Enhancing Vegetable Sales Forecasting with A CNN-LSTM-Transformer Hybrid Model

Aoxiang Tian *

Shandong University of Science and Technology, Shandong, China, 271001

* Corresponding Author Email: 1669703044@qq.com

Abstract. Due to the short shelf life of vegetable products, a significant portion of the inventory cannot be resold the following day. To facilitate more informed procurement decisions in superstores and minimize vegetable wastage, this study proposes a hybrid prediction model based on CNN-LSTM-Transformer for enhancing the accuracy of forecasting vegetable sales volumes. Firstly, an LSTM model is incorporated to account for the recurring and seasonal variations in vegetable sales. Secondly, a CNN model is introduced to address the limitations of LSTM in capturing spatial data components. The convolutional and pooling layers of CNN help establish spatial relationships among different feature values in the dataset. Finally, a Transformer model is integrated to tackle the issue of long-term dependencies, which LSTM alone struggles to resolve. The Transformer model employs a parallel attention mechanism, eliminating temporal dependencies and effectively addressing LSTM's long-term dependency challenge. This integration also accelerates model training. The evaluation metric employed in this paper is RMSE (Root Mean Square Error). The three-year sales volume of 11 vegetable types is predicted using seasonal ARIMA, XGBoost, LSTM, CNN-LSTM, and CNN-LSTM-Transformer models. The results indicate that the CNN-LSTM-Transformer model achieves the lowest RMSE at 0.0758, followed by ARIMA at 0.0792, CNN-LSTM at 0.0830, XGBoost at 0.0858, and LSTM at 0.0913. These findings demonstrate that the CNN-LSTM-Transformer model exhibits superior accuracy in forecasting vegetable sales volumes, yielding more precise predictions. This accurate forecasting not only aids superstores in reducing waste and enhancing profitability but also assists government authorities in rationalizing vegetable subsidy policies and optimizing the vegetable production and marketing system.

Keywords: CNN-LSTM-Transformer, vegetable sales prediction, seasonal, cyclical.

1. Introduction

With the development of agriculture, all aspects of vegetables and fruits are gradually being studied [1]. Vegetables are seasonal produce and there is a relationship between demand and supply. Therefore, it is necessary to forecast the demand for vegetables in the wholesale market to avoid wastage of vegetables and to increase the profits of all stakeholders. This paper investigates the problem of sales forecasting in the fruit and vegetable industry by analyzing the vegetable sales data in depth and extracting the features of the data to predict the number of vegetables to be sold in the coming days. Because of the wide variety of vegetable dishes, there are roughly two types of existing sales forecasts, one is the traditional time-series forecasting model, and the other is using machine learning to forecast. The former has high requirements on the smoothness of the data, and therefore is not accepted by the majority of scholars. The latter can generally be categorized into linear and nonlinear influences, because consumers are easily affected by external factors when they go to the market to buy vegetables, for example, the weather is good or bad, whether it is a holiday or not, and whether the vegetable sales are bundled or not can affect the consumers' ability to buy vegetables. He Zhang et al. [2] proposed a combined model based on LightGBM and LSTM for vegetable sales volume prediction, which utilizes the error inverse method to weight the array of the two models for sales forecasting, and experiments show that the combined model is more accurate than a single model. Sankaran S. et al. [3] proposed a Seasonal Autoregressive Integrated Moving Average (SARIMA) model for forecasting in vegetable sales forecasting, taking into account the significant impact of seasonal changes on the demand for vegetables. Yunhao Cui [4] proposed a fresh vegetable sales prediction model based on CNN-PSO-LSTM. The obtained feature data were selected as feature vectors through fuzzy processing and grey correlation analysis to select the influencing factors that are more related to vegetable sales as feature vectors, which were inputted into a CNN model for feature extraction, and the PSO-optimized LSTM model was used for sales prediction, which proves the feasibility of the prediction. Feng Chen et al. [5] et al. proposed the XGboost-LSTM combination model, the weighted combination of the LSTM model and the XGboost model prediction results obtains more accurate results than the single model. Wang Xu et al. [6] proposed a CNN-LSTM network sales prediction model, which extracts important time series data features from the upper layer CNN network and inputs them into the lower layer LSTM network to further extract complex irregular features to validate the feasibility of the model. XGboost and seasonal ARIMA models have better prediction effect on samples with small data volume, but their prediction effect is significantly reduced for large data samples of vegetable sales volume. The prediction effect of XGboost model for seasonal and periodic sample data is very good but the prediction effect is very poor, CNN-LSTM time series prediction model is still subject to the problem of long-term dependence to a certain extent, and the combined model of LightGBM and LSTM is relatively large in computation, and the prediction effect is affected by a single model. In order to address some of the above problems this paper proposes a hybrid model of CNN-LSTM-Transformer time series prediction, the LSTM model has a long-term memory so that it can be a good prediction of seasonal and cyclical data, and the addition of the self-attention mechanism of the CNN-LSTM prediction model can accelerate the speed of training to improve the accuracy of the prediction.

