous demand patterns. In practice, retailers manage both high-volume, continuously sold items and low-volume, sporadically purchased products [1, 2, 3]. These differences complicate model selection: while continuous demand is well captured by models that exploit seasonality and trend, intermittent demand requires specialized techniques that can

handle sparse and zero-inflated series. Applying a single forecasting approach across all products often results in systematic bias, excess inventory, or frequent stockouts [1][4].

The current literature on demand forecasting can be broadly divided into three main categories: (1) continuous demand forecasting, (2) intermittent demand forecasting, and (3) machine learning and deep learning approaches. Although each category has shown methodological advancements, the literature remains fragmented, with few cross-comparative

evaluations conducted under unified experimental conditions [5][6].

Continuous demand forecasting studies primarily utilize statistical and regression-based models, such as the Naïve approach, Moving Average, Exponential Smoothing, ARIMA, and SARIMA [7, 8, 9]. These models effectively capture trends and seasonal patterns in dense time series data [6]. However, recent research (e.g. [2][4]) indicates that these models have structural limitations when applied to zero-inflated or sparse datasets. This can result in biased forecasts and diminished robustness in diverse demand environments.

In contrast, research on intermittent demand forecasting focuses on specialized techniques such as Croston, SBA, TSB, ADIDA, and IMAPA. These methods specifically account for the random nature of both the occurrence and size of demand, making them more suitable for sparse and irregular data series [10, 11, 12]. Recent studies [13][14] have shown that aggregation-based approaches, such as ADIDA and IMAPA, outperform traditional methods by enhancing signal stability.

Machine learning and deep learning techniques, such as Arimax and neural networks [15], Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks have garnered significant attention for their capacity to model complex nonlinear dependencies and temporal dynamics [16][17]. Recent advancements include the use of the Facebook Prophet model, which effectively handles multiple seasonality, holiday effects, and trend changepoints using an additive regression framework [18, 19, 20]. Additionally, Abolghasemi et al. [21] explored the impact of promotions on demand volatility, noting that they often distort historical patterns and decrease the accuracy of forecasts. Chen and Lu [22] integrate clustering and machine learning algorithms for demand forecasting in multichannel fashion retailers. They concluded that forecasting models need to be customized for varying levels of variability, such as the coefficient of variation, to effectively capture underlying behaviors.

A critical synthesis of existing research reveals that forecasting methods are usually evaluated in isolated demand settings, often using different datasets, validation strategies, and performance metrics. This fragmentation limits the ability to compare findings and makes it challenging to draw consistent conclusions about model performance across mixed-demand portfolios. While statistical models effectively capture trends and seasonality in continuous demand, and intermittent-demand methods tackle

sparsity through aggregation or decomposition, these approaches are seldom assessed under the same experimental conditions. Similarly, machine learning and deep learning models are often tested independently with varying validation schemes, which further diminishes comparability.

This limitation is particularly relevant in retail environments where product portfolios consist of diverse SKUs displaying both continuous and intermittent demand patterns. Consequently, there is a lack of empirical evidence regarding how different forecasting approaches perform in relation to one another within a single operational context. The absence of a unified benchmarking framework hinders systematic comparisons and limits the ability to provide practical guidance for model selection.

To address this gap, this study proposes a comprehensive empirical benchmarking framework that evaluates statistical, intermittent-demand, and deep learning models under consistent experimental conditions. All models are assessed using the same dataset and identical validation strategies, which allows for direct and fair comparisons across various demand types. Furthermore, the framework explicitly links demand characteristics with model performance and incorporates a real-world operational benchmark to determine whether advanced models yield significant improvements over existing forecasting practices. This approach offers both empirical insights and practical guidance for model selection in mixed-demand environments.

This framework is illustrated using historical transaction data from Indonesia's Bicycle Parts, Accessories, and Apparel (PAA) sector, covering the years 2019 to 2025. It assesses various models, including Prophet, ADIDA, IMAPA, TSB, SES-Optimized, RNN, and LSTM, within a consistent validation setting.

METHOD

This study proposes an analyst-in-the-loop framework for forecasting mixed-demand products. It combines data preparation, demand classification, model selection, and iterative validation to ensure that the methods align with the specific characteristics of the demand. The framework is illustrated using data from the Bicycle Parts, Accessories, and Apparel (PAA) sector and is designed to be adaptable to other industries that exhibit similar demand variability.

Literature Analysis: State of the Art in Demand Forecasting

This section provides a systematic overview and critical evaluation of demand forecasting

methods by categorizing previous studies into continuous-demand, intermittent-demand, and machine learning approaches, while also highlighting their respective strengths and limitations.

Model Selection Rationale

The selection of forecasting models is based on the structural characteristics of demand data and the strengths of each method. There are three categories of models to consider: statistical models (such as Prophet), intermittent-demand methods (like Croston-based, ADIDA, IMAPA, and TSB), and deep learning models (including RNN and LSTM).

Prophet is particularly effective for continuous demand because it can model trends and seasonality. Intermittent-demand methods are used for sparse and zero-inflated data, utilizing decomposition or aggregation to enhance stability. Deep learning models are incorporated to capture nonlinear temporal patterns.

This combination allows for a structured comparison across different demand types, ensuring that model selection aligns with the specific characteristics of the demand rather than relying on a single method.

Forecasting Workflow

Figure 1 illustrates the overall research workflow, which consists of four main stages: (1) data preprocessing and cleaning, (2) classification of demand into continuous and intermittent categories, (3) selection of models and forecasting based on the type of demand, and (4) evaluation and validation of the models using various performance metrics and statistical tests. This structured workflow ensures methodological consistency and enhances reproducibility by clearly defining each step in the forecasting

process, from preparing raw data to comparing final models.

Exploratory data analysis identifies trends, seasonality, and zero-demand periods, guiding model selection. Performance is validated using sliding and expanding window strategies, which replicate rolling forecast scenarios. Accuracy is measured primarily through Mean Absolute Error (MAE), enabling comparison across products with different sales volumes. The process is iterative, with parameters adjusted until optimal accuracy is achieved. The summaries of the models are provided in the following sections.

Forecasting Continuous Demand

Research on continuous demand forecasting has traditionally focused on statistical models such as ARIMA and SARIMA, which capture trend and seasonality under regular demand conditions. However, their performance may decline in the presence of structural changes or promotional effects [23]. More recently, regression-based models such as Facebook Prophet have been introduced to handle multiple seasonality, changepoints, and calendar effects.

Prophet Forecasting

Prophet model focuses on three main components of decomposable time series: trend, seasonality, and holidays.

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (1)$$

where, $g(t)$ is the trend function, which models non-periodic changes in the time series, $s(t)$ is periodic change, $h(t)$ is the effect of holidays, and ϵ_t is error for any unusual changes not accommodated by the model (see [19][20] for the details).

