



Adaptive Inventory Replenishment Forecasting Using Time-Series Machine Learning for Multi-Branch Bookstore Operations

Jesicha Natasha Mashel^{1,*}, Immanuel Britain Purba²

^{1,2}Department of Information Systems, Faculty of AI and Data Science, Universitas Pelita Harapan, Indonesia

ABSTRACT

Agile inventory management requires timely, accurate, and stable short-term forecasts to support rapid yet reliable replenishment decisions. In retail distribution networks, particularly those with multiple branches and a large number of stock keeping units (SKUs), forecasting errors and excessive forecast volatility can lead to operational inefficiencies, frequent replanning, and increased costs. This study investigates the use of time-series machine learning to enhance agile replenishment planning by improving next-month readiness in a multi-branch bookstore context. Using a publicly available dataset of monthly inventory movements, the replenishment problem is formulated as a one-step-ahead forecasting task at the branch–SKU level. Inbound inventory transfers are treated as a proxy for replenishment demand, reflecting practical planning conditions in the absence of direct sales or inventory-on-hand data. A classical Exponential Smoothing (ETS) model is employed as a baseline and compared against a global Light Gradient Boosting Machine (LightGBM) regression model. The machine-learning approach integrates lagged demand features, seasonal indicators, pricing and margin information, and branch-level context, and is trained using a strictly time-aware validation strategy. Experimental results show that the LightGBM model consistently outperforms the ETS baseline in terms of forecasting accuracy, achieving lower Mean Absolute Error and Root Mean Squared Error on a held-out test set. In addition, an agility-oriented stability metric reveals that machine-learning-based forecasts are substantially smoother than those produced by the classical model, indicating reduced sensitivity to short-term fluctuations. Visual analysis further confirms that the global machine-learning model generalizes better across SKUs and branches, producing predictions that are more tightly aligned with actual replenishment quantities. The findings demonstrate that time-series machine learning can effectively support agile replenishment planning by balancing predictive accuracy and planning stability. The study also highlights a practical trade-off between responsiveness and smoothness, suggesting that machine-learning forecasts are well suited as a baseline planning tool, complemented by classical methods for exception handling. Overall, this work provides a reproducible and data-driven framework for integrating machine learning into agile inventory decision-making.

Keywords Agile Inventory Management, Replenishment Forecasting, Time-Series Machine Learning, Lightgbm, Retail Supply Chain

Introduction

Retail supply chains have grown increasingly complex due to multi-branch distribution structures, centralized planning mechanisms, and the integration of multi-channel fulfillment across online and offline platforms. These developments significantly increase coordination requirements and amplify the number of operational exceptions that planners must address within short

Submitted: 15 March 2025
Accepted: 30 April 2025
Published: 2 November 2025

Corresponding author
Jesicha Natasha Mashel,
1081230010@student.uph.edu

Additional Information and
Declarations can be found on
[page 255](#)

© Copyright
2025 Mashel and Purba

Distributed under
Creative Commons CC-BY 4.0

How to cite this article: J. N. Mashel and I. B. Purba, "Assessing the Efficacy of Convolutional Neural Networks in Recognizing Handwritten Digits: Analysis of the Relationship Between Trading Volume and Bitcoin Price Movements Using Pearson and Spearman Correlation Methods," *Agile Manag.*, vol. 1, no. 4, pp. 242-257, 2025.

planning horizons [1], [2]. As a result, firms face mounting difficulties in maintaining alignment between supply-chain design and volatile demand conditions. In unstable environments, frequent design and operational adjustments are required, which complicates forecasting and execution and increases the resources needed to achieve supply-chain alignment [3].

In inventory-intensive retail sectors, the operational consequences of this complexity are most visible in short-term replenishment decisions. Timely and frequent replenishment directly influences product availability, lost sales, and inventory holding costs across large portfolios of SKUs and geographically dispersed branches [4]. Consequently, short-horizon replenishment planning has become a central mechanism for balancing responsiveness and cost efficiency in retail operations.

Classical replenishment processes often struggle under these conditions because planners must reconcile numerous exception signals while accounting for network-level constraints. When demand is volatile and SKU portfolios are large, local demand noise can overwhelm planning systems, degrading decision reliability and increasing manual intervention [5]. These limitations highlight the growing need for forecasting approaches that can support short-term, multi-SKU, and multi-branch replenishment decisions in complex retail environments. Agility in operations and supply-chain management is commonly defined as the capability to sense and respond rapidly to environmental changes while maintaining effective coordination across the supply network [3], [4]. It encompasses both cognitive agility, which relates to rapid information processing and decision-making, and physical agility, which concerns the flexible execution of operational responses. Together, these dimensions enable firms to sustain service levels and cost efficiency under dynamic demand conditions.

While agility emphasizes responsiveness and adaptability, it also requires a degree of planning stability. Excessive reactivity can introduce operational instability and transaction costs, particularly in multi-branch systems where frequent tactical adjustments propagate across the network [3]. Therefore, agility is not synonymous with constant change, but rather with balanced responsiveness supported by stable and coordinated planning mechanisms.

Information technology, big data analytics, and decision support systems are widely recognized as key enablers of cognitive agility. By improving situational awareness and automating corrective actions, these technologies support more frequent and distributed replenishment cycles while reducing the cognitive burden on planners [2], [6]. Forecasting systems are thus embedded within a broader digital infrastructure that enables agile operations.

