Research On Vegetable Replenishment and Pricing Based on LSTM Neural Network Prediction Modeling

Junmin Pan *

School Of Information and Enginnering, Jingdezhen Ceramic University, Jingdezhen, China, 333403

* Corresponding Author Email: pan270512@163.com

Abstract. Vegetables in fresh food supermarkets face challenges with pricing and replenishment since the majority of vegetable varieties have a short shelf life. Fresh food supermarkets must replenish goods based on sales history and demand. In this paper, the pricing and replenishment decisions of fresh supermarkets are predicted using the random forest and LSTM neural network prediction models. The findings indicate that the prediction model in this paper can have a good predictive effect and that the decisions made by fresh supermarkets regarding pricing and replenishment are related to the historical time series data. This paper's study can assist in resolving price and replenishment issues in fresh food stores, which will have some positive economic effects.

Keywords: Vegetable restocking, Vegetable pricing, LSTM neural network prediction model, Random Forest regression.

1. Introduction

A vast array of perishable vegetables, many of which are seasonal and have a limited shelf life, are available in fresh food stores. Their quality and appearance will decline with time, and most unsold varieties cannot be resold the next day [1]-[2]. Supermarkets must therefore establish a reasonable replenishment and pricing strategy based on historical sales volume and pricing of vegetables, as well as daily replenishment and pricing, in order to reduce costs and increase revenue efficiency [3]. However, because of the wide range of vegetables, the different origin environments, and the time constraints involved in the transaction, supermarkets are not aware of the exact single product and the purchase price of various single products [4]. Consequently, in order to meet demand while accurately projecting vegetable prices, supermarkets must thoroughly examine the relationship between the demand for vegetables and past data, pricing decisions, sales space, and other factors.

Many academics have dedicated their time to studying various forecasting models in order to produce more accurate results in the field of vegetable forecasting research at the retail stage in order to address the aforementioned issues, such as the Box–Jenkins-based autoregressive integrated moving average model and machine learning-based algorithms such as long short-term memory (LSTM) networks ,support vector regression (SVR), random forest regression, gradient boosting regression (GBR) and extreme GBR (XGBoost/XGBR) were proposed and applied (i.e. modeling ,training ,testing and predicting) at the retail stage for selected vegetables to forecast demand[5].The performance analysis (i.e. forecasting error analysis) was carried out to select the appropriate forecasting model at the retail stage for selected vegetables[5].

This paper suggests an LSTM neural network prediction model [6] and a random forest regression model to conduct a prediction study on the pricing and replenishment of vegetables using the flower and leafy category as an example, based on the issues raised by the current research. The research presented in this paper has some economic benefits as well as helping to resolve the issues with pricing and replenishment in fresh food supermarkets.

2. LSTM neural network modeling and random forest regression modeling

2.1. Research methodology and problem analysis

Using the assumption that the price of each item at the supermarket is what remains after applying the cost-plus pricing method, the pricing method in this paper uses the "cost-plus pricing method" to analyze the relationship between sales volume and pricing of each category. Using the six categories of vegetable data from a fresh food supermarket from July 1, 2020, to June 30, 2023, this paper uses random forest regression to investigate the relationship between the pricing of a particular category of vegetables on a given day, the sales of that day, and the sales of the days prior. The data of the first five days of each day and the data of the day are used for model training in order to predict the pricing of vegetables. The LSTM neural network prediction model is also used in this work to train the model using historical series data in order to forecast the daily replenishment quantity of each vegetable category. The issues with pricing and replenishment are resolved by the strategies.

2.2. Fundamentals of LSTM Neural Networks

2.2.1 Core ideology of LSTM neural networks

The long-term dependence problem is resolved by the LSTM neural network model, which is an adaptation of the traditional RNN neural network model, by changing the propagation structure. Because traditional RNNs use the chain rule, gradient multiplication eventually causes the gradient to disappear. Unlike the traditional RNN, the LSTM network transforms the gradient into cumulative form in order to avoid problems such as these. The most important thing in LSTM network is the state of the cell, in order to update the state of the cell, there is a series of operations in the LSTM network, and this cell state is carried throughout the entire structure of the LSTM network [7].