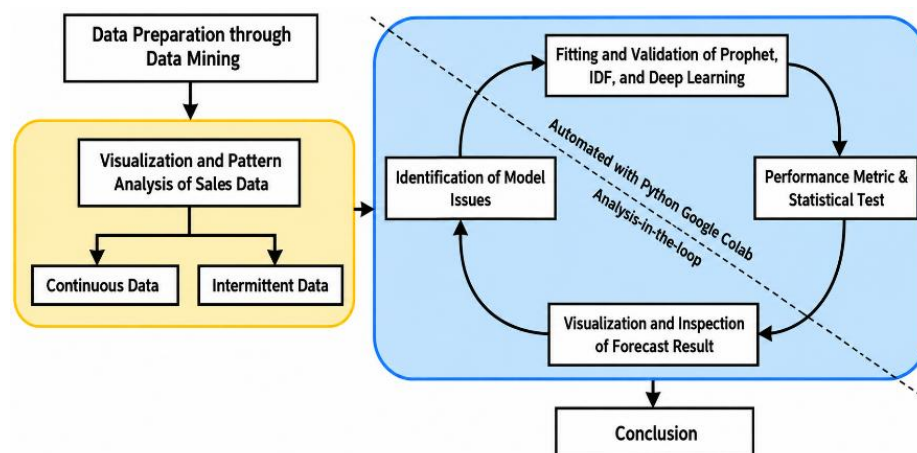


Figure 1. Data-driven Forecasting Workflow

Forecasting Intermittent Demand

Intermittent demand forecasting focuses on time series data characterized by sparse occurrences and a significant number of zero values.

Methods such as Croston, SBA, TSB, ADIDA, and IMAPA specifically distinguish between the size of demand and the occurrence of demand.

Intermittent Demand Forecasting (IDF)

Conventional forecasting models often struggle with intermittent demand, which refers to demand that occurs infrequently and irregularly. Several Intermittent Demand Forecasting (IDF) techniques have been created to handle intermittent demand patterns in time series data. These methods include the following:

a. Single Exponential Smoothing Optimized (SESOpt)

The Single Exponential Smoothing (SES) method can be expressed with the following equation [24].

$$\hat{d}_{t+1} = \hat{d}_t + \alpha(d_t - \hat{d}_t) \quad (2)$$

In equation (2), $d_t - \hat{d}_t$ is the forecast error, while t is the forecast period, and α is the smoothing factor ($0 < \alpha < 1$), where α can be optimized statistically.

b. Croston and Optimized Croston

Croston's method is expressed by the following equation [25]:

$$\hat{z}_t = \alpha y_t + (1 - \alpha)\hat{z}_{t-1} \quad (3)$$

$$\hat{p}_t = \alpha p_t + (1 - \alpha)\hat{p}_{t-1} \quad (4)$$

$$\hat{y}_{t+1} = \begin{cases} \frac{\hat{z}_t}{\hat{p}_t} & \text{if } y_t > 0 \\ \frac{\hat{z}_{t-1}}{\hat{p}_{t-1}} & y_t = 0 \end{cases} \quad (5)$$

where y_t represents the actual demand at time t and p_t is the number of periods since the last demand, α is the smoothing factor ($0 < \alpha < 1$), $\hat{z}_t$ denotes the estimated demand when it occurs, $\hat{p}_t$ is the estimated average interval between demands, and $\hat{y}_{t+1}$ is the forecast for the next period. Additionally, [26] developed an optimized version of Croston's method by evaluating various smoothing constant values to enhance prediction accuracy.

c. Teunter-Syntetos-Babai (TSB)

The TSB method updates Croston's two key components: the estimated demand size (z'_t) and the probability of demand (p'_t). The forecast value is expressed using an equation formulated by [14]:

$$\hat{y}_{t+1} = p'_t \cdot z'_t \quad (6)$$

if $y_t > 0$:

$$\begin{aligned} p'_t &= p'_{t-1} + \beta(1 - p'_{t-1}) \\ z'_t &= z'_{t-1} + \alpha(y_t - z'_{t-1}) \end{aligned}$$

If $y_t = 0$:

$$\begin{aligned} p'_t &= p'_{t-1} + \beta(0 - p'_{t-1}) \\ z'_t &= z'_{t-1} \end{aligned} \quad (7)$$

Where α and β are smoothing constants ($0 < \alpha, \beta < 1$)

d. Aggregate-Disaggregate Intermittent Demand Approach (ADIDA)

ADIDA is a method developed to address the shortcomings of Croston's approach. This method begins by aggregating demand data to a higher level using the following equation [13]

$$Y_1 = \sum_{t=1}^k y_t, Y_2 = \sum_{t=k+1}^{2k} y_t, \quad (8)$$

where, y_t refers to the demand for each period, k is the aggregation level, Y_t denotes the aggregate data. The forecast is performed on aggregate data using standard forecasting methods as represented by $\hat{Y}_{t+1} = f(Y_t, Y_{t-1}, \dots)$ and the disaggregation is calculated using $\hat{y}_{t+1} = \frac{\hat{Y}_{t+1}}{k}$.

e. Intermittent Multiple Aggregation Prediction Algorithm (IMAPA)

The Multiple Aggregation Prediction Algorithm (MAPA) aggregates demand data across various time levels. By combining the prediction results, it captures patterns that may not be apparent at a single level of aggregation [27]. [28] adapted MAPA for intermittent demand, leading to the development of the Intermittent Multiple Aggregation Prediction Algorithm (IMAPA), which is expressed by the following equation:

$$\hat{y}_{t+1} = \frac{1}{K} \sum_{k=1}^K \frac{\hat{Y}_{t+1}^{(a_k)}}{a_k} \quad (9)$$

k is the total number of aggregation levels used; a_k is the k th aggregation level.

Deep Learning Forecasting: Recurrent Neural Network (RNN) dan Long Short-Term Memory (LSTM)

A Recurrent Neural Network (RNN) is a type of artificial neural network specifically designed to process sequential data while maintaining information from previous time steps. However, RNNs face a significant limitation known as the vanishing gradient problem. This occurs when the gradient value becomes very small

during training, making it difficult for the network to recognize long-term patterns in the data [29].

To address these limitations in handling long-term dependencies, the Long Short-Term Memory (LSTM) architecture was developed. The LSTM architecture consists of three main gates: the forget gate, the input gate, and the output gate, which work in unison to dynamically regulate the flow of information.

Deep learning models (RNN and LSTM) were evaluated using a standard 80/20 train–test split rather than rolling window validation. This choice reflects the data requirements of neural network models, which rely on longer continuous sequences to learn temporal dependencies and may not perform reliably under small or fragmented training windows.

As a result, the evaluation of deep learning models is not directly comparable to that of statistical and intermittent-demand models, which are assessed using sliding and expanding window strategies. Instead, deep learning models are included as a complementary analysis to examine their potential in capturing nonlinear patterns under sufficient data conditions, rather than as a strictly equivalent benchmarking component.

Critical Synthesis of the State of the Art

The current literature on demand forecasting tends to emphasize individual

methodologies rather than comparative evaluation across diverse demand conditions. As a result, empirical evidence on how different forecasting approaches perform within mixed-demand environments remains limited. This highlights the need for a unified benchmarking framework that evaluates models for both continuous and intermittent demand under consistent conditions. Table 1 summarizes existing approaches and underscores the value of such a framework.

Windowing Strategy for Validating Time Series Forecasting Models

Validation follows cross-validation strategies tailored for time series windows [30]. In the sliding window approach, the training window size remains fixed, using only recent data. In the expanding window approach, the training window grows over time, incorporating all available historical data. Both approaches preserve chronological order, reducing look-ahead bias and ensuring realistic forecast evaluation. Figure 2 illustrates the differences between sliding windows, which focus on adapting to recent changes in demand, while expanding windows prioritize the stability of parameters. This difference in approach later clarifies the performance variations between continuous and intermittent demand models.

