

International Journal of Agriculture Extension and Social Development

Volume 7; SP-Issue 9; September 2024; Page No. 98-102

Received: 22-06-2024
Accepted: 30-07-2024

Indexed Journal
Peer Reviewed Journal

Application of long short term memory to forecast tomato prices

¹Ruqsar Khanum, ¹Siddayya, ²Ganavi NR and ¹Arjuman Banu

¹Institute of Agri-Business Management, University of Agricultural Sciences, GKVK, Bangalore, Karnataka, India

²Department of Agricultural Economics, University of Agricultural Sciences, GKVK, Bangalore, Karnataka, India

DOI: <https://doi.org/10.33545/26180723.2024.v7.i9Sb.1071>

Corresponding Author: Ruqsar Khanum

Abstract

Fluctuations in agricultural commodity prices affect the supply and demand of agricultural products and have a significant impact on consumers so accurate prediction of agricultural commodity prices would facilitate the reduction of risk caused by price fluctuations. This paper applies an Artificial Neural Network (ANN) method to forecast tomato prices. We showed how this new tool from machine learning, particularly Long-Short Term Memory (LSTM) models. Here we used LSTM model to implement price prediction as it is a best model for time series data. The research question investigated in this article is that whether and how the newly developed deep learning based algorithms for forecasting time series data, such as “Long Short-Term Memory (LSTM)”, are superior to the traditional algorithms. The empirical studies conducted and reported in this article showed that deep learning-based algorithms such as LSTM outperform traditional-based algorithms such as ARIMA and Holt-Winter Seasonal Method. From the result, LSTM is discovered to be the most accurate and efficient in handling increasing amounts of complex data. To evaluate the LSTM, weekly field prices of tomato for 10 years (20011-202) were used. The results showed that, compared to the ARIMA and the Holt-Winter Seasonal Method, the Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MBE) of LSTM showed lesser error value i.e., 267.66, 0.15 and 176.40 respectively. Thus, it provides a new idea for agricultural commodity price forecasting and has the potential to stabilize the supply and demand of agricultural products, thus help to understand the price fluctuation and also it provides an insight for the government in order to take necessary decisions for managing risk.

Keywords: ANN, deep learning, machine learning, LSTM, ARIMA, holt-winter seasonal method, RMSE, MAPE, MBE

Introduction

Agriculture holds immense significance in the Indian economy, with over two-thirds of the population directly or indirectly reliant on it, and over 70% of rural households deriving their livelihood from it. India ranks second globally in agricultural commodity production, contributing nearly 18% to the GDP. Price stability in agriculture is crucial for social sustainability, as sharp fluctuations impact producers, consumers, government policies, and national food security. (Guo *et al.*, 2022) ^[1].

Forecasting agricultural commodity prices is vital for farmers, policymakers, and administrative bodies to make informed decisions. Knowing future prices enables farmers to plan their crop selection and government agencies to implement schemes effectively. With increasing demand and diminishing agricultural land, forecasting is essential for the sustainability of future generations and the agricultural sector. (Mahto *et al.*, 2021) ^[4].

Agricultural price fluctuations pose challenges, affecting people's food expenditure and the entire supply chain from production to consumption. These fluctuations are influenced by various factors, both internal (production costs) and external (market environment, alternative prices), resulting in large, non-stationary, and nonlinear price changes. This uncertainty heightens risks for stakeholders, hindering the healthy development of agriculture and the

national economy. (Li *et al.*, 2010) ^[3].

Scientific data mining methods aid in understanding price patterns and predicting fluctuations, benefiting producers, consumers, processors, and rural development planners. However, despite various forecasting models, including statistical methods, no single efficient solution exists due to the complexity of price changes, which include both abrupt and gradual shifts. (Paul *et al.*, 2022) ^[5].

In addressing this complexity, machine learning models, particularly Long Short-Term Memory (LSTM) networks, have gained prominence. LSTMs, as recurrent neural networks, excel in learning sequence dependencies, offering the potential to predict time series data accurately by understanding contextual patterns. This departure from predefined assumptions in traditional models makes LSTMs particularly effective in nonlinear cases, offering promise in improving agricultural price forecasting methodologies for better decision-making and risk management. (Shin *et al.* 2018) ^[8].

Methodology

Data: For study purpose the prices of tomato were used for forecasting. The average weekly prices of tomato of Kolar market was collected from January 2011 to December 2020. The data was collected from the Krishimaratavahini. The monthly prices of tomato in the Kolar are predicted using

ARIMA, Holt-Winter's Seasonal method, and LSTM neural network. The performance of these models was evaluated for fitting the best model.

Proposed Techniques: Here, a model is fit for predicting the prices of Tomato. The performance of classical time-series models like Holt-Winter's Seasonal Method, ARIMA model and Machine learning model- LSTM were compared for finding out the most parsimonious model. Exploratory analysis indicated the presence of trend and seasonality in the time-series data of prices. ARIMA and Holt-Winter's seasonal method is the commonly used statistical method in forecasting data with trend and seasonality. These models are capable of producing forecasts with relatively well Performance. LSTM model was selected as it has the ability to capture the non-linear dependence of the data points that makes it favorable for time-series forecasting (Sabu and Kumar, 2020)^[6].