2.2.2 Structure of the LSTM neural network

In comparison to a traditional RNN, the LSTM added three gates: an input gate, an output gate, and a forget-the-gate to protect and control the unit's state. This allowed the LSTM unit to save useful data for a longer period of time and better capture the long-term dependence relationship. The structure of LSTM can improve the RNN training gradient disappearance and weight influence is too large problems [8]. The network structure is shown in Figure 1.

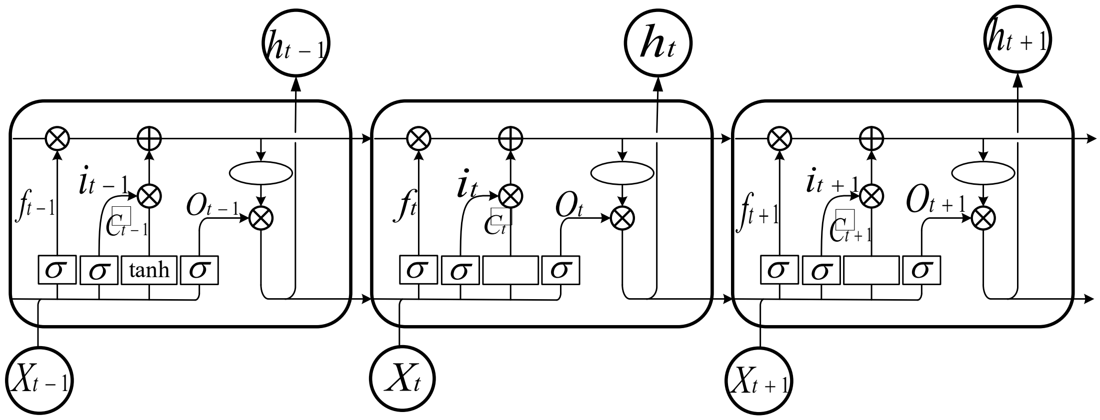

Figure 1 LSTM network structure

Information is permitted to flow through the LSTM network in a specific ratio when the "gate" accepts a value between 0 and 1. A multiplication operation and a sigmoid function make up the "gate" in this paper.

The LSTM network has three "gates" that are as follows.

(1) Forgetting gate determines how much information from the previous moment's state is forgotten.
(2) Input gate: determines how much information in the input of this moment is updated to the state.
(3) Output Gate: determines how much information has been output from the current state.

$$f_t = \sigma(W_f \bullet [h_{t-1}, x_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i g[h_{t-1}, x_t] + b_i) \tag{2}$$

$$\overset{\circ}{C}_i = \tanh(W_C g[h_{t-1}, x_t] + b_C) \tag{3}$$

$$C_t = C_{t-1} * f_t + \overset{\circ}{C}_t * i_t \tag{4}$$

$$O_t = \sigma(W_o g[h_{t-1}, x_t] + b_O) \tag{5}$$

$$h_t = O_t * \tanh(C_t) \tag{6}$$

$$y_t = f(W_p g h_t + b_p) \tag{7}$$

The sigmoid activation function, represented by the symbol σ in equation (1), has the forgetting function because it converts the input to the interval [0,1] and partially becomes 0 as a result of the activation function. Tanh is another type of activation function that converts the input to the interval [-1,1]. Figure 2 illustrates how these two types of functions appear.