Table 1. State of the Art Comparison of Demand Forecasting Approaches

Research Stream	Representative Methods	Main Strengths	Key Limitations	Evidence Gap	How This Study Addresses the Gap
Continuous demand forecasting	ARIMA, SARIMA, Prophet [7,13,14]	Captures trend and seasonality well; suitable for dense time series; interpretable	Performs poorly with zero-inflated or sparse data; limited applicability to intermittent demand	Rarely evaluated alongside intermittent-demand models within mixed portfolios	This study evaluates continuous-demand models alongside intermittent and deep learning models within a unified framework, using the same dataset and validation strategy.
Intermittent demand forecasting	Croston, SBA, TSB, ADIDA, IMAPA [8][9]	Designed for sparse and irregular demand; aggregation improves stability	Limited responsiveness to sudden changes; often evaluated in isolation	Lack of comparative benchmarking against continuous-demand and ML models	This study compares intermittent-demand methods to statistical and deep learning models under controlled conditions.
Machine / deep learning	RNN, LSTM [11][12]	Captures nonlinear patterns; low error metrics with sufficient data	Requires large datasets; low interpretability; overly smooth forecasts for sparse data	Limited evidence on operational suitability in mixed-demand settings	This study evaluates deep learning models for retail forecasting across continuous and intermittent demand.
Existing practice	Manual/heuristic forecasting	Incorporates managerial knowledge (lead time, safety stock)	Subjective; lower statistical accuracy	Not empirically compared with advanced models under consistent validation	This study compares manual forecasting with data-driven models to evaluate performance improvements and highlight the advantages of hybrid approaches.

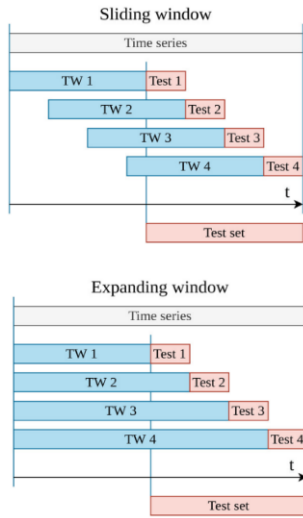


Figure 2. Windowing Strategy [31]

Performance Metric, Statistical Testing and Robustness Evaluation

Forecast accuracy is evaluated using multiple error metrics to ensure robustness across continuous and intermittent demand patterns. MAPE is reported for completeness but interpreted cautiously due to its sensitivity to low or zero actual values, especially in intermittent demand. MAE, MSE, and RMSE are also used to improve evaluation robustness. Relative performance is assessed using the ratio of model MAE to that of the company's current forecasting approach (RMAE). The "Current" model, based primarily on managerial judgment and operational experience, serves as a practical benchmark for evaluating improvements from data-driven models. All models are evaluated consistently across the same training and testing splits and validation strategies. Using multiple error metrics allows for cross-verification of model rankings, ensuring that conclusions are not overly influenced by the shortcomings of a single metric. Below are some key measurement metrics that can be utilized [32]:

Mean squared error: $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$

Mean absolute error: $MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$

Mean absolute percentage error:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Relative MAE: $RMAE = \frac{MAE_{model}}{MAE_{current}}$

In series with intermittent demand, where periods of zero demand are common, the Mean Absolute Percentage Error (MAPE) can exaggerate forecast errors and favor smoother forecasts in model comparisons.

Therefore, model evaluation is primarily based on MAE, which provides more stable and reliable measures of forecasting performance [6]. This approach of using multiple metrics helps minimize the risk of drawing misleading conclusions due to zero inflation and sparse data observations.

Statistical Significance Testing

To determine if the differences in forecasting accuracy are statistically significant, this study uses the Diebold–Mariano (DM) test to compare the predictive performance of competing models. The DM test assesses whether the difference in forecast errors between the two models is statistically significant [33].

The test is based on the loss differential series:

$$d_t = L(e_{1t} - e_{2t})$$

where $L(\cdot)$ represents the loss function (e.g., squared error), and e_{1t}, e_{2t} are forecast errors from two competing models.

Under the null hypothesis of equal predictive accuracy, $E[d_t] = 0$ [34]. A statistically significant DM statistic indicates that one model outperforms the other.

By integrating the DM test, this study improves the strength of model comparison beyond descriptive error metrics, such as MAPE, MAE, RMSE, and RMAE.

Computational Complexity

The computational complexity of forecasting models varies significantly across different categories. Statistical models, such as Prophet, exhibit relatively low computational complexity and fast training times, making them well-suited for large-scale retail applications.

Intermittent-demand methods, including ADIDA and IMAPE, involve additional steps for aggregation and disaggregation. However, they remain computationally efficient because they rely on simple statistical operations.

In contrast, deep learning models like RNN and LSTM have higher computational complexity due to their multi-layer architecture and iterative training processes. Their training time increases with the size of the dataset and the number of parameters, which makes them more resource-intensive.

This variation in computational requirements is an important factor to consider for practical implementation, particularly in real-time forecasting and large-scale inventory systems.

Software and Parameters

All models were implemented using Python. The Prophet model was created using the Prophet library, while intermittent-demand methods were developed with custom scripts and various time-series libraries. Deep learning models, specifically RNN and LSTM, were built using TensorFlow/Keras.

Model parameters were selected through a combination of default settings and empirical tuning. For the Prophet model, key parameters included trend specification (either linear or logistic), the changepoint prior scale, and seasonality mode (additive or multiplicative). In the case of intermittent-demand models, smoothing parameters were either fixed or optimized based on the training data. For deep learning models, hyperparameters such as the number of units, batch size, number of epochs, dropout rate, and regularization were tuned experimentally. All experiments were carried out with consistent data splits and validation strategies to ensure a fair comparison and reproducibility.

RESULTS AND DISCUSSION

Data Preparation and Demand Classification

The first step involved curating the dataset of 577 active SKUs from historical transactions between 2019 and 2025. The data from the COVID-19 period are not included, as there are many anomalies in the dataset during that time. The raw datasets were thoroughly cleaned and standardized, particularly by unifying vendor codes for physically identical products to reduce redundancy and enhance data quality. After cleaning and unifying product codes, the data were segmented into two categories based on sales frequency: continuous demand (products with stable, recurring sales patterns) and intermittent demand (products with irregular sales and zero-demand periods).

Outliers were identified using the Interquartile Range (IQR) method and adjusted to maintain data continuity without discarding valuable historical trends. The resulting datasets were then reformatted into a time-series structure, facilitating the robust training, validation, and evaluation of various forecasting models.

To enhance model performance, categorical variables like product size and promotional periods were handled differently across approaches. In the Prophet model, dummy variables acted as external regressors to account for size variations and promotions. For intermittent demand forecasting (IDF), SKU-level predictions included size variations, eliminating the need for

additional dummy variables. In deep learning models (RNN and LSTM), dummy variables were added as input features alongside historical demand data, allowing for nonlinear interactions. This tailored approach ensures each method effectively utilizes contextual information while remaining comparable.

Continuous Demand Models

Continuous demand refers to sales that follow a stable, consistent pattern, occurring nearly every month throughout the observation period. Although sales fluctuate sharply at the beginning and middle of the year, no prolonged gaps occur, indicating relatively consistent demand. Seasonal factors or promotions may cause monthly spikes. This pattern provides a solid basis for forecasting models to identify long-term trends and recurring seasonal patterns while minimizing overfitting.