2. Model introduction and establishment

2.1. Data sources and processing

This paper uses the sales data of each product from July 1, 2020, to June 30, 2023, of a fresh goods supermarket, and the sales data mainly includes sales volume, sales, individual items, categories and so on. According to the statistics a total of 251 dishes are included, a total of 6 categories. Among them, some varieties that do not have sales forecasts are removed, like cordyceps flowers, chanterelle mushrooms, lotus roots, monkey head mushrooms and so on. These dishes have less sales in these three years and some of them have been discontinued hence the choice to remove them from the forecast. In this paper, 11 individual items from 6 vegetable categories are selected for sampling and forecasting. Table 1 Departmental Vegetable Sales Volume Data for June 30,2023 below shows the daily sales volume on June 30, 2023, for the 11 vegetable items.

Table 1 Departmental Vegetable Sales Volume Data for June 30, 2023

Name of dish	kind	sales volume	unit (of measure)
Brassica pekinensis	philodendron	14.748	Kg
sweet potato tip	philodendron	4.837	Kg
cabbage (round cabbage most commonly found in Chinese medicine)	cauliflower (Brassica oleracea var. botrytis)	9.044	Kg
net lotus root	Aquatic rhizomes	6.44	Kg
cucumber	Aquatic rhizomes	3.78	Kg
eggplant (Solanum melongena L.)	eggplant	0.94	Kg
purple eggplant (Solanum melongena L.)	eggplant	14.365	Kg
king cobra or chili (Naga jolokia)	capsicum	8.47	Kg
chili	capsicum	1.586	Kg
enoki mushroom	edible mushrooms	2.494	Kg
Xixia Shiitake Mushroom	edible mushrooms	6.903	Kg

Considering that the vegetable sales volume has the characteristics of cyclicality and seasonality, and also affected by holidays and weather, we regard the zero value or the situation of larger sales volume on a certain day as a reasonable situation. By observing the three-year vegetable sales volume data of the superstore, by observe that there are some missing days, so in this paper make up the zeroes for these days to get the complete time-series data.

2.2. Feature selection

From the available data, XGboost is used to do correlation analysis to select the features with strong correlation with the set to be predicted as a subset of features to be input into the model. The main features are categorized as follows:

(1) Weather characteristics

It mainly involves precipitation and temperature.

(2) Characteristics of the dishes

Customers often buy multiple dishes together when purchasing dishes, and customers often consider the collocation between multiple dishes, so other dishes with strong correlation with the dishes to be predicted are selected as a feature subset.

(3) Historical sales figures for dishes

The data on the historical sales volume of the dishes for the last three years is used as a subset of the features.

2.3. Standardization

In order to remove the unit limitation of the data and transform the data into dimensionless pure values, and at the same time reduce the range of values to speed up the data calculation, so before inputting the data into the model, we have to standardize the data, and in this paper use the z-score standardization method.

z-score standardization, referred to as zero-mean standardization, is a common method of data processing, which is based on the principle of the original data to find out the mean and variance of the data, and then use the original data values to subtract the mean, and finally divided by the standard deviation, in order to get a mean of 0, the variance of 1, the standard distribution of too. This method is applicable to the premise that the maximum and minimum values in the data are unknown, or in cases where there are more outliers. The formula is as follows:

Z-sore normalized conversion formula:

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

Standard deviation formula:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{2}$$

where is the standard deviation andμ is the mean, and N is the sample size.