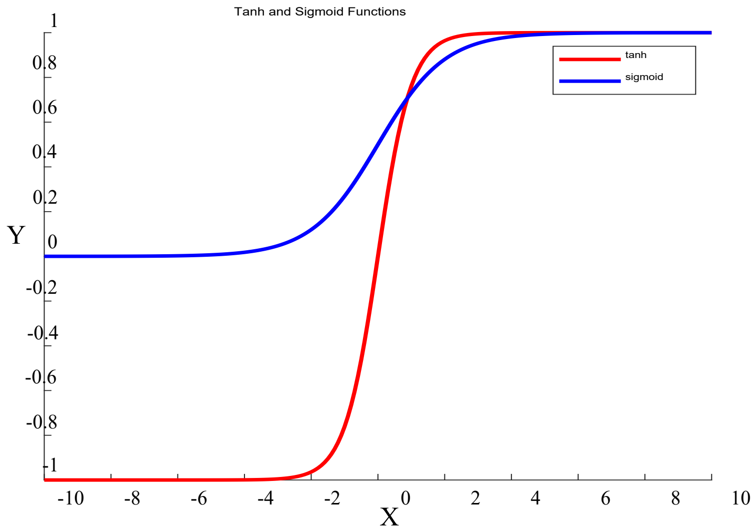

Figure 2 Activation function

When the number of hidden layer neurons is m_2 and the input features are m_1-dimensional, the weight parameter of the forgetful gate connecting the input and hidden layer neurons is represented by the symbol W_f in equation (1). The hidden layer of the forgetful gate is represented by a ($m_1 + m_2$)× m_2 matrix, and the bias of the neurons is equal to the number of hidden layer neurons. Equation (2)(3) uses W_i and W_C as the hidden layer neuron's parameters, b_i and b_C as the bias for the connection input.

2.3. Fundamentals of Random Forest Regression Modeling

Based on bagging techniques and CART (Classification and Regression Trees), Random Forest is an integrated learner [9]. Random forests are useful for regression and classification, Multiple regression trees that grow in parallel in the forest make up the random forest regression model; these decision trees are unrelated to one another. Because of its independence, the random forest model has strong parallelism and can efficiently make use of the multi-core processor found in contemporary computers. Every decision tree in the forest makes a prediction about the input samples during the prediction stage. The average of all the trees' predictions is the final prediction result. This integrated learning approach can successfully lower the model's variance and raise prediction accuracy [10].

This paper's implementation builds a random forest model using MATLAB's TreeBagger function. TreeBagger generates training data and feature subsets using feature random selection and bootstrap sampling, enabling the creation of multiple decision trees.

2.3.1 Random Forest Regression Model Construction

To create a random forest regression model, use the data from a fresh food supermarket's six vegetable categories between July 1, 2020, and June 30, 2023. The random forest model should be trained using the data from the prior five days as input features and the current day's data as target variables.

A few key parameters, like the number of trees and the maximum number of features to be taken into account when each node splits, must be changed during the training process. These variables have an impact on the model's functionality and forecast accuracy. This paper can determine the ideal set of parameters by using grid search and cross-validation techniques.

This study creates a random forest regression model that can be used to forecast vegetable prices in the future after the training is finished.

3. Model solving and results

3.1. Results of the relationship between daily pricing and total sales of leafy and flowering vegetables

As an example, the relationship between daily pricing and total sales of flower and leafy vegetables can be found by choosing data on daily average pricing and total sales from July 1, 2020, to June 30, 2023. Then, using the first five days of each day's data and the day's data to perform random forest regression, the results are displayed in Figures 3 and 4 and 5.