The Prophet model is structured: (1) Added dummy variables for size and promotion. A dummy variable representing size is included as an additional regressor in the Prophet model. This allows the model to account for the impact of size on demand patterns. This approach is applied to products in the apparel category, such as footwear, gloves, jerseys, shorts, pants, and socks. Additionally, dummy variables are included during significant promotional periods, such as the Ramadan Sale (May to July) and the Year-End Sale (December to January). These dummy variables serve as additional regressors to help the model account for the impact of promotional spikes in demand. By doing so, the model can distinguish between the effects of promotions and seasonal patterns or underlying trends. This leads to more accurate and responsive predictions. (2) Setting the floor and cap values, the floor is set at 0 to ensure that the forecast model does not predict negative values. The cap value is determined based on the average sales in one year, excluding the highest sales figure to prevent overestimating the forecasts.

Windowing strategy: Two windowing strategies were employed: the moving window and the expanding window. Each strategy was tested using three different combinations of window size and prediction horizon parameters (n -periods), which define the lengths of the training and testing periods. The combinations are as follows: (3,1) represents 3 months of training data and 1 month of testing, (6,1) consists of 6 months of training and 1 month of testing, and (12,3) includes 12 months of training and 3 months of testing.

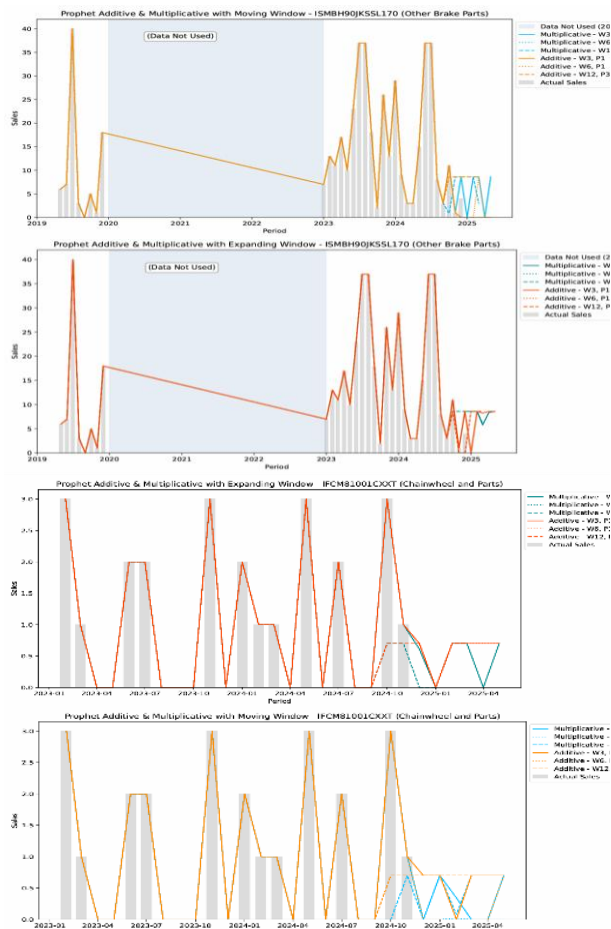


Figure 3. Examples of Prophet Forecasting for Continuous (upper figures) and Intermittent (lower figures) Data with All Possible Combinations of Windowing Strategies: Moving-Left and Expanding-Right, Window Sizes, and Horizons

These combinations were selected to evaluate how effectively the model can learn both short-term and long-term patterns. Each product will be tested with every combination.

Model Parameters: The Prophet model is developed using two approaches: additive and multiplicative. The additive model employs linear growth and a change-point prior scale of 0.01, which helps ensure more stable predicted trends and makes it less sensitive to minor fluctuations. The multiplicative model utilizes logistic growth and a changepoint prior scale of 0.03, providing greater flexibility in capturing rapidly changing trends. This model is also proportional to the data's level. During the model selection process, each product was tested using all possible combinations of windowing strategies, window sizes, and horizons. Figure 3 demonstrates how the multiplicative Prophet model captures proportional seasonal variations in continuous-

demand series. The comparison of different windowing strategies indicates that shorter moving windows enhance responsiveness to recent demand fluctuations, which explains the superior performance of continuous products.

The evaluation of continuous-demand series indicates that Prophet models provide competitive and relatively consistent performance compared to the company's manual forecasts. In several cases, the multiplicative form of Prophet yields slightly lower error values, suggesting its ability to capture proportional seasonal fluctuations and short-term demand shifts. The use of a moving window strategy with shorter training periods (for example, 3 months of training followed by 1 month of testing) appears to support adaptability to recent demand changes. In contrast, using longer window sizes and horizons reduces performance, as it tends to blur recent market signals in favor of older data. These findings suggest that continuous-demand forecasting may benefit from models capable of capturing trend and seasonality within shorter timeframes, although no single approach demonstrates a clear advantage across all cases.

IDF Models

Intermittent data is marked by irregular sales frequency, often featuring extended periods without sales (zero value) that occur inconsistently. The intermittent pattern is characterized by several months of sales gaps and small, irregular sales volumes, resulting in a graph that appears random. The intermittent pattern reflects products that are slow-moving or may only be required under specific conditions. Therefore, understanding this type of pattern is crucial not only for modeling but also to aid decision-making processes.

Six IDF models, Croston Classic, Croston Optimized, ADIDA, IMAPA, TSB, and SES Optimized, were evaluated. In its implementation, the model incorporates a dummy size variable within the apparel product category (which includes footwear, gloves, jerseys, shirts/t-shirts/polo, shorts/pants, and socks) to account for variations in product size. This allows the model to recognize the impact of size on demand patterns during the forecasting process.

The IDF model structured: Dummy size variables were incorporated into Apparel product categories, including Footwear, Gloves, Jerseys, Shorts/Pants, and Socks, to represent size variations. This enhancement enables the model to recognize how size influences demand patterns during the forecasting process.

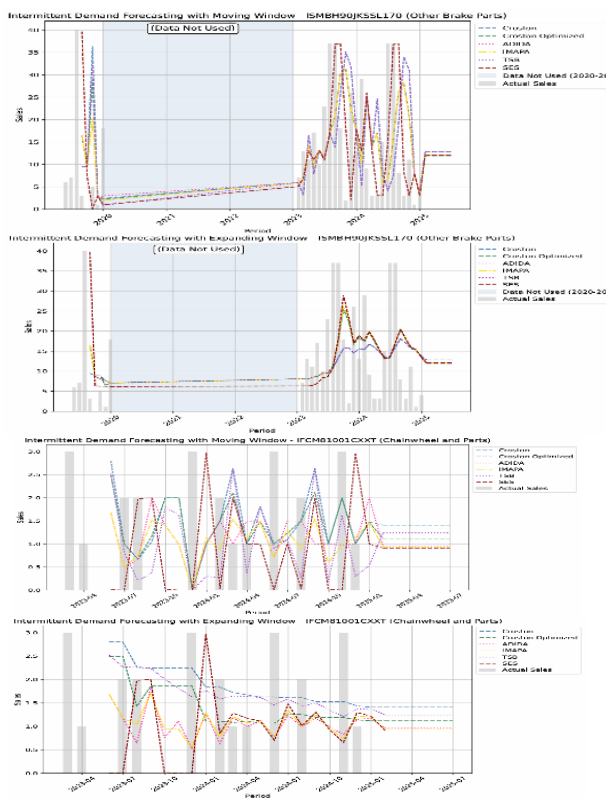


Figure 4. Examples of IDF Forecasting for Continuous (upper) and Intermittent (lower) Data with All

Windowing strategies: The fitting and validation process was carried out using two strategies: a moving window and an expanding window, both with a window size of 3 and a horizon of 1. The intermittent model relies significantly on the actual sales data within each training window. If the window size is too large, there is an increased risk that one or more windows will contain only zeros. This can ultimately lead to the model failing to make predictions or producing misleading estimates. To maintain stability and minimize the risk of prediction failure due to data irregularities, a fixed combination and a smaller window size were selected for the entire fitting, validation, and evaluation process of the model.

