European Journal of Business, Economics & Management

Vol. 1 No.1 2025

Article Open Access

Enhancing Inventory Forecasting Accuracy and Optimization Using Machine Learning

Xetong Liu [1,*]

[1] Ira A. Fulton School of Engineering, Arizona State University, Tempe, Arizona, 85281, United States
[*] Correspondence: Xuetong Liu, Ira A. Fulton School of Engineering, Arizona State University, Tempe, Arizona, 85281, United States

Abstract: As enterprises increasingly prioritize refined inventory management, the application of machine learning for inventory forecasting and optimization has become increasingly critical. This paper systematically discusses the application of machine learning models in inventory management systems, and focuses on the analysis of ETL integration, model deployment, real-time learning, and other key technologies in the system. This paper explores methods to enhance the accuracy of inventory forecasting from three key perspectives: constructing time series model, exogenous influence and comparison of important methods. The forecast results are applied to variable safety inventory, intelligent replenishment strategy and inventory structure optimization respectively, and an inventory optimization plan based on the forecast results is constructed. The research findings offer valuable theoretical and technical insights for enterprises to enhance inventory management efficiency.

Received: 07 April 2025
Revised: 11 April 2025
Accepted: 25 April 2025
Published: 28 April 2025

Keywords: machine learning; inventory forecasting; inventory optimization; time series model; artificial intelligence

1. Introduction

In the context of intelligent supply chain development, inventory forecasting and optimization has become one of the key links for enterprises to improve operational efficiency. However, traditional inventory optimization methods exhibit significant limitations when dealing with complex, non-linear, and dynamically changing data. With the development of artificial intelligence and computer technology, the advantages of machine learning in the application of inventory optimization have been gradually revealed, it can not only improve the accuracy of prediction, but also put forward new solutions for the optimization of inventory structure. Based on the inventory management system, this paper proposes an inventory optimization management path, studies the intervention path and application method of machine learning model, and aims to build an effective and intelligent inventory forecasting and decision-making process to promote the intelligent data-based management of enterprises.

2. Core Technology of Machine Learning Model Integrated into the Inventory Management System

2.1. Integrate ETL and API to Realize Efficient Data Flow

An inventory management system encompasses a vast array of data, including sales records, purchase records, inventory fluctuations, and supply chain allocations, among others. In order for machine learning models to easily and quickly acquire and process these data, it is essential to establish a robust numerical integration system for these data. ETL serves as a critical tool for data extraction and preprocessing. Using the ETL process, the source data distributed in various systems such as ERP and WMS can be processed in a unified manner to clear anomalies and extract useful information, providing structured and standardized input data for subsequent model training and prediction. In addition, in order to make the model in the inventory management system can be applied in real time, but also need to introduce API interface, to achieve the model and service two-way communication. RESTful API (Representational State Transfer Application Programming Interface) is a mainstream form of communication technology, which can package the trained model into an independent service, allowing seamless integration for real-time predictions within the inventory management system [1]. After the prediction, the feedback can be fed back to the model through the API for the optimization and adjustment of the model. The docking of ETL and API can ensure the timeliness and reliability of model output, establish the whole process data path between the inventory system and the machine learning model, and help to establish a data-driven real-time closed-loop architecture and build an intelligent inventory management system. Figure 1 below shows the data flow structure of ETL and API integration in an inventory management system:

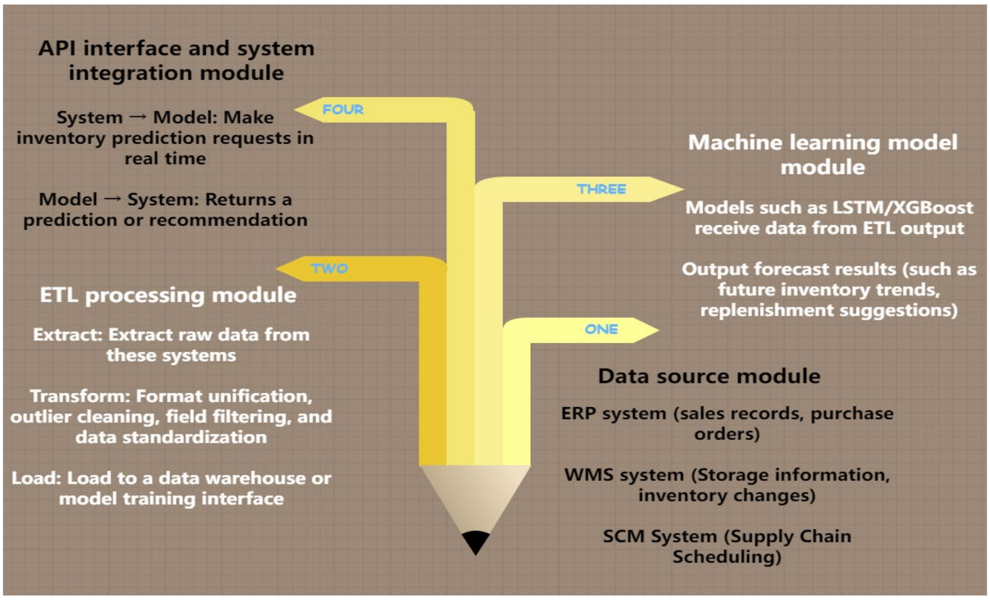

Figure 1. Data Flow Structure of ETL and API Integration in Inventory Management System.

2.2. Deploy Docker and Serving to Support Stable Model Calls

Inventory forecasting, during the deployment process, has high requirements for the response speed of the model and the quality of the integration of the business application system. In order to ensure that the predictive model can run stably in the actual deployment environment and can be quickly utilized, containerized deployment strategy should be adopted. Among them, Docker is a very small container platform, which can integrate the trained machine learning model and related environment into an image, which can prevent the unstable problem caused by environmental differences, and achieve the requirements of high system flexibility and fast deployment speed. In the specific deployment process, you can implement the integrated support of model version management, API invocation, and parallel processing through the TensorFlowServing model service framework [2]. TensorFlowServing has strong reasoning ability, which is very suitable for deploying the inventory forecasting model based on neural network. It can connect with the inventory management system through RestfulAPI and gRPC, realizing the requirement of real-time data input and predicted value output. Meet the business system's need for real-time accuracy [3].

In addition, by using the Kubernetes container scheduler, self-scaling and load balancing of model services can be achieved to ensure the system operates smoothly and without interruption during peak seasons. The deployment architecture is designed to effectively support and maximize the performance of the AI model, providing stable and controllable technology for the AI model to be used in inventory forecasting. Figure 2 below shows the containerized deployment architecture of the inventory forecasting model:

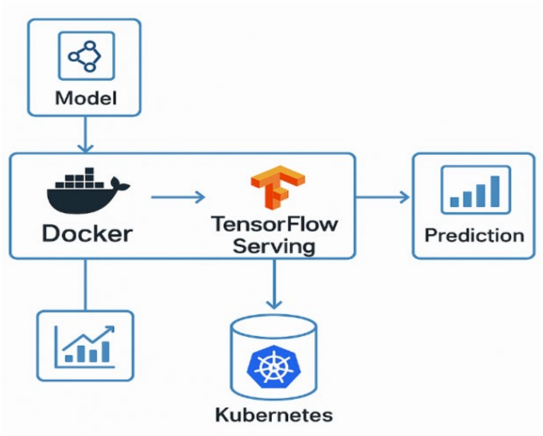

Figure 2. Schematic Diagram of Containerized Deployment Architecture for Inventory Forecasting Model.