Forecasting plays a foundational role in agile replenishment by transforming raw demand signals into anticipatory decisions. Accurate short-term forecasts enable planners to trigger replenishment actions in a timely manner, reduce reliance on excessive safety stock, and stabilize replenishment frequencies across SKUs and branches [4]. As such, forecasting directly supports the responsiveness and efficiency objectives of agile inventory management.

At the same time, forecasts must be integrated with operational rules and exception management processes to translate predictions into robust replenishment actions within centralized or decentralized planning architectures [1]. When forecasts are noisy or excessively volatile, planners face a trade-off between responsiveness and stability. Rapid reactions to short-term signals can improve service levels but may induce variability in orders and operations, while

excessive smoothing can reduce operational churn at the cost of increased stockouts or obsolescence [3], [4].

Empirical studies highlight that planners often struggle with this trade-off in retail contexts. Inaccurate or unstable forecasts force planners to choose between frequent manual interventions and rigid decision rules that dampen volatility but may be suboptimal for service or cost performance [5]. These frictions underscore the need for forecasting approaches that support both accuracy and stability in agile replenishment systems.

Classical univariate time-series models, such as moving averages, exponential smoothing, and simple ARIMA variants, remain widely used in retail practice due to their simplicity, interpretability, and computational efficiency [7]. These models adapt quickly to changes within individual series and provide high local responsiveness, making them attractive for operational planners managing large numbers of SKUs with limited data.

Their transparency and straightforward parameterization further support their adoption in practice, as planners can easily understand and trust model outputs [7]. However, these same characteristics also limit their effectiveness in modern retail environments. Univariate models typically fail to exploit cross-SKU, cross-branch, or auxiliary information such as pricing, promotions, or shared seasonality patterns [8].

As a result, classical approaches underutilize signals that are increasingly available in data-rich retail systems. This limitation becomes particularly problematic in multi-branch and multi-SKU settings where correlated demand patterns, substitution effects, or regional trends exist [9]. Independent per-series forecasts cannot borrow strength across the network, leading to inconsistent accuracy and poor performance for sparse or low-volume SKUs.

The growing adoption of Machine Learning (ML) in inventory forecasting is driven by increased data availability, advances in computational infrastructure, and the maturation of ML toolchains [2], [6]. Retailers now routinely collect detailed point-of-sale data, digital channel interactions, and operational telemetry, enabling richer representations of demand dynamics.

ML models offer greater expressive power than classical methods, allowing them to capture non-linear relationships, interactions, and complex feature dependencies that are difficult to model using traditional time-series techniques [10]. These capabilities make ML particularly suitable for modern retail environments characterized by heterogeneity and high dimensionality.

ML approaches such as gradient boosting, tree ensembles, and neural networks can integrate cross-sectional information across SKUs and branches, enabling low-signal series to benefit from patterns learned from high-signal counterparts [11]. This represents a conceptual shift from independent local models toward global forecasting models that learn shared structures across many time series.

Empirical studies demonstrate that global ML models often deliver more accurate and robust short-term forecasts in retail datasets, particularly when demand is sparse or noisy [11]. These advantages make ML approaches promising candidates for supporting agile replenishment planning in complex retail networks.

Despite their advantages, ML-based forecasting systems face practical challenges. Data governance constraints, limited inventory transparency, noisy labels, and data quality issues can restrict model performance [4]. Additionally, the need for interpretability and integration with operational exception

management limits the feasibility of purely black-box deployments [2]. Highly reactive ML models may also exacerbate order volatility if not combined with smoothing mechanisms, decision constraints, or human-in-the-loop processes. Consequently, effective ML adoption requires careful system design that balances accuracy, responsiveness, and stability in line with agile management principles [3].

Despite the growing adoption of ML techniques in demand and inventory forecasting, existing studies rarely link ML-based forecasting explicitly to agile replenishment readiness. Much of the prior literature evaluates forecasting performance primarily through accuracy metrics, such as MAE or RMSE, without situating these improvements within the operational logic of agile planning cycles. As a result, it remains unclear how forecasting models contribute to short-term planning stability, responsiveness, and decision reliability in multi-branch, multi-SKU retail environments.

In addition, forecast stability—the extent to which predictions fluctuate over time relative to actual demand—has received limited attention in inventory research. While accuracy is essential, excessively volatile forecasts can undermine agile operations by triggering frequent replanning and increasing operational noise. Furthermore, many existing ML-based studies rely on proprietary datasets or experimental settings that are difficult to reproduce, limiting their practical relevance and comparability. There is therefore a need for reproducible studies that evaluate forecasting methods using operationally realistic datasets and validation protocols that reflect real planning conditions.

To address these gaps, this study investigates how time-series machine learning can support agile replenishment planning. The research has three main objectives: (1) to empirically compare machine-learning-based global forecasting models with classical time-series approaches for branch-SKU replenishment forecasting; (2) to assess the trade-offs between forecast accuracy and stability as an indicator of agility; and (3) to demonstrate an evaluation framework that integrates traditional error metrics with agility-oriented measures. The key contributions of this work are threefold: methodologically, it proposes a reproducible ML-based forecasting pipeline with time-aware validation; empirically, it provides evidence of accuracy and stability improvements in a multi-branch retail context; and managerially, it offers actionable insights into how forecasting models can be positioned as decision-support tools for agile inventory management.

Literature Review

Fundamentals of Inventory Control and Replenishment Systems

Inventory control and replenishment systems translate demand information into procurement or transfer actions through explicit policies such as order-up-to, base-stock, and ((s,S)) rules. These policies are designed to balance service levels, holding costs, and ordering costs under uncertainty [12]. In inventory-intensive retail contexts, short-term replenishment operating on daily or weekly cadences is particularly critical, as it directly affects in-stock availability, lost sales, and the amount of working capital tied up across numerous SKUs and locations [1], [5]. Demand forecasts serve as a central input to these replenishment rules, determining order quantities and safety stock levels.

