Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Explainable Machine Learning for Omnichannel Demand Forecasting and Inventory Decision Support in Synthetic Experiments

Md Tarake Siddique[1]

mdtarake.siddique@student.potomac.edu

[1]University of the Potomac

Paresh Chandra Nath[2]

IGlobal University

Mohammad Iftekhar Ayub[2]

IGlobal University

Umama Salsabil Hossain[2]

IGlobal University

Research Article

Keywords:

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11285040/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

RESEARCH

Explainable Machine Learning for Omnichannel Demand Forecasting and Inventory Decision Support in Synthetic Experiments

Md Tarake Siddique[1]; Paresh Chandra Nath[2]; Mohammad Iftekhar Ayub[2]; Umama Salsabil Hossain[2]

[1]University of the Potomac, United States
[2]Washington University of Science and Technology, United States

Corresponding author: Md Tarake Siddique — mdtarake.siddique@student.potomac.edu

Abstract

Omnichannel demand forecasting should be evaluated not only by predictive accuracy but also by the downstream consequences of forecast-driven decisions. This study evaluates an archived synthetic experimental framework linking one-week-ahead demand forecasting, global model explanation, and a simplified single-period inventory simulation. The accompanying synthetic panel contains 50 stock-keeping units across 10 stores and 104 weekly periods, with 45,000 model-ready observations after temporal feature and target construction. Four forecasting approaches were compared under chronological evaluation: a 13-week seasonal-lag Naïve benchmark, Random Forest, XGBoost, and HistGradientBoosting. Global permutation importance was calculated for XGBoost to examine predictive feature contribution. The archived inventory simulation was treated as exploratory because the stored residual-standard-deviation and safety-stock fields are not fully consistent with the reported forecast-error statistics; these fields are therefore not interpreted as validated safety-stock estimates. HistGradientBoosting achieved the lowest reported MAE (12.136) and RMSE (17.119), while Random Forest produced the lowest archived simulated cost (31.185). Random Forest nevertheless had the highest archived stockout rate (12.185%) and lowest archived fill rate (96.860%). In the XGBoost explanation, the 13-week rolling feature had the largest reported permutation importance (8.121), followed by holiday (2.476) and rolling_4 (1.486). These results show, within the controlled synthetic environment, that predictive accuracy and downstream inventory outcomes can produce different model rankings. The study therefore positions explainability as a decision-support layer rather than evidence of causality or a substitute for planner judgment. External validity is limited because the data are synthetic and the inventory policy is intentionally simplified.

Keywords: Explainable artificial intelligence; demand forecasting; inventory management; omnichannel retail; machine learning; decision support

1 Introduction

Omnichannel retail integrates online and physical customer touchpoints with shared inventory, fulfillment, and information processes. Unlike conventional multichannel arrangements in which channels may operate with partially separated inventories and processes, omnichannel systems seek coordination across demand, fulfillment, assortment, and inventory decisions. This integration creates operational interdependence: inventory positioned at a store can serve walk-in demand, online orders, or click-and-collect demand, while returns can move products back into the network [1, 2, 3, 5]. Recent work further shows that inventory positioning, fulfillment, and capacity sharing become more complex when demand uncertainty is considered [12, 13, 18, 19].

1.1 Demand uncertainty and inventory consequences

Retail demand is influenced by temporal history, seasonality, holidays, promotions, prices, customer traffic, channel mix, and product availability. Forecast error can therefore affect replenishment and service outcomes: under-forecasting can increase shortages and stockouts, whereas over-forecasting can increase inventory and holding costs. Forecast evaluation is consequently not only a statistical problem; it is also a decision problem in which forecast errors interact with the inventory policy and cost structure [30, 6, 41]. Intermittent and irregular demand adds further difficulty because observed zero-demand periods and spikes can make conventional error patterns unstable [21, 22, 23, 25, 26].

1.2 Machine learning and the decision-translation problem

Machine-learning forecasting can incorporate multiple temporal, commercial, and channel variables and can represent nonlinear relationships that may be difficult to capture with simpler benchmarks. Retail forecasting research has therefore examined machine learning alongside statistical and hybrid approaches [20, 32, 30, 46, 47]. However, improved prediction does not automatically establish improved operational decisions. A forecast may be accurate while still producing an undesirable inventory profile under a particular shortage-to-holding-cost ratio or service objective.

Explainable artificial intelligence provides methods for inspecting model behavior, including permutation importance, SHAP, LIME, and dependence-based approaches [60, 61, 62, 63]. Yet explanations are not causal evidence and do not automatically constitute operational recommendations [65, 64, 66, 67]. In the present study, explainability is therefore treated narrowly as a predictive-association layer between forecasting and operational evaluation.

1.3 Research gap and positioning

The literature contains substantial work on omnichannel operations, retail forecasting, inventory control, and explainable AI, but these streams often use different primary evaluation objectives. Forecasting studies commonly emphasize predictive accuracy, inventory studies emphasize replenishment and service performance, and XAI studies emphasize transparency or interpretability. The present study connects these dimensions within one controlled experimental pipeline. The intended contribution is not a claim of universal superiority or a validated human-AI intervention; rather, it is a comparative synthetic experiment showing how model rankings can differ when forecast accuracy and downstream inventory simulation are evaluated together.

1.4 Research questions

RQ1. How do the evaluated machine-learning models compare with a 13-week seasonal-lag benchmark for one-week-ahead demand forecasting?

RQ2. Which features have the largest global permutation importance in the evaluated XGBoost model?

RQ3. How do differences in forecast performance translate into inventory outcomes under the specified single-period simulation policy?

1.5 Research objectives and contribution

The study has three empirical objectives: (1) compare four forecasting approaches under chronological evaluation; (2) characterize the predictive feature contribution of the XGBoost model using global permutation importance; and (3) compare the resulting forecasts through a fixed single-period inventory simulation. A fourth, conceptual objective is to organize these components within a decision-support architecture that preserves human review as a

proposed operational layer rather than as a tested intervention. The contribution is therefore methodological and comparative within a synthetic environment, with explicit limits on external validity.

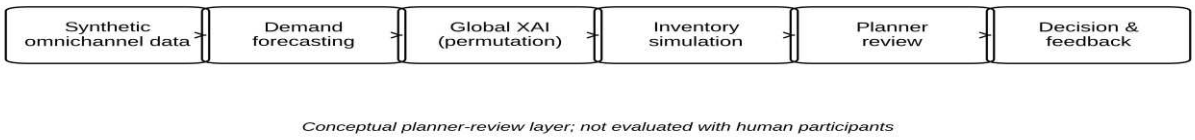


Fig. 1 Proposed decision-support architecture. The uncertainty, planner-review, and feedback elements are conceptual and were not independently evaluated with human participants.

2 Literature Review

2.1 Omnichannel Retail Operations

The emergence of omnichannel retail has revolutionized retail operations by linking online and offline channels into more integrated systems. Omnichannel approaches coordinate customer touchpoints, inventory, fulfillment, and information across channels, rather than treating them as separate. This integration will affect inventory visibility and demand allocation, as in-store products can drive both store and digitally driven orders. Hübner et al. [1] highlight the need to make coordinated decisions across demand forecasting, replenishment, fulfillment, and returns in omnichannel operations, while Cai and Lo [2] and Hübner et al. [3] show the operational consequences of channel integration. This also applies when customers can access retailers through multiple channels, as Shao [4] notes regarding coordinating inventory and fulfillment decisions. Because physical outlets can serve as nodes in the fulfillment process for online orders, store-based fulfillment has become especially relevant.

