Construction and Application of a Big-Data-Driven Demand Forecasting Model for Fresh-Produce E-Commerce

Wanying Zhang [1,*]

[1] Macau University of Science and Technology, Macau Taipa, China

* Corresponding Author Email: 1405003116@qq.com

Abstract. This paper addresses the prominent operational challenges faced by China's fresh-produce e-commerce sector, including high demand uncertainty, large product loss rates, and coarse inventory management. From a big-data-driven perspective, we construct a lightweight demand forecasting model tailored to small- and medium-sized fresh-produce e-commerce enterprises and propose a set of accompanying dynamic inventory optimization strategies. The paper first reviews the current state and persistent problems of demand forecasting in fresh e-commerce. It then designs a four-dimensional multi-source data system covering sales, environment, scenario, and supply, together with a three-layer "data-algorithm-application" model architecture. Three inventory optimization strategies are then proposed: dynamic safety stock with precise ordering, differentiated category management, and end-to-end data collaboration. Finally, with Hema (Freshippo) as a case, this paper analyzes how its demand-forecasting and inventory-management practices can inform improvements to the proposed framework. The analysis shows that a big-data-driven lightweight forecasting framework can effectively raise the forecast accuracy and inventory turnover of small- and medium-sized fresh e-commerce enterprises without significantly increasing their technical threshold, offering useful references for the industry's digital transformation.

Keywords: Big data, Fresh-produce e-commerce, Demand forecasting, Inventory optimization, Multi-source data fusion.

1. Introduction

The rapid development of the digital economy is driving profound transformation across the global retail industry, and e-commerce has become a key engine for growth in consumer markets [1]. In China, the fresh-produce e-commerce market has expanded rapidly and become an important component of the new-retail ecosystem [2]. In retail supply chain management, accurate demand forecasting is the key link for reducing operating costs and improving service levels, and the rise of big-data technology has brought new methodological tools to this long-standing problem [3], [4].

With the deepening of internet-business globalization and informatization, the fresh-produce e-commerce industry in China has developed rapidly, with increasingly intense market competition. Compared with general fast-moving consumer goods, fresh products are perishable, have short shelf lives, and are consumed at high frequency; their demand is therefore strongly affected by external variables such as weather conditions, holidays, and consumer habits, which makes demand forecasting in the supply chain particularly complex [5]. Inaccurate forecasts can lead to inventory backlog, stock-outs, and even product losses, severely undermining enterprise efficiency and competitiveness [6].

In the new-retail era, the integration of online and offline channels has become a clear trend, and big-data analytics has had a significant impact on the traditional marketing model of fresh products [7]. Informatization has become an inevitable trend for the development of fresh e-commerce: big data can penetrate every link of the supply chain—procurement, storage, cold-chain logistics, and sales—and, through the integrated analysis of multiple types of data, support accurate demand forecasting, reduce losses, improve efficiency, and provide both cost savings and revenue gains for enterprises [8].

However, looking at the current state of the industry, demand forecasting in fresh-produce e-commerce still faces three major pain points: (i) traditional manual-experience-based forecasting carries large errors and cannot meet real-time requirements; (ii) most enterprises fail to effectively integrate multi-source information, so information across the entire process cannot be sufficiently coordinated, severely degrading forecasting accuracy and reliability; and (iii) existing complex mathematical models are difficult to align with operational realities, while their high implementation costs and steep entry barriers prevent these advanced techniques from being put into practice [9]. The underlying causes are that small- and medium-sized fresh e-commerce platforms lack sufficient investment in technology, their informatization is relatively slow, dedicated data analysts are scarce, and proper database systems are largely absent, so that forecasting cannot effectively support supply chain operations.

To address these issues, this paper, from a big-data-driven perspective, constructs a lightweight demand forecasting model tailored to small- and medium-sized fresh e-commerce enterprises and proposes accompanying inventory optimization strategies. The main contributions are: (i) reviewing research progress on demand forecasting in fresh e-commerce and on the application of big data; (ii) designing a "sales-environment-scenario-supply" four-dimensional multi-source data system together with a three-layer "data-algorithm-application" model architecture; (iii) proposing three inventory optimization strategies based on dynamic safety stock, category-differentiated control, and end-to-end coordination; and (iv) using Hema (Freshippo) as a case to analyze how its demand-forecasting and inventory-management practices can inform improvements to the proposed framework.

