**JURNAL LOCUS: Penelitian & Pengabdian**
Volume 5 No. 11 November 2026, pp 9652-9671
E-ISSN 2829-7334
DOI: 10.58344/locus.v5i11.6328

# Intermittent Demand Forecasting and Inventory Service Level Evaluation Using The TSB Method and Machine Learning: A Case Study on PT X in The Document Solutions Industry

**Fitrotun Hasanah, Melia Eka Lestiani* , Maniah**
Universitas Logistik dan Bisnis Internasional, Indonesia
Email: mlestiani@ulbi.ac.id* , maniah@ulbi.ac.id

**Keywords:**
Intermittent Demand; Demand Forecasting; Random Forest; TSB; Service Level.

This is an open access article under the CC BY-SA license.

Copyright © 2026 by Author. Published by Riviera Publishing

**Abstract**

Intermittent demand presents significant challenges in demand forecasting and inventory management, particularly in the document solutions industry, where demand is driven by both rental and sales activities. Forecasting accuracy alone does not necessarily translate into improved inventory performance. This study evaluates the forecasting performance of the Teunter–Syntetos–Babai (TSB) method and machine learning models, including Random Forest, XGBoost, and LightGBM, across total, rental, and sales demand categories and examines their implications for inventory service levels. A quantitative analysis was conducted using supply demand data for 134 stock-keeping units (SKUs) of Brand Z at PT X from August 2023 to November 2025. Demand patterns were classified using the Average Demand Interval (ADI) and the squared coefficient of variation ($CV^2$). Forecasting models incorporated historical demand, lead times, new printer installations, and new printer sales. Forecasting performance was evaluated using root mean squared error (RMSE) and mean absolute error (MAE), while inventory performance was assessed using an order-up-to-level (S) policy based on the probability of no stockout (S1), fill rate (S2), and ready rate (S3). Random Forest demonstrated superior overall performance in terms of forecasting accuracy and inventory service levels, particularly for intermittent and lumpy demand patterns. However, greater forecasting accuracy did not consistently translate into better inventory performance. The TSB method performed better for smooth demand patterns and on several service-level indicators for erratic demand. Demand decomposition also revealed distinct characteristics between rental and sales demand. These findings indicate that forecasting accuracy alone is insufficient for selecting an appropriate forecasting method. Demand characteristics and inventory service-level performance must also be considered to support effective inventory management.

## INTRODUCTION

Industries that rely heavily on global imports and supply chain activities are particularly vulnerable to supply chain uncertainty. Production yield uncertainty arises from defective products, which reduce the quantity of goods available for supply. Capacity uncertainty also creates discrepancies in supply quantities due to suppliers' limited

production capacities. Furthermore, lead time uncertainty results in variations in the time required to receive ordered goods due to stock shortages, changes in production schedules, and disruptions in deliveries from suppliers (Snyder & Shen, 2019).

Uncertainty can also arise from demand variability resulting from increased product variety, new product launches, and short product life cycles, including product discontinuation. Demand patterns may differ considerably across products. Some products exhibit stable and predictable demand, while others experience infrequent and irregular demand characterized by varying intervals between demand occurrences and fluctuations in demand quantities (Sheffi, 2004).

The document solutions industry is particularly vulnerable to supply chain uncertainty. Company X procures goods from its principal supplier's warehouses located in several countries. The lead times specified by the principal supplier vary across products. For example, toner A has a lead time of one month, whereas toner B has a lead time of two months. However, actual lead times are also influenced by the availability of raw materials and finished goods. Deliveries may be completed earlier when the principal supplier maintains sufficient inventory. Conversely, disruptions in raw material supply or shortages of finished goods may extend actual lead times beyond the initially specified periods.

**Table 1.** Demand Distribution

| Demand Categories | Demand Size | Percentage (%) |
|---|---|---|
| Rental | 0 | 68.36 |
| Rental | > 0 | 31.64 |
| **TOTAL** | | **100.00** |
| Sales | 0 | 71.16 |
| Sales | > 0 | 28.84 |
| **TOTAL** | | **100.00** |
| Aggregate (Total) | 0 | 58.66 |
| Aggregate (Total) | > 0 | 41.34 |
| **TOTAL** | | **100.00** |

An analysis of demand distribution data for 134 Brand Z supply stock-keeping units (SKUs) at Company X from August 2023 to November 2025 revealed that no demand was recorded during 58.66% of the observation periods for aggregate demand. However, decomposing demand by source revealed that zero-demand periods accounted for 68.36% of the observation periods in the rental category and an even higher proportion of 71.16% in the sales category. These differences indicate that aggregate demand analysis can obscure the distinct characteristics of each demand category by reducing the apparent degree of demand intermittency, as rental and sales demand may occur in different periods and offset each other's zero-demand intervals when aggregated.

Demand classification is essential for identifying demand characteristics and selecting appropriate forecasting methods. The classification framework proposed by Syntetos et al., (2005), which uses the Average Demand Interval (ADI) and the squared coefficient of variation of nonzero demand sizes ($CV^2$), categorizes demand patterns into four types: smooth, intermittent, erratic, and lumpy (Syntetos et al., 2005). Disaggregating demand by category or operational source is also important because different demand sources may exhibit distinct patterns and variability, necessitating separate analyses to

improve forecasting accuracy and support inventory management decisions (Altug & Ceryan, 2022).

**Table 2.** Demand Type Classification (Syntetos et al., 2005)

| Demand Type | Demand Intermittency (ADI) | Demand Variability ($CV^2$) | Key Characteristics |
|---|---|---|---|
| Smooth | ADI ≤ 1.32 | $CV^2 \leq 0.49$ | Demand occurs in almost every period with low variability |
| Intermittent | ADI > 1.32 | $CV^2 \leq 0.49$ | Demand occurs infrequently but demand sizes are relatively stable when they occur |
| Erratic | ADI ≤ 1.32 | $CV^2 > 0.49$ | Demand occurs frequently but demand sizes are highly variable |
| Lumpy | ADI > 1.32 | $CV^2 > 0.49$ | Demand occurs infrequently and demand sizes are highly variable |

Intermittent demand is characterized by irregular demand occurrences, with zero demand in some periods and varying demand quantities in others (Teunter et al., 2011). These characteristics make demand forecasting particularly challenging, especially when prolonged periods of zero demand occur. Forecasting becomes even more complex for products at risk of obsolescence, as the probability of demand may decline over time. Under these conditions, traditional forecasting methods may fail to adequately capture changes in demand patterns, resulting in reduced forecasting accuracy and challenges in inventory management (Sanguri et al., 2024).