Forecast errors therefore propagate directly into inventory outcomes, causing stockouts, overstocks, and amplified variability across the supply chain,

commonly referred to as the bullwhip effect [12], [13]. As a result, forecasting performance is typically evaluated using both point-error metrics (e.g., MAE, RMSE, MAPE) and inventory-centric indicators such as fill rate, days of supply, stockout frequency, and inventory turnover [14], [15]. Forecasting at fine granularity, particularly at the SKU–location level, remains challenging due to short histories, intermittent demand, and high noise, which reduce signal-to-noise ratios and destabilize per-series estimation in large retail assortments [13], [16].

Classical Time-Series Models in Retail Forecasting

Classical retail forecasting methods are dominated by univariate time-series models, including exponential smoothing state-space (ETS), ARIMA, and Seasonal ARIMA (SARIMA). These approaches remain popular due to their parsimony, interpretability, and computational efficiency, enabling rapid calibration across large numbers of independent series [13], [17]. When demand exhibits stable trend and seasonal structure and sufficient historical data are available, classical models can provide reliable forecasts with transparent components that practitioners can inspect and audit [17].

However, limitations emerge under volatile or intermittent demand conditions. Frequent promotions, short SKU life cycles, and assortment changes introduce non-stationarities and structural breaks that classical identification procedures struggle to capture without manual tuning or additional exogenous variables [13], [18]. Consequently, while classical models perform well for stable medium- to long-term patterns, their effectiveness for short-term operational replenishment at the SKU–location level is often inferior unless augmented with cross-series information or hybrid structures [13], [19].

Machine Learning for Time-Series Forecasting

ML reframes time-series forecasting as a supervised regression problem, where engineered features—such as lagged demand, rolling statistics, calendar variables, prices, and promotional indicators—are mapped to future demand using flexible function approximators [20], [21]. Feature-based global forecasting models pool data across many series, allowing sparse or noisy SKU–location series to borrow strength from richer counterparts through shared parameters or representations [11], [21], [22].

Tree-based ensemble methods, including Gradient Boosting Machines, XGBoost, LightGBM, and CatBoost, are widely adopted in retail forecasting due to their ability to model nonlinear relationships, handle heterogeneous features, and scale to large datasets while offering feature-importance diagnostics for operational interpretability [20], [11], [23]. Empirical comparisons consistently show that ML ensembles and hybrid architectures outperform classical univariate models on short-term retail horizons, particularly when rich covariates and cross-series learning are available [11], [24]. Nevertheless, ML performance depends critically on feature engineering, data governance, and sufficient cross-sectional coverage to support pooled learning [10], [14], [25].

Global versus Local Forecasting Models

Local forecasting strategies fit one model per SKU–location series and are attractive for their simplicity and interpretability when long historical records exist. However, these approaches do not exploit cross-sectional commonalities

across products, stores, or channels, limiting their effectiveness in large-scale retail systems [17], [13]. In contrast, global forecasting strategies train a single model—or a small set of shared models—across many series, enabling automatic learning of shared seasonality, promotion effects, and store-level behavior [11], [21], [22].

Empirical evidence from retail and e-commerce datasets indicates that global ML models deliver more robust short-term forecasts and better generalization than independent univariate approaches, especially for low-volume, intermittent, or short-history series [11], [24], [21]. For multi-branch, multi-SKU retail networks, global models are particularly advantageous because they capture consistent cross-store behavior and substantially reduce the manual tuning burden associated with maintaining thousands of local models [26], [11].

Agile Management and Forecasting Support

Agility in operations and supply chains refers to the ability to rapidly detect changes and reconfigure processes to maintain alignment between supply and demand. This capability combines cognitive elements—such as information processing and forecasting—with physical execution capabilities in logistics and replenishment [3], [2], [4]. Predictive analytics and ML-based forecasting systems are key enablers of cognitive agility, improving situational awareness and supporting faster, more informed replenishment decisions [6], [10], [27].

However, agile systems also require planning stability. Overly reactive forecasts can induce order volatility, increase exception handling, and generate operational friction across the network. As a result, forecasting solutions must explicitly manage the responsiveness–stability trade-off through smoothing mechanisms, decision constraints, hierarchical cadencing, or integrated forecasting–inventory control frameworks [3], [28]. While much of the forecasting literature emphasizes accuracy improvements, fewer studies quantify operational implications such as order volatility, exception workload, or inventory performance, highlighting the need for research that connects forecasting methods to agile replenishment execution under realistic constraints [14], [15], [25].

Practical Implications for Research and Practice

Prior studies collectively suggest that forecasting models should be evaluated against inventory-centric objectives rather than point-error metrics alone. Incorporating measures such as fill rate, stockout frequency, order volatility, and bullwhip indicators aligns forecasting evaluation more closely with replenishment outcomes and managerial objectives [12], [15], [14]. Such multi-criteria evaluation frameworks are particularly relevant in agile environments where decision stability and execution reliability are as important as predictive accuracy.