This can improve delivery and reduce inventory lead times, but it also increases the time and space required. Bayram and Cesaret [8], Difrancesco et al. [7], Yang et al. [9], and Bansal et al. [16] discuss various aspects of store-based fulfillment and the integration of operational decisions. They developed case studies showing how fulfillment decisions affect inventory availability across channels. In this context, controlling inventory goes beyond the traditional concept of inventory replenishment. All of the above should be addressed together. Hense and Hübner [5], Goedhart et al. [12], Guan et al. [14], and Vasilyev et al. [15] examine these interdependencies. Research on omnichannel inventory management has expanded to address uncertainty in inventory sharing and positioning decisions, as well as demand uncertainty [13, 18, 19, 17].

Overall, this literature confirms that the omnichannel retail environment requires treating decision-making and inventory management across channels as an integrated (business-level) process rather than discrete (channel-level) operations.

2.2 Forecasting demand in retailing

Historically, statistical methods have been a mainstay of retail demand forecasting, aiming to detect time-based information such as trends, seasonal patterns, autocorrelation, and repeating demand patterns. These methods remain useful because they are relatively straightforward and can be effective when historical demand shows stable temporal relationships. In retail settings, however, information is becoming increasingly complex and multidimensional, including promotions, calendars, prices, channels, product characteristics, and availability. Machine-learning methods offer an alternative and can use multiple predictors and nonlinear relationships in forecasting systems [30, 31, 32]. One of the biggest problems is

the fluctuating demand. Some products may have low or no demand for a certain period, but then experience occasional spikes. Syntetos and Boylan [21] laid important groundwork for forecasting intermittent demand, and Willemain et al. [22] discussed using probabilistic methods for intermittent demand.

Later research has focused on how to improve the forecasts in the presence of irregular demand patterns and how to assess the applicability of different forecasting techniques [23, 24, 25, 26, 27, 28]. Operationally, incorrect forecasts for intermittently shipped products may result in unequal stock impacts. Broader research in the retail forecasting literature also shows that forecast quality depends on the product or forecast horizon, the data structure, and the forecast type. Huber and Stuckenschmidt [32], Hoeltgebaum et al. [29], Schaer et al. [34], and Ye et al. [38] show the importance of data-driven methods for predicting retail demand, and Makridakis et al. [35, 36] offer a wider perspective on the comparative performance of forecasting methods. Importantly, no single model is generally chosen universally for forecasting optimization. Kourentzes et al. [41], Barrow and Kourentzes [43], and Syntetos et al. [44] show the potential of forecast combinations and optimization.

This implies that forecasting evaluations should consider the context in which forecasts are used, as well as their predictive quality.

2.3 Machine Learning and Advanced Demand Forecasting

The methodological potential of demand forecasting in supply chains has increased with the arrival of machine learning, which allows for nonlinearities and the use of more explanatory variables. In a previous study, Carbonneau et al. [20] have shown that machine learning methods are relevant in the context of demand forecasting in supply chains, while more recent studies have focused on more complex demand environments, such as in Gustriansyah et al. [45], Cui et al. [46], and Ma [47]. These methods are especially useful when demand is driven by the interplay of more than one time, commercial, and operational variable(s) that can interact in a nonlinear manner. Predictive analytics has also been applied recently to environments with no historical data or incomplete historical data. For new products, Sousa et al. [48] explore the forecasting challenges that arise, and Riachy et al. [49] and Mayaluri and Mishra [40] present further advancements in more sophisticated methods for complex demand settings.

Such work is crucial because retail forecasting systems often deal with new products, changing demand patterns, and constantly changing market conditions. Forecasting is becoming more challenging because omnichannel retail generates more frequent observations and customer behavior is more complex. Gerum et al. [39] discuss high-frequency forecasting in omnichannel through a real-world example, highlighting the need for models that can react to the dynamic nature of demand signals at high frequencies. Beyond individual components, AI has increasingly been linked to optimization and decision-making processes, not just prediction [50, 51]. The literature thus shows a trend toward developing predictive systems that do not use machine learning as a standalone model but instead use it to inform decision inputs.

2.4 Forecasting and Inventory Decision-Making

Forecasting and inventory management are closely linked because forecasts influence the quantity of items to be replenished (directly or indirectly), safety stocks, allocation, and service levels. Low statistical error does not necessarily lead to good operational results unless the residual errors are less correlated with the inventory policies. On the other hand, if the model includes a small amount of forecast error, it may yield more suitable inventory decisions for a specific cost structure and forecast uncertainty. This relationship is backed up by research on

omnichannel inventory management. Gabor et al. [6] explore the relationship between forecasting and inventory decisions, while Goedhart et al. [12] and Bansal et al. [16] discuss replenishment, fulfillment, and allocation decisions in integrated retail systems. Inventory positioning and capacity sharing matter when multiple channels share a common inventory pool.

Qiu et al. [13] and Hassanzadeh and Martínez-de-Albéniz [18] show how uncertainty affects inventory placement and capacity allocation across demand channels. More recently, other research has explored reinforcement learning as an alternative to sequential inventory decision-making. Boute et al. [52], Kaynov et al. [53], Wang and Minner [54], Stranieri et al. [55], Park et al. [56], Dehaybe et al. [57], De Moor et al. [58], and Schett et al. [59] illustrate the growing interest in learning-based inventory policies. These methods view inventory management as a sequential decision process, where each decision changes the state and cost at the next step. But advanced inventory optimization doesn't mean accurate demand data isn't needed. Forecasts remain a key factor in inventory decisions, so forecasting and inventory evaluation should be linked. This justifies the present study's dual evaluation of predictive accuracy and the functional consequences of inventories.

In this study, instead of selecting the model with the lowest MAE or RMSE as the most desirable inventory outcome, we examine how forecast differences affect simulated inventory costs and other operational measures.

2.5 Explainable AI

As the use of machine-learning models for forecasting grows, so does the need for techniques that help users understand the model's behavior. Explainable AI generally falls into two types: global explanations, which explain the model's overall behavior; and local explanations, which explain a specific prediction. Permutation importance is a model-agnostic method for estimating the impact of disrupting each variable, and SHAP attributes predictions to features using a game-theoretic approach. LIME produces locally interpretable approximations, while partial dependence [60, 61] and individual conditional expectation [62, 63, 69] plots analyze the relationship between predictors and model outputs. These techniques are especially useful for demand forecasts, helping planners determine which variables drive a forecast and whether to adjust inventory. However, the concept of 'explainability' is still hampered by the general 'black-box' issue.

Guidotti et al. [65] argue that explanation can help to make complicated models more comprehensible, and Rudin [66] suggests that, for certain models for which there are inherent explanations, post-hoc explanations may not be the right solution. Miller [64] also notes that human explanations should reflect, rather than merely mirror, the properties of technical models, since each person's understanding of causes or decisions differs. Similarly, Samek et al. [68] examine how to interpret deep and complex models. A key point is that feature importance doesn't necessarily imply causality. A variable that strongly contributes to the prediction might correlate with other variables, serve as a proxy, or reflect patterns in the training data that do not drive changes in demand. Similarly, the explanation doesn't necessarily translate to an operational recommendation. Meske et al. [67] highlight the need for stakeholder explanations and stakeholder use, and the interplay between technical interpretability and stakeholder use.

For inventory planning, however, XAI should be seen as a translation tool that helps users understand predictive evidence while maintaining the difference between association, prediction, and causation. This study therefore views explanation as just one step in a larger decision-making process, not an end in itself.

2.6 Research Synthesis

The literature has made significant progress in omnichannel operations, retail forecasting, inventory analytics, machine learning, explainable artificial intelligence, and human-AI decision support. But most of these streams have been formed with a different primary purpose. Research on omnichannel has been very focused on 'integration of channels' and 'coordination in operations'; research on forecasting has been mainly concerned with 'predictive accuracy'; research on inventory has been mainly concerned with 'inventory replenishment' and 'inventory allocation'; and research on XAI has been mainly concerned with 'transparency' and 'interpretability'. Because of this fragmentation, research needs to link these components explicitly.