2.3. Apply Online Learning Mechanism to Promote Continuous Optimization of Prediction

In the case of frequent inventory changes, the static mode of non-networked learning is difficult to adapt to the trend characteristics of continuous changes and the requirements of immediate data refresh. In order to strengthen the flexibility and practicability of the inventory forecasting model, it is necessary to introduce the online learning method, so that the model constantly draws the latest data from the running process and modifies its parameters, so as to ensure the stability and foresight of the forecasting effect. Online learning is one of the key technologies in the field of artificial intelligence, with real-time and incremental updating capabilities, suitable for continuous output of various data, and distributed in frequently changing work scenarios. In the inventory forecasting model, online learning method is adopted. When the system receives a new set of data of sales, inventory or environmental factors, it can automatically adjust the weight of the model without retraining the model, which can effectively reduce the update time. In addition, in order to prevent cumulative errors and bias adjustment problems, rolling window technology and dynamic weight adjustment mode can be used to ensure that the model is both time-efficient and anti-interference [4].

With the help of the Internet learning platform, the inventory forecasting system can also constantly adapt to changes in market demand, improve the foresight of prediction, improve the agility of business decision-making, and highlight the application value of AI in the intelligent transformation of inventory management. The inventory forecasting system framework is shown in Figure 3 below:

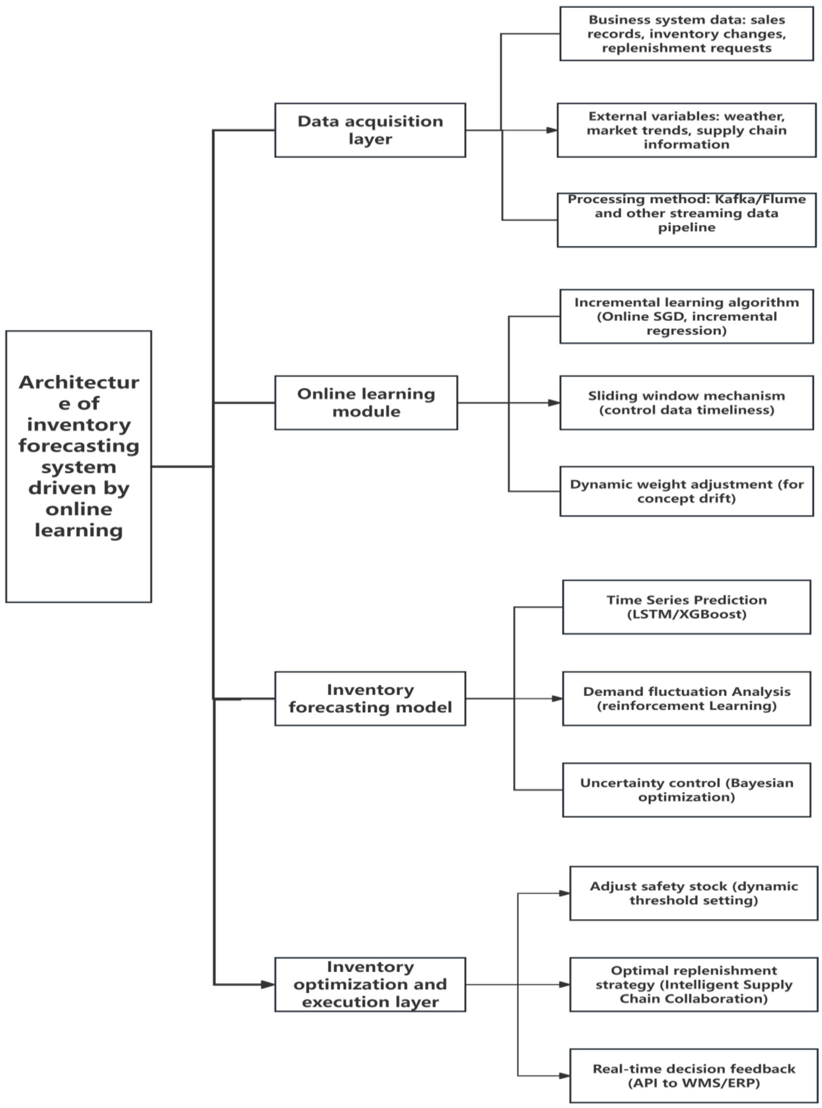

Figure 3. Architecture of Inventory Forecasting System Driven by Online Learning.