For practice, the literature supports the use of global, feature-based ML models when sufficient pooled data and cross-sectional covariates are available, combined with operational smoothing rules or action filters to preserve planning stability [11], [21], [28], [3]. Hybrid architectures that integrate classical models for interpretable seasonal dynamics with ML components for nonlinear and promotional effects have also proven effective [19], [23], [11]. Finally, phased implementation strategies with A/B testing, explainability dashboards, and

human-in-the-loop exception management are recommended to ensure successful integration of ML forecasting into agile replenishment workflows [14], [10], [4].

Method

Research Design and Problem Formulation

This study adopts a quantitative predictive modeling design to support agile replenishment planning in a multi-branch retail environment. The central objective is to improve next-month readiness by forecasting inbound replenishment quantities at the branch–SKU level. Rather than treating inventory demand as a static or purely reactive process, the approach frames replenishment as a short-horizon planning problem where forecast quality directly affects operational agility.

The prediction task is formulated as a one-step-ahead regression problem, where the target variable represents the quantity of items transferred to a destination branch in the following month. These inbound transfers are used as a proxy for replenishment demand, which is appropriate given the absence of direct point-of-sale or inventory-on-hand data. This formulation aligns with real-world replenishment cycles, which typically operate on monthly planning intervals.

Two forecasting paradigms are evaluated to assess their suitability for agile operations: a classical univariate time-series baseline and a global machine-learning-based time-series model. The classical model provides a transparent benchmark that captures local temporal dynamics, while the machine-learning model leverages shared patterns across products and branches to enhance generalization and planning stability.

Model performance is assessed not only through standard forecasting accuracy metrics but also through an agility-oriented stability metric that captures the trade-off between responsiveness and forecast smoothness. This dual evaluation reflects the practical requirements of agile inventory management, where overly reactive forecasts can be as detrimental as inaccurate ones.

Data Source, Preprocessing and Aggregation

The dataset consists of 21,566 monthly inventory movement records collected from a bookstore distribution network spanning multiple branches and stock keeping units (SKUs). Each record includes the transfer month, origin and destination locations, SKU identifier, quantity moved, unit cost, unit price, and associated extended cost and retail value. All records are timestamped at a monthly resolution, making the dataset suitable for time-series analysis.

To ensure consistency and avoid duplication effects, raw transaction records are aggregated at the branch–SKU–month level. Specifically, quantities and extended values are summed, while unit-level cost and price variables are averaged within each aggregation unit. This aggregation reflects operational planning units, where replenishment decisions are typically made per SKU for each branch monthly.

Several derived financial features are computed during preprocessing to enrich the modeling space. These include per-unit margin (unit price minus unit cost) and extended margin (total retail value minus total cost). Such variables provide

economic context that may influence replenishment behavior, as higher-margin products are often prioritized in restocking decisions.

Data preprocessing strictly follows time-aware principles to prevent information leakage. All records are sorted chronologically, and missing values introduced by lag feature construction are removed. This results in a reduced but methodologically valid dataset that preserves the temporal causality required for forecasting evaluation.

Feature Engineering and Temporal Representation

Feature engineering is designed to capture historical demand patterns, seasonality, economic signals, and branch-level context. Lagged demand features are constructed for 1, 2, 3, 6, and 12 months, enabling the models to learn short-term inertia, medium-term adjustments, and annual seasonal effects. These lag horizons are standard in retail forecasting and reflect common planning cycles.

To represent seasonality explicitly, the month of the year is encoded as a numerical feature, allowing models to learn recurring monthly effects such as holiday seasons or academic cycles that may affect book demand. In addition, a monotonically increasing time index is included to capture long-term trends and structural changes in replenishment behavior.

Branch-level context is incorporated by computing the total inbound quantity received by each branch in the previous month. This variable captures localized demand pressure and capacity effects, allowing the model to distinguish between isolated SKU fluctuations and branch-wide replenishment patterns. Importantly, this feature is lagged to ensure that only information available prior to the forecast month is used.

All features are engineered using strictly historical data relative to the prediction point. By avoiding the use of contemporaneous or future information, the feature set maintains the integrity of a real-world forecasting scenario and supports fair model comparison.

Time-Aware Data Splitting and Validation Strategy

The dataset is partitioned into training, validation, and test sets using a chronological split rather than random sampling. Approximately 70% of the earliest observations are used for model training, followed by 15% for validation and 15% for final testing. This approach mirrors operational deployment conditions, where models are trained on past data and applied to future periods.

The validation set serves two purposes: parameter tuning and early performance assessment. By evaluating candidate models on temporally subsequent data, the validation process provides a realistic estimate of generalization performance and reduces the risk of overfitting to historical idiosyncrasies.

For the classical baseline, forecasting is performed independently for each branch–SKU series using only historical observations up to the forecast month. In contrast, the machine-learning model is trained as a global model across all branch–SKU combinations, enabling it to learn shared structures and cross-sectional regularities.

All reported performance metrics are computed exclusively on the held-out test set, which represents the most recent time period in the dataset. This ensures that the results reflect genuine predictive capability rather than retrospective fitting.

Classical Baseline: Exponential Smoothing (ETS)

Exponential Smoothing is selected as the classical baseline due to its widespread use in retail forecasting and its ability to model level, trend, and seasonality with minimal assumptions. An additive formulation is employed for both trend and seasonal components, reflecting the approximately linear scale of monthly replenishment quantities.

The seasonal period is fixed at 12 months to capture annual patterns, which are common in book retail due to academic calendars and holiday cycles. Model parameters are estimated automatically through optimization of the in-sample likelihood, ensuring consistency across thousands of individual time series.

Forecasting is conducted in a one-step-ahead rolling manner, where each test observation is predicted using only data available prior to that month. In cases where a time series is too short or optimization fails, a fallback strategy based on historical mean is applied to ensure robust execution across all branch–SKU combinations.