ARIMA: ARIMA (Auto-Regressive Integrated Moving Average) model is a variation of Box Jenkins models used for predicting or understanding stationary or non-stationary time-series data. The Box Jenkins model plays a vital role in the area of time-series analysis and forecasting. ARIMA consists of three parts- the AR part, lag-difference (integrated), and the MA part, denoted as ARIMA (p, d, q) where p is the number of AR terms, d is the order of lag difference, and q is the number of MA terms (Jadhav *et al.*, 2017)^[2].

Autoregressive process of order (p) is,

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t$$

Moving Average process of order (q) is,

$$Y_t = \mu - q_1 \epsilon_{t-1} - q_2 \epsilon_{t-2} - \dots - q_q \epsilon_{t-q} + \epsilon_t$$

and the general form of ARIMA model of order (p, d, q) is

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \mu - q_1 \epsilon_{t-1} - q_2 \epsilon_{t-2} - \dots - q_q \epsilon_{t-q} + \epsilon_t$$

Where, Y_t is tomato prices, ϵ_t 's are independently and normally distributed with zero mean and constant variance for $t = 1, 2, \dots, n$; d is the fraction differenced while interpreting AR and MA and ϕ_p and q_q are coefficients to be estimated.

The seasonal variation model of ARIMA is the SARIMA denoted as SARIMA (p, d, q) (P, D, Q), used for data with trend and seasonality. Also, it consists of seasonal AR, seasonal lag difference, and seasonal MA terms. Here P is the number of seasonal AR terms, D is the order of seasonal lag difference, and Q is the number of seasonal MA terms. SARIMA model equation is written as follows: $(B)(B)^H D d X_t = + \theta(B)\theta(B)Z_t$

Holt-Winter Seasonal Method: Holt-Winter Seasonal Method is one of the variations of Exponential Smoothing used for seasonal data. It consists of a forecast equation and three smoothing equations:

$$y_{t+h/t} = hbt + st+h- (k+1)$$

- Trend

$$bt = * (lt - lt-1) + (1 - *) bt-1$$

- Level

$$l_t = + (1 -) (l_{t-1} + b_{t-1})$$

- Seasonal

$$= (y_t - l_{t-1} - b_{t-1}) + (1 -) s_{t-m}$$

Where, * and are smoothing parameters and m is the frequency of seasonal (Choon and Chuin 2008)

Long Short Term Memory (LSTM): LSTM Introduced by Hochreiter & Schmidhuber. LSTM is a type of RNN that can learn the order dependence between items in a sequence. It is able to learn the context required to make predictions in time series forecasting problems, rather than having this context pre-specified and fixed. Its capable of learning long-term dependencies. It is trained using Back propagation through time and overcomes the vanishing gradient problem. (Sherstinsky, 2020)^[7].

Stationary analysis: Augmented-Dickey Fuller test was used to check the stationarity. It is a hypothesis test in which the null hypothesis is that the series is non-stationary, and the alternate hypothesis is set to series is stationary. Based on the test, the p-value was found to be, the alternate hypothesis implying the stationarity is rejected.

Model Fitting: The dataset was divided into training and test set. The training set consists of first two third of observations and the remaining observations were taken as the test set.

Loss function

Root Mean Square Error (RMSE): is a metric commonly used to measure the accuracy of a forecasting model. It calculates the square root of the average of the squared differences between the predicted values and the actual (observed) values. RMSE is a popular choice because it penalizes larger errors more heavily, making it sensitive to outliers. The formula for calculating RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Where:

- n is the number of data points.
- y_i is the actual value of the dependent variable for the i th data point.
- $\hat{y}_i$ is the predicted value of the dependent variable for the i th data point.

Mean Absolute Error (MAE): Mean Absolute Error (MAE) is another metric used to assess the accuracy of a forecasting model. It computes the average of the absolute differences between the predicted values and the actual (observed) values.

The formula for calculating MAE is as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Where:

- n is the number of data points.
- y_i is the actual value of the dependent variable for the i th data point.
- $\hat{y}_i$ is the predicted value of the dependent variable for the i th data point.

Mean Absolute Percentage Error (MAPE): Mean Absolute Percentage Error (MAPE) is a metric used to evaluate the accuracy of a forecasting model in terms of percentage errors. It measures the average absolute percentage difference between the predicted values and the actual (observed) values.

The formula for calculating MAPE is as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Where:

- n is the number of data points.
- y_i is the actual value of the dependent variable for the i th data point.
- $\hat{y}_i$ is the predicted value of the dependent variable for the i th data point.

Results and Discussion

The implementation of proposed LSTM model, SARIMA and Holt Winters using python. Figure 1 presents a detailed visualization of the weekly price distribution of tomatoes in the Kolar market spanning the years 2011 to 2020. The graph vividly illustrates the substantial fluctuations observed in the weekly prices of tomatoes over this period. Figure 2 presents the decomposition of the time series components of tomato prices in the Kolar market from 2011 to 2022, showcasing the individual contributions of the trend, seasonal, and residual factors.