3. Machine Learning Serves as a Key Enabler for Enhancing the Accuracy of Inventory Forecasting

3.1. Build a Time Series Model to Improve the Accuracy of Trend Recognition

For inventory forecasting, the time series model can accurately grasp the change speed and track of the inventory sales volume, order quantity and other needs, as well as the order replenishment and other relevant data. Efficient methods such as ARIMA, Prophet and LSTM can be selected to build inventory forecasting models. LSTM is a kind of deep learning method, which can solve the problems of long correlation and nonlinear fluctuation well, and can establish the LSTM model when facing the complex demand in the process of inventory forecasting. Before building the model, it is necessary to monitor the stability of the previous inventory and sales data, and conduct trend analysis and cycle analysis to form a multi-step forecasting model. Then the sliding window method is used to process the input sequence data, and the reasonable time step number and prediction step length are set, so that the model has the ability to detect seasonal changes and sudden fluctuations [5].

The predictive performance of different time series models varies significantly. In order to verify the performance of each model in inventory forecasting practice, the historical sales dataset can be utilized as both the training set and the test set to evaluate and compare multi-step forecasting errors. Table 1 below shows the prediction accuracy evaluation results of the three models (ARIMA, Prophet, LSTM) under the same data set (evaluated by MAPE):

Table 1. Comparison of Accuracy of Different Time Series Models in Inventory Forecasting.

Model type	Model feature description	Average MAPE (%)
ARIMA	Linear model for stationary data	12.4
Prophet	Add holiday effect and fit flexibly	10.1
LSTM	Deep neural network, dealing with strong nonlinearity	6.8

As can be seen from Table 1, the LSTM model can predict and analyze complex inventory data more accurately, making it suitable for real-time and accurate inventory management systems.

3.2. Integrate External Influencing Factors to Strengthen the Demand Perception Effect

In inventory forecasting, using only historical sales data may not fully reflect changes in real demand. Incorporating exogenous variables such as promotions, festivals, and weather into the forecasting model can deepen the understanding of dynamic changes in demand, improve the model's rapid adjustment capability, and enhance the ability to cope with sudden and trend changes. The use of comprehensive information fusion modeling combined with machine learning feature engineering can effectively improve the adaptability of the model to cope with complex situations. In the process of feature modeling, the dimension of external variables is divided according to different time dimension attributes (such as holiday or promotion cycle), spatial dimension attributes and environmental dimension attributes (such as temperature and rainfall). After normalization and digital symbolization, the model is added as external features, and a multidimensional input vector is constructed together with the basic time series characteristics. It can greatly improve the information transmission efficiency. When applying deep learning models such as XGBoost and LSTM, external variables as auxiliary inputs can make the model better fit the complex consumption behavior model. Table 2 below reflects the prediction errors corresponding to different combinations of different input dimensions on the basis of LSTM, and MAPE (mean absolute percentage error) is selected to measure its quality:

Table 2. Improvement Effect of Fusion of External Influencing Factors on Prediction Accuracy.

Input feature combination	Description	Average MAPE (%)
Single historical volume	Only past sales data are used	10.6
Historical sales + promotional information	Add event variables such as product promotions and discounts	8.9
Historical sales + Promotions + Holidays	Add holidays and marketing cycles	7.5
All-variable fusion	Add weather, store location and other external factors	6.1

As illustrated in Table 2, the external environment exerts a measurable impact on inventory estimation. With the increase of the number of external parameters, the model error shows a gradually decreasing trend, indicating that the intelligent algorithm has the ability to integrate various information. Through this channel, the inventory system can grasp the demand environment and improve its coping ability.

