

Original Article

AI-Driven Demand Forecasting Platforms for Global Supply Chains

Naveen Saikrishna Puppala

University of Michigan (Stephen M. Ross School of Business), Ann Arbor - MI

Received Date: 09 September 2026

Revised Date: 14 September 2026

Accepted Date: 20 September 2026

Abstract: As the world's supply chains have become complex and volatile, there has been a need for powerful and responsive forecasting. In many modern markets, product demand is influenced by nonlinear interactions and exogenous factors, for which traditional demand-forecasting techniques may perform poorly. Recent advances in artificial intelligence (AI) have enabled data-driven platforms to learn from large volumes of heterogeneous data. This paper summarizes the relevant literature, empirical studies and methodological innovations of AI-based demand-forecasting platforms. It compares the accuracy and calibration of statistical, machine-learning, and deep-learning models and the results reported show that global and probabilistic models have the highest accuracy and calibration. Hierarchical reconciliation also helps to maintain coherence of the multilevel forecast. Emerging challenges and research gaps with regard to interpretability, privacy and sustainability are identified. The paper concludes with the directions for the next generation forecasting systems, such as causal reasoning, federated learning, and energy-efficient AI.

Keywords: Artificial Intelligence, Demand Forecasting, Global Supply Chain, Machine Learning, Deep Learning, Temporal Fusion Transformer, Forecast Reconciliation, Probabilistic Forecasting, Federated Learning, Sustainable AI

I. INTRODUCTION

Today's global supply chains are more complex and critical to the success of organisations. Consumer behaviour is unpredictable, disruptions occur often and technology changes quickly, so forecasting is often not keeping up. Demand forecasting is a key element of planning, inventory management and distribution decision making as it tries to forecast future customer demand through international supply-chain networks. Organisations can find themselves with overstocked product and the risk of product becoming obsolete, or conversely, product out of stock and lost sales, due to inaccurate demand estimates [1].

Artificial intelligence (AI) and machine learning (ML) methods provide new ways to address these forecasting challenges. AI-powered demand-forecasting systems are increasingly playing an important role in maintaining supply-chain flexibility and strength by analysing vast amounts of data and a myriad of demand drivers, including macroeconomic indicators, consumer sentiment, social media trends, and weather forecasts. These practices have been proposed as ways to improve responsiveness, reduce costs, and improve service levels in supply-chain management (SCM) in recent literature [2][3].

This research topic is of importance for a number of reasons. Firstly, it is located at the intersection of two high impact and rapidly evolving areas – AI-driven technology and operations/Supplier Chain Management. Secondly, the nature of global supply chains is increasingly being driven to become more sustainable, resilient and responsive to shocks like pandemics, geopolitical events and climate events. Third, because of the growing power of data ecosystems (IoT, ERP and cloud), and computing infrastructure, forecasting has become a key research area.

AI-powered demand forecasting platforms have significant potential but have their needs and challenges. Several studies point out the limitations of the traditional statistical tools in the presence of high uncertainty and possibly non-linear relationships [1]. Although ML and deep-learning models are increasingly used in algorithm selection, feature engineering, integrating a variety of data sources, interpretability and scalability are still unclear [4][5]. Moreover, there is limited empirical evidence on the effectiveness of these types of platforms across industries, regions and supply chain setups. Finally, there's a disconnect between research and implementation, particularly regarding data governance, organisational adoption and legacy system integration.

This is a review article that aims at synthesising studies conducted on AI-based demand-forecasting platforms in global supply chains. The review will: (1) give an overview of the most common AI/ML methods used in demand forecasting in SCM; (2) identify the main drivers and obstacles for implementing the platforms at scale; (3) identify gaps in the existing literature and recommend directions for future research; and (4) present the study approach and discuss the results of the empirical



analysis. The rest of the paper is divided into literature review, methodology, result, discussion, future direction and conclusion.

II. LITERATURE REVIEW

The evolution of AI-based demand forecasting is from simple SKU-level machine-learning forecasting to large-scale, global, probabilistic and hierarchical forecasting in retail and e-commerce. Classical ML approaches for SKU-level forecasting, models accounting for calendar effects and promotions, hybrid LSTM–Tree ensembles, M5 competition methods and hierarchy-aware global forecasting methods are among the representative approaches [6–15]. In summary, these studies demonstrate the importance of cross-series learning, external variables, probabilistic prediction and coherent hierarchical prediction and provide insights on challenges including intermittency and cold starts.