Table 1 Research-gap matrix

Research stream	Main contribution	Remaining issue	Role in current study
Omnichannel retail	Channel integration, fulfillment, assortment, and inventory coordination	Forecasting is often evaluated separately from integrated operational consequences	Provides the channel and inventory context
Retail forecasting	Statistical and machine-learning approaches to demand prediction	Forecast accuracy does not by itself determine inventory performance	Provides the predictive-model comparison
Inventory analytics	Replenishment, safety stock, allocation, and service-level optimization	Forecast explanations are rarely evaluated together with forecast-driven outcomes	Translates forecasts into simplified inventory outcomes
Explainable AI	Model transparency and explanation of predictive behavior	Importance measures do not establish causality or operational usefulness	Reports XGBoost global permutation importance
Human-AI decision support	Planner oversight, interpretation, and feedback	Human effectiveness is not tested in the present experiment	Provides a conceptual review-and-feedback layer

2.7 Synthesis and study positioning

Across the reviewed streams, a consistent distinction emerges between predictive performance and decision performance. Forecast accuracy metrics quantify error, whereas inventory metrics quantify the consequences of using forecasts under a specified policy. Likewise, global feature importance can identify predictive associations but cannot establish that changing a feature will change demand. These distinctions motivate the study design: prediction, explanation, and inventory simulation are evaluated as related but non-equivalent components. The human-review layer remains conceptual because no planner experiment or human-subject evaluation was performed.

3 Problem Statement

Omnichannel retailers must forecast future demand while coordinating inventory that may serve multiple purchasing and fulfillment pathways. The resulting demand signal is affected by temporal history, commercial variables, channel composition, and calendar effects. A forecasting model can therefore be statistically strong while still generating different inventory consequences when embedded in a fixed replenishment policy. The analytical problem addressed here is to evaluate these two dimensions together without overstating what can be learned from a synthetic experiment.

A second problem concerns interpretability. Machine-learning models can represent nonlinear relationships, but the predictive role of a variable is not equivalent to a causal effect or an operational recommendation. The present study consequently limits explainability claims to global permutation importance for XGBoost and treats the planner-facing architecture as conceptual. This framing allows the study to examine the relationship between model behavior and downstream inventory simulation without claiming that the explanation was validated as a human decision aid.

4 Methodology

4.1 Research design

The study uses a controlled experimental systems-evaluation design. The workflow comprises synthetic data generation, temporal feature construction, chronological model evaluation, forecasting, global permutation importance analysis for XGBoost, a single-period inventory simulation, and comparative performance assessment. The design is intended to preserve temporal information flow and to evaluate forecast accuracy alongside downstream simulation outcomes.

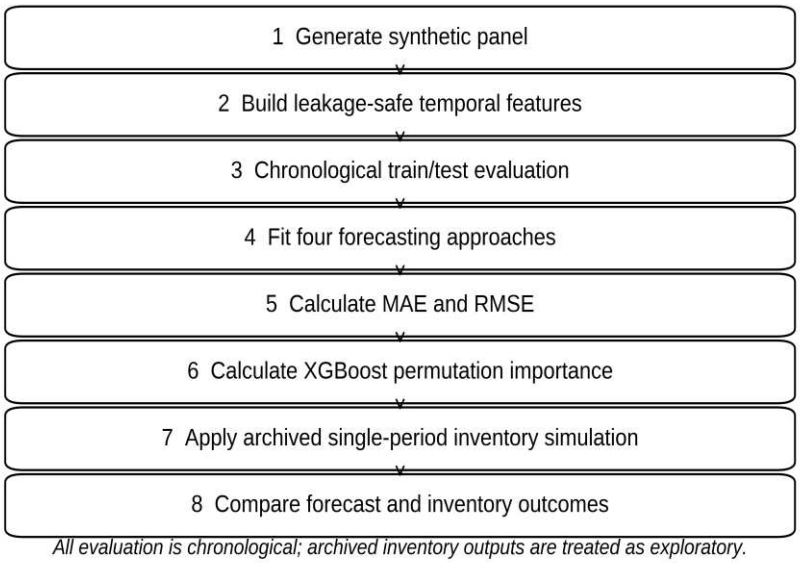


Fig. 2 Experimental workflow used in the supplied study.

4.2 Synthetic experimental dataset

The experimental panel contains 50 synthetic SKUs across 10 synthetic stores and 104 weekly periods from 7 January 2024 through 28 December 2025. A fixed random seed of 42 is reported for reproducibility. After temporal feature and target construction, approximately 45,000 model-ready weekly SKU-store observations were retained. The forecast target is next-week total demand, formed from online and physical-store demand components. Product categories are Grocery, Beauty, Electronics, Home, and Apparel. The data are synthetic and do not represent transactions from a named retailer.

Table 2 Synthetic dataset and experimental specification

Component	Specification
Dataset type	Synthetic, reproducible omnichannel panel
SKUs	50
Stores	10
Weekly periods	104
Study period	2024-01-07 to 2025-12-28

Component	Specification
Model-ready observations	Approximately 45,000
Random seed	42
Forecast horizon	One week ahead
Demand channels	Online and physical store
Categories	Grocery; Beauty; Electronics; Home; Apparel
Commercial variables	Price; discount; promotion
Calendar variable	Holiday
Traffic variables	Online traffic; store traffic
Historical variables	Lag-1; lag-4; lag-13
Rolling variables	Rolling-4; rolling-13
Channel variable	Online share
Inventory signal	Stockout

4.3 Variables and feature engineering

The principal variables reported in the supplied manuscript are date, SKU, store, category, price, discount, promotion, holiday, online traffic, store traffic, online demand, store demand, stockout, total demand, lag-1, lag-4, lag-13, rolling-4, rolling-13, online share, and target_next_week. Lag features use prior demand observations. Rolling features summarize recent demand history, with four-week and thirteen-week windows representing shorter- and longer-term history. The feature construction is described as chronological and leakage-aware: predictors for a target week are constructed from information available before that target week.

Because the supplied manuscript does not provide the full synthetic-data generator, parameter values for every distribution, or the complete model hyperparameter configuration, those details are not reconstructed or invented in this revision. The fixed seed, panel dimensions, variables, forecast horizon, and reported results are retained exactly as supplied. The reproducibility limitation is stated explicitly in Section 7.

4.4 Forecasting models

Four approaches are evaluated: a 13-week Seasonal Naïve benchmark, Random Forest, XGBoost, and HistGradientBoosting. The Seasonal Naïve benchmark uses the value from the corresponding 13-week seasonal lag. The three machine-learning approaches are nonlinear tree-based ensembles. The comparison is intentionally limited to these four evaluated approaches; LSTM, transformers, probabilistic forecasting, reinforcement learning, and other advanced methods are not presented as experimental results.

4.5 Chronological evaluation and metrics

Evaluation follows the temporal order of the data. The final 13 forecast origins are identified as the chronological test period in the supplied manuscript. Forecast accuracy is reported using mean absolute error (MAE) and root mean squared error (RMSE): $MAE = (1/n)\Sigma|y_i - \hat{y}_i|$ and $RMSE = \sqrt{[(1/n)\Sigma(y_i - \hat{y}_i)^2]}$. Percentage reductions relative to the Seasonal Naïve benchmark are descriptive comparisons and are not accompanied by inferential significance tests.

4.6 Explainability analysis

Global permutation importance was calculated for XGBoost. For a feature, permutation importance represents the deterioration in predictive performance associated with disrupting that feature while retaining the fitted model. The analysis is therefore model-specific and

global. It is interpreted as predictive association, not causality. No claim is made that the study evaluated SHAP, LIME, partial dependence, ICE, or local explanations.