2.4. CNN-LSTM-Transformer Hybrid Modeling

CNN-LSTM-Transformer is based on the CNN-LSTM model introduces the attention mechanism, mainly divided into four layers input layer, CNN layer, LSTM layer, attention mechanism layer, and output layer, the specific structure and schematic diagram is shown in Figure 1 CNN-LSTM-Transformer structure diagram.

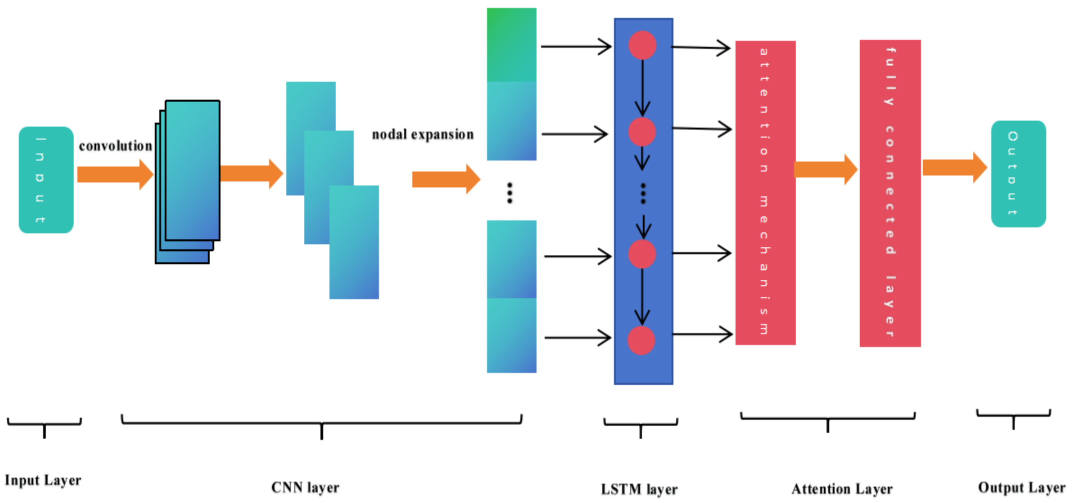

Figure. 1 CNN-LSTM-Transformer structure diagram

2.4.1 CNN layer

Convolutional neural network (CNN) is a kind of feed-forward neural network with convolutional structure, internal weight sharing and local connection structure, which can reduce the complexity of the model and at the same time can effectively extract the deep information contained in the data.CNN consists of an input layer, convolutional layer, pooling layer, fully connected layer, and output layer. Among them, the convolutional layer performs convolutional computation on the data and extracts local features, and the pooling layer downsamples the data and compresses the volume[7-8] , In this paper, we extract the spatial connection between different feature values in the data of vegetable sales volume through convolution and pooling to make up for the shortcomings of LSTM that cannot capture the spatial component of the data, and at the same time, it extracts the data that still has temporal sequential nature. The one-dimensional CNN structure is shown in Figure 2 CNN structure diagram.

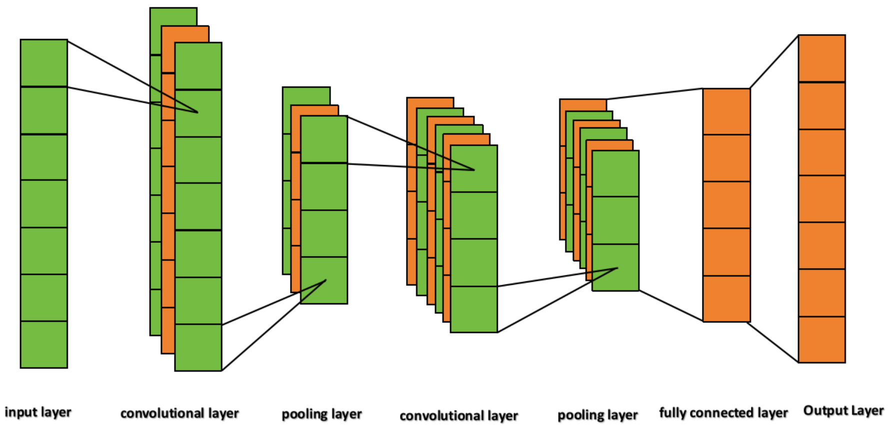

Figure 2 CNN structure diagram

2.4.2 LSTM layer

LSTM is a kind of recurrent neural network (RNN), RNN can only have short-term memory due to gradient disappearance, LSTM network combines short-term memory with long-term memory through gate control, which alleviates the problem of gradient decline to some extent. It can root the memory cells in the neural nodes of the hidden layer of the recurrent neural network to record historical information; by adding three gate structures (input, forget and output), historical information can be realized. As shown in Figure 3 Structure of LSTM memory cell, when setting the input sequence as x (x1,x2,x3.... ,xt), the implicit layer state is h(h1,h2,h3... ,ht), which is given by the formula:

$$i_t = \sigma(W_{hi}h_{t-1} + W_{xi}X_t) \tag{3}$$

$$f_t = \sigma(W_{hf} + W_{xf}X_t) \tag{4}$$

$$C_t = f_t \odot C_{t-1} + i_c \odot \tanh(W_{xc}W_{hc}Xh_{t-1}) \tag{5}$$

$$O_t = \sigma(W_{xo} + W_{ho}X_{t-1} + W_{co}C_t) \tag{6}$$

$$h_t = O_t \odot \tanh(C_t) \tag{7}$$

Where "⊙" denotes the Adama product; i_t, f_t and ot are the outputs of different gates; c_t is the vector of cell states; h_{t-1} is the output information of the hidden cell at the previous moment; h_t is the new state of the memorized cell; W_h , W_x , W_c are the weights of the corresponding gates; sigmoid and tanh are two different activation functions respectively[9] . In this paper, we utilize the characteristics of the long-term memory function of LSTM to extract the temporal change information of the vegetable sales volume data.LSTM can better deal with a large amount of vegetable sales volume data, and also can better deal with the characteristics of the vegetable sales volume of the cyclical and seasonal nature.

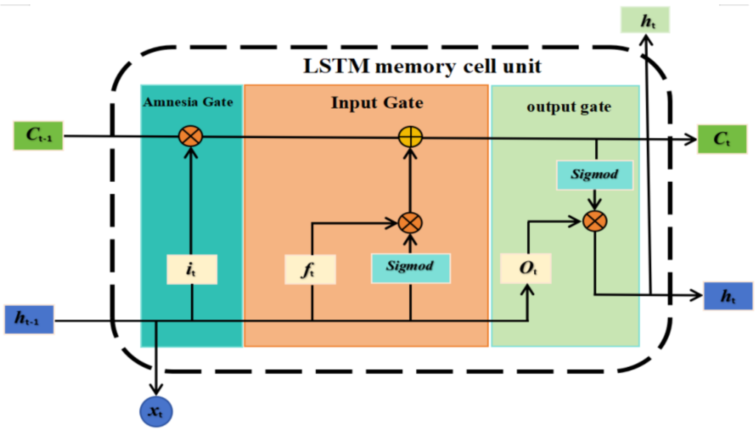

Figure 3 Structure of LSTM memory cell

2.4.3 Attention mechanism layer

Attention mechanism is a part of Transformer model, the other parts are positional encoding, residual connection and layer normalization, encoding-decoding framework, etc. Transformer uses multi-head attention mechanism to learn the dependencies in the sequence, in this paper, we mainly use the multi-head attention to learn the dependencies in different subspaces at the same time, so as to In this paper, we mainly utilize multi-head attention to simultaneously learn the dependencies in different subspaces, so as to capture the global context information in the sequence. Since LSTM still exists a certain long-term dependency in nature, we add parallel attention mechanism to effectively solve the problem of long-term dependency of LSTM to improve the important time of LSTM, so as to improve the accuracy of vegetable sales prediction.

2.5. Experimental parameterization

CNN-LSTM-Transformer model experiment parameters generally include the following aspects, the size of the convolution kernel of the CNN layer, the activation function, the number of hidden layers of the LSTM layer, the number of encoder layers of the Transformer layer, the learning rate, the number of training rounds and so on. The setting of experimental parameters affects the learning efficiency and accuracy. Through a large number of experimental comparisons, the experimental parameters selected for vegetable sales volume prediction are shown in the table 2 Model Parameter Configuration below. Among them, the activation function is an important element of the neural network model, the activation function makes the neurons in the neural network have the ability to learn and adapt independently, and the Relu function is similar to the structure of biological neurons, and it is chosen as the activation function with faster convergence and lower computational complexity than the traditional activation function, therefore, Relu is chosen as the activation function for the CNN layer and the LSTM layer.