Teunter et al. (2011) developed the Teunter–Syntetos–Babai (TSB) method to address the challenges associated with intermittent demand forecasting. This method separately estimates the probability of demand occurrence and the expected nonzero demand size. The demand occurrence probability is updated in every period, including periods with zero demand, while the estimated demand size is updated only when positive demand occurs. This approach enables the forecasting model to remain responsive to changes in demand occurrence, particularly during prolonged periods of zero demand (Teunter et al., 2011).

Previous studies have provided substantial evidence that machine learning (ML) methods can achieve greater forecasting accuracy than traditional statistical approaches under certain conditions. For example, a comparison between autoregressive integrated moving average (ARIMA) and Extreme Gradient Boosting (XGBoost) demonstrated that, although ARIMA can be effective in modeling temporal patterns and seasonal variations when appropriately specified, XGBoost may perform better when forecasting complex and nonlinear relationships (Bai, 2024). Another study comparing XGBoost and Light Gradient Boosting Machine (LightGBM) revealed that forecasting performance varied depending on the evaluation metrics employed. LightGBM achieved lower root mean squared logarithmic error (RMSLE), whereas XGBoost demonstrated better performance in terms of mean absolute error (MAE) and the coefficient of determination ($R^2$) (Nguyen et al., 2025). Similarly, a comparison between ARIMA and Random Forest indicated that Random Forest achieved greater forecasting accuracy by capturing both linear and nonlinear relationships among variables (Kumar et al., 2025). Random Forest may be particularly effective when demand forecasting incorporates explanatory variables,

whereas ARIMA is generally more suitable for modeling temporal dependencies in univariate time series data (Amnay et al., 2025). Furthermore, traditional intermittent demand forecasting methods, including Croston's method, the Syntetos–Boylan Approximation (SBA), and TSB, have been reported to exhibit lower forecasting accuracy than Random Forest and gradient-boosted trees in certain daily stock-keeping unit (SKU)-level forecasting applications (Spiliotis et al., 2022).

Recent advances in computing technology have substantially expanded the application of ML techniques across various predictive tasks, including demand forecasting. Compared with conventional statistical approaches, ML methods can capture complex patterns in historical data and identify nonlinear relationships and interactions among variables that may be difficult to model using traditional techniques (Hastie et al., 2009; James et al., 2013). These capabilities have created new opportunities in supply chain analytics, enabling organizations to improve forecasting processes and support more informed, data-driven decision-making. By integrating historical demand, operational information, and logistical variables, ML models can identify complex relationships that contribute to improved forecasting performance and inventory management (Maniah, 2026). Furthermore, SHapley Additive exPlanations (SHAP) provide a model-agnostic interpretability framework based on Shapley values, enabling the assessment of individual feature contributions to model predictions. This approach enhances the transparency and interpretability of ML models and supports the evaluation of their predictive behavior (Lundberg & Lee, 2017).

However, forecasting accuracy should not be the sole criterion for determining business policies and selecting forecasting methods. Forecasting performance should also be evaluated in terms of operational outcomes, particularly inventory service levels and the ability to meet customer demand (Theodorou et al., 2025). Accordingly, forecasting methods should be assessed based on their practical utility in supporting business decision-making rather than solely on statistical accuracy measures. A forecasting method with slightly lower predictive accuracy may still be preferable if it produces more stable inventory decisions, reduces operational costs, or improves overall inventory performance (Petropoulos et al., 2019).

The order-up-to-level (S) policy is an inventory control policy in which the replenishment quantity is determined by the difference between a predetermined target inventory position (S) and the current inventory position at the time of review. Under this policy, the inventory position, defined as on-hand inventory plus outstanding orders minus backorders, is restored to the target level at each scheduled review. Consequently, the replenishment quantity varies according to the inventory position when the ordering decision is made. Inventory service-level performance can be evaluated using three indicators: the probability of no stockout (S1), the fill rate (S2), and the ready rate (S3) (Axsäter, 2015).

**Table 3.** Inventory Service Level (Axsäter, 2015)

| Service Level | Definition |
|---|---|
| $S_1$ (Probability of No Stockout) | Quantifies the likelihood that a stockout will not transpire during a single ordering cycle, which is assessed by the occurrence of any stockout within that period. |
| $S_2$ (Fill Rate) | The fraction of total demand satisfied directly from existing inventory, measured by the volume of customer demand successfully met. |
| $S_3$ (Ready Rate) | The proportion of time inventory levels remain positive, determined by the frequency of inventory being above zero at the observation period's conclusion. |

This study created an analytical framework for demand forecasting by classifying demand on the decomposition of demand sources using TSB and ML methods, and evaluating the results with service level indicators using data of Company X from the document solutions industry. Therefore, this study did not only compare accuracy, but also to find out whether the results of forecasting with the best accuracy also resulted in a better service level.

## METHOD

This study used secondary data obtained from Company X's historical records and Enterprise Resource Planning (ERP) system, including rental and sales demand for supplies, Brand Z printer sales and installation records, and supplier and delivery lead times. The dataset covered the period from August 2023 to November 2025. Data processing was performed using Python and included preprocessing, demand classification based on the Average Demand Interval (ADI) and squared coefficient of variation ($CV^2$), and demand forecasting using the Teunter–Syntetos–Babai (TSB) method and machine learning (ML) models, including Random Forest, XGBoost, and LightGBM. The ML modeling process involved feature engineering, hyperparameter optimization, and performance evaluation. The TSB method and the best-performing ML model were subsequently used to evaluate inventory performance under an order-up-to-level (S) policy using inventory service-level indicators.