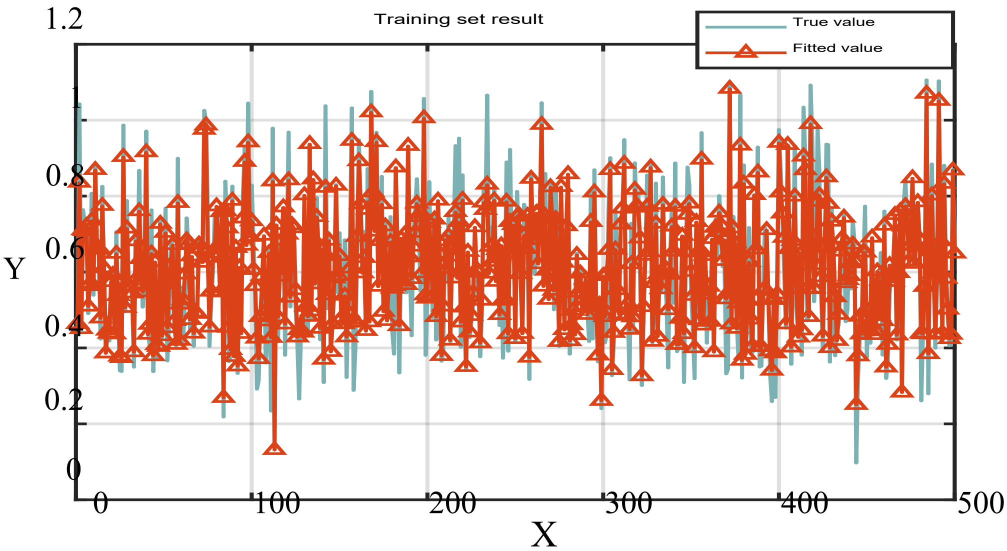

Figure 3 Random Forest regression training set

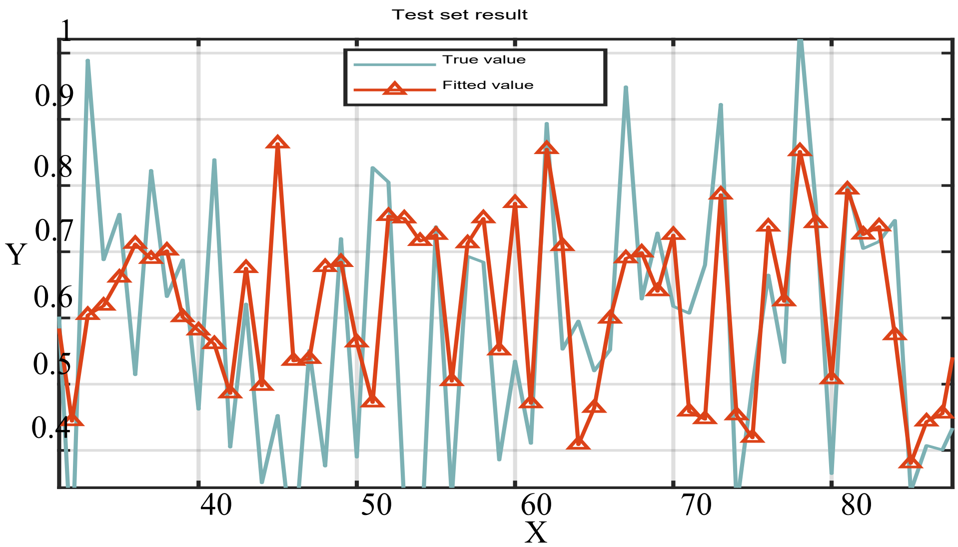

Figure 4 Random Forest regression test set

Good prediction is indicated by the random forest model's MAE of 0.050872 on the training set. Despite having a 0.13521 MAE on the test set, it can still make some predictions. As a result, this model is the one that this study will employ. Figure 3 illustrates the model's excellent prediction performance on the training set. Figure 4 shows that although while the test set's prediction impact is somewhat smaller than the training set's, it is still acceptable.