ETS serves as a transparent and interpretable benchmark, providing a point of comparison for evaluating whether machine-learning-based approaches offer tangible benefits for agile replenishment planning.

Machine-Learning Model: LightGBM Regression

The machine-learning approach employs LightGBM regression, chosen for its efficiency, scalability, and strong performance on structured tabular data. LightGBM builds an ensemble of decision trees using gradient boosting, allowing it to capture nonlinear relationships and complex feature interactions.

Model hyperparameters are selected to balance predictive power and generalization. The number of estimators is set to 500, providing sufficient boosting rounds to learn stable patterns, while the learning rate is fixed at 0.05 to prevent overly aggressive updates. The maximum tree depth is limited to 8 to reduce overfitting and maintain interpretability. Training verbosity is disabled to ensure reproducibility and clean execution.

Categorical variables representing branch and SKU identifiers are encoded using one-hot encoding within a preprocessing pipeline. This encoding allows the model to learn entity-specific effects while preserving compatibility with gradient-boosted trees. Numerical features are passed through without scaling, as tree-based models are insensitive to feature magnitude.

The LightGBM model is trained as a global forecasting model, meaning a single model learns from all branch–SKU histories simultaneously. This design enables knowledge transfer across series, improving performance for sparse or low-volume items and enhancing overall planning stability.

Evaluation Metrics and Agile Performance Indicators

Forecast accuracy is evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). MAE provides a straightforward measure of

average deviation in replenishment units, while RMSE penalizes larger errors more strongly, reflecting the higher operational cost of significant misallocations.

To complement traditional accuracy metrics, an agility-oriented stability index is introduced to assess forecast smoothness relative to actual demand volatility. This index is computed as the ratio between the variance of forecast month-to-month changes and the variance of actual changes. Values closer to zero indicate smoother, more stable forecasts.

The stability metric captures an important operational trade-off in agile systems: overly reactive forecasts can lead to frequent replanning and operational noise, whereas excessively smooth forecasts may lag behind genuine demand shifts. By reporting both accuracy and stability, the evaluation provides a more holistic view of model suitability.

All metrics are computed on the test set to ensure unbiased assessment. Results are further visualized through error comparison plots, scatter plots of actual versus predicted quantities, and illustrative time-series examples, supporting both quantitative and qualitative interpretation.

Result and Discussion

Dataset Characteristics and Experimental Outcome Overview

The experimental evaluation was conducted on a dataset comprising 21,566 monthly inventory movement records, which was reduced to 17,966 observations after time-series feature engineering and lag construction. This reduction is an expected consequence of enforcing temporal causality and ensuring that each observation contains sufficient historical context for forecasting. The remaining dataset maintains substantial coverage across branches and SKUs, supporting statistically meaningful model evaluation.

The time-aware data partitioning resulted in 12,868 training observations, 2,698 validation observations, and 2,400 test observations. This split preserves the chronological order of events and reflects a realistic deployment scenario in which forecasting models are trained on historical data and evaluated on future periods. The test set therefore represents genuine out-of-sample performance rather than retrospective fitting.

Two forecasting approaches were evaluated: a classical ETS model and a machine-learning-based LightGBM regression model. Both models were trained and evaluated using identical data splits and target definitions, ensuring a fair and controlled comparison. Performance was assessed using standard error metrics as well as an agility-oriented stability indicator.

Overall, the results demonstrate that both approaches can capture the general structure of bookstore replenishment behavior. However, measurable differences emerge in accuracy, error dispersion, and forecast smoothness, which have direct implications for agile replenishment planning.

Forecast Accuracy Comparison

Table-level evaluation shows that the LightGBM model achieves a Mean Absolute Error (MAE) of 4.23 and a Root Mean Squared Error (RMSE) of 5.39, outperforming the ETS baseline, which records an MAE of 4.55 and an RMSE

of 5.77, as shown in [figure 1](#). Although the numerical improvements may appear moderate, they represent a consistent reduction across both error metrics, indicating robust performance gains rather than metric-specific artifacts.

The relative MAE reduction of approximately 7% suggests that, on average, the machine-learning model predicts monthly replenishment quantities more accurately across a wide range of SKUs and branches. In operational terms, even small reductions in per-SKU error can accumulate into substantial improvements when aggregated across hundreds of products and multiple locations.

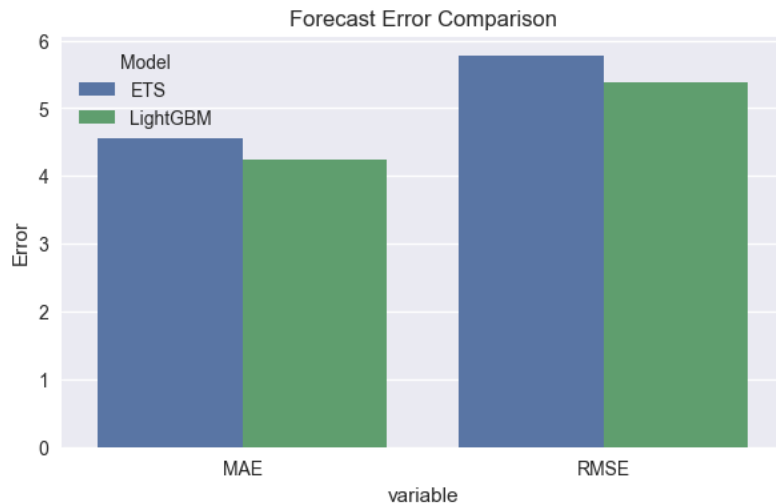


Figure 1 Forecast Error Comparison

RMSE improvements further indicate that LightGBM reduces the occurrence of large forecasting errors. This is particularly important in replenishment contexts, where extreme over- or under-forecasting can lead to stockouts, excess inventory, or emergency reorders. The lower RMSE therefore reflects improved risk management in planning decisions.

