International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 6797-6810

https://doi.org/10.70517/ijhsa463583

Machine learning implements a dual-objective coordination optimization model for enterprise supply chain demand and inventory costs.

Teng Zhang[1,*], Guoqiang Hao[1], Zhenhua Zhang[2], Chenyu Song[2] and Chenxin Cui[2]
[1] Economics and Management School, Taiyuan University of Technology, Taiyuan, Shanxi, 030024, China
[2] Software School, Taiyuan University of Technology, Taiyuan, Shanxi, 030024, China
Corresponding authors: (e-mail: 15031123257@163.com).

Abstract Insufficient accuracy of demand forecasting in current enterprise supply chain management leads to inefficient inventory management and high cost. In this study, we constructed an enterprise supply chain demand forecasting model based on the random forest algorithm and designed an inventory cost control system combined with the particle swarm optimization algorithm to solve the cost control problem caused by inaccurate forecasting in enterprise inventory management. The optimal feature subsets are screened by two feature selection algorithms, MCMR and rMCMR, and Random Forest is used to forecast the demand for three types of FMCG products in Company K. PSO-RF is used to optimize the inventory cost control. The results show that the random forest prediction model has the best prediction accuracy of 93.3% in the comparison of multiple classification algorithms, and the training time is only 26.148 seconds; in terms of inventory cost control, the fulfillment cost rate of Company K continues to decline after the adoption of big data technology, and reaches a historical low of 6.286% in 2021, with a significant increase in inventory turnover rate. Company K's user satisfaction score reaches 4.548, with a platform feedback rate of 100%, a comprehensive rating of over 0.9, and a “recommended order” rating. The study proves that Random Forest algorithm combined with PSO optimization can effectively improve the accuracy of enterprise supply chain demand forecasting, optimize inventory cost control, and enhance the operational efficiency and user satisfaction.

Index Terms random forest algorithm, supply chain, demand forecasting, PSO-RF model, inventory cost, feature selection

I. Introduction

Modern enterprise competition is more of a talent competition, technology competition, innovation competition, supply chain competition. The efficiency and capacity of the supply chain has a huge impact role in the process of enterprise business development, is the fundamental of its business development, and improve the competitiveness of enterprises [1], [2]. Supply chain management is an important part of current engineering management, covering all aspects of enterprise production, transportation, inventory and sales, which directly affects the efficiency and profit of enterprises [3]. Among them, supply chain demand forecasting is the core link of supply chain management, which directly affects the operational efficiency and cost control of the supply chain. Accurate demand forecasting can more accurately and quickly reflect the relationship between supply and demand under the market, so as to reduce the cost in the supply chain and improve the final cost-effectiveness [4], [5]. However, different supply chain nodes are usually managed by different enterprises. As a result, demand information is not shared among them, and even if the upstream manufacturer has a sales agent, the agent may send orders to the manufacturer that exceed the actual demand for its own profit and to avoid out-of-stock considerations [6]-[8]. This leads to frequent distortion and mutation of demand information in the process of transferring from the end to the first end of the supply chain, which amplifies demand fluctuations and forms the bullwhip effect [9]. The bullwhip effect reflects the asynchrony of demand in the supply chain, which reveals a common phenomenon in supply chain inventory management: “what is seen is not actual”. This effect leads to distortion of demand information, and the distorted information makes the members in the supply chain deviate from the prediction of the market and customers, which results in problems such as inventory backlogs, overproduction or underproduction, and increase in transportation costs in each link of the supply chain, which reduces the responsiveness, and it is not conducive to the establishment of cooperative partnerships within the supply chain, which leads to a decline in the profit of the whole supply chain, and causes a huge loss to the enterprise [10]-[14].

In addition, with the integration of the global economy, the competition of single product has evolved into the competition of supply chain. The profit space of enterprises has been squeezed continuously, and the space for digging out the potential of traditional management methods is limited, and more and more enterprise managers have begun to pay attention to the problem of inventory management [15]. The safety range of enterprise inventory exceeds the enterprise liquidity, which is easy to cause capital breaks, and under the demand of urgent or demand uncertain orders, the inventory and production coordination efficiency is low, and the response speed of supply chain is more than 20 days, which leads to the decline of enterprise supply chain efficiency [16]-[18]. With the rapid development of information technology and the application of the Internet, digitalization has become the trend of future enterprise development, and enterprise digital transformation has become an inevitable choice. Since then, supply chain management has also entered into digital management, through digital information, Internet of Things and artificial intelligence and other technologies to achieve the whole process of monitoring and automatic adjustment of the supply chain, not only to improve the management efficiency, transparency and controllability, but also to reduce the cost [19]-[21]. Therefore, how to utilize digital technology to effectively address the deficiencies in demand forecasting and inventory cost control, and improve the synergy and flexibility of the supply chain is an important research topic in the field of supply chain management.

Enterprise supply chain management as the core link of modern enterprise operation, its efficiency directly affects the overall competitiveness of enterprises. In the current e-commerce environment, enterprises are faced with massive data, rapid changes in user demand, increasingly fierce market competition and other challenges, the traditional supply chain management model has been difficult to adapt to the development needs of modern enterprises. Accurate demand forecasting is the basis of efficient supply chain management, while inventory cost control is a key part of enterprise cost reduction and efficiency. In recent years, the application of big data technology and artificial intelligence algorithms in the field of supply chain management has been deepening, providing a new technical path to solve the problems of insufficient demand forecast accuracy and improper inventory control. From the practice of e-commerce enterprises such as Company K, the use of advanced algorithmic technology to improve the accuracy of demand forecasting has become an important means for enterprises to reduce inventory costs and improve operational efficiency. As an integrated learning algorithm, Random Forest's insensitivity to multivariate covariance, outliers, and partially missing values gives it an obvious advantage in dealing with complex e-commerce data. In addition, current research mostly focuses on the application of a single algorithm, and lacks research on the systematic integration of feature selection, model optimization and inventory cost control. Practice shows that inventory cost is the main component of the operating cost of e-commerce enterprises, including procurement cost, in-stock maintenance cost, ordering cost and out-of-stock cost. Traditional management methods have led to frequent occurrence of overstocking or understocking due to insufficient forecasting accuracy, which in turn increases the overall operating costs of the enterprise.The financial data of Company K from 2015 to 2021 show that through technological innovation to optimize supply chain management, its revenue grows from 1,864 to 9,548, its gross profit grows from 249 to 1,295, and inventory turnover continues to improve, and the proportion of fulfillment cost to net revenue decreased significantly, indicating the important value of technological innovation in supply chain management.