The remainder of this paper is organized as follows. Section 2 reviews related research. Section 3 introduces the construction of the big-data-driven demand forecasting model. Section 4 presents the inventory optimization strategies based on demand forecasting. Section 5 analyzes Hema as a case study. Section 6 discusses the implications of the case for the proposed model as well as research limitations. Section 7 concludes the paper.

2. Related Work

Existing research on big-data-driven retail supply chain demand forecasting has been carried out from multiple perspectives, including data modeling, algorithm optimization, and application scenarios.

Research on retail supply chain demand forecasting originated overseas. Existing reviews have noted that overseas research has focused on multi-source information fusion and machine-learning model optimization to improve forecasting accuracy, with successful applications in large retail enterprises [4]. Boone et al. [8] emphasized that consumer analytics in the big-data era provides new opportunities for retail demand forecasting; Fildes et al. [5] systematically compared the advantages and applicable scenarios of machine-learning methods versus traditional statistical methods in retail forecasting. At the methodological level, Bandara et al. [10] built a sales forecasting model for e-commerce using a long short-term memory (LSTM) network; Punia et al. [6] further proposed a data-driven multi-item newsvendor model that extends from predictive to prescriptive analytics; and Feizabadi [11] empirically demonstrated the positive impact of machine-learning-based demand forecasting on supply chain performance.

Concerning the integration of big-data analytics with operations management, Choi et al. [7] discussed a big-data analytics framework for operations management, emphasizing the importance of real-time analytics and dynamic adjustment for fresh-produce supply chains. Hofmann and Rutschmann [12] conducted a conceptual analysis of how big-data analytics enhances demand forecasting in supply chains. Pereira and Frazzon [13] proposed a data-driven adaptive synchronization approach for demand and supply in omni-channel retail supply chains. In addition, Lau et al. [14] improved sales-forecasting accuracy through big-data sentiment analysis, illustrating the potential value of unstructured data in forecasting modeling.

Regarding research within China, existing studies on demand forecasting in fresh-produce e-commerce mainly focus on advanced machine-learning techniques such as LSTM and Random Forest, while lightweight and easy-to-use models for small fresh e-commerce companies have received relatively little attention [9]. Although many Chinese fresh-produce enterprises have begun exploring big-data applications, traditional forecasting methods still rely heavily on manual computation and manual entry of historical data, which leads to sluggish market response, susceptibility to error, and a lack of multi-source data fusion capabilities.

In summary, the existing literature has provided rich theoretical foundations and methodological tools for fresh-produce demand forecasting; however, two gaps remain. First, overly complex models impose high requirements on enterprises' data-processing capabilities and the cost of skilled personnel, making them difficult for small- and medium-sized e-commerce enterprises to adopt. Second, the coordination mechanism between forecasting models and downstream functions such as inventory optimization and replenishment decisions remains relatively under-researched [6], [13]. Building on these gaps, this paper combines the characteristics of the fresh-produce industry to apply big-data-driven optimization to existing models, constructing a lightweight forecasting model that both fits real operations and is easy to implement, and providing end-to-end coordinated design with inventory management.

3. Construction of a Big-Data-Driven Demand Forecasting Model

3.1. Modeling Objectives

Combining the characteristics of the fresh e-commerce industry with the actual application context of small- and medium-sized enterprises, this paper sets four objectives for the demand-forecasting model. (i) Accuracy: using multi-source data to achieve higher forecasting accuracy and smaller forecasting bias. (ii) Lightness: easy to understand without requiring complex theoretical background, suitable for the operating scenarios of small fresh e-commerce enterprises. (iii) Coordinability: capable of working in coordination with warehousing, replenishment, and cold-chain logistics to support supply chain management tasks. (iv) Flexibility: capable of responding to demand fluctuations caused by seasonality, weather, and holidays.