3.3. Enhance Result Accuracy through Comparison with Leading Prediction Algorithms

Significant performance variations exist among different machine learning methods used for inventory forecasting, and model selection plays a crucial role in determining forecast accuracy. At present, the most commonly used representative algorithms mainly include linear regression, random forest, XGBoost, short-duration memory neural network, etc., each of which has its own applicability and generalization characteristics. By comparing the prediction performance of different models under the same data set and similar characteristics, it can provide a direction reference for selecting the best algorithm and improving the prediction ability of the inventory system. The linear regression model features a straightforward structure, making it particularly well-suited for scenarios where a clear linear relationship exists between the data and the range of variation is limited. On the basis of decision tree model, random forest improves the anti-interference ability of the model through integration, and can better deal with the environment with more complex features. XGBoost (eXtreme Gradient Boosting) is an enhanced tree model, which has advantages in nonlinear fitting, fast learning, and training, and has been widely used in structured data prediction tasks. Table 3 below shows the comparison of the ability of four typical algorithms to predict inventory tasks in the same historical sales volume and exogenous variable data set. The evaluation indicators used are MAPE, RMSE and model training time:

Table 3. Performance Comparison of Mainstream Forecasting Algorithms in Inventory Forecasting.

Algorithm type	**Model characteristics**	**MAPE (%)**	**RMSE**	**Training time (seconds)**
Linear regression	Linear fitting, simple structure	12.8	21.4	1.2
Random forest	Integrated tree structure to deal with nonlinearity	9.6	17.3	4.8
XGBoost	Enhanced tree model with strong fitting ability	7.2	14.9	6.1
LSTM	Deep learning, dealing with time dependence	6.4	13.5	18.7

As shown in Table 3, LSTM, a deep learning model, has the highest prediction accuracy, especially in inventory situations that require timely responses and processing of massive relationships. XGBoost can balance high prediction accuracy with faster training speed, benefiting large and medium-sized enterprises with more structured historical data.

4. Inventory Optimization Strategy Based on Machine Learning

4.1. Establish Dynamic Safety Stock Levels to Mitigate the Risks of Out-Of-Stock and Overstocking

In the process of storage management, it is difficult to deal with actual demand fluctuations and uncontrollable supply by setting a constant safety inventory, which may lead to stock shortages or delays. The machine learning method is used to predict the historical consumption and delivery cycle, which can accurately estimate the degree of future demand and the uncertainty of supply. In this way, the forecast error is combined with supply fluctuation through modeling, which is used as the reference data for dynamic safety stock setting.

Based on the forecast results of LSTM and XGBoost, the mean, standard deviation, and periodic change are used to calculate the dynamic safety inventory as follows:

$$SS = Z \times \sqrt{\delta \times {}^{2}_{D} \times L + {}^{-2}_{D} \times \delta^{2}_{L}} \quad (1)$$

In the formula, SS represents the safety inventory, Z represents the standard deviation of the normal distribution corresponding to the service level, σ_D represents the standard deviation of the demand forecast error, $\bar{D}$ represents the average demand, L

represents the average supply time, and σ_L represents the standard deviation of the average supply lead time. The model can revise the system safety inventory at any time according to environmental changes, enhance the stability of the system and adapt to risks.

4.2. Optimize Replenishment Decision-Making Strategy and Improve Inventory Turnover Efficiency

Replenishment decisions have a direct impact on the effectiveness of inventory management and inventory holding costs. The replenishment rule based on experience and established experience boundary cannot meet the demand under the condition of large demand fluctuation and high supply chain risk. Machine learning algorithms are needed to predict the changes of factors such as future demand, arrival time, and sales volume. In addition, the dynamic adjustment of replenishment time and quantity can be realized, and the predicted data of artificial intelligence model can be used to build a replenishment scheme based on service level and economic volume, which can greatly improve the replenishment efficiency.