Based on the above background, this study proposes a research idea of optimization of enterprise supply chain demand forecasting and inventory cost control based on random forest algorithm. First, the supply chain data are feature screened by two feature selection algorithms, MCMR and rMCMR, so as to determine the key factors affecting demand forecasting. Second, a random forest demand forecasting model is constructed based on the screened feature data and compared with XGBoost, LightGBM and CatBoost algorithms to evaluate the model performance. Again, particle swarm optimization algorithm is introduced to optimize the parameters of the random forest model, and PSO-RF model is constructed for inventory cost control. Finally, the practical application effect of the model is evaluated from four dimensions: purchasing cost, in-stock cost, ordering cost and total inventory cost. This research framework not only improves the accuracy of enterprise demand forecasting, but also optimizes inventory cost control at the system level, providing theoretical support and practical guidance for enterprise supply chain management.

II. Enterprise supply chain demand forecasting model based on random forest algorithm

II. A.Random Forest Algorithm

Random Forest (RF) is a flexible and highly accurate integrated learning algorithm proposed by combining mutually independent decision trees based on the self-sampling technique in Bagging algorithm [22]. Compared with the traditional algorithm, Random Forest has many advantages, which is insensitive to multivariate covariance, individual outliers, some missing values, and can also solve the problems of overfitting of the tree model, and can

still maintain a high accuracy rate in dealing with complex data. In practice, Random Forest is mainly used for classification and prediction, which is divided into Random Forest Regression Model (RFR) and Random Forest Classification Model (RFC), and now the Random Forest algorithm is widely used in many fields such as economics and medicine.

The essence of a random forest is to aggregate multiple tree models $\{h(x,\theta_k),k=1,2,3,\cdots\}$, with x as the given input vector and $\{\theta_k\}$ as an independently distributed random vector. The idea of constructing a random forest is specified as follows:

Step 1: For the training set $D=\{(x_i,y_i),x_i\in R^d,y_i\in R,1\le i\le n\}$ use bootstrap self-help sampling to extract K self help sets $D_1,D_2,\cdots,D_k$, where the self-help sets satisfy mutual independence and are of the same size as D.

Step 2: Each self-help set is trained to grow into a single tree model. A tree is generated by randomly picking d_0 features from d feature variables at each node of the tree and selecting the feature with the smallest impurity as the split node.

Step 3: Generate K decision trees $T_1,T_2,\cdots,T_k$ in K self-help sets, predict the required data according to the generated tree model, and then the majority vote integrates the Random Forest model to output the final prediction results.

Random forest algorithm commonly used Gini coefficient and information gain two methods to select the optimal split features.

(1) Gini coefficient selects the optimal classification features [23]:

$$Gini(D,S)=\sum_{i=1}^{n}\frac{|D_i|}{|D|}Gini(D_i) \tag{1}$$

$$Gini(D_i)=1-\sum_{j=1}^{p}\left(\frac{|D_{ij}|}{|D_i|}\right)^2 \tag{2}$$

where D is the dataset, $|C_i|$ is the number of samples, S is the features, n is the number of features, and p is the number of categories.

(2) Selection of optimal categorization features by information gain

Calculate the information entropy of dataset D, where $|C_i|$ is the number of samples corresponding to category C_i:

$$H(D)=-\sum_{i=1}^{p}\frac{|C_i|}{|D|}\log\frac{|C_i|}{|D|} \tag{3}$$

Calculate the conditional entropy of d_0 randomly selected features with respect to D, where $H(D_i)$ is the information entropy of D_i:

$$H(D\mid S)=-\sum_{i=1}^{n}\frac{|D_i|}{|D|}H(D_i) \tag{4}$$

$$H(D_i)=-\sum_{j=1}^{p}\frac{|D_{ij}|}{|D_i|}\log\frac{|D_{ij}|}{|D_i|} \tag{5}$$

Calculate the information gain of each of the d_0 features and then select the feature with the largest information gain for splitting [24]:

$$g(D,S)=H(D)-H(D\mid S) \tag{6}$$

II. B.Random Forest-based Supply Chain Demand Forecasting

II. B. 1) Selection of supply chain characteristics

For the massive data of e-commerce enterprises, the data often has the characteristics of high dimensionality, sparse data, low value density, etc. In doing classification regression problems, the regression problem often requires the data set to have the characteristics of high density, however, classification usually does not need to have a high-density data set in the feature space. The main reason is that the regression problem needs a curve to fit all the data points as much as possible, so the density of the training data in the feature space is more

demanding than the classification problem. The multi-classification problem seeks several hyperplanes to differentiate the data as much as possible, so high-dimensional data are often more suitable for classification problems.

(1) Regression problem transformed into classification problem

Based on the above analysis, the regression problem of commodity demand prediction needs to be transformed into a classification problem. According to the business scenario analysis of reasonable data set category label transformation, the specific ideas are shown below:

Assume that there exists a formula T with commodity demand labeled a.

Step1: Define $b = T(a)$.

Step2: Set the threshold, split the dataset in some way, divide the dataset into n classes, and transform the labels into: $1:n$.

(2) MCMR feature selection

The dataset obtained by transforming the label in the previous step is subjected to feature selection, and the specific steps of MCMR algorithm for feature selection of demand forecast data are shown below:

1) Normalize the data to eliminate the influence of the magnitude between the features, assuming that the feature variable is H, H_i represents the i th value of the feature variable H, and the normalization formula is: $H^* = \frac{H_i - \min H}{\max H - \min H}$, and the normalized dataset is noted as D.

2) Calculate the linear correlation coefficient r and nonlinear correlation coefficient R_n^* between features in the dataset D using Pearson and COS methods, calculate the full correlation coefficient w between features using exponential scaling, and calculate the correlation $sim(X,Y)$ between features and categories using the information gain rate method.