3.2. Big-Data-Driven Modeling Approach

Given the limitations of traditional forecasting methods—narrow applicability, weak flexibility, and a lack of tight integration with the supply chain—this paper builds improvements at three layers of "data-model-application" with emphasis on real-world deployment. First, multi-source heterogeneous data are used as a foundation to strengthen fusion: the model breaks free from the limitation of single sales-data sources by introducing multi-dimensional data covering historical sales, weather conditions, holidays, consumer preferences, and capacity levels, and applies big-data cleaning and normalization to address data dispersion. Next, in view of the strong volatility and burstiness of fresh-product demand, single-shot prediction is replaced with a three-stage rolling forecast: a 72-hour short cycle, a one-week medium cycle, and a one-month long cycle. Furthermore, without pursuing computational complexity—and on the premise of operational convenience—the approach addresses accuracy concerns so that small e-commerce enterprises can better adapt to current technical conditions. Finally, the forecasting results are linked across procurement, replenishment, warehousing, and cold-chain logistics, achieving the vision of "ordering by forecasts and managing warehouses by data" and reducing fresh-product losses.

3.3. Multi-Source Data System

Since a big-data forecasting model is built on the integration of diverse data types, this paper categorizes the data into four groups. Sales data include historical sales, order volume, inventory turnover, and the proportion of the sales cycle. Environmental data include weather conditions, humidity, season, precipitation, and meteorological-disaster early warnings. Consumer-scenario data include holidays, promotional activities, weekend disturbance factors, and regional consumer characteristics. Supply-chain data cover production capacity, cold-chain transportation performance, replenishment cycle, and procurement price. Together they provide comprehensive data support for the forecasting task.

3.4. Data Preprocessing

Because the collected raw data contain redundancy, missing values, and anomalies, direct use will distort forecasting. A rigorous standardization pipeline is therefore needed. The first step is data cleaning, in which redundant and anomalous records are removed and missing values are imputed, ensuring the authenticity and reliability of the foundational data. The second step is standardization, which transforms data formats and units to eliminate dimensional differences between sources. Z-score standardization is adopted, with the formula:

$$X^* = \frac{X - \bar{X}}{\sigma} \tag{1}$$

Where X* is the standardized value, X is the original value, X with overline is the data mean, and σ is the standard deviation. The final step is data fusion, which integrates multi-source heterogeneous data into a single database as effective input to the model.

3.5. Model Architecture

The model improves a traditional statistical approach with big-data techniques. The full architecture consists of three layers and is simple enough for small- and medium-sized e-commerce platforms. The first layer is the data layer, which collects and cleans data from heterogeneous sources to provide high-quality inputs. The second layer is the algorithm layer, where a concise multivariate linear regression model is built:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \tag{2}$$

Where Y is future product demand; X_1 is historical sales data; X_2 contains weather, holiday, and promotion information; X_3 contains supply-side information; X_4 represents consumer-scenario factors; β_0 through β_4 are regression coefficients; and ε is a random disturbance term. Through these input variables, the model extracts the most influential factors from large volumes of data and feeds the results to the application layer—procurement, inventory, and logistics—to support operations.

3.6. Implementation Path

Once improvement measures are chosen, a clear methodology must support deployment. First, an information system covering customers, products, environment, and logistics is built, with all data classified and updated under a unified standard. Second, the original model is iteratively upgraded: on top of traditional time-series analysis and linear regression, big-data feature processing and dynamic parameter tuning are introduced, refining the model without altering its core structure. Third, results are fed back into a loop, forming a "forecast-operation-feedback-correction" closed loop in which model parameters are continuously updated according to actual sales. Finally, scenario-specific forecasting algorithms are deployed for normal weather, holidays, promotion days, and abnormal weather, to improve accuracy in these special situations.

3.7. Implementation Procedure

The model covers the entire workflow from initial data collection to result presentation. Stage 1 is data collection, including sales, environmental, scenario, and supply data. Stage 2 is the cleaning, integration, and consolidation of information. Stage 3 selects influential factors as model inputs from the large pool of data. Stage 4 applies the improved model to produce short-, medium-, and long-term forecasts. Stage 5 translates forecasts into operational decision-support information. Stage 6 continually adjusts model parameters in practice to enhance forecasting accuracy.