From an agile management perspective, these accuracy gains translate into enhanced predictive agility. More accurate forecasts allow planners to define more reliable short-term replenishment backlogs, reducing the need for frequent replanning and corrective actions during execution.

Forecast Dispersion and Bias Analysis

The scatter plot comparing actual versus predicted quantities ([figure 2](#)) provides deeper insight into model behavior beyond aggregate metrics. Both ETS and LightGBM predictions align broadly along the diagonal, indicating that the overall demand structure is well captured by both models. This confirms that bookstore replenishment patterns exhibit a degree of regularity and predictability.

However, the LightGBM predictions display noticeably tighter clustering around the diagonal compared to ETS. This reduced dispersion suggests that the machine-learning model generalizes more effectively across different SKUs and branches, benefiting from its global training strategy and richer feature set. ETS, in contrast, shows wider vertical spread, particularly at mid-range demand

levels.

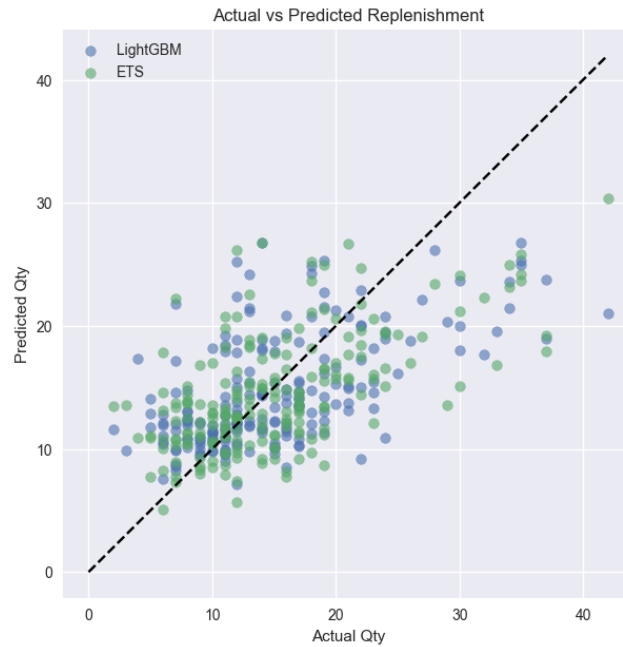


Figure 2 Actual vs Predicted Replenishment Plot

Systematic under-prediction at higher demand levels is observed for both models, although it is more pronounced in ETS. This behavior reflects the conservative nature of historical extrapolation in the presence of volatile peaks and highlights the inherent difficulty of forecasting extreme replenishment events using limited contextual information.

These observations support the interpretation that LightGBM better balances bias and variance, offering improved generalization without excessive sensitivity to noise. For agile planning, this translates into more dependable forecasts that reduce uncertainty in capacity and allocation decisions.

Temporal Behavior and Stability–Responsiveness Trade-off

Time-series inspection of individual SKU–branch examples reveal highly volatile historical replenishment patterns, with frequent spikes and irregular fluctuations, as visualize in [figure 3](#). Such behavior challenges classical time-series models, which rely heavily on recent trends and seasonal components. In this context, ETS forecasts tend to follow short-term fluctuations more closely.

In contrast, the LightGBM model produces forecasts that are closer to the long-term average behavior of the series. By incorporating cross-sectional information from other SKUs and branches, the machine-learning approach effectively regularizes predictions, dampening transient fluctuations that may not reflect sustained demand changes.

This difference is quantitatively reflected in the stability index results. The ETS model records a stability index of 0.29, while LightGBM achieves a substantially lower value of 0.12. Lower values indicate smoother forecasts relative to actual demand volatility, confirming that LightGBM emphasizes planning stability over short-term reactivity.

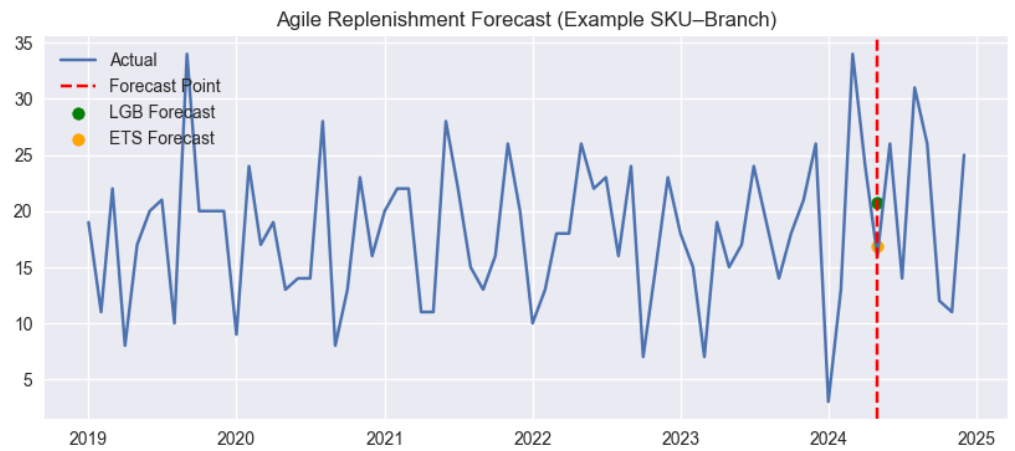


Figure 3 Agile Replenishment Forecast

From an agile management standpoint, this trade-off is significant. While responsiveness is essential for handling exceptional events, excessive volatility in forecasts can disrupt sprint-level planning. The LightGBM model's stability therefore supports consistent execution, while ETS may be better suited for detecting abrupt changes requiring managerial intervention.