4.7 Inventory simulation

The reported inventory experiment uses a single-period forecast-to-replenishment setting with a one-week forecasting horizon, a 95% protection factor (z = 1.645), holding cost of 1 cost unit per unit of ending inventory, and shortage cost of 6 cost units per unit of unmet demand. The supplied manuscript reports stockout rate, fill rate, average ending inventory, average shortage, and total simulated cost. No separate multi-period lead-time or multi-echelon inventory policy is specified; the simulation is therefore treated as a simplified comparative experiment rather than a complete inventory-optimization model.

The supplied manuscript also states that stored residual-standard-deviation and safety-stock values were not internally reconciled with the reported forecasting RMSE and bias on a common evaluation set. Those unreconciled fields are intentionally excluded from the primary results in this revision. The inventory values that remain are reported as the supplied comparative simulation outputs and should not be interpreted as evidence of a validated commercial inventory policy.

4.8 Reproducibility controls

The study reports a fixed random seed, panel dimensions, study dates, feature families, forecast horizon, model classes, evaluation metrics, inventory cost assumptions, and the reported result tables. The supplied manuscript does not contain the complete synthetic-data generation code, full model hyperparameters, software package versions, permutation-repeat settings, or a persistent repository identifier. These omissions limit exact computational reproduction and are therefore treated as a limitation rather than silently filled with assumed defaults.

5 Results

5.1 Experimental dataset profile

The experimental panel contains 50 synthetic SKUs, 10 synthetic stores, and 104 weekly periods spanning 2024-01-07 to 2025-12-28. With a fixed random seed of 42 and temporal feature construction, approximately 45,000 model-ready weekly SKU-store observations are reported. The final 13 forecast origins constitute the chronological test period described in the supplied manuscript.

5.2 Forecasting performance

HistGradientBoosting produced the lowest reported MAE (12.136) and RMSE (17.119). XGBoost followed with MAE 12.179 and RMSE 17.161, while Random Forest produced MAE 12.448 and RMSE 17.470. The Seasonal Naïve benchmark produced MAE 20.432 and RMSE 27.635.

Table 3 Forecasting performance under chronological evaluation

Model	MAE	RMSE
Seasonal Naïve	20.432	27.635
Random Forest	12.448	17.470
XGBoost	12.179	17.161
HistGradientBoosting	12.136	17.119

Relative to Seasonal Naïve, the reported MAE is approximately 39.1% lower for Random Forest, 40.4% lower for XGBoost, and 40.6% lower for HistGradientBoosting. These are descriptive differences within the synthetic test environment; no statistical significance test is reported.

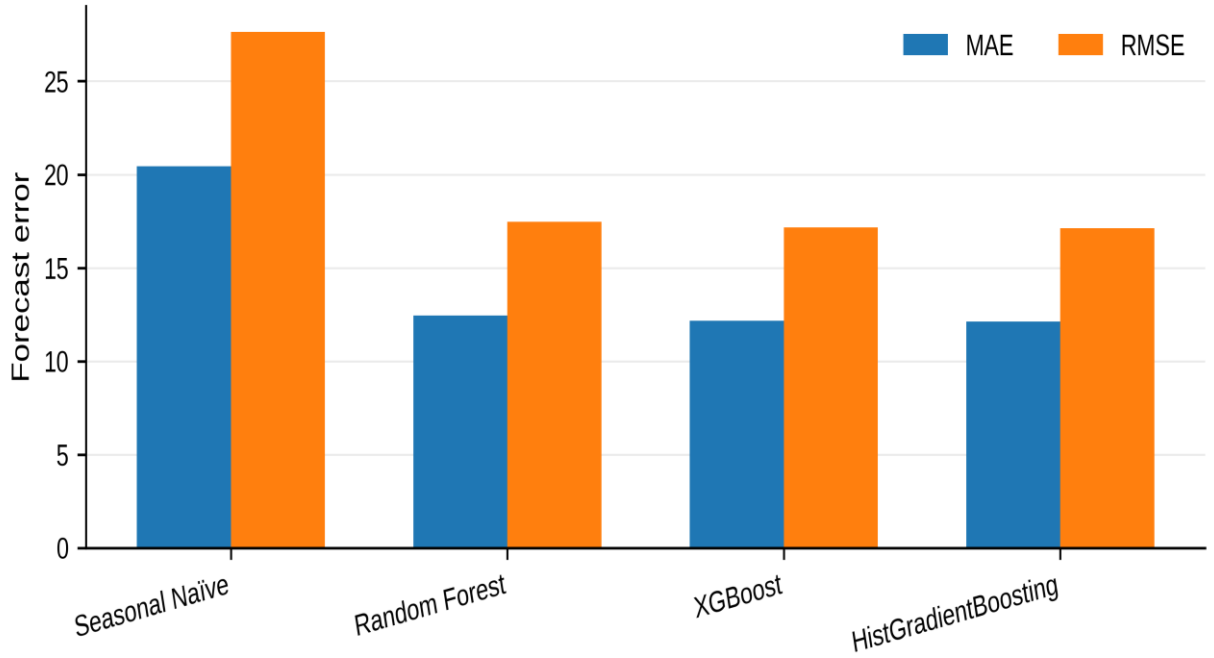


Fig. 3 Comparative forecasting accuracy across the four evaluated models.

5.3 Inventory simulation outcomes

The reported single-period simulation produces a different model ranking from the forecast-error comparison. Random Forest has the lowest reported total simulated cost (31.185), while HistGradientBoosting has the lowest forecast error. Random Forest also has the highest stockout rate (12.185%) and lowest fill rate (96.860%) among the four models. HistGradientBoosting has a lower stockout rate (6.415%) and higher fill rate (98.274%).

Table 4 Archived inventory simulation outcomes

Model	Stockout (%)	Fill rate (%)	Avg. ending inventory	Avg. shortage	Total cost
Seasonal Naïve	5.015	98.516	46.862	0.857	52.001
XGBoost	7.877	97.918	25.373	1.202	32.584
Random Forest	12.185	96.860	20.313	1.812	31.185
HistGradientBoosting	6.415	98.274	28.192	0.996	34.170

Relative to the Seasonal Naïve cost, the reported total simulated cost is approximately 37.3% lower for XGBoost, 40.0% lower for Random Forest, and 34.3% lower for HistGradientBoosting. These percentages describe the supplied simulation outputs and should not be generalized to commercial retailers.