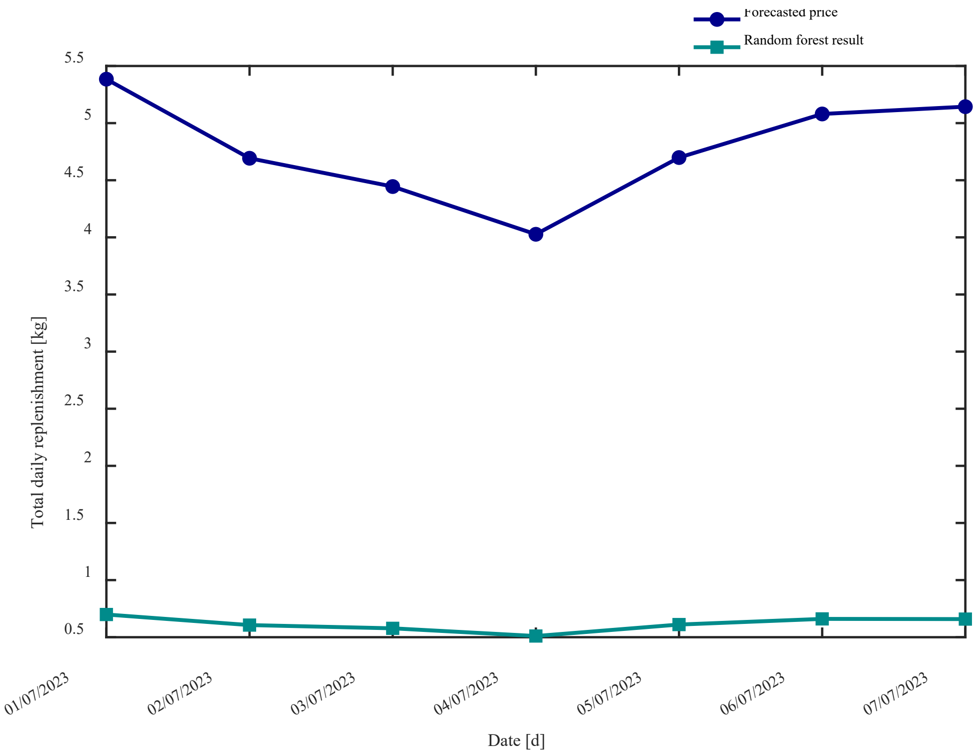

Figure 5 Random Forest results and predicted prices.

For every day between 2023/07/01 and 2023/07/07, Figure 5 displays the pricing prediction results and random forest results. The random forest regression results are multiplied by the total number of vegetable pieces and divided by 13 because in this paper, a day is measured in terms of 13 hours. The final pricing is the average price of leafy vegetables for a given day, and pricing is positively correlated with sales volume and somewhat regular. In general, pricing will be higher when there is a high demand, and overall pricing will be relatively smooth with time.

3.2. Forecast results of total daily replenishment of leafy and flowering vegetables

Using the sales volume prediction of foliage as an example, the sales volume prediction of foliage was completed in this LSTM article using sales volume data selected from August 5, 2021, to June 30, 2023, to forecast the total daily sales volume. This article uses 691 leafy flower sales samples in an LSTM neural network. The training, test, and validation sets are divided into 8:1:1 ratio. The LSTM model has 32 hidden neurons with a 0.001 learning rate and a maximum of 350 training cycles. Figures 6, 7, 8, and 9 present the findings.