Table 2 Model Parameter Configuration Table

CNN-LSTM-Transformer Modeling			
CNN layer	convolutional layer	convolution kernel (math.)	3
	Maximum pooling layer	Pool size	2
	spreading level	Data dimensions	3
	activation function		Relu
LSTM layer	storey		10
	Hidden state dimensions		64
	activation function		Relu
Transformer layer	Num_heads		12
	Attention_dropout		0.2

In this paper, adam (Adaptive Moment Estimation) is chosen as the optimizer, mean square error (MSE) is chosen as the loss function, the initial learning rate is set to 0.01, and a dropout layer is also set to reduce the number of training parameters to prevent overfitting, respectively, the four cases of 20, 30, 50, 70 are chosen as the training period, in which the cabbage root mean square error (RMSE) value is shown in the table 3 Comparison of results for 4 iterations of cabbage below. Root Mean Square Error (RMSE) values are shown in the table below, from the table we can see that when the epochs are 50 the model begins to converge, the accuracy of the forecast of the sales volume of the dishes is better, so the epochs are selected as 50 times.

Table 3 Comparison of results for 4 iterations of cabbage

Predicted effect of cabbage	
Number of iterations (epochs)	RMSE
20	0.0646
30	0.0617
50	0.0643
70	0.0646

3. Comparison of the effect of different models

In order to verify the accuracy of the CNN-LSTM-Transformer model proposed in this paper in predicting the sales volume of vegetables, different models are used to predict the sales volume of sampled dishes, 80% of the data of the sales volume of vegetables are used as the training set, and the remaining 20% of the data are used as the test set to verify the model's performance, the evaluation function used in this paper is the root-mean-square-error (RMSE). The smaller its value indicates the more accurate the prediction result, and its mathematical expression is:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{m}}\sum_{i=1}^{\mathrm{n}} \left(\mathrm{y}_i - \hat{y}_{\mathrm{i}}\right)}^{\,2} \tag{8}$$

Where y_{i}、$\hat{y}_{\mathrm{i}}$ are the actual and predicted values, respectively, and m represents the number of samples.

In this paper, we have used ARIMA model, XGboost model, LSTM model, CNN-LSTM model, CNN-LSTM-Transformer model to predict the 11 kinds of dishes such as Chinese cabbage, sweet potato tips, and tall gourd, etc., and the following table 4 RMSE values for different models shows the values of the RMSE of the five prediction models for the 11 kinds of dishes.

Table 4 RMSE values for different models

Kind mould	LSTM	CNN-LSTM	CNN-LSTM-Transormer	ARIMA	XGboost
Brassica pekinensis	0.0731	0.0612	0.0617	0.0658	0.0700
cucumber	0.1034	0.0929	0.0839	0.0893	0.1120
sweet potato tip	0.0796	0.0730	0.0773	0.0764	0.0980
enoki mushroom	0.0680	0.0614	0.0593	0.0593	0.0670
net lotus root	0.1083	0.0982	0.0958	0.0973	0.1010
king cobra or chili (Naga jolokia)	0.0984	0.0872	0.0843	0.0874	0.0817
Xixia Shiitake Mushroom	0.1296	0.1096	0.1091	0.0982	0.1110
cabbage (round cabbage most commonly found in Chinese medicine)	0.1282	0.1171	0.1050	0.1130	0.1190
eggplant (Solanum melongena L.)	0.0780	0.0735	0.0727	0.0732	0.0800
chili	0.0600	0.0695	0.0468	0.0470	0.0430
purple eggplant (Solanum melongena L.)	0.0781	0.0695	0.0383	0.0649	0.0620
average value	0.0913	0.0830	0.0758	0.0792	0.0858

By comparing Tables 4 RMSE values for different models, it can be seen that the RMSE value of CNN-LSTM-Transformer model has the smallest average value in sales volume prediction of different dishes, so CNN-LSTM-Transformer model is better than the other models mentioned above in vegetable sales volume prediction, where the seasonal ARIMA model is also more effective than the XGboost, LSTM, and CNN-LSTM models, and the LSTM model is the least effective in dealing with vegetable sales volume prediction. Figure 4 Effect of LSTM model on cabbage prediction, Figure 5 Effect of CNN-LSTM model on cabbage prediction, Figure 6 Effect of CNN-LSTM-Transformer model on cabbage prediction, Figure 7 Effect of ARIMA model on cabbage prediction, and Figure 8 Effectiveness of XGboost model for cabbage prediction below show the prediction effect of LSTM model, CNN-LSTM model, CNN-LSTM-Transformer model, ARIMA model, and XGboost model on Chinese cabbage, respectively, and it can also be clearly seen through the figure that the prediction effect of CNN-LSTM-Transformer model is significantly better than other models.