Implications for Agile Replenishment Planning

The combined results suggest that machine-learning-based global forecasting enhances agile replenishment planning by improving both accuracy and forecast stability. More accurate predictions reduce planning errors, while smoother forecasts support stable short-term execution cycles. Together, these properties improve next-month readiness across branches.

Importantly, the results do not imply that classical models are obsolete. Instead, ETS demonstrates stronger responsiveness to local fluctuations, which can be valuable for anomaly detection or exception management. This distinction supports a complementary rather than competitive view of forecasting methods within agile systems.

The observed trade-off aligns with agile principles that emphasize stable plans with rapid feedback loops. Machine-learning forecasts can provide the baseline plan, while classical or rule-based mechanisms can trigger adjustments when deviations exceed acceptable thresholds.

Overall, the findings demonstrate that time-series machine learning can function as an effective decision-support tool for agile inventory management, improving planning quality without sacrificing adaptability. The results also highlight the importance of evaluating forecasting models through both accuracy and agility-oriented lenses, rather than relying solely on traditional error metrics.

Conclusion

This study demonstrated that time-series machine learning can effectively enhance agile replenishment planning in a multi-branch bookstore context. By framing inbound inventory movements as a next-month forecasting problem and comparing a classical Exponential Smoothing baseline with a global LightGBM model, the results showed consistent improvements in forecasting

accuracy and risk reduction. The machine-learning approach achieved lower MAE and RMSE while producing more stable forecasts, indicating its suitability for supporting short-term planning cycles where reliability and consistency are critical.

Beyond accuracy gains, the study highlighted an important agility-related trade-off between responsiveness and stability. While classical models reacted more strongly to short-term fluctuations, the machine-learning model favored smoother predictions that support predictable execution and reduced replanning effort. These findings suggest that machine-learning-based forecasting can serve as a robust baseline for agile inventory planning, complemented by classical or rule-based methods for handling exceptional demand shifts. Overall, the proposed approach offers a practical and reproducible pathway for integrating predictive analytics into agile supply chain decision-making.

Declarations

Author Contributions

Conceptualization: J.N.M. and I.B.P.; Methodology: I.B.P.; Software: J.N.M.; Validation: J.N.M. and I.B.P.; Formal Analysis: J.N.M. and I.B.P.; Investigation: J.N.M.; Resources: I.B.P.; Data Curation: I.B.P.; Writing Original Draft Preparation: J.N.M. and I.B.P.; Writing Review and Editing: I.B.P. and J.N.M.; Visualization: J.N.M.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] A. Alftan, R. Kaipia, L. Loikkanen, and K. Spens, "Centralised grocery supply chain planning: Improved exception management," *Int. J. Phys. Distrib. Logist. Manag.*, vol. 45, no. 3, pp. 237–259, 2015, doi: 10.1108/ijpdlm-02-2014-0017.
- [2] M. Giannakis and M. Louis, "A multi-agent based system with big data processing for enhanced supply chain agility," *J. Enterprise Inf. Manag.*, vol. 29, no. 5, pp.