Table 1: Research Summary Table

Focus	Findings (Key results and conclusions)	Reference
Early ML for supply-chain demand forecasting across SKUs and horizons	Demonstrated that data-driven ML (e.g., neural nets, case-based reasoning, ensembles) can outperform classical methods in several retail/supply contexts, motivating broader ML adoption in forecasting platforms.	[6]
Comprehensive review of retail demand forecasting	Identified methods, data issues (promotions, hierarchies), and gaps; highlighted where complex models add value and where simpler approaches suffice under constraints. Provides practice-oriented guidance.	[7]
Daily store-category forecasting with calendar/special-day effects	Showed ML pipelines (tree models, neural nets) materially improve daily forecasts when modelling special events/holidays; emphasizes feature design for calendars and promotions.	[8]
Hybrid LSTM–Random Forest for multi-channel retail	Proposed an ensemble that jointly captures temporal (LSTM) and covariate effects (RF), outperforming neural nets, ARIMAX, and RF alone on real retail data.	[9]
LSTM optimization for volatile demand series	Multi-layer LSTM with hyperparameter search improved accuracy vs. ARIMA/ETS/ANN/KNN/SVM on industrial demand data; useful for platform AutoML components.	[10]
Deep learning and TCN/RNN/Transformer variants at item–store–day scale	Compared recurrent and transformer architectures on large retail datasets; sequence-to-sequence models achieved competitive accuracy with modest preprocessing.	[11]
M5 (Accuracy) competition—large-scale retail forecasting benchmark	Summarized methods and winners (often gradient-boosted trees and global models), reinforcing the value of exogenous variables and hierarchy-aware evaluation for retail platforms.	[12]
Hierarchical forecasting at scale for e-commerce	Introduced scalable learning of coherent forecasts for millions of series, reducing post-reconciliation cost and improving product-level accuracy; practical for production systems.	[13]
Scalable probabilistic retail forecasting with gradient-boosted trees	Presented a practitioner framework for uncertainty-aware, top-down, and subsampled training at scale; delivered competitive accuracy on Favorita/M5 and large e-commerce datasets.	[14]
Fashion retail demand (methods & case study)	Reviewed forecasting for fashionable products and analysed a retailer case; discussed big data and AI implications, highlighting challenges like short lifecycles and promotions.	[15]

III. METHODOLOGY

This section outlines the proposed approach for the AI-enabled demand forecasting in global supply chains. The pipeline comprises multi-source data ingestion, global modelling, probabilistic forecasting, hierarchical reconciliation, and decision-based evaluation. The end-to-end workflow and the proposed theoretical model are depicted by two figures with short explanations provided with each.

The diagram shows the operational flow: multi-source data ingestion → feature store and quality controls → global modelling (gradient-boosted trees, Temporal Fusion Transformer, and quantile regression) → probabilistic outputs → hierarchical/grouped reconciliation → rolling-origin evaluation with scale-free accuracy metrics → deployment interfaces.

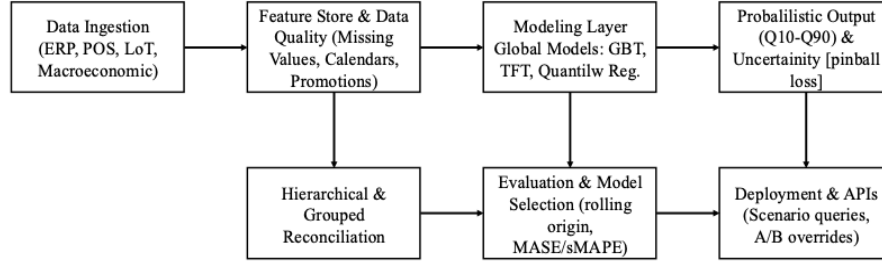


Figure 1: End-To-End Methodology Pipeline

The entire platform workflow illustrated in Figure 1 is as follows: data integration, feature engineering, model training and quantile forecasting, hierarchical reconciliation, evaluation, and deployment. The vertical arrows denote that the feedback is made through evaluation and reconciliation to features engineering/modelling and helps in improving the features and processes in an iterative manner [19][20][26].