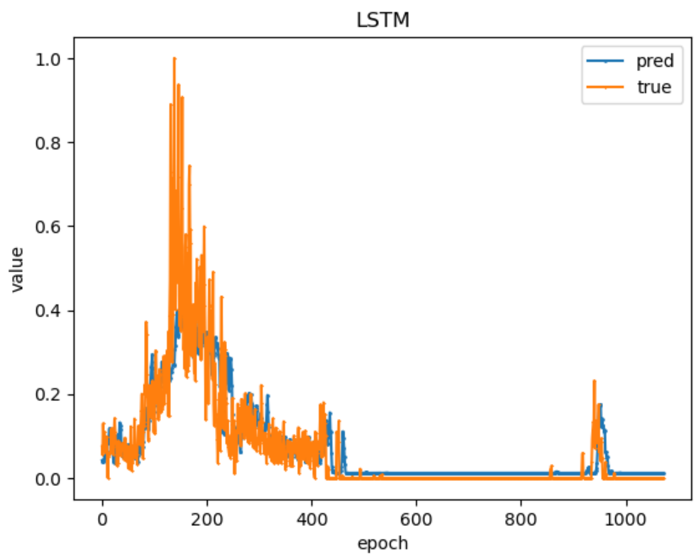

Figure 4 Effect of LSTM model on cabbage prediction

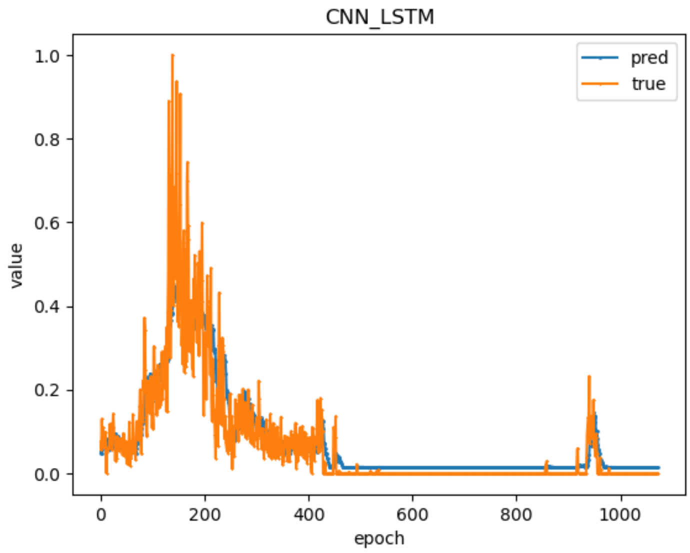

Figure 5 Effect of CNN-LSTM model on cabbage prediction

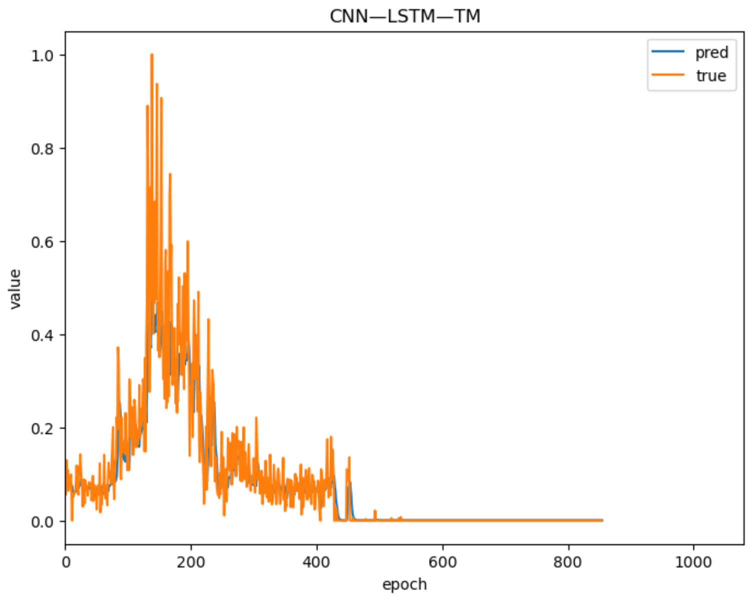

Figure 6 Effect of CNN-LSTM-Transformer model on cabbage prediction

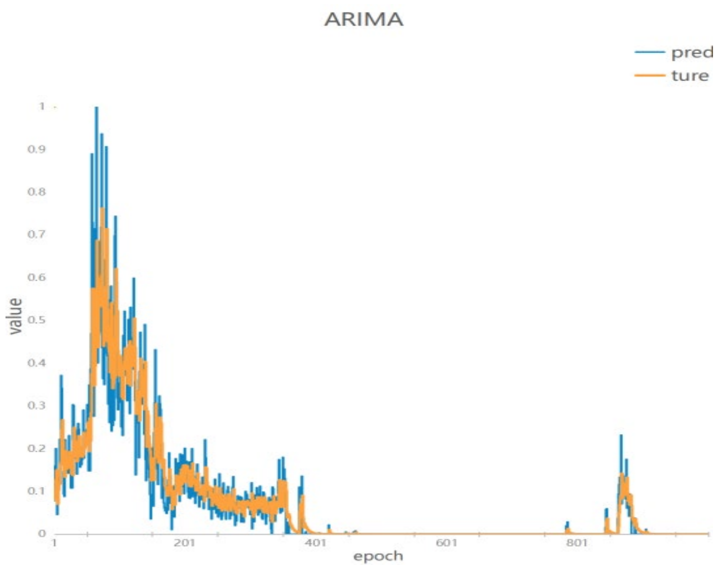

Figure 7 Effect of ARIMA model on cabbage prediction

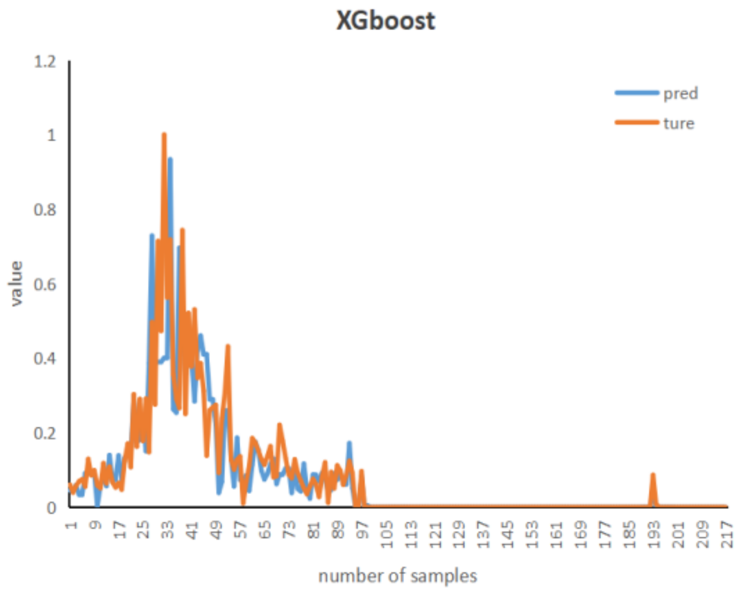

Figure 8 Effectiveness of XGboost model for cabbage prediction

LSTM model performance usually improves as the length of the time series increases. If the time series is very short, the LSTM may not be able to take full advantage of its internal memory mechanism. In this paper, it is possible that the relatively small amount of vegetable data makes the LSTM poorer at predicting sales volume than the seasonal ARIMA model, the XGboost model.The seasonal ARIMA is designed specifically to deal with seasonal and cyclical time series data, and it includes differential and seasonal component modeling. This makes it inherently suitable for dealing with these types of time series compared to the XGboost model is better in the prediction of vegetable sales volume, the CNN-LSTM-Transformer model, which integrates the CNN and Transformer mechanisms, can be a good solution to the long-term dependence of the LSTM on time, so even if the number of samples is relatively small, it can also make the prediction accuracy maintained at a high level, while LSTM is a deep learning model suitable for nonlinear relationships, many seasonal and cyclical time series contain nonlinear trends, LSTM can better capture and model these nonlinear relationships, while the seasonal ARIMA model may be limited by the assumption of linearity, so the CNN-LSTM-Transformer model has a higher accuracy in vegetable sales volume prediction effect is more accurate.