Models Parameters:

The Croston Classic method utilizes the basic Croston approach without optimizing parameters, applying a smoothing factor (α) of 0.1, which is the default value from the programming library. The Croston Optimized Single Exponential Smoothing (SES) incorporates an internal optimization process. The smoothing factor is determined automatically based on the performance of the training data, allowing the model to adapt to dynamic variations in demand patterns.

The Aggregate-Disaggregate Intermittent Demand Approach (ADIDA) requires no manual parameter adjustments. It involves aggregating sales data over a longer time period (e.g., quarterly) to reduce randomness due to sparsity, followed by forecasts using a conventional model, which enhances noise resistance.

The Intermittent Multiple Aggregation Prediction Algorithm (IMAPA) is also based on aggregation but operates at multiple levels in parallel (e.g., monthly, bimonthly, quarterly, etc.). The prediction results from each level are averaged, eliminating the need for manual parameter adjustments. This method aims to stabilize prediction outcomes under conditions of highly fluctuating and infrequent demand.

The Teunter-Syntetos-Babai (TSB) method employs two smoothing factors: α demand (α_d) = 0.1 for smoothing the demand estimate, and α demand probability (α_p) = 0.1 for smoothing the estimated probability of demand occurrence. These parameter values are explicitly set to maintain a balance between responsiveness to new patterns and the stability of the prediction output.

For products with intermittent demand, specialized IDF models demonstrate relatively stable and competitive performance compared to both the Prophet model and the company's manual forecasting method. Among these, ADIDA, IMAPA, and SES Optimized, show slightly lower MAE and RMSE in several cases, reflecting their ability to reduce variability through temporal aggregation.

The expanding window strategy appears to support more stable predictions by incorporating a larger set of historical data, which can be beneficial in sparse demand settings. In contrast, Croston-based methods and TSB were less responsive to irregularities, often resulting in biased or static forecasts. These findings underscore the importance of aggregation-based approaches and long-memory strategies to effectively manage the challenges of sparsity and volatility in intermittent demand. Figure 4 illustrates the stabilizing effect of aggregation-based intermittent demand forecasting methods. By minimizing the impact of periods with zero demand, aggregation enhances forecast stability and reduces bias. This explains why ADIDA and IMAPA show relatively more stable performance than Croston-based methods when dealing with highly intermittent series.



Figure 5. Examples of Deep Learning

Deep Learning Models

In this study, we utilized RNN and LSTM as our main deep learning methods. Like the Prophet and IDF models, the RNN and LSTM approaches incorporate dummy size factors for various garment products, including footwear, gloves, jerseys, shirts/t-shirts/polos, shorts/pants, and socks. This inclusion ensures that size is considered during the training process.

Due to their necessity for large amounts of training data to capture complex patterns and long-term dependencies, RNN and LSTM models do not utilize windowing, moving, or expanding techniques. These models consist of multiple layers and memory cells, requiring substantial historical data to understand temporal patterns. Large window sizes can reduce the sample size, particularly for items with limited data, and increase the risk of overfitting. Thus, a window size of three and a horizon strike a balance by ensuring sufficient training data while maintaining relevance for short-term predictions in retail operations.

To ensure consistency, the training hyperparameters for the RNN and LSTM models include the ReLU (Rectified Linear Unit) activation function and L2 regularization set at 0.01 to help mitigate overfitting. A dropout rate of 0.4 is employed to prevent neurons from becoming overly dependent on each other, and an early stopping mechanism based on validation forecasting for Continuous (left) and Intermittent (right) data loss is applied, with a patience of 12 epochs. The dataset is split into 80% for training and 20% for validation.

Three combinations of hyperparameters are explored to assess how model complexity impacts prediction performance: (units = 16, batch size = 16, epochs = 10), (units = 32, batch size = 20, epochs = 12), and (units = 48, batch size = 24, epochs = 15). To avoid overfitting complex models and underfitting in simpler ones, historical data restrictions are considered when selecting the

combination. This incremental approach aims to find the optimal balance between model capacity and data sufficiency.

Figure 5 demonstrates that RNN and LSTM models generate smoother forecasts compared to statistical methods. While smoothing decreases error metrics in dense series, it also reduces responsiveness to short-term demand fluctuations, especially in sparse or zero-inflated contexts. This is attributed to the relatively short historical data span and the inherent complexity of deep learning models, which require large datasets to capture non-linear temporal dependencies effectively.

The RNN and LSTM models achieved relatively low error metrics, in some cases outperforming statistical methods on continuous series. However, their predictions often appeared overly smooth, failing to capture sudden fluctuations that are critical for short-term retail planning. This limitation is partly due to the limited data horizon and the complexity of deep learning models, which generally require larger datasets to reach full potential. While promising, deep learning was less interpretable for managers and less operationally practical than Prophet or IDF methods. These results suggest that deep learning can complement, but not replace, statistical models in mixed-demand environments.

Discussion

Table 2 reports forecast accuracy across MAPE, MAE and RMSE, providing a check for model comparison. Additionally, an important observation from Table 2 is that error levels remain relatively high across all models, particularly when assessed using MAPE. In practical terms, MAPE values exceeding 100% indicate that forecast errors are, on average, comparable to or larger than the actual demand, reflecting substantial forecasting difficulty rather than strong model performance. This does not solely indicate shortcomings in the models; rather, it highlights the inherent complexity of forecasting in mixed-demand environments. The coexistence of continuous and intermittent demand patterns introduces variability, which limits the models' ability to consistently capture the underlying dynamics across different products.

Given the limitations of MAPE when dealing with zero-inflated demand, this analysis focuses on MAE and RMSE as more reliable measures. A closer examination of Table 2 reveals that while differences in forecast performance exist, they are generally moderate. For intermittent demand, aggregation-based methods such as ADIDA, IMAPA, and SES Optimized reduce the MAE from 4.16 (the current method) to approximately 3.54-

3.58, corresponding to an improvement of around 14-15%. In the case of continuous demand, there is less variation among the models, with MAE values ranging from 7.90 to 10.51, indicating only incremental gains.

The Relative Mean Absolute Error (RMAE) provides further context by comparing each model against the company's current forecasting approach. Several models achieve RMAE values below 1, indicating improved performance; however, these improvements are not uniform across all demand types. In many instances, RMAE values are close to 1, suggesting that performance is comparable rather than clearly superior. To assess the statistical significance of these findings, the Diebold–Mariano (DM) test was conducted at the individual time-series level. Representative results can be found in Appendix

1. The findings indicate that while some models show statistically significant improvements over the baseline in certain cases ($p < 0.05$), these results are not consistent across all products. For instance, in intermittent-demand cases like "Chainwheel and Parts," aggregation-based methods demonstrate significant improvements. However, in other cases, such as "BB Set and Part," the differences between models are not statistically significant, despite variations in MAE. This variability suggests that statistical superiority is context-dependent and cannot be generalized across all SKUs. Overall, forecasting performance is influenced not only by the choice of model but also by the structural characteristics of demand, including intermittency and variability.