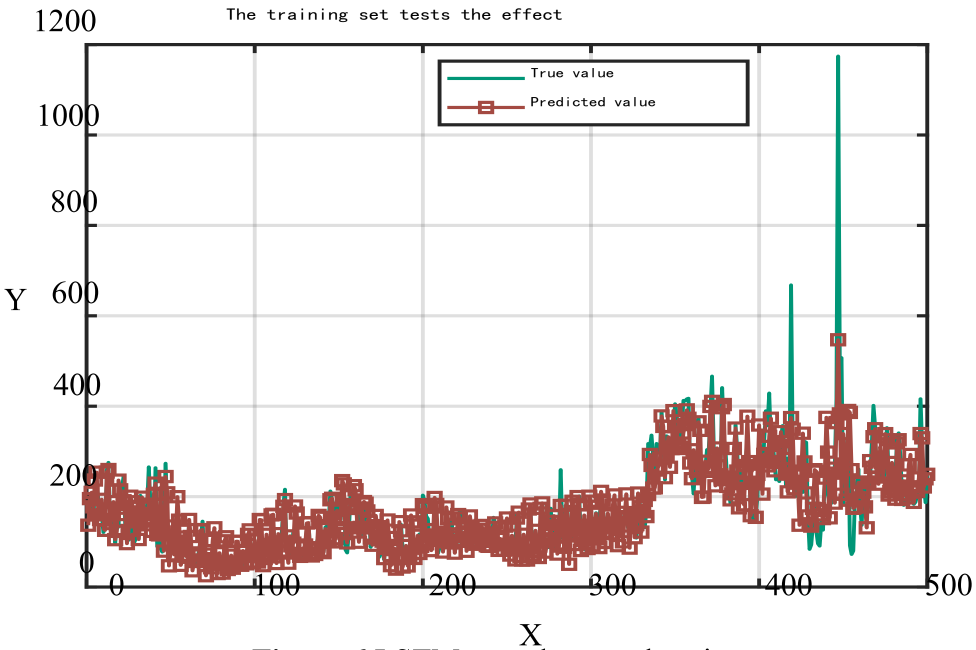

Figure 6 LSTM neural network train set

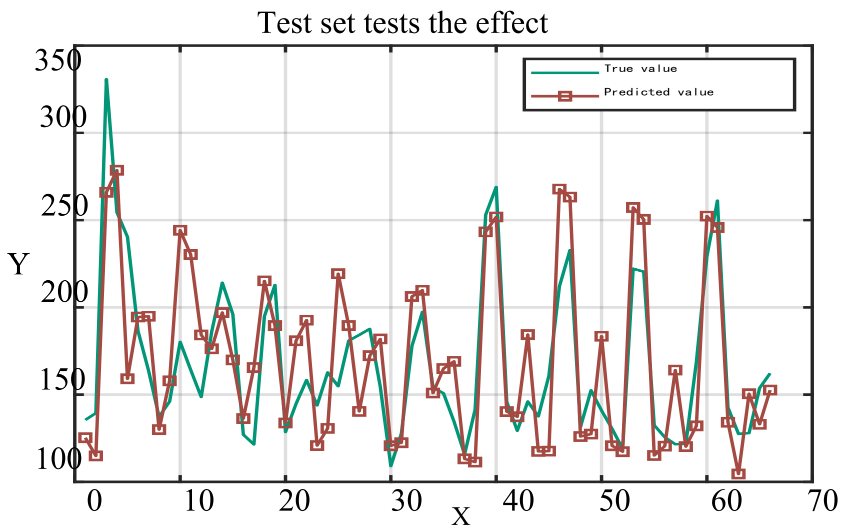

Figure 7 LSTM neural network test set

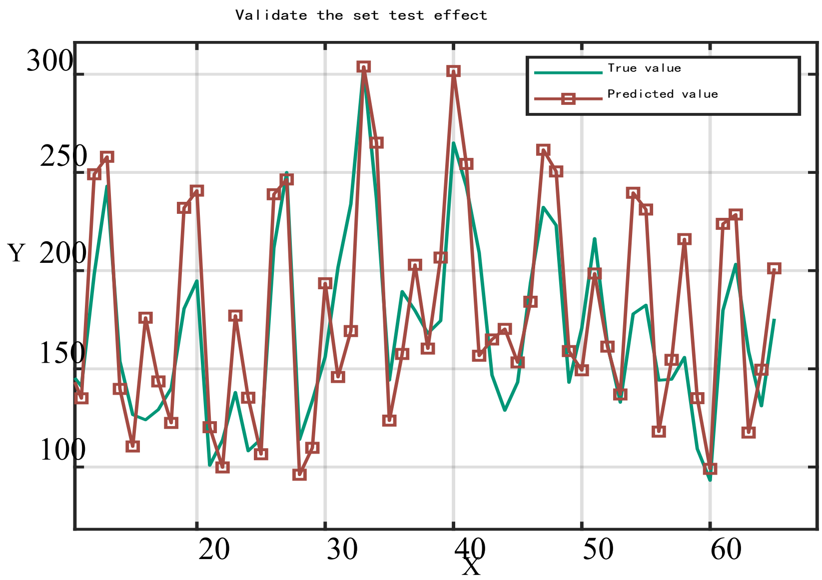

Figure 8 LSTM neural network Validation set.

As can be seen from Figure 6, the LSTM neural network model performs better on the training set, and all the error metrics are within the acceptable range. Figure 7 demonstrates that while the model's performance on the validation set is somewhat worse than on the training set, all error measures remain within a reasonable bound, suggesting some degree of generalizability. All of the error metrics on the test set drop in Figure 8, particularly the average absolute error of 27.2866 and the average relative error of 0.1702, indicating that the model performs well in terms of prediction on unobserved data. This suggests that the model is capable of correctly projecting future trends. Therefore, this model is chosen to be used in this paper.