Replenishment can be optimized using an EOQ model, taking full account of demand forecast, ordering costs, and carrying costs:

$$Q = \sqrt{\frac{2\times D_f\times S}{H\times(1-\alpha)}} \quad (2)$$

Where, Q is the optimal order quantity, D_f is the sum of future demand predicted by machine learning, S is the fixed cost per purchase expense, H is the cost per unit of item storage, and α is the compensation factor for the uncertainty based on the forecast, derived from the formulation of the forecast range. This control method can adjust the frequency and scale of replenishment and purchasing, which not only ensures stable supply, but also greatly improves inventory utilization.

4.3. Optimizing Inventory Structure for Efficient Resource Allocation

The reasonableness of an enterprise's inventory structure will affect the efficiency of resource utilization and the cash circulation of the enterprise. For different kinds of goods, multi-channel sales products, inventory in different regions, etc., the traditional storage mode cannot meet the storage requirements of different goods. Machine learning can be used to establish models that analyze sales, gross profit, cycles, etc., for intelligent inventory processing. Through cluster analysis, association rule discovery, and other AI techniques, product categories are subdivided, frequently purchased high-end products, seasonal products, unsalable products, etc., are identified, and then hierarchical management is implemented to improve the inventory structure.

In actual configuration, a multi-factor weighted inventory priority scoring system can guide resource allocation and storage planning and deployment.

$$P_i = w_1 \cdot \frac{S_i}{max(S)} + w_2 \cdot \frac{R_i}{max(R)} + w_3 \cdot (1 - \frac{C_i}{max(C)}) \quad (3)$$

In the formula, P_i is the i the product configuration priority, S_i represents sales proportion, R_i represents turnover, C_i is cost of ownership, and w_1 w_2 w_3 are weight coefficients. This model can dynamically configure the inventory structure and improve the scientific and operational efficiency of inventory resource allocation.

5. Conclusion

In the process of continuous improvement of inventory intelligence, it is an inevitable trend to use machine learning to accurately forecast demand and match inventory. This paper analyzes three aspects of AI: the demand prediction model, the optimization of the demand prediction model, and the construction of an accurate optimization ratio, fully demonstrating the specific application of AI in demand prediction. It provides both theoretical and technical support for improving the accuracy of inventory forecasting and enhancing inventory optimization and adjustment using machine learning. In the future, there is potential to combine online computing and deep reinforcement learning to further enhance adaptive and self-learning inventory decision-making.

References

1. T. Demizu, Y. Fukazawa, and H. Morita, "Inventory management of new products in retailers using model-based deep reinforcement learning," *Expert Syst. Appl.*, vol. 229, p. 120256, 2023, doi: 10.1016/j.eswa.2023.120256.
2. Y. M. Tang, K. Y. Chau, Y. Y. Lau, and Z. Zheng, "Data-intensive inventory forecasting with artificial intelligence models for cross-border e-commerce service automation," *Appl. Sci.*, vol. 13, no. 5, p. 3051, 2023, doi: 10.3390/app13053051.
3. B. Yang, X. Xu, Y. Gong, and Y. Rekik, "Data-driven optimization models for inventory and financing decisions in online retailing platforms," *Ann. Oper. Res.*, vol. 339, no. 1, pp. 741–764, 2024, doi: 10.1007/s10479-023-05234-4.
4. W. Ye, "E-commerce logistics and supply chain network optimization for cross-border," *J. Grid Comput.*, vol. 22, no. 1, p. 22, 2024, doi: 10.1007/s10723-023-09737-z.
5. C. Saha, D. K. Jana, and A. Duary, "Enhancing production inventory management for imperfect items using fuzzy optimization strategies and Differential Evolution (DE) algorithms," *Franklin Open*, vol. 5, p. 100051, 2023, doi: 10.1016/j.fraope.2023.100051.