3) Using Random Forest (RF) as a classifier, introduce the weight parameters α, β to set the initialization parameter $\alpha = 0$, $\beta = 1\{\alpha \in [0,0.4]、\beta \in [0.6,1]\}$, the step size of each iteration is 0.1, and the optimal feature subset d_1 is selected by RF classification accuracy.

4) Output the optimal feature subset d_1.

MCMR algorithm for demand forecast data feature selection specific process is shown in Figure 1:

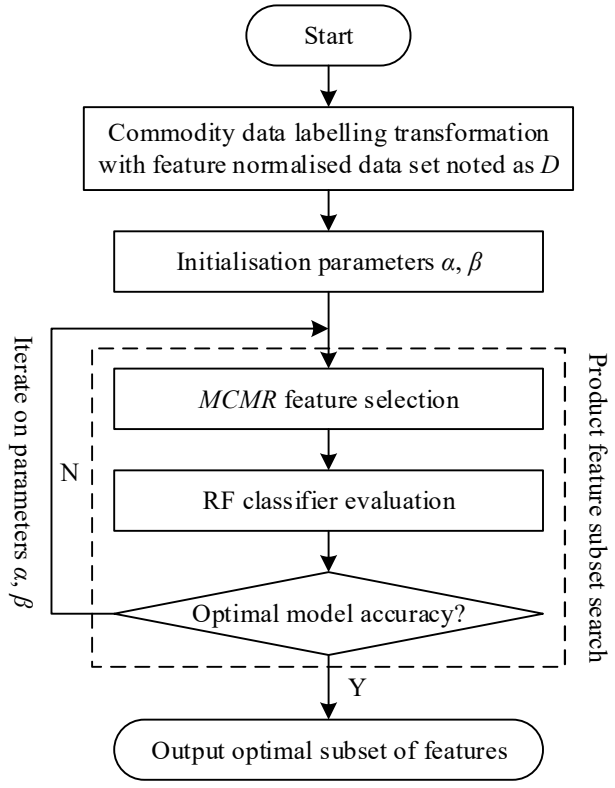

Figure 1: The MCMR product feature selection process

(3) rMCMR feature selection

The rMCMR feature selection method is utilized for the selection of commodity demand data, the first 2 steps are the same as the MCMR algorithm feature selection, and the different steps after that are shown below:

1) Use Random Forest (RF) as a classifier, set the initialization parameter $\eta = 0\{\eta \in [0.6,1]\}$, the step size of each iteration parameter 0.1, and select the optimal subset of features d_2 by RF classification accuracy.

2) Output the optimal feature subset d_2.

(4) Comparison of feature selection methods

The optimal feature subset d_1 selected by MCMR algorithm and the optimal feature subset d_2 selected by rMCMR algorithm are used in steps 2 and 3, and the accuracy of the optimal feature subset selected by the two algorithms is compared on the random forest in this section, and the one with larger accuracy is selected as the final feature subset. Meanwhile the comparison of the two algorithms verifies the performance ability on large datasets.

II. B. 2) Supply chain demand forecasting models

(1) Model building

Using the optimal feature subset selected by the feature selection algorithm, the optimal feature subset is divided into a test set and a training set, and the training set data is used as an input to the XGBoost, RF, GBRT, Rule, and SVR supply chain demand prediction models, and the five optimal training models are obtained by adjusting the parameters of the models iteratively [25]. The test set data is fed into the trained models to output the value of the future demand of the commodity.

(2) Model Fusion

In integrated learning, multiple sets of weak learners are integrated to obtain a model that performs better than any single model through some transformation. The better the generalization ability of the individual model, then the integrated model generalizes better. Using the idea of integrated learning models, the final prediction model is obtained by fusion of multiple sets of disparity models. The model fusion formula is shown in Equation (7), where the parameters x_1, x_2, x_3, x_4 are the model fusion coefficients:

$$M1 = x_1 * XGBoost + x_2 RF + x_3 GBRT + x_4 SVR \tag{7}$$

II. C.Analysis of Random Forest-based Supply Chain Demand Forecasting Results

II. C. 1) Food and Beverage FMCGs

This subsection of the paper utilizes the random forest model to forecast the demand for three categories of FMCG products of Company K. The optimized parameters of the random forest model are utilized to train and predict the FMCG products of Company K in the categories of food and beverages, personal care FMCG products, and other daily necessities. After the optimization of the random forest parameters, the test set data of the three categories of FMCG products of Company K are brought into the model to obtain the prediction results. Figure 2 shows the prediction effect of three types of FMCG products in Company K.

Calculating the average absolute error and the mean square error of all kinds of FMCG products in Company K, the average absolute error of FMCG products in the food and beverage category is 1807.548, and the mean square error is 5241484.249. At present, there is not much difference between the actual sales and the demand predicted by the random forest, and the mean values are 16085.715 and 13613.216, respectively.

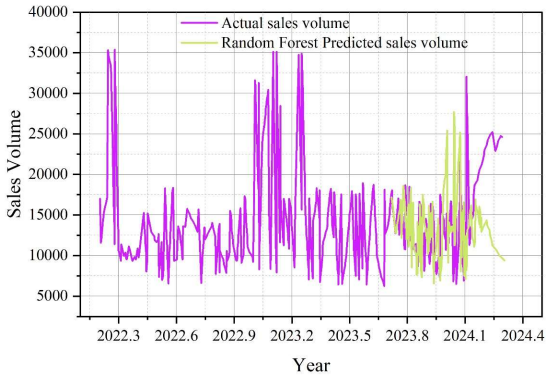

Figure 2: The prediction effect of three kinds of fast elimination products in K companies

II. C. 2) Personal Care

Figure 3 shows the effect of random forest prediction for personal care FMCG, the actual sales mean is 7515.458 and the predicted demand is 7345.964. The standard deviation is 2974.496,2954.425, respectively.