4. Conclusions

Vegetables are essential daily commodities, but their pricing and supply are vulnerable to seasonal fluctuations, introducing a degree of uncertainty. Supermarkets must adapt their purchasing and restocking strategies based on sales forecasts to avert economic losses arising from stockouts or overstocking. This proactive approach not only prevents issues like surplus, discounted, or discarded vegetables but also ensures that sought-after vegetables do not elude consumers due to shortages. Therefore, to enhance the prediction of vegetable sales volumes, reduce waste, and boost superstore profits, this paper employs the CNN-LSTM-Transformer model. This model excels in forecasting the sales volumes of vegetables with seasonal and cyclical variations.

The Convolutional Neural Network (CNN) adeptly captures spatial data features, while the Long Short-Term Memory network (LSTM) possesses memory functions for extracting temporal variations in vegetable sales data. Additionally, the Transformer model employs a parallel attention mechanism, effectively eliminating temporal dependencies. As a result, the integration of the Transformer model efficiently addresses the long-term dependence issues inherent in LSTM.

In this study, this paper focused on predicting the sales volumes of 11 common supermarket vegetables. In this paper utilized historical sales data, precipitation, and interdependencies between various vegetables as relevant feature inputs. So evaluated the prediction accuracy using RMSE as the performance metric and compared the CNN-LSTM-Transformer model to LSTM, CNN-LSTM, ARIMA, and XGBoost models. Experimental results conclusively demonstrate the superior predictive capabilities of the CNN-LSTM-Transformer model, underscoring its accuracy and practicality in forecasting vegetable sales volumes.

References

[1] Lian, J., Ma, L., Wu, X., Zhu, T., Liu, Q., Sun, Y., ... & Lou, X. (2023). Visualized pattern recognition optimization for apple mechanical damage by laser relaxation spectroscopy. International Journal of Food Properties, 26(1), 1566-1578.

[2] He Zhang & Yu Sun. (2020). Application of LightGBM and LSTM combined model in vegetable sales forecast.Journal of Physics: Conference Series(1) ,012110-.

[3] Sankaran S. Demand forecasting of fresh vegetable product by seasonal ARIMA model [J]. International Journal of Operational Research, 2014, 20(3). 315-330.

[4] Yunhao Cui. Research on fresh vegetables sales prediction based on CNN-PSO-LSTM combination model. 2023.Anhui Agricultural University, MA thesis.

[5] Feng Chen,AND Chen Zhide." Application of weighted combination model based on XGBoost and LSTM for sales forecasting." Computer Systems Applications 28.10 (2019):226-232. doi: 10.15888/j.cnki.csa.007091.

[6] Wang, X., Liao, T.,and Zhang, Shunxiang." An online multi-task sales prediction model based on CNN-LSTM network." Journal of Fuyang Normal University (Natural Science Edition) 38.02(2021):85-91. doi: 10.14096/j.cnki.cn34-1069/n/2096-9341(2021)02-0085-07.

[7] Ren Jianji, Bit Huihui, Zou Zhuolin, et al. Ultra-short-term power load forecasting based on CNN-BiLSTMAttention [J]. Power System Protection and Control, 2022, 50(8): 108-116. REN Jianji, WEI Huihui, ZOU Zhuolin, et al. Ultra-shortterm power load forecasting based on CNN- BiLSTMAttention[J]. Power System Protection and Control, 2022, 50(8): 108-116.

[8] Tang Ge, Yu Yiping, Qin Chuan, et al. Day-ahead interval prediction of bus load based on CNN-LSTM quantile regression [J]. Electric Power Engineering Technology, 2021, 40(4): 123-129. TANG Ge, YU Yiping, QIN Chuan, et al. Day-ahead interval prediction of bus load based on CNN-LSTM quantile regression [J]. Electric Power Engineering Technology, 2021, 40(4): 123-129.

[9] Juxiang Zhu, et al. "CEEMDAN-IGWO-CNN-LSTM based air quality predictive modeling." Computer Simulation.