### Demand Classification

Demand classification using ADI and $CV^2$ were calculated by these formulas:

$$ADI = \frac{N}{\sum_{t=1}^{N} I(y_t > 0)}$$

$N$ = total number of periods
$t$ = time period index
$y_t$ = actual demand in period $t$
$I(y_t > 0)$ = demand occurrence indicator in period $t$, 1 when demand occurs and 0 otherwise

$$CV^2 = \left(\frac{\sigma}{\mu}\right)^2$$

$\sigma$ = standard deviation of demand
$\mu$ = mean of demand

**TSB Method Forecasting**

The TSB method was calculated with this mathematical formula:

$$\hat{p}_t = \begin{cases} \hat{p}_{t-1} + \beta(0 - \hat{p}_{t-1}), \text{if } p_t = 0 \\ \hat{p}_{t-1} + \beta(1 - \hat{p}_{t-1}), \text{if } p_t > 0 \end{cases}$$
$$\hat{z}_t = z_{t-1}, \text{if } p_t = 0$$
$$\hat{z}_t = \hat{z}_{t-1} + \alpha(z_t - \hat{z}_{t-1}), \text{if } p_t > 0$$
$$\hat{y}_t^{TSB} = \hat{p}_t \times \hat{z}_t$$

$\hat{y}_t$ = forecasted demand in period $t$
$\hat{z}_t$ = estimated demand size in period $t$
$\hat{p}_t$ = estimated non-zero demand interval in period $t$
$z_t$ = actual demand size in period $t$
$p_t$ = actual interval between non-zero demands
$\hat{p}_{t-1}$ = previous period's estimate of the demand interval
$\hat{z}_{t-1}$ = previous period's estimate of demand size
$q_t$ = number of periods since the last non-zero demand
$\alpha$ = smoothing parameter for demand size
$\beta$ = smoothing parameter for demand probability occurrence

The smoothing parameters in the TSB method were determined using a search grid in the search space α = 0.05–0.30 and β = 0.01–0.30 with an increment of 0.01, following the range of values used in the previous TSB method research (Teunter et al., 2011). The combination with the lowest error value is selected as the optimal parameter. For Stock Keeping Units (SKUs) with a demand frequency of no more than once during the observation period, the default parameter (α = 0.05; β = 0.01) was used to make the model less responsive to very rare demand events, so that the forecast results remained stable.

**Machine Learning Forecasting**

**Feature engineering**

The predictive variables formed consisted of rolling demand, lag features of sales and printer installation, rolling lead time principal and delivery, calendar variables, and SKU identities that were converted into dummy variables through one-hot encoding. Missing values are marked with a value of −1 after the rolling feature formation process, to avoid confusion with a zero value that can be interpreted as the availability of goods on the same day as the booking period.

**Table 4.** Feature Engineering Summary

| Feature Categories | Variable |
|---|---|
| Rolling demand | avg_demand_3m, std_demand_3m, avg_rent_3m, std_rent_3m, avg_sales_3m, std_sales_3m |
| Sales lag | sales_lag_1, sales_lag_2, sales_lag_3, sales_lag_6 |
| Install lag | install_lag_1, install_lag_2, install_lag_3, install_lag_6 |
| Lead time | avg_LT_principal_6m, std_LT_principal_6m, avg_LT_delivery_6m, std_LT_delivery_6m |
| Calendar | month_number, year |
| Product identity | one-hot encoding SKU |

**Hyperparameter optimization**

Hyperparameter optimization was carried out using a random search combined with a three-fold Time Series Cross Validation (TSCV). TSCV was chosen because it is specifically designed for time-series data, where the temporal order must be preserved since each observation depends on preceding ones (Owen, 2022). The parameter range determined based on the literature, characteristics of each algorithm, and test results on the research data. Each combination of hyperparameters was evaluated using the RMSE and MAE on all three folds, and the model with lowest error was selected as the best configuration. The model's performance was compared to determine the best model, which was further interpreted using SHAP.

**Table 5.** Split Fold Time Series Cross Validation

| Fold | Training period | Testing Period |
|---|---|---|
| 1 | August 2023 – February 2024 | March 2024 – September 2024 |
| 2 | August 2023 – September 2024 | October 2024 – April 2025 |
| 3 | August 2023 – April 2025 | May 2025 – November 2025 |

**Table 6.** Machine Learning Model Hyperparameter Search Space

| Models | Hyperparameter | Rental | Sales | Total |
|---|---|---|---|---|
| Random Forest | n_estimators | 100–1000 | 100–800 | 100–1000 |
| | max_depth | 3–30 | 3–30 | 3–30 |
| | min_samples_split | 2–20 | 2–20 | 2–20 |
| | min_samples_leaf | 1–10 | 1–10 | 1–10 |
| | max_features | sqrt, log2, None, 0.5, 0.7, 1.0 | sqrt, log2, None, 0.5, 0.7, 1.0 | sqrt, log2, None, 0.5, 0.7, 1.0 |
| XGBoost | n_estimators | 100–500 | 100–800 | 100–1000 |
| | max_depth | 3–20 | 3–30 | 3–30 |
| | learning_rate | 0,01–0,30 | 0,01–0,30 | 0,01–0,30 |
| | subsample | 0,6–1,0 | 0,6–1,0 | 0,6–1,0 |
| | colsample_bytree | 0,6–1,0 | 0,6–1,0 | 0,6–1,0 |
| LightGBM | n_estimators | 100–800 | 100–800 | 100–1000 |
| | max_depth | 3–20 | 3–30 | 3–20 |
| | learning_rate | 0,01–0,20 | 0,01–0,20 | 0,01–0,20 |
| | num_leaves | 31–150 | 31–150 | 31–128 |
| | min_child_samples | 10–50 | 10–50 | 10–50 |
| | feature_fraction | 0,7–1,0 | 0,7–1,0 | 0,7–1,0 |

**Forecasting Accuracy Evaluation**

The forecasting result of each method was evaluated using the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics, with below calculation:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|y_t - \hat{y}_t|$$

$t$ = time period index
$T$ = total number of periods
$y_t$ = actual demand
$\hat{y}_t$ = forecasted demand

**Inventory Performance Evaluation**

The simulation used forecasting results from the TSB method and the best ML model as the basis for replenishment decisions. All SKUs are assumed to have an initial inventory of zero units to equalize the initial conditions between SKUs, so that performance comparisons between forecasting methods could be done objectively. To reduce the impact of initial inventory initialization, a warm-up period until the first replenishment was applied to each SKU (Goldsman, 1992). The period before the first replenishment received was not included in the calculation of inventory performance indicators (Waters, 2003). The order up-to S was calculated by this formula.

$$Q = S - IP$$
$$IP = OH + OO - BO$$
$$S = Forecast_{(T+LT)} + SS$$
$$SS = Z \times \sigma \times \sqrt{(T + LT)}$$