Table 2. Forecast Accuracy Means Value Measurement

Model	MAPE	Intermittent Demand		RMAE
		RMSE	MAE	
Current - Current	114.7945	4.933145	4.164384	1
IDF - ADIDA	110.2522	4.209665	3.581131	0.887365
IDF - CrostonClassic	107.2378	4.504217	3.880021	1.012423
IDF - CrostonOptimized	110.2726	4.331175	3.720462	0.934937
IDF - IMAPA	111.2501	4.217801	3.585267	0.890428
IDF - SESOpt	113.6025	4.272918	3.54086	0.898370
IDF - TSB	100.9758	4.368075	3.744965	0.909163
Prophet - Additive	94.01314	5.025749	4.15945	1.030033
Prophet - Multiplicative	93.19837	5.055649	4.137515	1.027803
LSTM	55.53213	4.316909	3.488821	0.856494
RNN	71.85266	5.215044	4.275282	1.082215

Model	MAPE	Continuous Demand		RMAE
		RMSE	MAE	
Current - Current	109.7527	10.25953	9.320513	1
IDF - ADIDA	143.212	23.26737	22.18019	1.42818
IDF - CrostonClassic	217.6589	18.98033	17.79369	1.953021
IDF - CrostonOptimized	215.7029	24.97344	24.02189	2.023226
IDF - IMAPA	143.7687	23.27386	22.18761	1.431892
IDF - SESOpt	134.1255	26.78171	25.6214	1.413087
IDF - TSB	192.4053	18.35369	17.15922	1.744524
Prophet - Additive	93.13619	11.78094	10.4397	1.113778
Prophet - Multiplicative	93.05705	11.78570	10.51204	1.076093
LSTM	53.67539	9.35288	7.898826	0.877358
RNN	83.44715	13.29766	11.18101	1.130780

Table 3. Summary of Model Performance by Demand Type

Demand Type	Models with Relatively Lower MAE	MAE Range	RMSE Range	Key Observation
Continuous Demand	Prophet (Additive & Multiplicative), LSTM	7.90 – 10.51	9.35 – 11.78	Prophet models show consistent performance; LSTM achieves lower MAE in some cases.
Intermittent Demand	ADIDA, IMAPA, SES Optimized, LSTM	3.48 – 3.58	4.21 – 4.32	Aggregation-based methods exhibit stable error; LSTM shows slightly lower MAE. However, their predictions often appeared overly smooth.
Baseline (Current)	Company (Manual)	4.16 (Intermittent) 9.32 (Continuous)	4.93 – 10.26	Serve as operational reference; performance comparable in some cases

Note: The table summarizes relative performance trends based on aggregated MAE and RMSE values reported in Table 2. Differences across models are moderate and no single model consistently dominates across all demand types.

From a methodological standpoint, the findings indicate that models designed for different demand structures exhibit performance patterns that are generally in line with their intended applications. For instance, Prophet models provide consistent results for continuous demand by effectively capturing trends and seasonal components. On the other hand, aggregation-based methods such as ADIDA and IMAPA demonstrate relatively stable performance for intermittent demand. Deep learning models, including Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks, can also model temporal patterns, but their error rates are comparable to those of the other methods studied.

It should be noted that the performance of deep learning models is evaluated under a different validation scheme compared to other models. Therefore, their results should be interpreted cautiously and viewed as indicative rather than directly comparable. This distinction further supports the conclusion that differences in forecasting performance are context-dependent and influenced by both model structure and evaluation design.

Overall, the absence of a consistently superior forecasting model is not a coincidence; it highlights the structural complexity of mixed-demand environments. The performance of a model depends more on how well the forecasting method aligns with the characteristics of the demand than on the sophistication of the algorithm itself. This emphasizes the significance of demand-based segmentation. By choosing the right models for different types of demand, we can achieve more meaningful and operationally useful forecasting results.

Comparative Synthesis

Table 3 presents a comparison of forecasting accuracy across different demand types. It shows that the performance of forecasting models is significantly influenced by the structure of the demand rather than the complexity of the models themselves. Specifically, aggregation-based methods tend to produce lower errors for intermittent demand, while the Prophet model consistently performs better with continuous demand series.

The comparative positioning of models across multiple performance criteria shows that Prophet balances accuracy, responsiveness, and interpretability, IDF methods provide high stability, deep learning prioritizes accuracy at the expense of interpretability and responsiveness, and manual forecasting remains strong on practical responsiveness and low data requirements. These findings

are broadly consistent with previous studies, which reported that aggregation-based methods such as ADIDA and IMAPA can improve forecasting stability for intermittent demand [13], [14], while model performance may vary according to the underlying demand characteristics [22]. The present study extends these findings by demonstrating their applicability within a mixed-demand retail portfolio under a unified empirical benchmarking framework.

Managerial Implications

From a managerial perspective, the findings indicate that forecasting in mixed-demand environments is constrained more by demand heterogeneity than by model choice alone. This limits the effectiveness of applying a single forecasting approach across product portfolios and highlights the need for adaptive forecasting systems.

Improvements from advanced forecasting models are often incremental, highlighting the need to focus on both accuracy and operational suitability, including interpretability and ease of implementation.

The comparison with the company's current forecasting approach emphasizes the value of combining data-driven models with managerial judgment. Statistical and machine learning models offer structured forecasts, but managerial insights are essential for considering contextual factors like promotions and lead times.

Overall, effective forecasting in mixed-demand environments depends less on identifying a superior model and more on developing flexible, context-aware forecasting processes that enhance inventory planning and supply chain responsiveness.

Integrating Data-Driven Models with Managerial Judgment

The study emphasizes the importance of combining human expertise with advanced forecasting methods. While the company's manual approach was statistically less accurate, it factored in critical operational elements such as safety stock, lead times, and product value, which purely data-driven models often overlook. This indicates that the most effective forecasting practice is a hybrid approach: models provide strong, data-driven predictions, while managers adjust these forecasts considering real-world operational factors. This integration creates a more comprehensive decision-making framework that balances predictive accuracy with organizational needs.

Broader Implications

The insights gained from the Bicycle Parts, Accessories, and Apparel (PAA) sector are applicable to other industries that handle mixed-demand portfolios, such as spare parts distribution, fashion retail, and service parts logistics. The proposed framework is versatile and can be adapted to situations where both continuous and intermittent demand exist. By clarifying when and why different models are most effective, this study not only enhances the academic understanding of forecasting but also provides practitioners with a replicable methodology for managing demand variability in complex supply chains.

CONCLUSION

This study presents a comprehensive empirical framework for benchmarking forecasts in mixed-demand environments, where both continuous and intermittent demand exist within the same product portfolio. The main contribution of this research is twofold: it not only compares statistical, intermittent-demand, and deep learning models under consistent validation conditions, but also illustrates that forecasting performance is fundamentally limited by the unique characteristics of demand and the quality of the data.