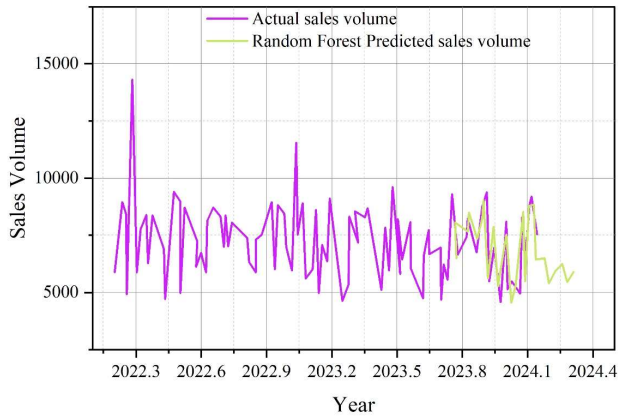

Figure 3: The results of the random forest prediction of personal care

II. C. 3) Other daily necessities

Figure 4 shows the predicted effect of FMCG products in the category of other daily necessities, the mean value of actual sales is 8702.558, and the predicted demand of Random Forest is 7409.408, and the standard deviation of the two is 3072.707 and 3004.536, respectively, and the mean squared error is 93,536,645.551.

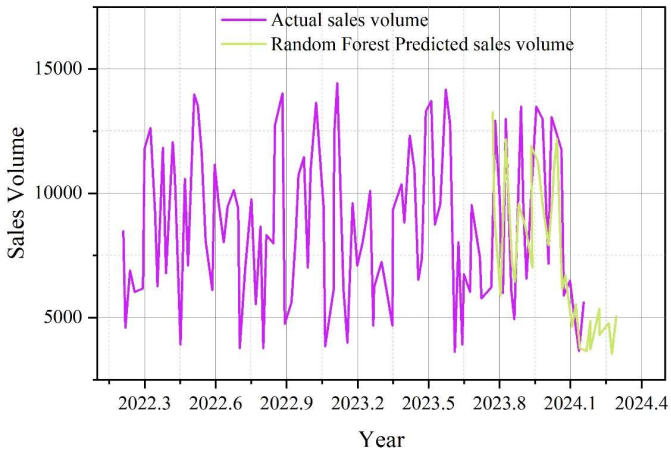

Figure 4: Other daily necessities are predicted

II. D.Evaluation and analysis of the effectiveness of demand forecasting

II. D. 1) Selection of indicators for evaluating predictive effects

In dichotomous model evaluation, a series of metrics are often used to measure model performance. These indicators reflect the overall prediction accuracy of the model and facilitate intuitive comparison and analysis.

Accuracy rate: its formula is shown in equation (8), by calculating the ratio of the number of correctly classified samples to the total number of samples, the accuracy rate of the model on the classification task is obtained, which in turn assesses the model performance advantages and disadvantages:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

Accuracy rate: its formula is equation (9), which quantitatively assesses the accuracy rate of the model by calculating the ratio of the number of true cases to the total number of samples predicted to be positive cases:

$$\Pr ecision = \frac{TP}{TP + FP} \tag{9}$$

Recall: recall whose formula is shown in equation (10), the recall of the model can be obtained by calculating the ratio of the number of samples of actual positive cases predicted as positive cases to the total number of samples of actual positive cases:

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \tag{10}$$

II. D. 2) Accuracy analysis

The experiment does classification prediction of regional supply chain demand orders by using classification algorithms of Random Forest, XGBoost, LightGBM and CatBoost, and then compares them according to the model effect, according to the setting in the order data, supply chain order is set to 1 and supply chain's order is set to 0, and then this column of data is set to be predicted, and through the model learns the relationship between other feature columns and relationship between this target column, and finally get the prediction results and evaluation index results for the test set prediction.

The prediction accuracy of each region is shown in Fig. 5, from which it can be observed that the two models, RF and CatBoost, perform better in terms of prediction accuracy, and the effect between them is similar, with a high degree of overlap of the data points, in which the RF model fluctuates between 0.85 and 1, with an average value of 0.933. This phenomenon suggests that both models are able to effectively capture the key information in the data and generate more accurate predictions when dealing with the prediction task in a specific region. Further analysis reveals that the accuracy trends in each region also show some similarity. This phenomenon may be related to the amount of supply chain demand data in each region. The size of the supply chain data volume directly affects the amount of information that the model can acquire during the training process, which in turn affects the prediction performance of the model. When the supply chain data volume is large, the model can learn more features and laws from it, thus improving the prediction accuracy. Conversely, when the supply chain data volume is small, the model may not be able to fully learn the intrinsic structure of the data, leading to a decrease in prediction performance.

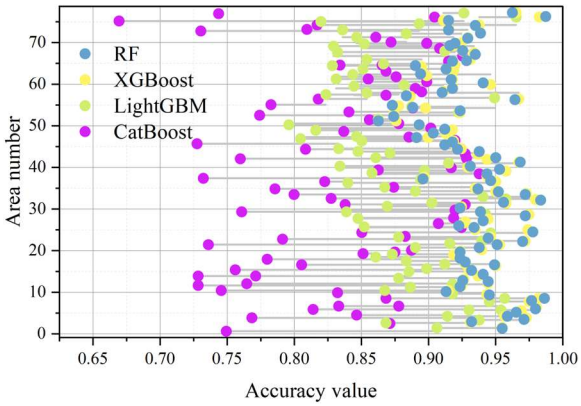

Figure 5: Accuracy of various models in each district domain

II. D. 3) Analysis of model prediction results

Table 1 shows the analysis of the model prediction results, the prediction accuracy of RF reaches 93.3%, the best effect, and the training time is also shorter, only 26.148s. XGBoost training accuracy is not much different from RF, and the model effect is also relatively good.

Table 1: Model prediction analysis

Model name	Accuracy	Precision	Recall	Training time
RF	0.933	0.999	0.934	26.148s
XGBoost	0.922	0.984	0.778	26.898s
LightGBM	0.881	0.914	0.934	33.469s
CatBoost	0.839	0.948	0.189	108.798s

III. Inventory cost control based on PSO-RF modeling

III. A. PSO-RF modeling approach

III. A. 1) Particle Swarm Algorithm

(1) Introduction of the algorithm

The PSO algorithm initializes the fitness function and then randomly places the particles in the search space, and then finds the optimal solution by constantly trying to iterate. In each iteration, the particles consider both the individual's own best known position and the best known position in the search space when moving, and eventually the population of particles will be pushed to the optimal solution.