$Q$ = Order Quantity
$S$ = Target inventory level (order-up-to level)
$IP$ = Inventory Position
$OH$ = On-Hand Inventory
$OO$ = *Outstanding Order*
$BO$ = Unfulfilled Customer Order (backorder)
$SS$ = Safety Stock
$Z$ = Z-*score* based on service level target
$\sigma$ = standard deviation of demand
$T$ = periodic review
$Forecast_{(T+LT)}$ = Forecasted Demand T+LT
$LT$ = Lead Time

Inventory performance was measured using service level indicators $S_1$, $S_2$, and $S_3$ that were compared between the TSB method and the best ML model in each category and demand classification. This approach was used to evaluate differences in inventory performance on various demand characteristics, as well as to identify whether the method with the best forecasting performance also results in better inventory performance. Service level was calculated by this fromula (Axsäter, 2015):

$$S_1 = \frac{\sum_{t=1}^{T} I(Stockout_t = 0)}{T} \times 100\%$$

$$S_2 = \left(\frac{\sum_{t=1}^{T} Demand_t - \sum_{t=1}^{T} Stockout_t}{\sum_{t=1}^{T} Demand_t}\right) \times 100\%$$

$$S_3 = \frac{\sum_{t=1}^{T} I(Ending\ Stock_t > 0)}{T} \times 100\%$$

$T$ = total observation period
$I$ = flag indicators
$t$ = observation period index
$Demand_t$ = total actual demand in period $t$
$Stockout_t$ = total unfulfilled demand in period $t$
$Ending\ Stock_t$ = ending inventory in period $t$

## RESULTS AND DISCUSSION

### Research Results

#### a. Demand Classification

Table IV.5 showed the results of the classification of demand based on ADI and $CV^2$ values.

**Table 7.** Demand Classification Result

| Demand Type | Total | | Rental | | Sales | |
|---|---|---|---|---|---|---|
| | SKU | % | SKU | % | SKU | % |
| Smooth | 18 | 13.43 | 26 | 19.41 | 9 | 6.71 |
| Intermittent | 68 | 50.75 | 93 | 69.40 | 76 | 56.72 |
| Erratic | 21 | 15.67 | 0 | 0 | 6 | 4.48 |
| Lumpy | 27 | 20.15 | 15 | 11.19 | 43 | 32.09 |
| **Total Amount** | **134** | **100.00** | **134** | **100.00** | **134** | **100.00** |

Company X's demand characteristics were predominantly intermittent type which was shown with more than 50% across all demand categories classified as this type. These results showed that many SKUs have irregular demand patterns, but fairly stable demand variability.

The difference in demand classification results in each demand category showed that rental categories tended to have a rare but more stable demand pattern, which was shown by the intermittent type dominating, but followed by smooth and no erratic type. The sales category dominated by intermittent and lumpy proved that demand was not only rare, but also had high variability. While the total category had the lowest percentage of intermittent types due to data aggregation that caused data to overlap to cover the demand gap in the rental and sales categories.

#### b. TSB Forecasting

#### 1. Grid Search Results

**Table 8.** Grid Search Result

| Demand Categories | α Dominant | Number of SKUs | β Dominant | Number of SKUs | Default SKUs |
|---|---|---|---|---|---|
| Rental | 0.05 | 97 | 0.01 | 82 | 50 |
| Sales | 0.05 | 78 | 0.30 | 57 | 18 |
| Total | 0.05 | 62 | 0.01 | 59 | 8 |

The table showed that the value of α = 0.05 was the most frequently selected parameter in all demand categories. This small α value made the estimate of demand size less sensitive to demand fluctuations. In contrast, the β value that governed the update of the probability of the demand occurring resulted in a different dominant value in each demand category. In the total and rental categories, the value of β = 0.01 was more dominant because demand was more stable than sales with a value of β = 0.30 which required greater smoothing to capture the pattern of demand variability due to the lumpy type that was quite dominant. There were a number of SKUs that use default parameters because the frequency of demands was very limited during the observation period.

**2. Evaluation of TSB Method Forecasting**

**Table 9.** TSB Method Evaluation Result

| Demand Type | Total | | Rental | | Sales | |
|---|---|---|---|---|---|---|
| | **RMSE** | **MAE** | **RMSE** | **MAE** | **RMSE** | **MAE** |
| Smooth | 17.05 | 13.24 | **6.38** | **4.91** | 9.82 | 8.07 |
| Intermittent | 1.77 | 1.36 | **0.96** | **0.78** | 1.88 | 1.48 |
| Erratic | **11.75** | **8.53** | - | - | 23.29 | 18.73 |
| Lumpy | 6.92 | 4.56 | **5.50** | **4.33** | 8.41 | 5.89 |

A comparison of results by demand category showed that rental demands resulted in lower error values than total and sales demands. These results indicated that the data structure in rental activities was more stationary so that it could provide more optimal predictions on the TSB method. Meanwhile, sales demands showed higher error values in all types and categories of demands, especially in erratic and lumpy types. These results indicated that sales demand had a greater level of uncertainty than rental demand.

In total demand, the smooth type produced the highest error value compared to other demand types. This condition indicated that TSB method which was made specifically for intermittent demand showed low-performing capability to capture demand patterns that occur regularly. Evaluation of the overall results showed that the TSB method tended to provide decreased prediction accuracy in demand types with higher variability in demand size and timing such as smooth and erratic.

**c. Machine Learning Forecasting**

**1. Results of Total Demand Evaluation of Machine Learning Model**

**Table 10.** Total Demand Machine Learning Evaluation Result

| Demand Type | LightGBM | | Random Forest | | XGBoost | |
|---|---|---|---|---|---|---|
| | **RMSE** | **MAE** | **RMSE** | **MAE** | **RMSE** | **MAE** |
| Smooth | 20.15 | 15.79 | **18.53** | **14.29** | 20.47 | 15.97 |
| Intermittent | 1.59 | 1.01 | **1.50** | **0.95** | 1.52 | 0.96 |
| Erratic | 10.82 | 8.13 | **10.70** | **8.48** | 10.99 | 8.68 |
| Lumpy | 4.21 | **2.23** | 4.22 | 2.45 | **4.09** | 2.25 |

The results of the ML model evaluation showed that the forecasting of the total demand category with the best level of accuracy is in the Random Forest model compared to other ML models. Of the four types of demands, only the lumpy type had a lower accuracy level but with a small difference, namely 0.13 compared to RMSE XGBoost and 0.22 compared to MAE LightGBM.