The findings indicate that no single forecasting method consistently outperforms the others for all types of demand. This does not suggest that any model is inadequate; instead, it highlights the structural complexity of mixed-demand systems. Factors such as heterogeneity, zero-inflation, and irregular demand patterns can hinder predictive accuracy. Therefore, the effectiveness of forecasting relies more on aligning the chosen models with the specific demand patterns rather than simply increasing the complexity of the models.

From a practical standpoint, the proposed framework offers a structured method for selecting forecasting techniques based on demand characteristics. This enables better inventory planning and supply chain decision-making. The findings support the use of segmented and hybrid forecasting strategies that integrate data-driven models with managerial judgment to enhance operational relevance.

Limitations and Future Research

This study has several limitations that should be considered when interpreting the results. First, the analysis relies on a single dataset from the Bicycle Parts, Accessories, and Apparel (PAA) sector in Indonesia, which may

restrict the generalizability of the findings to other industries or geographic regions. Second, the study focuses solely on selected forecasting models and validation strategies, omitting alternative methods such as ensemble approaches or probabilistic forecasting.

Additionally, the evaluation is based on historical demand data without considering other explanatory variables that could affect forecasting performance. Despite these limitations, the study offers a structured and general framework for assessing forecasting models in mixed-demand environments.

Future research can expand on this work in several ways. First, integrating hybrid forecasting approaches—by combining statistical models, machine learning, and deep learning techniques—could enhance predictive performance. Second, adding more explanatory variables, such as pricing, promotions, and external signals, may improve the model's responsiveness to demand variability.

AI USAGE STATEMENT

In preparing this paper, the authors utilized ChatGPT and Grammarly to enhance readability and language. Subsequently, they reviewed and edited as necessary, accepting full responsibility for the paper's content.

REFERENCES

- [1] Jeeva Mohana Chandran and M. Reyesudin Basir Khan, "A strategic demand forecasting: Assessing methodologies, market volatility, and operational efficiency," *Malaysian J. Buss. Econ. Mgmt*, pp. 150–167, Aug. 2024, doi: 10.56532/mjbem.v3i2.71.
- [2] K. Swaminathan and R. Venkitasubramony, "Demand forecasting for fashion products: A systematic review," *International Journal of Forecasting*, vol. 40, no. 1, pp. 247–267, Jan. 2024, doi: 10.1016/j.ijforecast.2023.02.005.
- [3] P. A. Marques and A. M. De Carvalho, "Reducing shrinkage and stockouts in a food retail store: A Portuguese case study," *J. Tek. Ind. J. Keilmuan dan Apl. Tek. Ind*, vol. 23, no. 2, pp. 93–100, Dec. 2021, doi: 10.9744/jti.23.2.93-100.
- [4] S. Mejía and J. Aguilar, "A demand forecasting system of product categories defined by their time series using a hybrid approach of ensemble learning with feature engineering," *Computing*, vol. 106, no. 12, pp. 3945–3965, Dec. 2024, doi: 10.1007/s00607-024-01320-y.
- [5] F. Petropoulos *et al.*, "Forecasting: Theory and practice," *International Journal of*

- Forecasting*, vol. 38, no. 3, pp. 705–871, Jul. 2022, doi: 10.1016/j.ijforecast.2021.11.001.
- [6] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “M5 accuracy competition: Results, findings, and conclusions,” *International Journal of Forecasting*, vol. 38, no. 4, pp. 1346–1364, Oct. 2022, doi: 10.1016/j.ijforecast.2021.11.013.
- [7] Y. Ensafi, S. H. Amin, G. Zhang, and B. Shah, “Time-series forecasting of seasonal items sales using machine learning – A comparative analysis,” *International Journal of Information Management Data Insights*, vol. 2, no. 1, p. 100058, Apr. 2022, doi: 10.1016/j.ijimei.2022.100058.
- [8] T. Sari and S. Sakti, “Forecast of sugar demand in retail using SARIMA and decomposition models case study: a retail store in Indonesia,” *Sinergi*, vol. 29, no. 2, p. 331, Apr. 2025, doi: 10.22441/sinergi.2025.2.006.
- [9] R. Fildes, S. Ma, and S. Kolassa, “Retail forecasting: Research and practice,” *International Journal of Forecasting*, vol. 38, no. 4, pp. 1283–1318, Oct. 2022, doi: 10.1016/j.ijforecast.2019.06.004.
- [10] Y. Yang, C. Ding, S. Lee, L. Yu, and F. Ma, “A modified Teunter-Syntetos-Babai method for intermittent demand forecasting,” *Journal of Management Science and Engineering*, vol. 6, no. 1, pp. 53–63, Mar. 2021, doi: 10.1016/j.jmse.2021.02.008.
- [11] P. Rafieisangari and L. Bisaglia, “Advances in Intermittent Demand Forecasting,” in *Methodological and Applied Statistics and Demography III*, A. Pollice and P. Mariani, Eds., in Italian Statistical Society Series on Advances in Statistics., Cham: Springer Nature Switzerland, 2025, pp. 202–207. doi: 10.1007/978-3-031-64431-3_34.
- [12] K. Sanguri and K. Mukherjee, “Forecasting of intermittent demands under the risk of inventory obsolescence,” *Journal of Forecasting*, vol. 40, no. 6, pp. 1054–1069, Sep. 2021, doi: 10.1002/for.2761.
- [13] R. Sarlo, C. Fernandes, and D. Borenstein, “Lumpy and intermittent retail demand forecasts with score-driven models,” *European Journal of Operational Research*, vol. 307, no. 3, pp. 1146–1160, Jun. 2023, doi: 10.1016/j.ejor.2022.10.006.
- [14] K. Sanguri, S. Patra, K. Nikolopoulos, and S. Punia, “Intermittent demand, inventory obsolescence, and temporal aggregation forecasts,” *International Journal of Production Research*, vol. 62, no. 5, pp. 1663–1685, Mar. 2024, doi: 10.1080/00207543.2023.2199435.
- [15] J. Feizabadi, “Machine learning demand forecasting and supply chain performance,” *International Journal of Logistics Research and Applications*, vol. 25, no. 2, pp. 119–142, Feb. 2022, doi: 10.1080/13675567.2020.1803246.
- [16] I. Izzati, I. K. Sriwana, and S. Martini, “Drug forecasting and supply model design using Artificial Neural Network (ANN) and Continuous Review (r,q) to minimize total supply cost,” *SINERGI*, vol. 28, No. 2, pp. 219–230, 2024, doi: 10.22441/sinergi.2024.2.002.
- [17] A. C. Türkmen, T. Januschowski, Y. Wang, and A. T. Cemgil, “Forecasting intermittent and sparse time series: A unified probabilistic framework via deep renewal processes,” *PLoS ONE*, vol. 16, no. 11, p. e0259764, Nov. 2021, doi: 10.1371/journal.pone.0259764.
- [18] P. S. Hartono, S. Halim, and H. Sutrisno, “Navigating uncertainty: A multivariate forecasting approach to Shanghai containerized freight index trends during the COVID-19 and Red Sea shipping crises,” *Journal of Modelling in Management*, pp. 1–26, Mar. 2026, doi: 10.1108/JM2-09-2025-0480.
- [19] J. Ahmad, I. Ahmad, Q. Miao, and Z. Su, “Forecasting the human cost of disasters under Sustainable Development Goal: A time series analysis using Facebook Prophet model,” *Technology in Society*, vol. 83, p. 102992, Dec. 2025, doi: 10.1016/j.techsoc.2025.102992.
- [20] P. Manandhar, H. Rafiq, E. Rodriguez-Ubinas, and T. Palpanas, “New forecasting metrics evaluated in Prophet, Random Forest, and Long Short-Term Memory models for load forecasting,” *Energies*, vol. 17, no. 23, p. 6131, Dec. 2024, doi: 10.3390/en17236131.
- [21] M. Abolghasemi, G. Tarr, and C. Bergmeir, “Machine learning applications in hierarchical time series forecasting: Investigating the impact of promotions,” *International Journal of Forecasting*, vol. 40, no. 2, pp. 597–615, Apr. 2024, doi: 10.1016/j.ijforecast.2022.07.004.
- [22] I.-F. Chen and C.-J. Lu, “Demand forecasting for multichannel fashion retailers by integrating clustering and machine learning algorithms,” *Processes*, vol. 9, no. 9, p. 1578, Sep. 2021, doi: 10.3390/pr9091578.
- [23] 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," *Future Internet*, vol. 15, no. 8, p. 255, Jul. 2023, doi: 10.3390/fi15080255.
- [24] W. Junthopas and C. Wongoutong, "Setting the initial value for single exponential smoothing and the value of the smoothing constant for forecasting using Solver in Microsoft Excel," *Applied Sciences*, vol. 13, no. 7, p. 4328, Mar. 2023, doi: 10.3390/app13074328.
- [25] J. M. Rožanec, B. Fortuna, and D. Mladenčić, "Reframing Demand Forecasting: A Two-fold approach for lumpy and intermittent demand," *Sustainability*, vol. 14, no. 15, p. 9295, Jul. 2022, doi: 10.3390/su14159295.
- [26] S. D. Prestwich, S. A. Tarim, and R. Rossi, "Intermittency and obsolescence: A Croston method with linear decay," *International Journal of Forecasting*, vol. 37, no. 2, pp. 708–715, Apr. 2021, doi: 10.1016/j.ijforecast.2020.08.010.
- [27] L. Li, Y. Kang, F. Petropoulos, and F. Li, "Feature-based intermittent demand forecast combinations: Accuracy and inventory implications," *International Journal of Production Research*, vol. 61, no. 22, pp. 7557–7572, Nov. 2023, doi: 10.1080/00207543.2022.2153941.
- [28] R. C. Ganzevoort and J. H. Van Vuuren, "A two-phased cluster-based approach towards ranked forecast-model selection," *Machine Learning with Applications*, vol. 13, p. 100482, Sep. 2023, doi: 10.1016/j.mlwa.2023.100482.
- [29] M. Waqas and U. W. Humphries, "A critical review of RNN and LSTM variants in hydrological time series predictions," *MethodsX*, vol. 13, p. 102946, Dec. 2024, doi: 10.1016/j.mex.2024.102946.
- [30] G. Papageorgiou and C. Tjortjis, "Adaptive sliding window normalization," *Information Systems*, vol. 129, p. 102515, Mar. 2025, doi: 10.1016/j.is.2024.102515.
- [31] D. Markudova *et al.*, "Preventive maintenance for heterogeneous industrial vehicles with incomplete usage data," *Computers in Industry*, vol. 130, p. 103468, Sep. 2021, doi: 10.1016/j.compind.2021.103468.
- [32] Md. I. A. Efat *et al.*, "Deep-learning model using hybrid adaptive trend estimated series for modelling and forecasting sales," *Ann Oper Res*, vol. 339, no. 1–2, pp. 297–328, Aug. 2024, doi: 10.1007/s10479-022-04838-6.
- [33] L. Coroneo and F. Iacone, "Testing for equal predictive accuracy with strong dependence," *International Journal of Forecasting*, vol. 41, no. 3, pp. 1073–1092, Jul. 2025, doi: 10.1016/j.ijforecast.2024.11.003.
- [34] A. Grant, M. Mrazik, and S. Satchell, "Evaluating forecasts at multiple horizons: an extension of the Diebold–Mariano approach," *Journal of Forecasting*, p. for.70150, Mar. 2026, doi: 10.1002/for.70150.