(2) Algorithm parameters and formulas

There are N particles in D-dimensional space. The particle i position is $x_i=(x_{i1},x_{i2},\cdots x_{iD})$, the adaptation value can be found by substituting x_i into the adaptation function $f(x_i)$, the particle i velocity is $v_i=(v_{i1},v_{i2}\cdots v_{iD})$, and the best position that the particle i individual has experienced is $pbest_i=(p_{i1},p_{i2},\cdots p_{iD})$, and the group experiences the best position as $gbest=(g_1,g_2,\cdots g_D)$.

The d th dimension velocity update formula for particle i is shown below:

$$v_{id}^{k}=wv_{id}^{k-1}+c_1r_1\left(pbest_{id}-x_{id}^{k-1}\right)+c_2r_2\left(gbest_d-x_{id}^{k-1}\right) \tag{11}$$

The formula for updating the d th dimension position of particle i is shown below:

$$x_{id}^{k}=x_{id}^{k-1}+v_{id}^{k-1} \tag{12}$$

where v_{id}^{k} is the d-dimensional component of the velocity vector of the i-flight of the particle in the k th iteration, x_{id}^{k} is the d-dimensional component of the position vector of the particle in the k th iteration, c_1, c_2 are acceleration constants to regulate the maximum step of the learning, and r_1, r_2 are two stochastic functions with the value range of $[0,1]$ to increase the search randomness. w is the inertia weight, which regulates the search range for the solution space, and the inertia weight w indicates to what extent the particle retains its original velocity. Larger w has strong global convergence and weak local convergence, smaller w has strong local convergence and weak global convergence.

(3) Steps of the algorithm

Step 1: Set up the particle population randomly in the space, and the main parameters are particle swarm population size, particle dimension, particle random initial position and velocity, and the maximum number of iterations $G_{\max}$.

Step 2: According to the fitness function, traverse all particles and evaluate the fitness of the initial particles.

Step 3: Compare the fitness value calculated for the current position of each particle with the corresponding fitness value of its own best known position $(pbest)$, and if the current fitness is higher, update the regional optimal solution $pbest$ to the current particle position.

Step 4: Compare the adaptation value calculated from the current position of each particle with the corresponding adaptation value of the global best known position $(gbest)$, and if the current adaptation is higher, update the global optimal solution $gbest$ as the current particle position.

Step 5: Update the velocity and position of each particle according to Eq. (11) and Eq. (12).

Step 6: If the final end condition is not reached, return to step 2 to continue the iterative calculation, usually the algorithm stops when the algorithm meets the target error criterion or when the maximum number of iterations is reached.

The specific algorithm flow is shown in Figure 6:

III. A. 2) Random forest cost control

(1) Decision Tree

Decision tree is the building block of RF algorithm. Decision tree is a decision support tool that forms a tree-like structure. Decision tree consists of three parts: decision node, opportunity node and final node. Its opportunity nodes represent tests of data attributes, each branch represents a test result, and each final node represents a final result marker. The decision tree algorithm divides the training dataset into branches which are further divided into other branches. The sequence continues until the final node is obtained. The final node cannot be further separated to yield the final result.

(2) RF Algorithm

Bagging algorithm is one of the typical integration methods for handling large and high dimensional data, which uses self-sampling technique to obtain multiple different versions of the original sample set by random sampling. RF algorithm i.e., Bagging method is used to generate the desired prediction. The specific steps are as follows:

Step 1: Using Bagging method sampling, generate T training sets randomly from the sample set respectively $S_1, S_2, \cdots, S_T$.

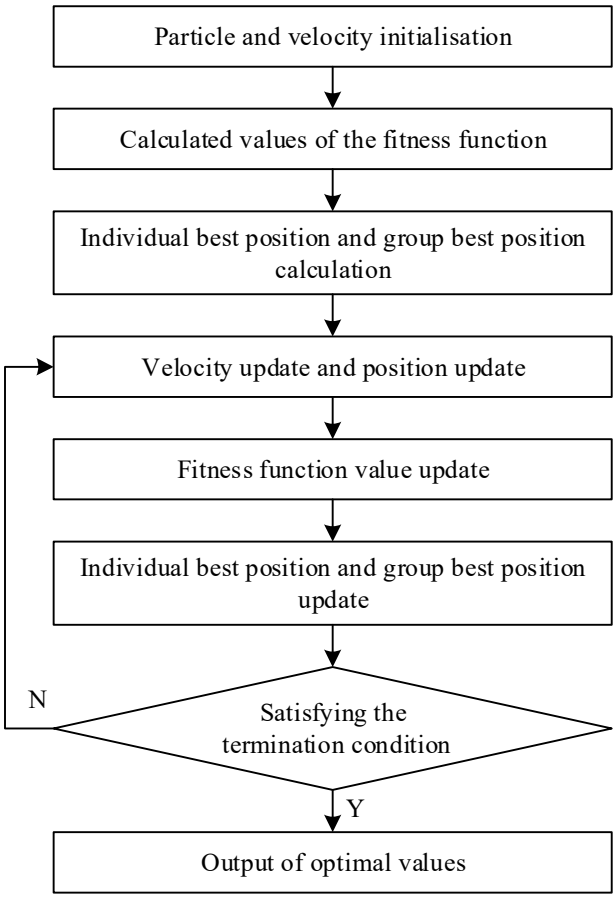

Figure 6: Particle swarm optimization algorithm process

Step 2: Grow the corresponding decision trees $C_1, C_2, \cdots, C_T$ on each training set. Before selecting attributes to split and grow on each opportunity node, m attributes in the M attribute set are randomly selected as the set of split attributes for the current opportunity node in order to split the node using the best set of attributes on each node.

Step 3: Import the test set samples x into the trained decision tree for testing and obtain the corresponding results $C_1(x), C_2(x), \cdots, C_T(x)$.

Step 4: The average of all the outputs of T decision trees is used as the predicted value of the test set sample x.

(3) Model Hyperparameters

The RF model contains a number of important hyperparameters, including: the number of trees $n\_estimators$ constructed by the algorithm before averaging the predictions, the maximum number of features $max\_features$ for the RF to consider splitting the nodes, and the minimum number of leaves $mini\_sample\_leaf$ required to determine the split internal nodes. Some of the samples are not used in the training data, but are used to evaluate its performance, these samples are called out-of-bag samples oob_score, and the error calculated from the out-of-bag samples is called out-of-bag error. These parameters are mainly used to evaluate the model performance, enhance the model prediction ability, and improve the model calculation speed.