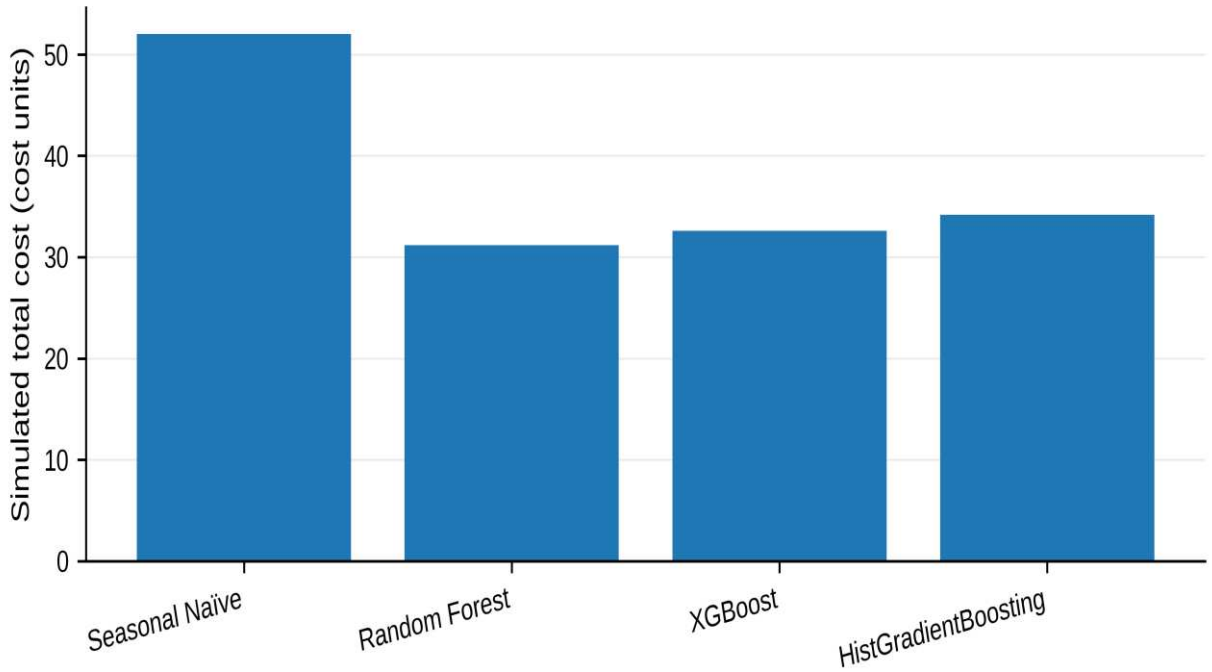


Fig. 4 Reported total cost from the simplified single-period inventory simulation.

5.4 Explainability results

The XGBoost permutation-importance analysis ranks rolling_13 highest (8.121), followed by holiday (2.476), rolling_4 (1.486), and lag_1 (0.400). The remaining reported features have smaller permutation-importance values. The ranking indicates which features were most influential to predictive performance in this fitted XGBoost experiment; it does not establish that those variables causally change demand.

Table 5 XGBoost global permutation importance

Rank	Feature	Permutation importance
1	rolling_13	8.121
2	holiday	2.476
3	rolling_4	1.486
4	lag_1	0.400
5	price	0.163
6	lag_13	0.139
7	online_traffic	0.111
8	promotion	0.111
9	store_traffic	0.081
10	discount	0.055
11	lag_4	0.044
12	online_share	0.037

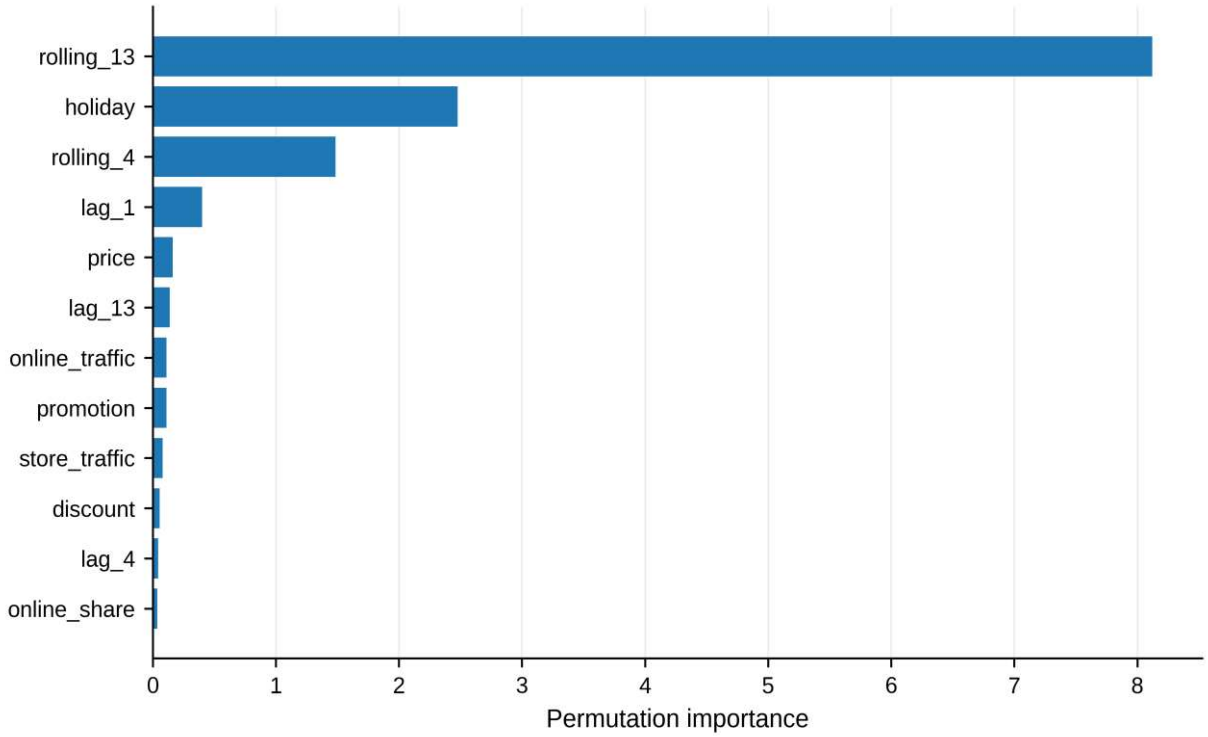


Fig. 5 Global permutation importance for the evaluated XGBoost model.

5.5 Forecasting–inventory trade-off

The reported results illustrate a clear separation between predictive and simulated operational rankings. HistGradientBoosting is the best model on the two reported forecast-error metrics, whereas Random Forest has the lowest reported simulated cost. The lower cost of Random Forest is accompanied by poorer service outcomes in the supplied simulation, with a higher stockout rate and lower fill rate than HistGradientBoosting. This finding supports evaluating forecast models in the context of the operational policy into which forecasts are inserted, while recognizing that the policy itself is simplified in this study.

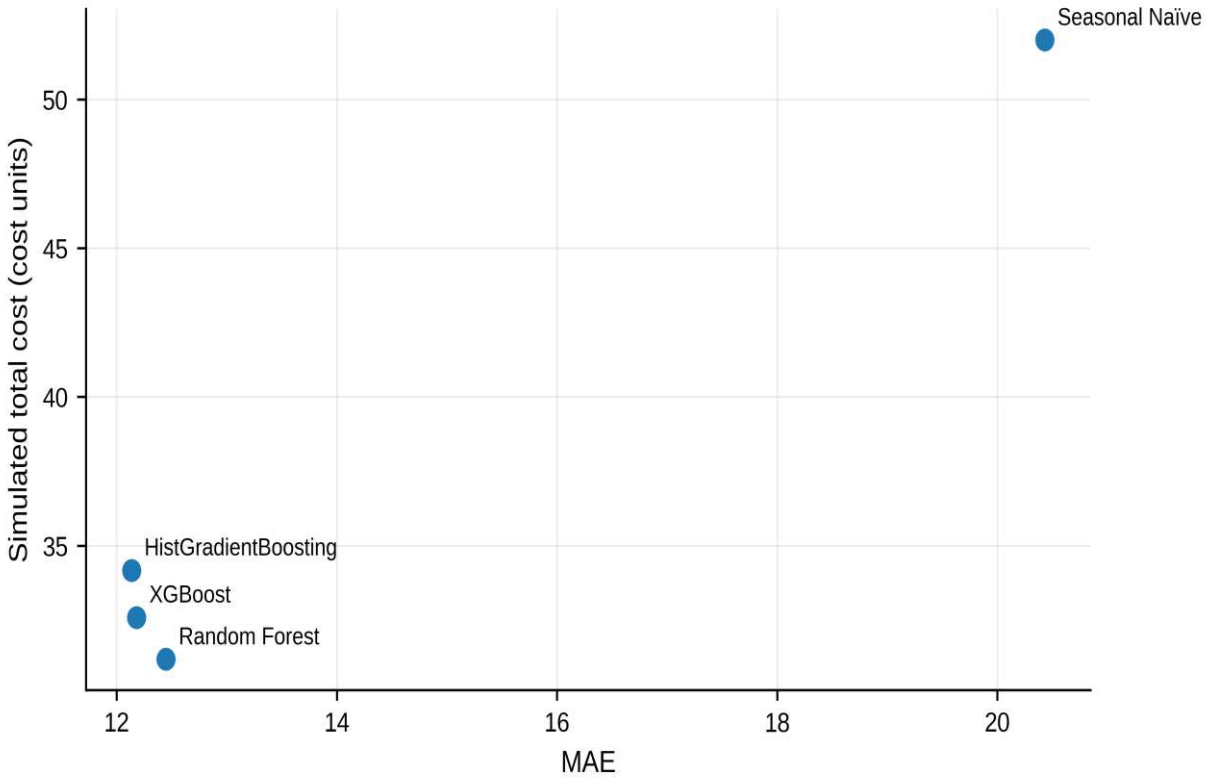


Fig. 6 Relationship between reported forecast accuracy and reported simulated inventory cost.