Appendix 1. Examples of DM-Test for Intermittent and Continuous Demand Models

Intermittent Demand: Chainwheel and Parts					
Model	MAE	RMSE	MAPE	RMAE	p_value
Current - Current	4.666667	4.760952	166.6667		
IDF - ADIDA	3.064471	3.206224	113.2601	0.656672	0.027418
IDF - CrostonClassic	7.4199	7.479559	258.4411	1.589979	0.569547
IDF - CrostonOptimized	3.070751	3.212227	113.4695	0.658018	0.062308
IDF - IMAPA	3.064471	3.206224	113.2601	0.656672	0.028752
IDF - SESOpt	2.850829	3.002684	106.1387	0.610892	0.04294
IDF - TSB	7.363983	7.424091	256.5772	1.577996	0.728355
Prophet - Additive	2.69697	2.850257	84.09091	0.577922	0.082611
Prophet - Multiplicative	2.847359	3.313563	105.2749	0.610148	0.449434
RNN-LSTM - LSTM	0.832648	1.272724	39.09804	0.178425	0.201778
RNN-LSTM - RNN	2.333333	2.516611	66.66667	0.5	0.162829
Intermittent Demand: Footwear					
Model	MAE	RMSE	MAPE	RMAE	p_value
Current - Current	3.333333	4.690416	88.88889		
IDF - ADIDA	3.315497	4.713481	89.48345	0.994649	0.023338
IDF - CrostonClassic	3.360388	4.656386	87.98707	1.008116	0.370141
IDF - CrostonOptimized	3.34035	4.681479	88.655	1.002105	0.284674
IDF - IMAPA	3.291648	4.745089	90.27841	0.987494	0.026443
IDF - SESOpt	3.266686	4.779094	91.11045	0.980006	0.055442
IDF - TSB	3.356573	4.661114	88.11423	1.006972	0.054926
Prophet - Additive	3.30303	4.729894	89.89899	0.990909	0.688013
Prophet - Multiplicative	3	4.700683	89.89899	0.9	0.918636
RNN-LSTM - LSTM	3.333333	4.690416	88.88889	1	0.029102
RNN-LSTM - RNN	3.333333	4.690416	88.88889	1	0.93645
Continuous Demand: BB Set and Part					
Model	MAE	RMSE	MAPE	RMAE	p_value
Current - Current	5.333333	5.944185	99.81481		
IDF - ADIDA	4.757205	5.433221	55.2805	0.891976	0.527602
IDF - CrostonClassic	3.939046	4.733389	42.70509	0.738571	0.85351
IDF - CrostonOptimized	3.860641	4.668344	41.49998	0.72387	0.820896
IDF - IMAPA	4.757205	5.433221	55.2805	0.891976	0.527602
IDF - SESOpt	6.781254	7.271471	86.39088	1.271485	0.391428
IDF - TSB	4.268029	5.010485	47.76169	0.800256	0.762362
Prophet - Additive	6.333333	6.557439	85.92593	1.1875	0.010411
Prophet - Multiplicative	7.666667	8.103497	100	1.4375	0.232
RNN-LSTM - LSTM	6.666667	7.164728	84.62963	1.25	0.412416
RNN-LSTM - RNN	6.666667	7.164728	84.62963	1.25	0.217799