**2. Evaluation Results of Machine Learning Model Rental Demands**

**Table 11.** Rental Demand Machine Learning Evaluation Result

| Demand Type | LightGBM | | Random Forest | | XGBoost | |
|---|---|---|---|---|---|---|
| | **RMSE** | **MAE** | **RMSE** | **MAE** | **RMSE** | **MAE** |
| Smooth | **6.96** | **5.49** | 7.21 | 5.60 | 7.95 | 6.30 |
| Intermittent | 0.72 | 0.40 | **0.67** | **0.37** | 0.69 | **0.37** |
| Erratic | - | - | - | - | - | - |
| Lumpy | **2.77** | 1.46 | 2.93 | 1.54 | 2.80 | **1.40** |

The intermittent type obtained the lowest RMSE value of 0.67 in the Random Forest model, while MAE obtained Random Forest and XGBoost with the same error value of 0.37. The best accuracy results for the smooth type were in the LightGBM model with an RMSE value of 6.96 and MAE 5.49, while the best lumpy RMSE type was in the LightGBM and MAE in XGBoost with a small error difference. All of these results showed that all three ML models have equivalent forecasting capabilities on rental category demand data.

**3. Results of Sales Demand Evaluation Machine Learning Model**

**Table 121.** Sales Demand Machine Learning Evaluation Result

| Demand Type | LightGBM | | Random Forest | | XGBoost | |
|---|---|---|---|---|---|---|
| | **RMSE** | **MAE** | **RMSE** | **MAE** | **RMSE** | **MAE** |
| Smooth | 10.81 | 8.84 | **10.18** | **8.39** | 10.38 | 8.49 |
| Intermittent | 1.55 | 0.98 | **1.41** | **0.86** | 1.46 | 0.94 |
| Erratic | 27.12 | 21.21 | **24.86** | **20.23** | 27.82 | 22.86 |
| Lumpy | **6.92** | **4.85** | 7.19 | 5.51 | 7.20 | 5.33 |

Random Forest produced the lowest RMSE and MAE values on all three types of demand, except for lumpy where the lowest error value is in the LightGBM model. In the sales demand category, the XGBoost model did not generate the lowest error value across all demand types. The Random Forest model was proven to provide better forecasting accuracy than other ML models.

**4. Machine Learning Model Selection**

Evaluation of all ML models by category and demand type showed that Random Forest was the most stable ML model compared to LightGBM and XGBoost. Although in the rental category all ML models provided equal performance, Random Forest was chosen as the best ML model because it produced the highest accuracy in all three demand categories compared to other ML models.

**d. SHAP Analysis *of the Random Forest Model***

**1. Total Demand**

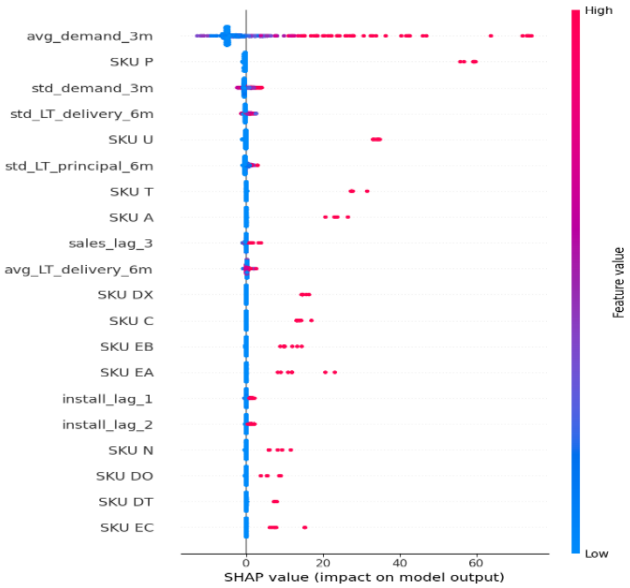



**Figure 1.** Total Demand SHAP Analysis

The feature with the most contribution was the historical demand feature, which was the average demand for the last 3 months (avg_demand_3m) and the standard demand deviation for the last 3 months (std_demand_3m). Other features such as printer sales in the last 3 months (sales_lag_3) and printer installations in the last 1 and 2 months (install_lag_1 and install_lag_2) also contributed to this model. Some features such as lead time and SKUs such as SKU P, U, and T also appeared as features that impacted the forecast results.

This analysis explained that the model relied heavily on short-term demand features (3m) and the characteristics of specific SKUs. The long-term lead time (6m) and the increase in printer sales up to the last 3 months also affect the model's forecasting results. The addition of new printer rentals has only a small impact on the model compared to printer sales.

**2. Rental Demand**

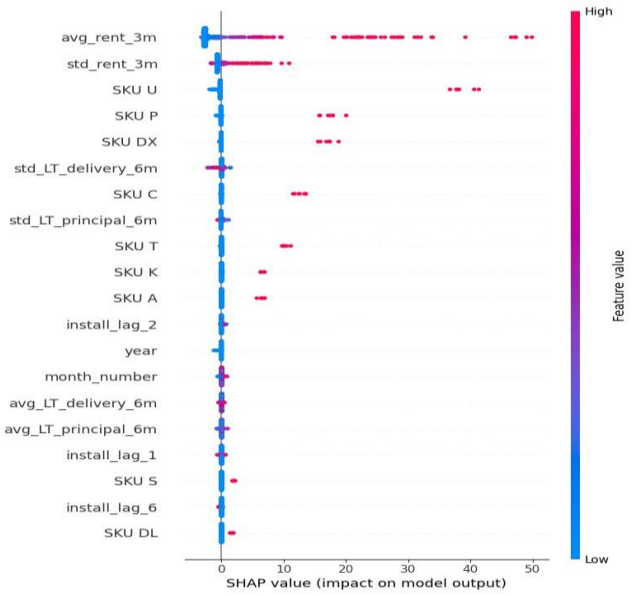



**Figure 2.** Rental Demand SHAP Analysis

Visualization of SHAP analysis results for rental demand in the Random Forest model showed that the historical rental demand feature was the feature that had the most impact on forecasting results through the average rental demand over the last 3 months (avg_rent_3m) and the standard rental demand deviation over the last 3 months (std_rent_3m).