(4) Feature Importance Calculation

In RF algorithm, the importance calculation of a feature M can be realized, the specific operation steps are as follows:

Step 1: In the RF framework, consider the out-of-bag error as the importance metric based on the feature M randomly ranked in the out-of-bag sample oob_score, and compute the out-of-bag data error, denoted as $errOOB_1$.

Step 2: Randomize the noise interference in all the out-of-bag samples oob_score of feature M, and again compute the out-of-bag error in the RF framework, denoted as $errOOB_2$.

Step 3: If there is an $Ntree$ tree in the RF, then the importance for feature M can be computed according to Equation (13):

$$I_M = \sum \left(errOOB_2 - errOOB_1 \right) / Ntree \tag{13}$$

III. A. 3) PSO-RF

Since the prediction accuracy of RF model will be affected by two parameters, the number of nodes at each split node mtry and the number of decision trees ntree, we consider applying PSO algorithm to optimize the two parameters of RF to improve the applicability and stability of the model. The specific steps are as follows:

Step 1: Data preprocessing. Randomly divide the data into training group and test group, according to the previous model constantly verified, the algorithm shows the highest accuracy when the ratio of training group and test group is 9:1, so 90% of the sample data is randomly selected as the training group, and 10% of the sample data as the test group.

Step 2: Initialize the PSO and RF parameters and construct the PSO-RF model according to the initialized data. In the PSO algorithm, set the maximum number of iterations, the number of populations, the range of values, the range of search speeds, the initial particle position and speed. In the RF algorithm, initialize the number of trees ntree and the number of nodes mtry.

Step 3: Iteratively update the velocity and coordinates of the particles, update the parameters ntree and mtry, and then calculate the corresponding fitness values and compare the optimal fitness values of the updated individuals and the whole to determine the optimal parameters ntree and mtry.

Step 4: The optimal training subset is extracted according to the feature importance and the optimized parameters are introduced into the RF model to constitute the PSO-RF model.

Step 5: The screened training group and test group are input into the PSO-RF model for training, and the error analysis is carried out by using the three indexes of RMSE, MAE, and MAPE, and stops when the model reaches the required accuracy.

With the development of machine learning algorithms in the field of prediction and engineering cost management, PSO-RF model is selected as the main prediction model in this paper. The model is suitable for diverse data sets from different sources and is good at dealing with elastic data structures. Supply chain price information from various sources, data information including different kinds of prices, various types of impact factor indicators, and many years of work experience can provide enough sample data, can meet the data requirements of the model, to achieve an effective combination of reasonable prediction.

III. B. Control based on PSO-RF inventory costs

In this paper, it is assumed that the cost of inventory based on the fixed interval replenishment strategy consists of purchasing cost, in-stock maintenance cost, ordering cost, and out-of-stock cost. The calculation of each cost is explained below.

(1) Purchasing cost

Purchasing cost is the capital expenditure incurred as a result of purchasing goods. The selling prices of the SKUs used in the numerical analysis are known, and in this paper, we assume that the cost of purchasing a unit of SKU at the purchase entry price is α times $(0 < \alpha < 1)$ its labeled unit price $price_{sign}$, and that the cost of purchasing is if the purchased entry quantity is Q:

$$C_{buy} = price_{sign} \times \alpha \times Q \tag{14}$$

(2) In-stock maintenance costs

Inventory maintenance cost is the cost of storage, maintenance and other costs incurred by the goods stored in the warehouse. At the beginning of the t period for the purchase of the number of SKUs into the warehouse, and the beginning of the period in the warehouse of the remaining inventory $Inventory_{surplus}$ constitutes the t week to maintain the inventory. Let the weekly in-stock maintenance cost per SKU be β times $(0 < \beta < 1)$ of its labeled unit price, and the in-stock maintenance cost is calculated as:

$$C_{stock} = \left[Inventory_{surplus} + Q \right] \times price_{sign} \times \beta \tag{15}$$

(3) Ordering costs

Ordering cost is the cost incurred because of inquiry, communication and negotiation, organization of logistics and transportation. Let the ordering cost be k times the purchase cost, $(0<k<1)$. The formula for ordering cost is:

$$C_{order} = price_{sign} \times \alpha \times Q \times k \tag{16}$$

(4) Out-of-stock cost

Out-of-stock cost is the loss of sales amount due to ordering too little quantity Q into the warehouse, so that the real demand in the t th week can not be met. Let the out-of-stock loss per unit SKU be ω times $(0<\omega<1)$ of its real selling price $price_{sale}$, so the formula for the out-of-stock cost for week t is:

$$C_{short} = \left[Y - Inventory_{surplus} - Q \right] \times price_{sale} \times \omega \tag{17}$$

where Y is the true demand in week t.

In summary, the total inventory cost for the i th SKU in week t is given by:

$$C(t,i) = C_{buy}(t,i) + C_{stock}(t,i) + C_{order}(t,i) + C_{short}(t,i) \tag{18}$$

Since there are N SKUs in total, each SKU has T weeks of data. So for each model M, the total cost of inventory based on replenishment against demand forecasts is:

$$C_{toal}(M) = \sum_{i=1}^{N} \sum_{t=1}^{T} \left[C(t,i) \right] \tag{19}$$

III. C. Evaluation of supply chain cost control effect

III. C. 1) Effectiveness of procurement cost control

This chapter mainly selects financial and non-financial data from 2015 to 2021 to analyze the effectiveness of internal and external supply chain inventory cost control of Company K from four aspects.

Based on the data of the annual corporate report published by Company K, its main sources of revenue are sales revenue and third-party platform service fee revenue. The operating costs mainly include purchasing costs and operating expenses. As a retail e-commerce enterprise, Company K's business model is different from that of the traditional manufacturing industry, and its cost of sales does not cover the production and processing process of goods, but refers specifically to the procurement cost of goods.

Table 2 shows some of Company K's profits from 2015-2021. although Company K's net profits are mostly negative, Company K's revenues and gross profits have been growing steadily over the past seven years, with revenues growing from 1,864 to 9,548 and gross profits from 249 to 1,295, an increase of 4.201 In the 2022 annual report, the disclosure shows that Company K's revenues have surpassed the “trillion mark”. This is all related to the fact that Company K responds to the Internet era and actively utilizes big data technology to control costs.