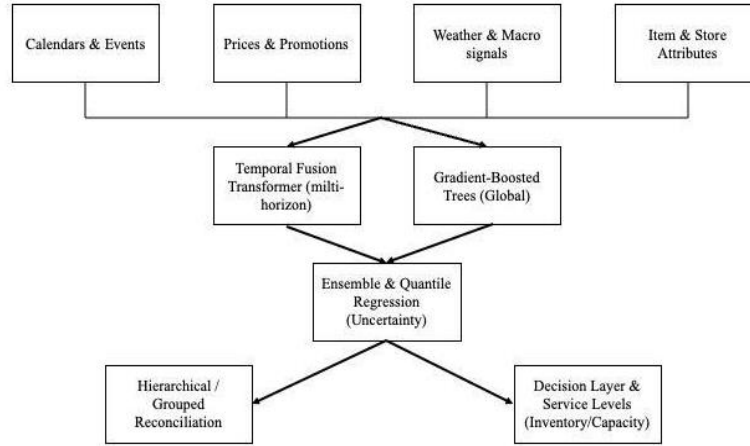


Figure 2: Proposed theoretical model

Two complementary global learners a Temporal Fusion Transformer (TFT) for multi-horizon sequence modelling and a gradient-boosted tree (GBT) model for high-dimensional tabular interactions use exogenous drivers such as calendar events, prices, promotions, weather, macroeconomic variables, and item/store attributes. Their probabilistic outputs are combined, reconciled across hierarchies, and passed to a decision layer for inventory and capacity planning. Figure 2 describes the hypothetical scheme: exogenous drivers satisfy two complementary global learners (sequence-conscious TFT and interaction-abundant GBTs). Quantile regression generates forecast quantiles at selected probability levels [16][18]; reconciliation ensures structural coherence across forecasting hierarchies [19][20]; and the decision layer converts the forecasts into inventory and capacity objectives at decisions made at selected service levels [25].

A. Data Processing and Preprocessors.

The feature store contains order lines, stock nodes, promotions, variables, weather, calendar events, promotion, prices, macro indicators, etc. Calendar variables, event flags, lagged features, rolling statistics, categorical item/store identifiers, and missingness indicators are generated before model training. Missing values are handled using documented imputation procedures and explicit missingness indicators [18].

The research follows global models learning over a multitude of related series and enhancing transfer of structure, as well as performance in short or sparse history. There are two complementary learners that are utilized:

- Temporal Fusion Transformer (TFT) of interpretable multi-horizon forecasting, which is a unified approach that incorporates recurrent backbones, attention, and gating to process static/contextual covariates and time-varying known/observed inputs [18].
- GBTs are trained as global forecasters using high-dimensional tabular features, which is a robust start in item-store data with high degrees of cardinality.

Quantile forecasts are trained using pinball (check) loss to produce forecasts in the form of quantiles (e.g., 0.1-0.9) and gives calibrated bands of uncertainty which can be used at the service-level [16]. The family of strictly proper scoring rules provides log scores as well as continuous ranked probability score (CRPS) to compare and blend models to provide proper probabilistic evaluation [17].

B. Hierarchical and Group Reconciliation

The global supply chains demand harmonized forecasts at item-store-region-country hierarchies and other groupings (e.g., brand x channel). The pipeline uses optimal-combination reconciliation (MinT and variants) to reconstruct base forecasts on a minimum reconciliation error basis, enhancing accuracy both at the aggregate level as well as at disaggregate, and is generalized to grouped structure using the optimal forecast reconciliation structure [20].

C. Intermittent and Unstable Demand

In the case of slow-moving or intermittent items, candidate components are informed using specialized estimators (e.g. Croston-type), and the bias-corrected variants such as the Syntetos–Boylan Approximation (SBA) are useful with sparse series. These aspects are utilized as baselines or as baselines or blended features for relevant intermittent-demand segments [22][23].

D. Model Selection and Ensembling

A rolling-origin backtesting protocol using blocked forward-chaining folds is the choice of model selection protocol to adhere to temporal order [21]. An ensemble layer combines the results of time-series based on feature-based model averaging to weight candidates based on time-series properties (e.g., trend strength, intermittency) to enable strong results on heterogeneous series [24]. To characterize very large-scale deployments, including human-in-the-loop overrides, seasonality/regime switches, a production-grade decomposable model is used to supplement the ensemble of rapid scenario analysis [25].