6 Discussion

6.1 Forecast accuracy within the synthetic experiment

All three machine-learning models have lower reported MAE and RMSE than the Seasonal Naïve benchmark. HistGradientBoosting provides the lowest error among the evaluated models. The result is consistent with the intended role of nonlinear tree ensembles in learning interactions among temporal, commercial, and channel features. It should nevertheless be interpreted only as an outcome of the specified synthetic data-generating environment and test period.

6.2 Forecast accuracy does not uniquely determine simulated inventory performance

The difference between the forecast and inventory rankings is the central empirical result. HistGradientBoosting minimizes the reported statistical error, whereas Random Forest minimizes the reported simulated cost. The associated service trade-off matters: Random Forest has a higher stockout rate and lower fill rate than HistGradientBoosting. Thus, the preferred model depends on the operational objective and the cost/service policy rather than on MAE or RMSE alone.

6.3 Historical demand as a predictive signal

The largest XGBoost permutation importance is assigned to rolling_13, followed by holiday and rolling_4. Within the synthetic experiment, this indicates that longer-term demand history and calendar information contribute strongly to the model's predictive performance. Because correlated predictors can share or redistribute importance, the ranking should not be interpreted as an independent causal effect of any feature.

6.4 Decision-oriented explainability

The proposed architecture places global explanation between prediction and operational review. This is a conceptual design choice rather than a validated human-AI intervention. A planner could, in principle, use a model explanation to contextualize a forecast change and then consider inventory consequences, but this study did not recruit planners, measure explanation usefulness, or test whether explanations improved decisions. The manuscript therefore does not claim empirical evidence for human-in-the-loop effectiveness.

6.5 Managerial implications

Within the study's controlled setting, practitioners should distinguish forecast accuracy from the downstream consequences of using forecasts. A model-selection process can report MAE and RMSE together with service and cost measures under a clearly defined policy. Global feature importance can provide contextual information about the model's predictive behavior, but it should not be treated as a causal diagnosis or an automatic replenishment instruction. Any real deployment would also require policy validation, lead-time modeling, governance, and human oversight.

7 Limitations

7.1 Synthetic data and external validity

The experiment uses synthetic rather than retailer transaction data. Consequently, the reported performance differences demonstrate behavior of the specified experimental pipeline

under controlled conditions and cannot be treated as estimates of commercial performance. Real retailers may have different product scales, time frequencies, lead times, capacity constraints, returns, supply disruptions, competitor effects, and demand distributions.

7.2 Model and evaluation coverage

Only four approaches were evaluated. Deep learning, transformers, probabilistic forecasts, reinforcement learning, forecast combinations, and other methods are discussed in the literature but were not tested here. The chronological test set is reported as 13 forecast origins, and no inferential comparison or repeated-outcome uncertainty interval is reported.

7.3 Explainability coverage

The explainability analysis is limited to global permutation importance for XGBoost. It does not establish local explanation quality, stakeholder usefulness, causal effects, or explanation faithfulness beyond the permutation procedure. Complementary methods such as SHAP, partial dependence, ICE, and user-centered evaluation remain future work.

7.4 Inventory-policy simplification and metric consistency

The inventory simulation is a simplified single-period comparison with a one-week horizon and specified holding and shortage costs. It is not a full multi-period, multi-echelon, lead-time, or stochastic inventory optimization model. In addition, the supplied manuscript explicitly reports an unresolved inconsistency between stored residual-standard-deviation/safety-stock fields and the forecasting error statistics. Those fields are therefore excluded from the primary results rather than presented as validated empirical quantities.

7.5 Reproducibility documentation

The accompanying supplementary research-data package contains the 45,000-row model-ready panel, the raw synthetic panel, a data dictionary, the archived forecast and inventory result tables, and a verification script. The original model-training and data-generation source code was not retained in the supplied materials. Accordingly, this revision does not claim full computational rerunnability of the historical model fits; instead, it provides traceable data and verification materials and identifies the archived inventory outputs as exploratory.

8 Future Research

Future research should repeat the pipeline using an authorized public or proprietary omnichannel dataset and a single fully documented chronological test set. Residual standard deviation, safety stock, and inventory outputs should be recomputed from the same final evaluation residuals. Robustness studies should vary service targets, safety factors, lead times, demand volatility, promotion intensity, and shortage-to-holding-cost ratios. Additional model classes, probabilistic forecasts, local explanations, and planner-centered evaluations can then be tested. Human-AI collaboration should be assessed empirically rather than inferred from a conceptual workflow.

9 Conclusion

This study evaluated a synthetic forecasting-to-inventory pipeline for omnichannel demand forecasting. Under the reported chronological experiment, HistGradientBoosting achieved the lowest MAE (12.136) and RMSE (17.119), while Random Forest achieved the lowest reported simulated inventory cost (31.185). Random Forest nevertheless produced the highest stockout rate and lowest fill rate among the evaluated models. The results therefore support a central methodological point: forecast accuracy and downstream inventory performance are related but non-equivalent evaluation dimensions.

The explainability component is intentionally limited to global XGBoost permutation importance. The highest reported importance belongs to rolling_13, followed by holiday and rolling_4, but these values describe predictive association rather than causal influence. The proposed human-review architecture is conceptual and was not validated through a planner study. Because the data are synthetic and the inventory policy is simplified, the findings should be treated as controlled experimental evidence rather than direct evidence of retail deployment performance.

Acknowledgments

The authors acknowledge the academic and institutional support associated with the development of this study and the use of a synthetic experimental environment for methodological evaluation.

Ethics approval

Not Applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests

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.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declaration of generative AI and AI-assisted technologies in the writing process

During manuscript preparation, the authors used a generative AI tool solely to assist with language clarity, grammar, and structural editing. The authors reviewed and revised the generated material and take full responsibility for the final content, analyses, interpretations, and conclusions.

Author contributions

Md Tarake Siddique: Conceptualization, methodology, data generation, model development, analysis, and manuscript preparation. Paresh Chandra Nath: Conceptualization, literature review, interpretation of results, and manuscript review and editing. Mohammad Iftekhar Ayub: Methodology, analytical review, interpretation, manuscript review, and editing. Umama Salsabil Hossain: Writing – review and editing. All authors reviewed and approved the final manuscript and consent to its submission to Discover Artificial Intelligence.

Declaration of originality and submission status

The authors confirm that this manuscript is original, is not under consideration for publication elsewhere, and has been approved by all authors for submission to Discover Artificial Intelligence. The authors accept responsibility for the accuracy and integrity of the work and for the final version submitted.