The top three SKUs in this rental model were slightly different from the total model, namely SKU U, P, and DX, which added evidence that data aggregation could disguise the true characteristics of each SKU. This analysis also provided information that the addition of printer rentals over the last 2 months contributed more to the forecast results than other new install horizons. In accordance with the characteristics of rental demand, where the need for supplies will arise after the installation of a new printer. Based on this SHAP analysis, it was known that the addition of new installations would have an impact on rental demand on average 2 months after the installation was carried out.

### 3. Sales Demand

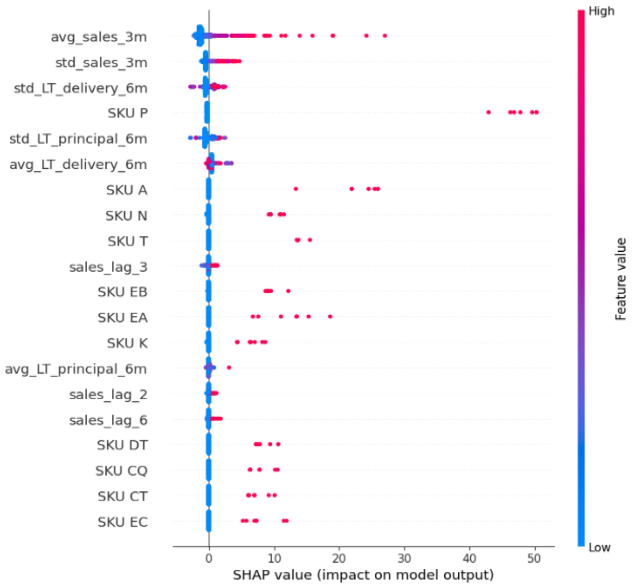



**Figure 3.** Sales Demand SHAP Analysis

The interpretation of the Random Forest model for the sales demand category presented information on the features that contribute a lot to the model's results, namely the historical sales demand feature followed by the lead time feature, several SKUs such as SKUS P, A, and N, as well as new sales features. Historical data on short-term demand (3m) contributed more than long-term (6m). The lead time feature (6m) also had more influence on forecasting results such as total demand and rental categories. The new printer sales feature for the last 3 months (sales_lag_3) was considered to have an effect on the increase in demand for supplies in the 3 months after the sale of printers occurred.

In contrast to rental demand where new install features in the last 1 and 6 months were in the lower rankings and tend to have negative values, in sales demand, all new sales features made a positive contribution even though they were not in the top position. The proportion of the formation of sales demand patterns was highly dependent on the sales of new printers. Meanwhile, in rentals, demand could arise from printers that had been rented more than the research horizon.

### e. Comparison of TSB and Random Forest Forecasting Performance

**Table 132.** Forecasting Accuracy of TSB and Random Forest in Total Demand

| Demand Type | TSB | | Random Forest | |
|---|---|---|---|---|
| | **RMSE** | **MAE** | **RMSE** | **MAE** |
| Smooth | **17.05** | **13.24** | 18.53 | 14.29 |
| Intermittent | 1.77 | 1.36 | **1.50** | **0.95** |
| Erratic | 11.75 | 8.53 | **10.70** | **8.48** |
| Lumpy | 6.92 | 4.56 | **4.22** | **2.45** |

**Table 143.** Forecasting Accuracy of TSB and Random Forest in Rental Demand

| Demand Type | TSB | | Random Forest | |
|---|---|---|---|---|
| | RMSE | MAE | RMSE | MAE |
| Smooth | **6.38** | **4.91** | 7.21 | 5.60 |
| Intermittent | 0.96 | 0.78 | **0.67** | **0.37** |
| Erratic | - | - | - | - |
| Lumpy | 5.50 | 4.33 | **2.93** | **1.54** |

**Table 154.** Forecasting Accuracy of TSB and Random Forest in Sales Demand

| Demand Type | TSB | | Random Forest | |
|---|---|---|---|---|
| | RMSE | MAE | RMSE | MAE |
| Smooth | **9.82** | **8.07** | 10.18 | 8.39 |
| Intermittent | 1.88 | 1.48 | **1.41** | **0.86** |
| Erratic | **23.29** | **18.73** | 24.86 | 20.23 |
| Lumpy | 8.41 | 5.89 | **7.19** | **5.51** |

Random Forest had proven to be more resilient on some types of demands across all demand categories, especially in intermittent types. TSB method which was specifically made to deal with intermittent patterns, could not outperform the accuracy of Random Forest. TSB showed its best performance in the smooth type which had a stable frequency demand pattern. Random Forest with its additional features could not capture smooth patterns any better than TSB which relied only on historical data on demands. This table also corroborated the evidence related to the potential disguise of demand characteristics as evidenced by erratic types. This type was outperformed by TSB for accuracy in the separation of the sales category demand sources, but the total was won by Random Forest due to data aggregation.

**f. Inventory Performance Evaluation**

**1. Total Demand Service Level**

**Table 16.** Total Demand Service Level

| Demand Type | TSB | | | Random Forest | | |
|---|---|---|---|---|---|---|
| | S1 | S2 | S3 | S1 | S2 | S3 |
| Smooth | 92.41 | **97.87** | 91.11 | **94.79** | 97.86 | **94.14** |
| Intermittent | 97.88 | 94.43 | 85.19 | **99.05** | **95.61** | **91.89** |
| Erratic | **96.90** | 96.13 | **95.63** | 96.36 | **97.54** | 94.90 |
| Lumpy | 95.06 | 81.63 | 89.27 | **98.16** | **95.96** | **92.37** |

Random Forest was proven as a forecasting technique that provided the best performance on service level indicators of total demand. TSB still had better performance in several types and indicators, but it was mostly won by Random Forest who was only less thin in $S_2$ smooth and erratic in $S_1$ and $S_3$. These erratic results were in contrast to Table XIII which pointed to the best Random Forest for accuracy. Inventory conditions through order up-to-S simulation provided information that existing inventory was indeed able to meet most customer demands. However, the stockout frequency was slightly more frequent than TSB. On the other hand, the smooth type, which was won by TSB in absolute accuracy, turned out to be only slightly better in the total category in meeting customer demand, but often stockout. Some customers had to wait a little longer to get their orders.