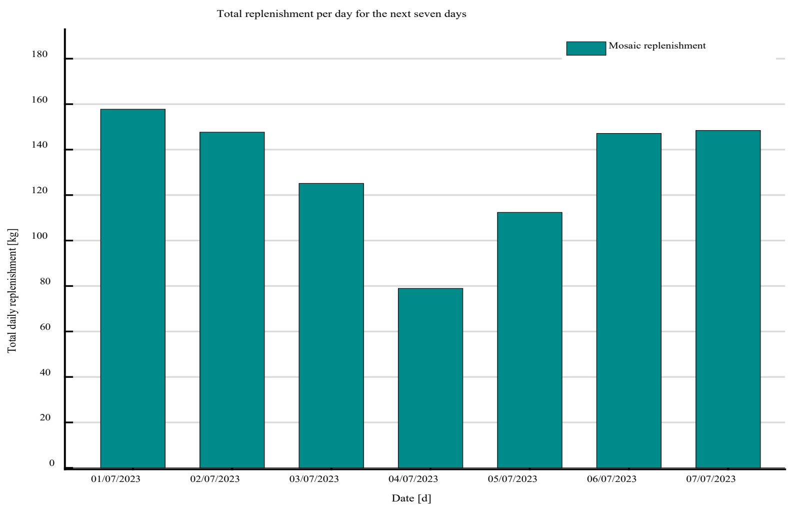

Figure 9 LSTM neural network Replenishment forecast.

The results of the following seven-day forecast are displayed in Figure 9. The demand was predicted by utilizing the trained LSTM neural network model to predict the sales of the category of flowers and leaves during the previous seven days, resulting in the total daily replenishment. The total daily replenishment falls precipitously from July 1, 2023, to July 4, 2023, rises precipitously from July 4, 2023, to July 6, 2023, and then rises somewhat from July 6, 2023, to July 7, 2023. The results are regular and time dependent.

4. Conclusions

The pricing and replenishment decisions of fresh food supermarkets are predicted in this paper using the random forest and LSTM neural network prediction models. The findings indicate a regularity and a certain relationship between the replenishment decisions of fresh food supermarkets and the historical time-series data. In general, pricing and sales volume have a positive correlation; that is, more sales volume of a certain type of vegetable will translate into higher pricing, and there is a link with the historical time series data. The study presented in this article has some economic benefits as well as helping to resolve the issues with pricing and replenishment in fresh food shops.

References

[1] Ramjan M D, Ansari M T. Factors affecting of fruits, vegetables and its quality [J]. J. Med. Plants, 2018, 6: 16-18.

[2] Mangaraj S, Goswami T K. Modified atmosphere packaging of fruits and vegetables for extending shelf-life-A review [J]. Fresh produce, 2009, 3(1): 1-31.

[3] Fan T, Xu C, Tao F. Dynamic pricing and replenishment policy for fresh produce [J]. Computers & Industrial Engineering, 2020, 139: 106127.

[4] Huber J, Stuckenschmidt H. Intraday shelf replenishment decision support for perishable goods [J]. International Journal of Production Economics, 2021, 231: 107828.

[5] Priyadarshi R, Panigrahi A, Routroy S, et al. Demand forecasting at retail stage for selected vegetables: a performance analysis[J]. Journal of Modelling in Management, 2019, 14(4): 1042-1063.

[6] Jin D, Yin H, Gu Y, et al. Forecasting of vegetable prices using STL-LSTM method[C]//2019 6th International Conference on Systems and Informatics (ICSAI). IEEE, 2019: 866-871.

[7] Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network [J]. Physica D: Nonlinear Phenomena, 2020, 404: 132306.

[8] Huang R, Wei C, Wang B, et al. Well performance prediction based on Long Short-Term Memory (LSTM) neural network [J]. Journal of Petroleum Science and Engineering, 2022, 208: 109686.

[9] Sun L, Ji Y, Zhu X, et al. Process knowledge-based random forest regression for model predictive control on a nonlinear production process with multiple working conditions[J]. Advanced Engineering Informatics, 2022, 52: 101561.

[10] Genuer R, Poggi J M, Genuer R, et al. Random forests[M]. Springer International Publishing, 2020.