There are many factors affecting the cost of commodity procurement, but a detailed report on procurement costs was not found in Company K's public financial reports. Therefore, according to the familiar operation mode of Company K, the pricing of its products is usually the purchase price of the product plus the gross profit, and Company K boasts “the lowest price on the net”, which means that they will try their best to provide the most competitive price among their peers. Therefore, when evaluating the cost control effect of Company K's purchasing process, the control effect of purchasing cost is analyzed by simply calculating the proportion of purchasing cost to revenue and then comparing it to the gross profit margin indicator.

Table 2: From January 2015 to 2021, the company is a member of the K company

/	2015	2016	2017	2018	2019	2020	2021
Operating income	1864	2648	3648	4568	5975	7456	9548
Operating cost	1578	2246	3169	3953	4926	6315	8236
Gross margin	249	348	503	648	853	1098	1295
Technical r&d cost	-	30	68	120	149	163	165
Management fee	35	65	120	185	200	224	286
Net profit	-95	-35	-1189	-30	120	496	-50

III. C. 2) Inventory costs

Figure 7 shows the inventory turnover ratio of Company K from 2015-2021, which clearly shows that the number of warehouse employees of Company K has been increasing year by year, and it has been increasing until 2018 when it reached a peak of 33,485 employees. However, compared with 2018, the number of employees decreased significantly in 2019 and gradually decreased in the following years. This is mainly due to the fact that in recent years, K has focused on technological development and investment to improve warehouse intelligence. In particular, the adoption of unmanned warehousing technology has significantly reduced the reliance on labor, which in turn has reduced labor costs and inventory management costs. At the same time, the adoption of such technology has also improved the efficiency of warehousing and reduced inventory costs, which has helped Company K to improve its competitiveness in the e-commerce market.Prior to 2019, Company K's inventory turnover ratio was lower than that of Company S. The inventory turnover ratio of the two companies was 9.615 in 2019, and has been higher than that of Company S since 2019.

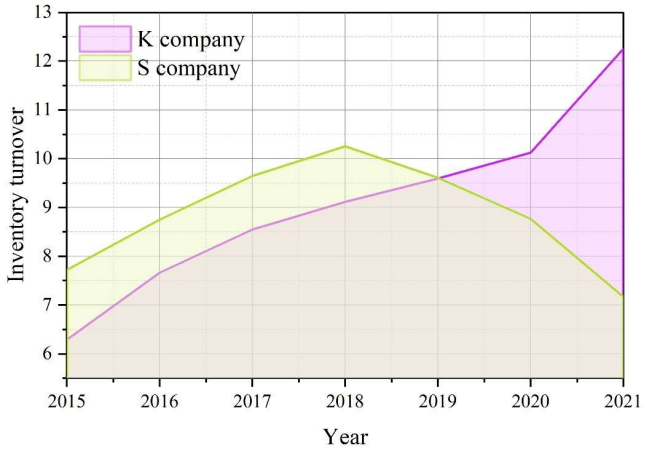

Figure 7: The turnover rate of the inventory of the K company 2015-2021

III. C. 3) Ordering costs

Figure 8 shows that Company K's fulfillment costs are mainly derived from logistics-related costs, and Company K's fulfillment costs as a percentage of net revenue from 2015 to 2021 have clearly shown a downward trend and gradually increased, with the fulfillment cost ratio reaching 6.286 in 2021, a historical low. The continued low level of the fulfillment cost ratio indicates that the application of big data technology in Company K has had a significant effect on the efficiency of logistics cost control.

Company K's address database consists of data provided by multiple sources, including merchants, suppliers, logistics companies and addresses added by users themselves. According to K Company's public data, as of 2021, the address database has covered the country's provinces, cities, counties, villages and other levels, and contains more than 1.4 billion pieces of address information.

Figure 8: The performance of K companies mainly comes from logistics related expenses

When a user places an order, Company K's logistics system, supported by big data technology, is able to respond quickly and intelligently match the user's order. Although K Company Logistics delivers more than 23 million parcels per day on average, its own logistics fleet size of more than 8,000 vehicles, with more than 200,000 delivery staff, big data technology can calculate the best route and time for the delivery staff and vehicles, thus improving the efficiency and quality of delivery. Company K Logistics has significantly improved its efficiency in warehousing, transportation, sorting and distribution through automated handling robots, sorting robots, intelligent express vehicles, drones, etc. These optimization measures have helped Company K to better control its logistics and distribution costs.

III. C. 4) Total inventory costs

Table 3 shows the rating indicators of e-commerce in 2021, in 2021, Company K's user satisfaction, platform feedback rate, response timeliness indicators are higher than other e-commerce companies in the same industry, the indicators were 4.548, 100%, 0.998, and the overall evaluation is more than 0.9. Company K is also evaluated by the consumers as a “recommended to place an order” rating. This shows that Company K's consumer experience is very good, and it also shows that Company K's “intelligent marketing”, which utilizes technological advances and consumer upgrades to satisfy more diversified shopping experiences of consumers, has achieved obvious results and reduced the cost of customer retention.

Table 3: The e-commerce rating index of 2021

/	E-commerce	User satisfaction	Platform refeed(%)	Response time	Composite index	Rating
1	K company	4.548	100	0.998	0.904	Suggested order
2	S company	3.924	100	0.975	0.869	Suggested order
3	H company	3.863	85.348	0.824	0.748	Suggested order
4	W company	4.139	100	0.993	0.896	Suggested order
5	T company	0.248	0.488	0.018	0.198	No order

IV. Conclusion

The research on the optimization of enterprise supply chain demand forecasting and inventory cost control based on random forest algorithm has achieved remarkable results. The supply chain data were optimally screened by MCMR and rMCMR feature selection algorithms, and a high-performance random forest demand forecasting model was constructed. The model prediction results show that in food and beverage FMCG, the mean value of actual sales is 16085.715, and the mean value of random forest predicted demand is 13613.216, and the prediction results are close to the actual values; in personal care FMCG, the mean value of actual sales is 7515.458, and the predicted demand is 7345.964, and the standard deviations are respectively 2974.496 and 2954.425, the prediction accuracy is high. Multi-model comparison analysis shows that the random forest model is significantly better than other models in terms of prediction accuracy, reaching 93.3%, and the training time is shorter. Inventory cost control through the PSO-RF model enabled Company K to consistently reduce its fulfillment expense ratio to an all-time low in 2021, and its inventory turnover ratio to surpass competitor Company S from 2019 onwards. In terms of user experience, Company K obtains a user satisfaction score of 4.548, a response timeliness of 0.998, and a comprehensive evaluation of more than 0.9, which is a leading position in the industry. Practice has proven that the random forest algorithm combined with PSO optimization technology can effectively improve the accuracy of enterprise supply chain demand forecasting, reduce inventory costs, and improve the overall operational efficiency and market competitiveness of enterprises. Future research will further explore the construction of algorithm optimization and multi-dimensional evaluation system to promote the continuous innovation and development of supply chain management technology.