- 706–727, 2016, doi: 10.1108/jeim-06-2015-0050.
- [3] D. Gligor, “The role of supply chain agility in achieving supply chain fit,” *Decision Sciences*, vol. 47, no. 3, pp. 524–553, 2016, doi: 10.1111/dec.12205.
- [4] B. Bai, “Acquiring supply chain agility through information technology capability: The role of demand forecasting in retail industry,” *Kybernetes*, vol. 52, no. 10, pp. 4712–4730, 2022, doi: 10.1108/k-09-2021-0853.
- [5] S. M. S. Rana, “Supply chain drivers and retail supply chain responsiveness: Strategy as moderator,” *Int. J. Manag. Pract.*, vol. 13, no. 1, pp. 1–22, 2020, doi: 10.1504/ijmp.2020.10025714.
- [6] Y. Yu, X. Wang, R. Y. Zhong, and G. Q. Huang, “E-commerce logistics in supply chain management,” *Ind. Manag. Data Syst.*, vol. 117, no. 10, pp. 2263–2286, 2017, doi: 10.1108/imds-09-2016-0398.
- [7] M. U. Ashraf, “A predictive analysis of retail sales forecasting using machine learning techniques,” *Lahore Garrison Univ. Res. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 4, pp. 39–49, 2022, doi: 10.54692/igurjcsit.2022.0604399.
- [8] S. D. Mourtas, E. Drakonakis, and Z. Bragoudakis, “Forecasting the gross domestic product using a weight direct determination neural network,” *AIMS Mathematics*, vol. 8, no. 10, pp. 24254–24273, 2023, doi: 10.3934/math.20231237.
- [9] X. Shao, D. Chang, and M. Li, “Optimization of lateral transfer inventory of auto spare parts based on neural network forecasting,” *J. Intell. Syst. Control*, vol. 1, no. 1, pp. 2–17, 2022, doi: 10.56578/jisc010102.
- [10] S. E. Aifuwa, “Predictive analytics models enhancing supply chain demand forecasting accuracy and reducing inventory management inefficiencies,” *Int. J. Multidiscip. Res. Growth Eval.*, vol. 1, no. 3, pp. 171–181, 2020, doi: 10.54660/ijmrge.2020.1.3.171-181.
- [11] A. Mitra, A. Jain, A. Kishore, and P. Kumar, “A comparative study of demand forecasting models for a multi-channel retail company: A novel hybrid machine learning approach,” *Oper. Res. Forum*, vol. 3, no. 58, pp. 1–24, 2022, doi: 10.1007/s43069-022-00166-4.
- [12] J. Ma, L. Zhu, Y. Ye, and S. Hou, “Study of the bullwhip effect in a multistage supply chain with callback structure considering two retailers,” *Complexity*, vol. 2018, no. 1, pp. 1–12, 2018, doi: 10.1155/2018/3650148.
- [13] P. Ramos, N. Santos, and R. Rebelo, “Performance of state space and ARIMA models for consumer retail sales forecasting,” *Robot. Comput.-Integr. Manuf.*, vol. 34, no. August, pp. 151–163, 2015, doi: 10.1016/j.rcim.2014.12.015.
- [14] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “Statistical and machine learning forecasting methods: Concerns and ways forward,” *PLOS ONE*, vol. 13, no. 3, e0194889, 2018, doi: 10.1371/journal.pone.0194889.
- [15] J. Chen, “Advanced analytics for retail inventory and demand forecasting,” *Theor. Econ. Bus. Manag. Res.*, vol. 10, no. October, pp. 1–7, 2024, doi: 10.62051/jme9b319.
- [16] Z. Fu, “Optimizing retail inventory management through time series analysis,” *Theor. Econ. Bus. Manag. Res.*, vol. 2024, no. 1, pp. 1–10, 2024, doi: 10.62051/dy4kaf37.
- [17] P. Ramos and J. M. Oliveira, “A procedure for identification of appropriate state

- space and ARIMA models based on time-series cross-validation,” *Algorithms*, vol. 9, no. 4, p. 76, 2016, doi: 10.3390/a9040076.
- [18] Y. Zhao, “E-commerce demand forecasting using SARIMA model and K-means clustering analysis,” *J. Innov. Develop.*, vol. 7, no. 1, pp. 1–6, 2024, doi: 10.54097/ctfb0379.
- [19] D. P. Abellana, D. M. Rivero, M. E. Aparente, and A. Rivero, “Hybrid SVR–SARIMA model for tourism forecasting using PROMETHEE II as a selection methodology,” *J. Tourism Futures*, vol. 7, no. 1, pp. 78-97, 2020, doi: 10.1108/jtf-07-2019-0070.
- [20] K. Ferreira, B. H. A. Lee, and D. Simchi-Levi, “Analytics for an online retailer: Demand forecasting and price optimization,” *Manuf. Serv. Oper. Manag.*, vol. 18, no. 1, pp. 1–175, 2016, doi: 10.1287/msom.2015.0561.
- [21] X. Lu, “A comparative study of machine learning-based regression models for supply chain management,” *Appl. Comput. Eng.*, vol. 53, no. March, pp. 1–8, 2024, doi: 10.54254/2755-2721/53/20241233.
- [22] Y. M. Tang, K. Y. Chau, Y. Lau, and Z. Zheng, “Data-intensive inventory forecasting with artificial intelligence models for cross-border e-commerce service automation,” *Appl. Sci.*, vol. 13, no. 5, Art. no. 3051, 2023, doi: 10.3390/app13053051.
- [23] A. Wang, Q. Yu, J.-Y. Wang, X. Yu, Z. Wang, and Z. Hu, “Electric load forecasting based on deep ensemble learning,” *Appl. Sci.*, vol. 13, no. 17, Art. no. 9706, 2023, doi: 10.3390/app13179706.
- [24] M. I. Ayub et al., “AI-driven demand forecasting for multi-echelon supply chains,” *Am. J. Manag. Econ. Innov.*, vol. 7, no. 7, pp. 74-85, 2025, doi: 10.37547/tajmei/volume07issue07-09.
- [25] C. Deng and L. Yong-ji, “A deep learning-based inventory management and demand prediction optimization method for anomaly detection,” *Wireless Commun. Mobile Comput.*, vol. 2021, no. 1, pp. 1-14, 2021, doi: 10.1155/2021/9969357.
- [26] Q. Du, “An overview on omnichannel inventory strategy under uncertain demand,” *Am. J. Ind. Bus. Manag.*, vol. 11, no. 5, pp. 523-535, 2021, doi: 10.4236/ajibm.2021.115033.
- [27] Y. Fu, D. Wu, and B. Boulet, “Reinforcement learning based dynamic model combination for time series forecasting,” *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 6, pp. 6639-6647, 2022, doi: 10.1609/aaai.v36i6.20618.
- [28] M. U. Sattar, V. Dattana, R. Hasan, S. Mahmood, H. W. Khan, and S. Hussain, “Enhancing supply chain management: A comparative study of machine learning techniques with cost–accuracy and ESG-based evaluation,” *Sustainability*, vol. 17, no. 13, Art. no. 5772, 2025, doi: 10.3390/su17135772.
- [29] B. Lim and S. Zohren, “Time-series forecasting with deep learning: A survey,” *Phil. Trans. R. Soc. A*, vol. 379, no. 2194, Art. no. 20200209, 2021, doi: 10.1098/rsta.2020.0209.