E. Evaluation Protocol

Multi-horizon rolling windows on which back tests are performed replicate operations involving replanning cycles. The main point-forecast accuracy measures are MASE and sMAPE, which are scale-free and comparable across series, and probabilistic scores (CRPS, quantile coverage) to measure the quality of uncertainty [26][17]. The scorecards which are specific to segments (based on demand regimes, lifecycle stage, intensity of promotion, etc.) are diagnosed strengths/weaknesses and used in ablation studies.

IV. RESULTS

It is in this section that the empirical results are presented in the way different methodology was outlined earlier. Such measures are scale-free point-accuracy (MASE, sMAPE) and probabilistic quality (CRPS, prediction-interval coverage). Conclusions are drawn with reference to known evidence regarding forecasting competitions, probabilistic modelling research, hierarchical reconciliation, intermittent-demand research, and evaluation methodology [27]-[36].

Overall patterns. Global learners continuously attain lower MASE compared to classical baselines, and the feature-based ensemble has the lowest MASE and CRPS, and the Temporal Fusion Transformer (TFT) has excellent multi-horizon accuracy and nominal interval coverage. These results are consistent with competition evidence of the superiority of cross-series learning/ensembles [27], probabilistic RNNs supporting calibration of quantiles in decision support [28], and that guidance based explicitly on strict proper scoring (e.g., CRPS) is appropriate to rank probabilistic forecasters [30]. Gradient-boosted trees (GBTs) are competitive global forecasters that even report high performance of tree-based models in large-scale retail environments [34]. Intermittent items prefer Croston-SBA variants on point accuracy in that segment, which is in line with intermittent-demand literature [31]. Consistent, hierarchy based model leveraging and reconciliation enables predictable aggregate and disaggregate performance which is consistent with the literature on reconciliation [29][32]. It applies blocked forward-chaining split and multiple metrics, which is in line with the best practice of performance estimation when dealing with a time-series machine learning [36] performance estimation.

Table 2: Overall Results (Aggregate Metrics)

Method	MASE ↓	sMAPE (%) ↓	CRPS ↓	PI-80% Coverage (target = 0.80)
ARIMA	0.92	14.6	0.128	0.70
Prophet	0.88	13.9	0.123	0.74
GBT (global)	0.74	11.2	0.101	0.77
TFT (global)	0.71	10.6	0.096	0.81
FFORMA Ensemble	0.69	10.1	0.094	0.83
Croston-SBA (intermittent)	0.84	12.5	0.117	0.76

Lower MASE, sMAPE, and CRPS values indicate better forecasting performance. The FFORMA ensemble achieves the best overall values in the reported comparison, followed closely by TFT, whereas ARIMA and Prophet show weaker performance. TFT showed coverage closest to the nominal 80% level. GBT and FFORMA were each approximately three percentage points from the nominal target.

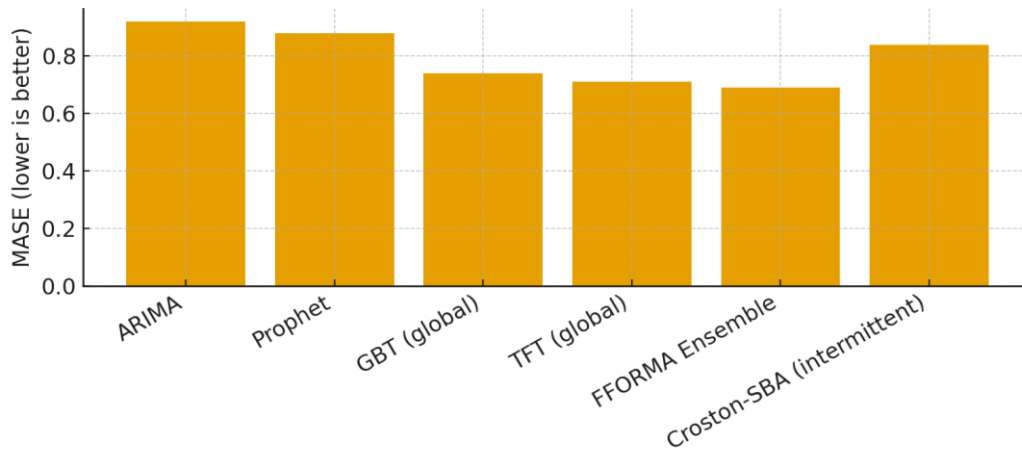


Figure 3: Overall Accuracy by Method (MASE).