References

[1] Deng, W., Feng, L., Zhao, X., & Lou, Y. (2020). Effects of supply chain competition on firms' product sustainability strategy. Journal of Cleaner Production, 275, 124061.

[2] Zheng, D., & Wang, T. (2025). Supply chain resilience, logistics efficiency, and enterprise competitiveness. Finance Research Letters, 79, 107335.

[3] Kot, S. (2018). Sustainable supply chain management in small and medium enterprises. Sustainability, 10(4), 1143.

[4] Hofmann, E., & Rutschmann, E. (2018). Big data analytics and demand forecasting in supply chains: a conceptual analysis. The international journal of logistics management, 29(2), 739-766.

[5] Liu, P., Hendalianpour, A., Hamzehlou, M., & Feylizadeh, M. (2022). Cost reduction of inventory-production-system in multi-echelon supply chain using game theory and fuzzy demand forecasting. International journal of fuzzy systems, 24(4), 1793-1813.

[6] Panahifar, F., Shokouhyar, S., & Mosafer, S. (2022). Identifying and assessing barriers to information sharing in supply chain-a case study of the automotive industry. International Journal of Business Information Systems, 41(2), 258-288.
[7] Katiyar, R., Barua, M. K., & Meena, P. L. (2018). Analysing the interactions among the barriers of supply chain performance measurement: an ISM with fuzzy MICMAC approach. Global Business Review, 19(1), 48-68.
[8] Aboolian, R., Berman, O., & Wang, J. (2021). Responsive make‐to‐order supply chain network design. Naval research logistics (nrl), 68(2), 241-258.
[9] Li, G., Yu, G., Wang, S., & Yan, H. (2017). Bullwhip and anti-bullwhip effects in a supply chain. International Journal of Production Research, 55(18), 5423-5434.
[10] Lu, J., Feng, G., Lai, K. K., & Wang, N. (2017). The bullwhip effect on inventory: a perspective on information quality. Applied Economics, 49(24), 2322-2338.
[11] Jeong, K., & Hong, J. D. (2019). The impact of information sharing on bullwhip effect reduction in a supply chain. Journal of Intelligent Manufacturing, 30, 1739-1751.
[12] Almeida, M. M. K. D., Marins, F. A. S., Salgado, A. M. P., Santos, F. C. A., & Silva, S. L. D. (2017). The importance of trust and collaboration between companies to mitigate the bullwhip effect in supply chain management. Acta Scientiarum: Technology, 39(2), 201-210.
[13] Chang, H., Chen, J., Hsu, S. W., & Mashruwala, R. (2018). The impact of the bullwhip effect on sales and earnings prediction using order backlog. Contemporary Accounting Research, 35(2), 1140-1165.
[14] Pournader, M., Narayanan, A., Keblis, M. F., & Ivanov, D. (2023). Decision bias and bullwhip effect in multiechelon supply chains: Risk preference models. IEEE Transactions on Engineering Management, 71, 9229-9243.
[15] Becerra, P., Mula, J., & Sanchis, R. (2022). Sustainable inventory management in supply chains: Trends and further research. Sustainability, 14(5), 2613.
[16] Rahman, F. (2018). A predictive model to determine the causes of safety stock requirements: An analytical approach to reduce working capital. Journal of Supply Chain Management, Logistics and Procurement, 1(2), 114-124.
[17] Tan, H., & Fu, X. (2023). Emergency order response strategy under supply chain collaboration. RAIRO-Operations Research, 57(4), 2239-2265.
[18] Dadaneh, D. Z., Moradi, S., & Alizadeh, B. (2023). Simultaneous planning of purchase orders, production, and inventory management under demand uncertainty. International Journal of Production Economics, 265, 109012.
[19] Chauhan, S., Singh, R., Gehlot, A., Akram, S. V., Twala, B., & Priyadarshi, N. (2022). Digitalization of supply chain management with industry 4.0 enabling technologies: a sustainable perspective. Processes, 11(1), 96.
[20] Chatterjee, S., Mariani, M., & Ferraris, A. (2023). Digitalization of supply chain and its impact on cost, firm performance, and resilience: technology turbulence and top management commitment as moderator. IEEE Transactions on Engineering Management, 71, 10469-10484.
[21] Udeh, E. O., Amajuoyi, P., Adeusi, K. B., & Scott, A. O. (2024). The role of IoT in boosting supply chain transparency and efficiency. Magna Scientia Adv. Res. Rev., 12(1), 178-197.
[22] Kai Yang,JiaMing Wang,GeGe Zhao,XuAn Wang,Wei Cong,ManZheng Yuan... & Jing Tao. (2025). NIDS-CNNRF integrating CNN and random forest for efficient network intrusion detection model. Internet of Things,32,101607-101607.
[23] Mengwei Zhao,Pan Xiao,Chao Liang,Chaoyuan Wang,Baoming Li & Weichao Zheng. (2025). A new index based on Gini coefficient for evaluating the distribution uniformity of environmental parameters in buildings. Building and Environment,278,112910-112910.
[24] Fang Liu,Ming Hui Chen,Xiaojing Wang & Roeland Hancock. (2025). Decomposition of WAIC for assessing the information gain with application to educational testing. The British journal of mathematical and statistical psychology,
[25] Junqing Zhu,Yiqun Yin,Tao Ma & Dan Wang. (2025). A novel maintenance decision model for asphalt pavement considering crack causes based on random forest and XGBoost. Construction and Building Materials,477,140610-140610.