4. Inventory Optimization Based on Big-Data Demand Forecasting

4.1. Objectives of Inventory Optimization

This section aims to achieve optimal inventory management through big-data-driven demand forecasting. The main task is to maintain real-time optimal inventory levels. Only with accurate forecasts can fresh-product surpluses and shortages be effectively mitigated; perishable losses reduced; warehousing and cold-chain transportation costs lowered; and the turnover efficiency and profitability of the entire fresh-produce supply chain improved.

4.2. Core Strategies of Big-Data-Driven Dynamic Inventory Optimization

Traditional inventory management suffers from information lag, rigid decisions, and high deviation dispersion. This section combines dynamic warehouse control, category management, and full-chain data coordination with rolling big-data forecasts to make full use of real-time pull signals.

4.2.1 Dynamic Safety Stock and Precision Ordering

This is the core of inventory control. Because safety stock and order quantity are tightly coupled, they are jointly treated as dynamic values driven by real-time big-data forecasts rather than fixed numbers. The dynamic safety stock level is determined by big-data-driven demand variability:

$$SS = Z \times \sigma \times \sqrt{L} \tag{3}$$

Where Z is the service-quality factor of real-time stream-data processing, σ is the big-data-derived demand standard deviation reflecting fluctuation, and L is a big-data-tracked replenishment cycle. Because big data can sense in real time the effects of holidays, seasonality, and weather on sales, safety stock is no longer a fixed quantity: it rises automatically during high-sales periods and falls during low-sales periods, balancing inventory cost against service level. The reorder point is computed in real time as:

$$ROP = \bar{D} \times L + SS \tag{4}$$

Where ROP is the reorder point, D with overline is the big-data-derived average daily demand, L is the real-time replenishment cycle, and SS is the dynamic safety stock. The "order to sales, plan to demand" principle thus runs across the entire supply chain, reducing over-purchasing and inventory backlog while ensuring continuous availability.

4.2.2 Category-Differentiated Inventory Control

Because the fresh category covers vegetables, fruits, meat, seafood, and other subcategories with widely different loss and turnover rates, big-data analytics is needed to separate these rates and develop category-specific controls. Highly perishable leafy vegetables use small-batch, high-frequency local delivery; moderately perishable fruits use small-batch, high-frequency replenishment; and less perishable meat and seafood use large-batch, low-frequency replenishment. Big-data technology also continuously monitors per-category inventory turnover and dynamically optimizes replenishment frequencies and inventory ceilings to reduce loss rates.

4.2.3 End-to-End Data Coordination

Traditionally, inventory management separates procurement, warehousing, cold-chain transportation, and store-sales information; big-data technology is required to integrate information across the supply chain and enable synchronous sharing. A big-data platform can capture procurement, inventory, cold-chain logistics, and store sales information in real time. Based on product-level demand forecasts, cold-chain routes, distribution plans, and replenishment timing can be planned in advance to accelerate inventory turnover. Combined with end-to-end data coordination, dynamic inventory transfer becomes possible: when a store faces a shortage, stock from nearby areas can be reallocated, avoiding stock-out risks and reducing overall inventory cost.

4.3. Evaluation of Inventory Optimization Effects

This paper establishes an evaluation system for big-data-driven demand forecasting and inventory optimization, with four main indicators: inventory turnover rate, loss rate, stock-out rate, and inventory cost ratio. Before-and-after comparisons under big-data support directly reveal the optimization effects and provide a basis for further refinement.

5. Case Study: Hema (Freshippo)

5.1. Company Overview

Hema (Freshippo) is a leading enterprise of China's new-retail industry. Its "store-warehouse-integrated" model combines online and offline operations on a self-built information-based supply chain to achieve high operating efficiency [2]. Its forecasting models and replenishment algorithms have substantial demonstration value for the industry. At present, demand-forecasting methods used by fresh e-commerce enterprises show a bimodal pattern: leading companies such as Hema and Dingdong Maicai have begun deploying big-data forecasting to raise accuracy, while most small and medium fresh e-commerce companies still rely on experience-based judgments or simple analyses, so that forecasting results cannot inform actual operations. Although Hema has already established a data-driven supply chain system that, in most cases, uses big-data models to manage replenishment for core products, smaller stores and minor categories in certain regions still rely on manual estimation, and losses have not yet been fully eliminated.