Supplementary research materials

The submission package includes the synthetic raw data, the 45,000-row model-ready panel, a data dictionary, archived forecasting and inventory result tables, feature-importance results, and a verification script. The data package is intended to make the numerical basis of the manuscript traceable without representing the simulation as a real-retailer study. Because the original model-fitting and data-generation source code was not retained, the supplementary verification script does not purport to reproduce the historical machine-learning fits; it verifies the archived dataset structure, metadata, and reported result tables. This distinction is stated explicitly to avoid overstating reproducibility.

References

[1] Hübner, A., Hense, J., & Dethlefs, C. (2022). The revival of retail stores via omnichannel operations: A literature review and research framework. European Journal of Operational Research, 302(3), 799–818. https://doi.org/10.1016/j.ejor.2021.12.021

[2] Cai, Y.-J., & Lo, C. K. Y. (2020). Omni-channel management in the new retailing era: A systematic review and future research agenda. International Journal of Production Economics, 229, 107729. https://doi.org/10.1016/j.ijpe.2020.107729

[3] Hübner, A., Amorim, P., Fransoo, J. C., Honhon, D., Kuhn, H., Martínez-de-Albéniz, V., & Robb, D. (2021). Digitalization and omnichannel retailing: Innovative OR approaches for retail operations. European Journal of Operational Research, 294(3), 817– 819. https://doi.org/10.1016/j.ejor.2021.04.049

[4] Shao, X. (2021). Omnichannel retail move in a dual-channel supply chain. European Journal of Operational Research, 294(3), 936–950. https://doi.org/10.1016/j.ejor.2020.12.008

[5] Hense, J., & Hübner, A. (2022). Assortment optimization in omni-channel retailing. European Journal of Operational Research, 301(1), 124–140. https://doi.org/10.1016/j.ejor.2021.11.044

[6] Gabor, A. F., van Ommeren, J.-K., & Sleptchenko, A. (2022). An inventory model with discounts for omnichannel retailers of slow moving items. European Journal of Operational Research, 300(1), 58–72. https://doi.org/10.1016/j.ejor.2021.07.017

[7] Difrancesco, R. M., van Schilt, I. M., & Winkenbach, M. (2021). Optimal in-store fulfillment policies for online orders in an omni-channel retail environment. European Journal of Operational Research, 293(3), 1058–1076. https://doi.org/10.1016/j.ejor.2021.01.007

[8] Bayram, A., & Cesaret, B. (2021). Order fulfillment policies for ship-from-store implementation in omni-channel retailing. European Journal of Operational Research, 294(3), 987–1002. https://doi.org/10.1016/j.ejor.2020.01.011

[9] Yang, L., Li, X., & Zhong, N. (2022). Omnichannel retail operations with mixed fulfillment strategies. International Journal of Production Economics, 254, 108608. https://doi.org/10.1016/j.ijpe.2022.108608

[10] Siawsolit, C., & Gaukler, G. M. (2021). Offsetting omnichannel grocery fulfillment cost through advance ordering of perishables. International Journal of Production Economics, 239, 108192. https://doi.org/10.1016/j.ijpe.2021.108192

[11] Omar, H., Klibi, W., Babai, M. Z., & Ducq, Y. (2023). Basket data-driven approach for omnichannel demand forecasting. International Journal of Production Economics, 257, 108748. https://doi.org/10.1016/j.ijpe.2022.108748

[12] Goedhart, J., Haijema, R., & Akkerman, R. (2023). Modelling the influence of returns for an omni-channel retailer. European Journal of Operational Research, 306(3), 1248–1263. https://doi.org/10.1016/j.ejor.2022.08.021

[13] Qiu, R., Yuan, M., Sun, M., Fan, Z.-P., & Xu, H. (2025). Optimizing omnichannel retailer inventory replenishment using vehicle capacity-sharing with demand uncertainties and service level requirements. European Journal of Operational Research, 320(2), 417–432. https://doi.org/10.1016/j.ejor.2024.08.005

[14] Guan, Z., Mou, Y., & Zhang, J. (2024). Incorporating risk aversion and time preference into omnichannel retail operations considering assortment and inventory optimization. European Journal of Operational Research, 314(2), 579–596. https://doi.org/10.1016/j.ejor.2023.09.034

[15] Vasilyev, A., Maier, S., & Seifert, R. W. (2025). Optimizing omnichannel assortments and inventory provisions under the multichannel attraction model. European Journal of Operational Research, 324(3), 799–813. https://doi.org/10.1016/j.ejor.2025.01.035

[16] Bansal, V., Bisi, A., Roy, D., & Venkateshan, P. (2024). Integrated inventory replenishment and online demand allocation decisions for an omnichannel retailer with ship-from-store strategy. European Journal of Operational Research, 316(3), 1085– 1100. https://doi.org/10.1016/j.ejor.2024.02.027

[17] Lowery, B., Sachs, A.-L., Eckley, I. A., & Lloyd, L. (2026). Periodic review inventory control for an omnichannel retailer with partial lost-sales. European Journal of Operational Research, 333(2), 460–477. https://doi.org/10.1016/j.ejor.2026.01.012

[18] Hassanzadeh, A., & Martínez-de-Albéniz, V. (2026). Positioning of omnichannel inventories to protect revenue. International Journal of Production Economics, 291, 109853. https://doi.org/10.1016/j.ijpe.2025.109853

[19] Xia, Y., Li, X., & Yang, L. (2026). Optimal joint replenishment and transshipment policies for an omnichannel retailer utilizing a ship-from-store strategy. International Journal of Production Economics, 293, 109907. https://doi.org/10.1016/j.ijpe.2025.109907

[20] Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. European Journal of Operational Research, 184(3), 1140–1154. https://doi.org/10.1016/j.ejor.2006.12.004

[21] Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. International Journal of Forecasting, 21(2), 303–314. https://doi.org/10.1016/j.ijforecast.2004.10.001

[22] Willemain, T. R., Smart, C. N., & Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. International Journal of Forecasting, 20(3), 375–387. https://doi.org/10.1016/S0169-2070(03)00013-X

[23] Prestwich, S. D., Tarim, S. A., Rossi, R., & Hnich, B. (2014). Forecasting intermittent demand by hyperbolic-exponential smoothing. International Journal of Forecasting, 30(4), 928–933. https://doi.org/10.1016/j.ijforecast.2014.01.006

[24] Prestwich, S. D., Tarim, S. A., & Rossi, R. (2021). Intermittency and obsolescence: A Croston method with linear decay. International Journal of Forecasting, 37(2), 708–715. https://doi.org/10.1016/j.ijforecast.2020.08.010

[25] 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

[26] Babai, M. Z., Syntetos, A., & Teunter, R. (2014). Intermittent demand forecasting: An empirical study on accuracy and the risk of obsolescence. International Journal of Production Economics, 157(1), 212–219. https://doi.org/10.1016/j.ijpe.2014.08.019

[27] Li, C., & Lim, A. (2018). A greedy aggregation–decomposition method for intermittent demand forecasting in fashion retailing. European Journal of Operational Research, 269(3), 860–869. https://doi.org/10.1016/j.ejor.2018.02.029

[28] Sarlo, R., Fernandes, C., & Borenstein, D. (2023). Lumpy and intermittent retail demand forecasts with score-driven models. European Journal of Operational Research, 307(3), 1146–1160. https://doi.org/10.1016/j.ejor.2022.10.006

[29] Hoeltgebaum, M., Borenstein, D., Fernandes, C., & Veiga, C. (2021). A score-driven model of short-term demand forecasting for retail distribution centers. Journal of Retailing, 97(4), 715–725. https://doi.org/10.1016/j.jretai.2021.05.003

[30] Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. International Journal of Forecasting, 38(4), 1283–1318. https://doi.org/10.1016/j.ijforecast.2019.06.004