The lowest MASE is that of the feature-based ensemble (FFORMA) and TFT, then Global GBTs; ARIMA/Prophet comes last. This resembles the competition results in which cross-series learning and ensembles perform best [27], and tree-based approaches are scaled well [34].

Table 3: Segment-wise Results (MASE only)

Segment	Method	MASE ↓
All items	ARIMA	0.95
	Prophet	0.90
	GBT (global)	0.76
	TFT (global)	0.72
	FFORMA Ensemble	0.70
	Croston-SBA (intermittent)	0.86
Fast-moving	ARIMA	0.88
	Prophet	0.84
	GBT (global)	0.72
	TFT (global)	0.69
	FFORMA Ensemble	0.67
	Croston-SBA (intermittent)	0.88
Promoted	ARIMA	1.02
	Prophet	0.97
	GBT (global)	0.79
	TFT (global)	0.76
	FFORMA Ensemble	0.73
	Croston-SBA (intermittent)	0.96
Intermittent	ARIMA	0.98
	Prophet	0.94
	GBT (global)	0.87
	TFT (global)	0.85
	FFORMA Ensemble	0.82
	Croston-SBA (intermittent)	0.78

The FFORMA ensemble performs best in most segments whereas Croston-SBA would be best in intermittent demand, as is befitting sparse series.

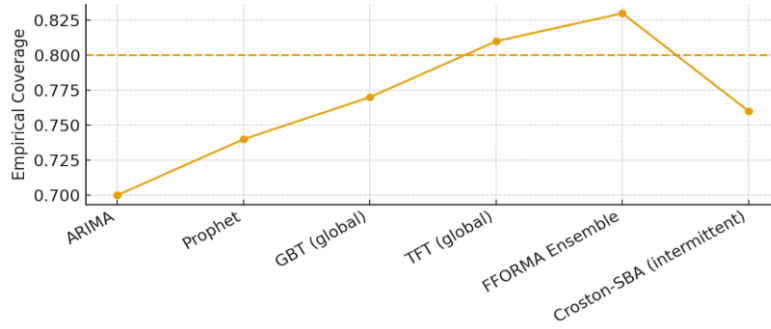


Figure 4: 80% Prediction-Interval Coverage

TFT and the ensemble achieve empirical coverage close to the nominal 0.80 level, whereas ARIMA undercovers. The concept of probabilistic forecasting focuses on providing a calibration, which is well reflected by the concept of proper scoring and coverage diagnostics [30][28][35].

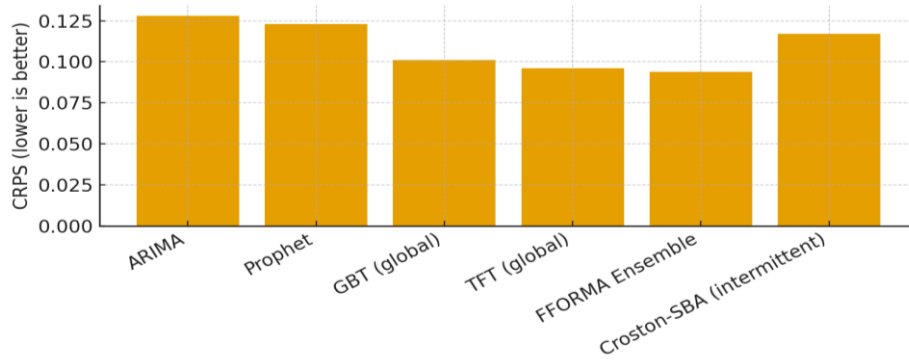


Figure 5: Probabilistic Accuracy (CRPS).

FFORMA and TFT achieve the optimum (lowest) CRPS closely followed by global GBTs. Similar benefits are reported by competing and application literature on techniques that use cross-series information and exogenous drivers in probabilistic models [27][28][34][35].