It should be emphasized that the purpose of this section is not to directly validate the model proposed in this paper. Instead, by selecting a leading enterprise in the industry as a reference, the section aims to demonstrate the reasonableness of the design philosophy and provide direction for model improvement.

5.2. Successful Practices of Hema's Demand Forecasting and Inventory Management

Based on the analysis above, Hema's successful practices can be summarized in three main points.

First, a comprehensive big-data combined forecasting approach. Unlike traditional experience-based forecasting, Hema integrates historical sales data, real-time shopping-cart data, weather and holiday effects, individual consumer behaviors, and trade-area foot traffic to form a comprehensive data ecosystem. This is one of the reasons Hema can deliver accurate demand forecasts [2].

Second, dynamic big-data forecasting models. Hema's models are simple yet continuously evolving, providing hierarchical forecasts at three horizons—72 hours short, one week medium, and one month long—with accuracy refined to the store-and-category level. The result is real-time forecasting information that directly supports procurement, replenishment, and storage decisions, rather than methods that are sophisticated but impractical.

Third, deep integration between forecasting and inventory management. Forecasting is tightly linked with procurement, warehousing, cold-chain distribution, and inventory control to realize "order by forecasts and monitor inventories digitally." Safety stock levels, order batches, and replenishment frequencies are dynamically adjusted, achieving industry-leading low loss rates for fresh products and improving inventory utilization.

6. Discussion

6.1. Implications of the Hema Case for the Proposed Model

Analysis of Hema's successful practices provides three improvement directions for the demand-forecasting and inventory strategies proposed in this paper.

First, multi-source data fusion can improve forecasting accuracy. The proposed model's data dimensions can be further extended. Hema's experience suggests that richer data sources tend to yield higher accuracy. Beyond the existing sales, environment, scenario, and supply dimensions, additional dimensions such as consumer-trend, trade-area foot traffic, and competitor sales data could be added to build a more comprehensive data ecosystem and improve data quality.

Second, a lightweight rolling-forecast scheme that aligns with industry practice. The proposed model can adopt a more sophisticated hierarchical rolling-forecast scheme to improve flexibility. Short-term (72-hour), medium-term (mid-month), and long-term (end-of-month) hierarchies can each be designed, with forecasting parameters dynamically adjusted to reflect short-term demand fluctuations, seasonal patterns, and holiday effects, substantially enhancing flexibility and real-time responsiveness.

Third, deeper integration between forecasting and inventory control. Big-data-driven forecasts can be used to dynamically adjust safety stock and reorder points, with category-tiered controls maintaining end-to-end coherence. A "forecast-inventory-execution-feedback" loop should be institutionalized so that forecasting directly drives inventory decisions, raising overall strategic effectiveness.

6.2. Research Limitations and Future Work

This study focuses on improvements to a general modeling framework; differentiated refinements for regions and category types remain to be explored. Moreover, due to research constraints, this paper has not conducted large-scale empirical validation, and the forecast accuracy and inventory optimization effects of the model still need to be verified using real operational data. Future work could include unstructured data, such as public opinion and consumer reviews, to improve generalizability [14]. AI large models and digital twins offer additional avenues to refine forecasting precision. Industry-wide data standardization will allow data sharing among producers, e-commerce platforms, and logistics providers. Lightweight, easy-to-use forecasting tools should also be developed to lower adoption barriers for smaller enterprises, ultimately advancing the comprehensive digitization and automation of the fresh-produce e-commerce industry.

7. Conclusion

This paper analyzed the current state of demand forecasting in the fresh-produce e-commerce industry and its main problems, and proposed a lightweight big-data-driven improvement together with a coordinated inventory strategy, using Hema (Freshippo) as a case to demonstrate the feasibility of the design philosophy. The principal weaknesses of current demand forecasting lie in insufficient data sources, models that are difficult to implement, and a lack of effective coordination with downstream operations; big-data technologies can address these weaknesses by raising forecasting accuracy to meet the specific characteristics of the fresh-produce sector. Demand forecasting and inventory optimization are tightly coupled, jointly supporting real-time safety stock, precise ordering, and category-tiered control, and thereby substantially reducing losses and operating costs. The Hema case further indicates that a big-data-driven demand-forecasting framework is both feasible and effective in practice, offering useful reference value for peer enterprises pursuing digital transformation. Although the present study is limited by the absence of large-scale empirical validation, the proposed framework provides a tractable starting point for small- and medium-sized fresh e-commerce enterprises and lays the groundwork for subsequent quantitative refinement and field testing.