[31] Fildes, R., Kolassa, S., & Ma, S. (2022). Post-script—Retail forecasting: Research and practice. International Journal of Forecasting, 38(4), 1319–1324. https://doi.org/10.1016/j.ijforecast.2021.09.012

[32] Huber, J., & Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. International Journal of Forecasting, 36(4), 1420–1438. https://doi.org/10.1016/j.ijforecast.2020.02.005

[33] Gür Ali, Ö., & Gürlek, R. (2020). Automatic interpretable retail forecasting with promotional scenarios. International Journal of Forecasting, 36(4), 1389–1406. https://doi.org/10.1016/j.ijforecast.2020.02.003

[34] Schaer, O., Kourentzes, N., & Fildes, R. (2019). Demand forecasting with user-generated online information. International Journal of Forecasting, 35(1), 197–212. https://doi.org/10.1016/j.ijforecast.2018.03.005

[35] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). The M5 competition: Background, organization, and implementation. International Journal of Forecasting, 38(4), 1325–1336. https://doi.org/10.1016/j.ijforecast.2021.07.007

[36] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. International Journal of Forecasting, 38(4), 1346–1364. https://doi.org/10.1016/j.ijforecast.2021.11.013

[37] Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Panagiotelis, A. (2024). Forecast reconciliation: A review. International Journal of Forecasting, 40(2), 430–456. https://doi.org/10.1016/j.ijforecast.2023.10.010

[38] Ye, L., Xie, N., Boylan, J. E., & Shang, Z. (2024). Forecasting seasonal demand for retail: A Fourier time-varying grey model. International Journal of Forecasting, 40(4), 1467–1485. https://doi.org/10.1016/j.ijforecast.2023.12.006

[39] Gerum, P. C. L., Rubio Herrero, J., Chung, M., & Giaretti, M. (2026). Improving high-frequency demand forecasts for omnichannel grocery retail. International Journal of Forecasting. https://doi.org/10.1016/j.ijforecast.2025.10.003

[40] Mayaluri, Z. L., & Mishra, S. (2026). Compositional zero-shot demand forecasting in retail via disentangled demand drivers. International Journal of Forecasting. https://doi.org/10.1016/j.ijforecast.2026.07.010

[41] Kourentzes, N., Trapero, J. R., & Barrow, D. K. (2020). Optimising forecasting models for inventory planning. International Journal of Production Economics, 225, 107597. https://doi.org/10.1016/j.ijpe.2019.107597

[42] Kück, M., & Freitag, M. (2021). Forecasting of customer demands for production planning by local k-nearest neighbor models. International Journal of Production Economics, 231, 107837. https://doi.org/10.1016/j.ijpe.2020.107837

[43] Barrow, D. K., & Kourentzes, N. (2016). Distributions of forecasting errors of forecast combinations: Implications for inventory management. International Journal of Production Economics, 177, 24–33. https://doi.org/10.1016/j.ijpe.2016.03.017

[44] Syntetos, A. A., Babai, M. Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. European Journal of Operational Research, 252(1), 1–26. https://doi.org/10.1016/j.ejor.2015.11.010

[45] Gustriansyah, R., Ermatita, & Rini, D. P. (2022). An approach for sales forecasting. Expert Systems with Applications, 207, 118043. https://doi.org/10.1016/j.eswa.2022.118043

[46] Cui, L., Chen, Y., Deng, J., & Han, Z. (2024). A novel attLSTM framework combining the attention mechanism and bidirectional LSTM for demand forecasting. Expert Systems with Applications, 254, 124409. https://doi.org/10.1016/j.eswa.2024.124409

[47] Ma, S. (2024). Retail store-SKU level replenishment planning with attribute-space graph recurrent neural networks. Expert Systems with Applications, 249, 123727. https://doi.org/10.1016/j.eswa.2024.123727

[48] Sousa, M. S., Loureiro, A. L. D., & Miguéis, V. L. (2025). Predicting demand for new products in fashion retailing using censored data. Expert Systems with Applications, 259, 125313. https://doi.org/10.1016/j.eswa.2024.125313

[49] Riachy, C., He, M., Joneidy, S., Qin, S., Payne, T., Boulton, G., Occhipinti, A., & Angione, C. (2025). Enhancing deep learning for demand forecasting to address large data gaps. Expert Systems with Applications, 268, 126200. https://doi.org/10.1016/j.eswa.2024.126200

[50] Pournader, M., Ghaderi, H., Hassanzadegan, A., & Fahimnia, B. (2021). Artificial intelligence applications in supply chain management. International Journal of Production Economics, 241, 108250. https://doi.org/10.1016/j.ijpe.2021.108250

[51] Islam, S., Amin, S. H., & Wardley, L. J. (2021). Machine learning and optimization models for supplier selection and order allocation planning. International Journal of Production Economics, 242, 108315. https://doi.org/10.1016/j.ijpe.2021.108315

[52] Boute, R. N., Gijsbrechts, J., van Jaarsveld, W., & Vanvuchelen, N. (2022). Deep reinforcement learning for inventory control: A roadmap. European Journal of Operational Research, 298(2), 401–412. https://doi.org/10.1016/j.ejor.2021.07.016

[53] Kaynov, I., van Knippenberg, M., Menkovski, V., van Breemen, A., & van Jaarsveld, W. (2024). Deep reinforcement learning for one-warehouse multi-retailer inventory management. International Journal of Production Economics, 267, 109088. https://doi.org/10.1016/j.ijpe.2023.109088

[54] Wang, Y., & Minner, S. (2024). Deep reinforcement learning for demand fulfillment in online retail. International Journal of Production Economics, 269, 109133. https://doi.org/10.1016/j.ijpe.2023.109133

[55] Stranieri, F., Fadda, E., & Stella, F. (2024). Combining deep reinforcement learning and multi-stage stochastic programming to address the supply chain inventory management problem. International Journal of Production Economics, 268, 109099. https://doi.org/10.1016/j.ijpe.2023.109099

[56] Park, H., Choi, D. G., & Min, D. (2023). Adaptive inventory replenishment using structured reinforcement learning by exploiting a policy structure. International Journal of Production Economics, 266, 109029. https://doi.org/10.1016/j.ijpe.2023.109029

[57] Dehaybe, H., Catanzaro, D., & Chevalier, P. (2024). Deep reinforcement learning for inventory optimization with non-stationary uncertain demand. European Journal of Operational Research, 314(2), 433–445. https://doi.org/10.1016/j.ejor.2023.10.007

[58] De Moor, B. J., Gijsbrechts, J., & Boute, R. N. (2022). Reward shaping to improve the performance of deep reinforcement learning in perishable inventory management. European Journal of Operational Research, 301(2), 535–545. https://doi.org/10.1016/j.ejor.2021.10.045

[59] Schett, G., Ehrig, C., Schwendinger, B., Neumann, U., Beck, N., Stadler, F., & Ansari, F. (2026). Integrated demand forecasting and reinforcement learning for order point optimization in inventory planning. International Journal of Production Economics, 299, 110111. https://doi.org/10.1016/j.ijpe.2026.110111

[60] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

[61] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[62] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774. https://doi.org/10.48550/arXiv.1705.07874

[63] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

[64] Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

[65] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), Article 93. https://doi.org/10.1145/3236009

[66] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1, 206–215. https://doi.org/10.1038/s42256-019-0048-x

[67] Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable artificial intelligence: Objectives, stakeholders, and future research opportunities. Information Systems Management, 39(1), 53–63. https://doi.org/10.1080/10580530.2020.1849465

[68] Samek, W., Wiegand, T., & Müller, K.-R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. ITU Journal: ICT Discoveries, 1(1), 39–48. https://doi.org/10.48550/arXiv.1708.08296

[69] Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2021). Explainable AI: A review of machine learning interpretability methods. Entropy, 23(1), 18. https://doi.org/10.3390/e23010018