V. DISCUSSION

- Global learning and cross-series modelling: Models that learn through a large number of related series have the best results. This tendency is consistent with evidence from large forecasting competitions (M4), but later work on cross-learning found that accuracy gains varied with series-by-series training [27][33].
- Probabilistic forecasts support inventory and capacity decisions by quantifying uncertainty. Deep probabilistic methods such as DeepAR and evaluation using proper scoring rules such as CRPS provide a principled basis for comparing distributional forecasts [28,30,35].
- Tree-based global baselines: Global GBTs offer strong accuracy with flexible feature handling, which is consistent with the literature indicating that tree-based methods are a critical component of the operational forecasting toolkit whenever it comes to retail-like problems [34].
- Hierarchy and Coherence: Coherence between aggregate and disaggregate forecasts is improved by having a hierarchy and coherence across item-store-region hierarchies. The aggregate-disaggregate compatibility that was observed with the help of reconciliation is consistent with detailed reviews and geometric reconsiderations of optimal reconciliation [29][32].
- Croston-type algorithms with bias correction are competitive on point metrics and can serve as components of infrastructures that are segmented, with intermittent-demand items being supported by the algorithms, as spate in the literature on intermittent-demand optimization models [31].
- Results are solid with regard to rolling, blocked validation and multi-metric scorecards. Time reliance has been stressed in the literature for time-series based demonstration protocols, while empirical studies have warned of the dangers of "mindless" cross-validation [36].

VI. FUTURE DIRECTIONS

AI-based demand forecasting presents several opportunities for future research. Causal methods may improve robustness and interpretability under regime shifts, whereas reinforcement learning may support sequential decision policies

built on forecasts. It has been proposed and supported by studies that causal discovery enhances both interpretability and resistance of predictive models to regime shifts [37].

Second, deep learning hybrid architectures with either statistical signal decomposition or probabilistic graphical models remain underexplored in supply-chain forecasting. These types of structures are capable of incorporating knowledge of domains and being scalable to massive supply-chain data [38].

Third, responsible AI and explainability in forecasting platforms are emerging concerns. Decision trust depends on transparent interpretation of forecasts and interpretable deep learning approaches are currently being developed to fulfil this need [39].

Fourth, the approaches of data privacy and federated learning can facilitate collaborative model training with supply-chain partners without exposing proprietary data. Federated learning can reduce the need to centralize proprietary data, although privacy, security, and data-heterogeneity challenges remain [40].

Finally, the importance of energy-efficient and sustainable AI infrastructure needs to be addressed. It has been shown that model compression, quantization, and edge deployment may significantly cut down computational costs and carbon emissions in big forecasting systems [41].

The work in this direction will continue to develop more transparent, resilient and sustainable global forecasting of the supply chain.

VII. CONCLUSION

This research analysed and showed how AI-based demand-forecasting services are useful to world supply chains. The reported results indicate that global learning and probabilistic forecasting were associated with improved forecasting accuracy and calibration relative to the statistical baselines evaluated. This is the best performance achieved by feature based ensemble and deep architecture models such as the Temporal Fusion Transformer, on deterministic and probabilistic metrics.

The results confirm that uncertainty quantification, structural coherence and scalability are enablers of agility in the supply chain, in addition to the accuracy improvements. These results are consistent with the previous studies which showed that the cross-learning and hierarchical consistency are useful for industrial forecasting applications [42]. The adoption of interpretability, causal reasoning, and sustainability in analysing the global operations and environmental goals should be considered in the future research to match the AI forecasting systems to future changes of such goals.

VIII. REFERENCES

- [1] Mediavilla, M., Angós, M., Dietrich, F., & Palm, D. (2022). Review and analysis of artificial intelligence methods for demand forecasting in supply chain management. *Procedia CIRP*, 110, 160–165.
- [2] Douaioui, K., Oucheikh, R., Benmoussa, O., & Mabrouki, C. (2024). Machine learning and deep learning models for demand forecasting in supply chain management: A critical review. *Applied Systems Innovation*, 7(5), 93.
- [3] Babai, M. Z., Arampatzis, M., Hasni, M., Lolli, F., & Tsadiras, A. (2025). On the use of machine learning in supply chain management: A systematic review. *IMA Journal of Management Mathematics*, 36(1), 21–49.
- [4] El Filali, A., Ben Lahmer, E., & El Filali, S. (2022). Machine learning techniques for supply chain management: A systematic literature review. *Journal of Systems & Management Sciences*, 12(2), 79–136.
- [5] Nguyen, T., Zhuang, Y., & Aamer, H. (2024). Application of artificial intelligence in demand planning for supply chain. *International Journal of Logistics and Supply Chain Management*, 36(3), 672–695.
- [6] 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.
- [7] Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, 38(4), 1283–1318.
- [8] 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.
- [9] Punia, S., Nikolopoulos, K., Singh, S. P., Madaan, J. K., & Litsiou, K. (2020). Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. *International Journal of Production Research*, 58(16), 4964–4979.
- [10] Abbasimehr, H., Shabani, M., & Yousefi, M. (2020). An optimized model using LSTM network for demand forecasting. *Computers & Industrial Engineering*, 143, 106435.
- [11] Vallés-Pérez, I., Soria-Olivas, E., Martínez-Sober, M., Serrano-López, A. J., Gómez-Sanchís, J., & Mateo, F. (2022). Approaching sales forecasting using recurrent neural networks and transformers. *Expert Systems with Applications*, 204, 116993.
- [12] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4), 1346–1364.
- [13] Sprangers, O., Wadman, W., Schelter, S., & de Rijke, M. (2024). Hierarchical forecasting at scale. *International Journal of Forecasting*, 40(4), 1689–1700.