**2. Rental Demand Service Level**

**Table 18.** Rental Demand Service Level

| Demand Type | TSB | | | Random Forest | | |
|---|---|---|---|---|---|---|
| | **S1** | **S2** | **S3** | **S1** | **S2** | **S3** |
| Smooth | 94.20 | 98.39 | 91.82 | **96.43** | **99.37** | **94.05** |
| Intermittent | 98.98 | 92.21 | 48.30 | **99.51** | **98.31** | **48.47** |
| Erratic | - | - | - | - | - | - |
| Lumpy | 95.88 | 89.30 | 88.92 | **99.48** | **98.93** | **92.53** |

All types of rental demands won by Random Forest. It turned out that TSB, which was a landslide winner in Table XIV, was not accompanied by its service level. The value of $S_3$ in the intermittent type was only below 50% for both forecasting methods according to the nature of the intermittent type. Although the percentage of replenishment and the proportion of customer needs met was high, stockouts were very frequent. This was a result of the infrequent emergence of demand while the variability was stable. So that the forecasting method made predictions not to maintain inventory because of the potential for its infrequent appearance.

**3. Service Level Sales Demand**

**Table 19.** Sales Demand Service Level

| Demand Type | TSB | | | Random Forest | | |
|---|---|---|---|---|---|---|
| | **S1** | **S2** | **S3** | **S1** | **S2** | **S3** |
| Smooth | 93.45 | 93.66 | 92.14 | **97.38** | **97.06** | **95.63** |
| Intermittent | 98.40 | **92.88** | 70.83 | **98.55** | 92.68 | **87.07** |
| Erratic | **96.79** | **98.20** | **95.51** | 94.23 | 95.78 | 92.95 |
| Lumpy | 96.54 | 92.84 | 92.19 | **97.87** | **95.41** | **95.21** |

The service level of sales demand made Random Forest's superiority clear. Most of the demand types were dominated by Random Forest. For the intermittent type, Random Forest was slightly inferior to TSB in fulfilling customer orders, but stockout conditions were more controlled than TSB. Just like rentals, the smooth type and service level were different from Table XIV, where smooth was won by TSB. Even so, this result was consistent with the erratic type, whose service level results were in the direction of forecasting accuracy.

**Discussion**

The overall results of the forecasting accuracy evaluation provided information related to the excellence of Random Forest over the TSB method. Especially in the intermittent and lumpy types that excelled in all demand categories. The TSB method was very resilient in the smooth type which swept away accuracy performance across all demand categories. Smooth type demand occurred in almost every period with low variation, making the update of the probability of demand occurrence and demand size on TSB worked effectively. Even though it was made specifically for intermittent, it turned out that TSB actually lacked of accuracy performance in this type.

Intermittent and lumpy, whose demand rarely appears, made TSB less good at capturing its patterns. Random Forest, which had additional features, such as rolling statistics, lag, lead time, new installs, and new sales, made the ML model recognized more complex

demand patterns and the results were more accurate. These findings proved that demand characteristics had an impact on forecasting results.

The erratic type indicated a continuity between accuracy and inventory performance. The service level provided by TSB was consistently better than Random Forest. However, the results of the evaluation for other types stated otherwise. The smooth type with the best accuracy in all demand categories, when tested for its inventory performance, it turned out that TSB only won in $S_2$ total demand. This indicated the best accuracy TSB was only able to do slightly better at fulfilling customer orders in total demand.

The results also indicated that no single model consistently outperformed the others across all service level indicators. For instance, while Random Forest dominated in handling intermittent sales demand, TSB held a slight edge in $S_2$. So even though customer orders from most data horizons could be fulfilled, the infrequent occurrence of demand could make inventory systems unable to read patterns for demand predictions.

The TSB method, which was several times better than its $S_2$, revealed that its forecast results were able to serve customer needs better than Random Forest, but stockouts often occurred. Random Forest, which could be helped by additional features, could better capture patterns on most of the demand data, which was only a very thin margin from TSB.

Demands with frequent zeros and small variations made the prediction stable at zero. When there was a demand, it often could not be fully met due to a lack of inventory. The lumpy type, although it had rare emergent of demand, such as intermittent, tends to had a high variability of demand. In cases like this, it was easier to read the demand pattern.

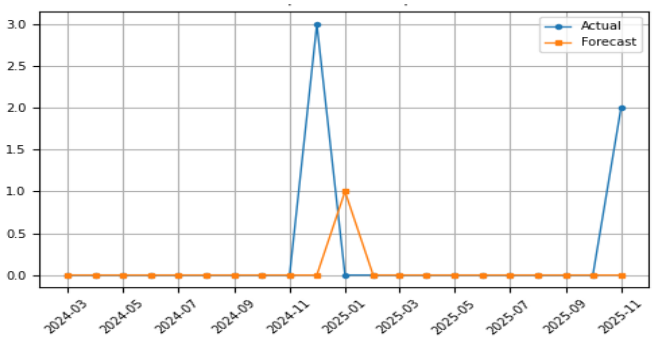



**Figure 4.** SKU AR From Intermittent Sales

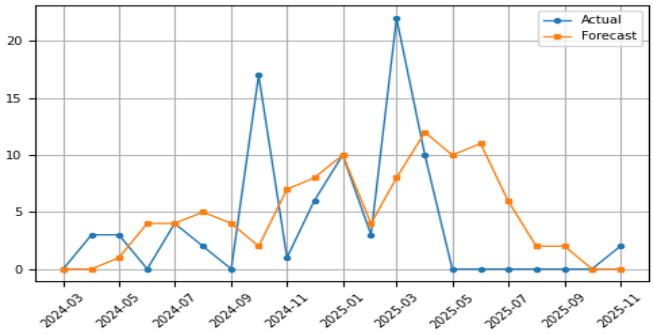



**Figure 5.** SKU F From Intermittent Sales

The accuracy of the forecasting provided by the RMSE and MAE values did not necessarily have any one-way implications to the service level. Error metrics were used in

aggregate for all SKUs in the demand type. While inventory simulations with order up-to-S policy were carried out in each period at the SKU level. As a result, large error values in some SKUs could be masked by small error values if those SKUs were on the same demand type. This debit-credit effect could not be applied to order up-to-S policy which calculated the shortage of inventory and fulfillment of customer orders in each period before then aggregating the value per type of demand. Demand forecasting evaluations should consider their impact on the company's operational performance. High forecasting accuracy did not always encourage effective inventory management.