References

[1] S. Akter and S. F. Wamba, "Big data analytics in E-commerce: a systematic review and agenda for future research," *Electronic Markets*, vol. 26, no. 2, pp. 173–194, 2016. DOI: 10.1007/s12525-016-0219-0.

[2] Y. Wang and N. M. Coe, "Platform ecosystems and digital innovation in food retailing: Exploring the rise of Hema in China," *Geoforum*, vol. 126, pp. 310–321, 2021. DOI: 10.1016/j.geoforum.2021.08.007.

[3] G. Wang, A. Gunasekaran, E. W. T. Ngai, and T. Papadopoulos, "Big data analytics in logistics and supply chain management: Certain investigations for research and applications," *International Journal of Production Economics*, vol. 176, pp. 98–110, 2016. DOI: 10.1016/j.ijpe.2016.03.014.

[4] M. Seyedan and F. Mafakheri, "Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities," *Journal of Big Data*, vol. 7, no. 1, art. 53, 2020. DOI: 10.1186/s40537-020-00329-2.

[5] R. Fildes, S. Ma, and S. Kolassa, "Retail forecasting: Research and practice," *International Journal of Forecasting*, vol. 38, no. 4, pp. 1283–1318, 2022. DOI: 10.1016/j.ijforecast.2019.06.004.

[6] S. Punia, S. P. Singh, and J. K. Madaan, "From predictive to prescriptive analytics: A data-driven multi-item newsvendor model," *Decision Support Systems*, vol. 136, art. 113340, 2020. DOI: 10.1016/j.dss.2020.113340.

[7] T.-M. Choi, S. W. Wallace, and Y. Wang, "Big Data Analytics in Operations Management," *Production and Operations Management*, vol. 27, no. 10, pp. 1868–1883, 2018. DOI: 10.1111/poms.12838.

[8] T. Boone, R. Ganeshan, A. Jain, and N. R. Sanders, "Forecasting sales in the supply chain: Consumer analytics in the big data era," *International Journal of Forecasting*, vol. 35, no. 1, pp. 170–180, 2019. DOI: 10.1016/j.ijforecast.2018.09.003.

[9] A. Aamer, L. P. Eka Yani, and I M. A. Priyatna, "Data analytics in the supply chain management: Review of machine learning applications in demand forecasting," *Operations and Supply Chain Management: An International Journal*, vol. 14, no. 1, pp. 1–13, 2021. DOI: 10.31387/oscm0440281.

[10] K. Bandara, P. Shi, C. Bergmeir, H. Hewamalage, Q. Tran, and B. Seaman, "Sales demand forecast in e-commerce using a long short-term memory neural network methodology," in *Proc. 26th Int. Conf. Neural Information Processing (ICONIP)*, Sydney, Australia, 2019, pp. 462–474. DOI: 10.1007/978-3-030-36718-3_39.

[11] J. Feizabadi, "Machine learning demand forecasting and supply chain performance," *International Journal of Logistics Research and Applications*, vol. 25, no. 2, pp. 119–142, 2022. DOI: 10.1080/13675567.2020.1803246.

[12] E. Hofmann and E. Rutschmann, "Big data analytics and demand forecasting in supply chains: a conceptual analysis," *The International Journal of Logistics Management*, vol. 29, no. 2, pp. 739–766, 2018. DOI: 10.1108/IJLM-04-2017-0088.

[13] M. M. Pereira and E. M. Frazzon, "A data-driven approach to adaptive synchronization of demand and supply in omni-channel retail supply chains," *International Journal of Information Management*, vol. 57, art. 102165, 2021. DOI: 10.1016/j.ijinfomgt.2020.102165.

[14] R. Y. K. Lau, W. Zhang, and W. Xu, "Parallel Aspect-Oriented Sentiment Analysis for Sales Forecasting with Big Data," *Production and Operations Management*, vol. 27, no. 10, pp. 1775–1794, 2018. DOI: 10.1111/poms.12737.