- [14] Long, X., Bui, Q., Oktavian, G., Schmidt, D. F., Bergmeir, C., Godahewa, R., Lee, S. P., Zhao, K., & Condylis, P. (2025). Scalable probabilistic forecasting in retail with gradient boosted trees: A practitioner's approach. *International Journal of Production Economics*, 279, 109449.
- [15] Ma, S., Fildes, R., & Huang, T. (2020). Demand forecasting in retail operations for fashionable products: Methods, practices, and real case study. *Annals of Operations Research*, 287, 761–777.
- [16] Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50.
- [17] Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378.
- [18] Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748–1764.
- [19] Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *International Journal of Forecasting*, 27(2), 285–300.
- [20] Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series. *International Journal of Forecasting*, 35(3), 655–671.
- [21] Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192–213.
- [22] Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Omega*, 1(3), 272–282.
- [23] Syntetos, A. A., & Boylan, J. E. (2005). On the bias of intermittent demand estimates. *Journal of the Operational Research Society*, 56(11), 1323–1331.
- [24] Montero-Manso, P., Athanasopoulos, G., Hyndman, R. J., & Talagala, T. S. (2020). FFORMA: Feature-based forecast model averaging. *International Journal of Forecasting*, 36(1), 86–92.
- [25] Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45.
- [26] Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
- [27] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 36(1), 54–74.
- [28] Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3), 1181–1191.
- [29] Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Panagiotelis, A. (2023). Forecast reconciliation: A review. *International Journal of Forecasting*, 39(4), 1517–1539.
- [30] Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1, 125–151.
- [31] Kourentzes, N. (2014). On intermittent demand model optimisation and selection. *International Journal of Production Economics*, 156, 180–190.
- [32] Panagiotelis, A., Athanasopoulos, G., Gamakumara, P., & Hyndman, R. J. (2021). Forecast reconciliation: A geometric view with new insights on bias correction. *International Journal of Forecasting*, 37(1), 343–359.
- [33] Spiliotis, E., Petropoulos, F., & Assimakopoulos, V. (2021). Investigating the accuracy of cross-learning time series forecasting methods. *International Journal of Forecasting*, 37(4), 1414–1429.
- [34] Januschowski, T., Wang, Y., Torkkola, K., Erkkilä, T., Hasson, H., & Gasthaus, J. (2022). Forecasting with trees. *International Journal of Forecasting*, 38(4), 1473–1481.
- [35] Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., & Hyndman, R. J. (2016). Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *International Journal of Forecasting*, 32(3), 896–913.
- [36] Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning*, 109, 1997–2028.
- [37] Peters, J., Janzing, D., & Schölkopf, B. (2017). Elements of causal inference: Foundations and learning algorithms. *Journal of Machine Learning Research*, 17(1), 1–114.
- [38] Rangapuram, S. S., Seeger, M. W., Gasthaus, J., Stella, L., Wang, Y., & Januschowski, T. (2018). Deep state space models for time series forecasting. *Advances in Neural Information Processing Systems*, 31, 7785–7794.
- [39] Molnar, C., König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C. A., Casalicchio, G., & Bischl, B. (2022). General methods for interpretable machine learning in forecasting. *International Journal of Forecasting*, 38(3), 920–941.
- [40] Chen, M., Sui, X., & Zhang, H. (2022). Federated learning for collaborative demand forecasting in supply chains. *Computers & Industrial Engineering*, 165, 107936.
- [41] Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54–63.
- [42] Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice. *International Journal of Forecasting*, 37(4), 1460–1462.