## CONCLUSION

The results showed that the Random Forest model achieved greater forecasting accuracy and higher inventory service levels than the Teunter–Syntetos–Babai (TSB) method, particularly for intermittent and lumpy demand patterns. Random Forest's superior performance was attributed to its ability to incorporate various predictive features, including rolling statistics, lagged demand variables, lead times, and business activities such as new printer sales and installations. Meanwhile, the TSB method demonstrated good forecasting accuracy for smooth demand patterns, although this advantage did not consistently translate into higher inventory service levels. Interestingly, TSB also achieved favorable service-level performance for erratic demand patterns, despite being primarily designed to address intermittent demand. For erratic demand, particularly in the sales category, forecasting accuracy was more closely aligned with inventory service-level performance, with TSB demonstrating superior results. For other demand patterns, Random Forest generally outperformed TSB in terms of both forecasting accuracy and inventory service levels. Furthermore, the findings indicated that demand aggregation could obscure the underlying characteristics of individual demand patterns. Disaggregating demand by source, specifically rental and sales, provided more detailed insights into demand characteristics and their variability, including differences in SKU classifications across demand categories. For smooth demand, Random Forest achieved the highest service levels in the rental and sales categories. However, for aggregate demand, TSB achieved a slightly higher fill rate (S2), with a marginal difference of only 0.01. Overall, the findings confirmed that improvements in forecasting accuracy did not necessarily correspond to improvements in inventory service levels. Models with greater forecasting accuracy did not consistently achieve higher service levels, and under certain conditions, improved forecasting accuracy was associated with lower service-level performance. Therefore, forecasting method selection should consider not only predictive accuracy but also inventory service-level outcomes and the characteristics of the underlying demand patterns. Future research could extend the evaluation of inventory performance by incorporating additional operational measures, particularly inventory holding costs, stockout costs, and total inventory costs. Moreover, ML model performance could be enhanced by incorporating additional information on supply chain disruptions and other relevant external factors. Further studies could also investigate alternative machine learning and deep learning forecasting approaches, apply additional performance evaluation metrics, and validate the findings using datasets covering different observation periods, industries, and demand characteristics.

**REFERENCES**

Altug, M. S., & Ceryan, O. (2022). Optimal dynamic allocation of rental and sales inventory for fashion apparel products. *IISE Transactions*, *54*(6), 603–617.

Amnay, I., Charani, E., & Tahiri, A. (2025). A Comparative Study of Machine Learning and Statistical Methods for Demand Forecasting in Supply Chains. Revue Internationale Des Sciences de Gestion, 8(2025), 509–534.

Axsäter, S. (2015). Inventory Control. In International Series in Operations Research & Management Science (Vol. 225). Springer International Publishing. https://doi.org/10.1007/978-3-319-15729-0

Bai, Y. (2024). Machine Learning Implementation for Demand Forecasting in Supply Chain Management. Proceedings of the 1st International Conference on E-Commerce and Artificial Intelligence, 77–84. https://doi.org/10.5220/0013206900004568

Goldsman, D. (1992). Simulation output analysis. Proceedings of the 24th Conference on Winter Simulation - WSC '92, 97–103. https://doi.org/10.1145/167293.167310

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer New York. https://doi.org/10.1007/978-0-387-84858-7

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning (Vol. 103). Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

Kumar, V. D., Maheswari, S., Raman, Y. S., Iniyavan, S., & Hashim, S. A. (2025). Inventory Optimization and Demand Forecasting Using Machine Learning. In Proceedings of the International Conference on Intelligent Systems and Digital Transformation (ICISD 2025) (pp. 760–773). https://doi.org/10.2991/978-94-6463-866-0_62

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. 31st Conference on Neural Information Processing Systems (NIPS 2017), 11(3), 4768–4777. https://doi.org/https://dl.acm.org/doi/10.5555/3295222.3295230

Maniah. (2026). Performance Evaluation of Machine Learning Algorithms for Supply Chain Data Classification. Engineering Science Letter, 5(01), 34–40. https://doi.org/10.56741/IISTR.esl.001524

Nguyen, N. P. T., Dang, T. T., & Le, D. D. (2025). Inventory Demand Forecasting Using XGBoost and LightGBM Algorithms: A Case Study of Grupo Bimbo. In Advances in Transdisciplinary Engineering (Vol. 73, pp. 46–57). IOS Press BV. https://doi.org/10.3233/ATDE250516

Owen, L. (2022). Hyperparameter Tuning With Python. Packt Publishing. https

Petropoulos, F., Wang, X., & Disney, S. M. (2019). The inventory performance of forecasting methods: Evidence from the M3 competition data. International Journal of Forecasting, 35(1), 251–265. https://doi.org/10.1016/j.ijforecast.2018.01.004

Sanguri, K., Patra, S., Nikolopoulos, K., & Punia, S. (2024). Intermittent demand, inventory obsolescence, and temporal aggregation forecasts. International Journal of Production Research, 62(5), 1663–1685. https://doi.org/10.1080/00207543.2023.2199435

Sheffi, Y. (2004). Demand Variability and Supply Chain Flexibility. In Entwicklungspfade und Meilensteine moderner Logistik (pp. 85–117). Gabler Verlag. https://doi.org/0.1007/978-3-322-89044-3

Snyder, L. V., & Shen, Z.-J. M. (2019). Fundamentals of Supply Chain Theory (2nd ed.). John Wiley & Sons, Inc.

Spiliotis, E., Makridakis, S., Semenoglou, A. A., & Assimakopoulos, V. (2022). Comparison of statistical and machine learning methods for daily SKU demand forecasting. Operational Research, 22(3), 3037–3061. https://doi.org/10.1007/s12351-020-00605-2

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

Teunter, R. H., Syntetos, A. A., & Babai, M. Z. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. European Journal of Operational Research, 214(3), 606–615. https://doi.org/10.1016/j.ejor.2011.05.018

Theodorou, E., Spiliotis, E., & Assimakopoulos, V. (2025). Forecast accuracy and inventory performance: Insights on their relationship from the M5 competition data. European Journal of Operational Research, 322(2), 414–426. https://doi.org/10.1016/j.ejor.2024.12.033

Waters, D. (2003). Inventory Control and Management (2nd ed.). John Wiley & Sons Ltd.