As the variability in tomato prices increases, the task of forecasting becomes more challenging. To address this, the study employs three different forecasting methods: SARIMA, Holt Winters, and LSTM. The primary objective is to compare these methods and determine which one yields the most accurate predictions.

To conduct the comparison, the tomato price data is split into two segments: a training set comprising 60% of the data and a testing set containing the remaining 40%. This division allows for assessing the accuracy of predictions made by SARIMA, Holt Winters, and LSTM on unseen data. Due to the presence of seasonality in tomato prices, the Holt Winters method proves to be the most suitable approach for effectively capturing both trends and season

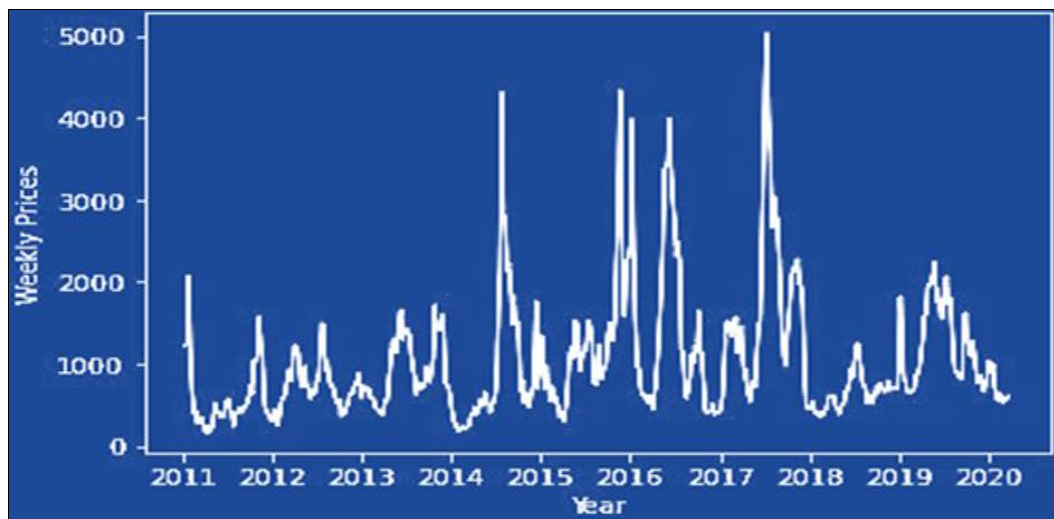


Fig 1: Weekly average prices of tomato in Kolar market from 2011 to 2020.

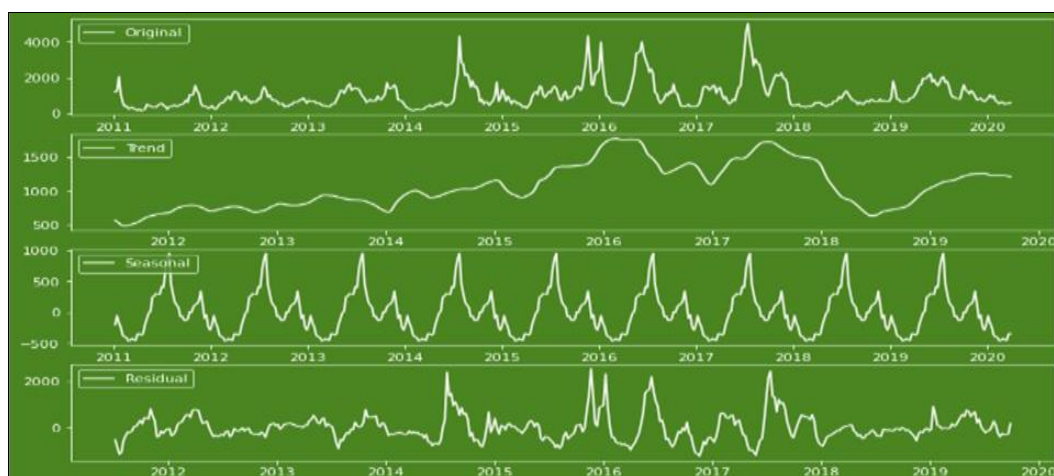


Fig. 2: Decomposition of time series component

In Figure 3, we observed the forecast of tomato prices using the Holt Winters method. The blue line represents the training data, while the orange line represents the data used for testing the model, and the green line represents the predicted data. It is noticeable that there is a considerable variation between the tested and predicted data, indicating potential areas for improvement in the forecasting model. (Yin *et al.* 2020) ^[9] Considering the seasonality in tomato

prices, the SARIMA method emerges as the most appropriate approach for effectively capturing trends and seasonal patterns. In Figure 4, we can examine the forecast of tomato prices generated using the SARIMA method. The blue line corresponds to the training data, the orange line reflects the data utilized for testing the model, and the green line represents the predicted data. Notably, there exists a significant variation between the tested and predicted data.

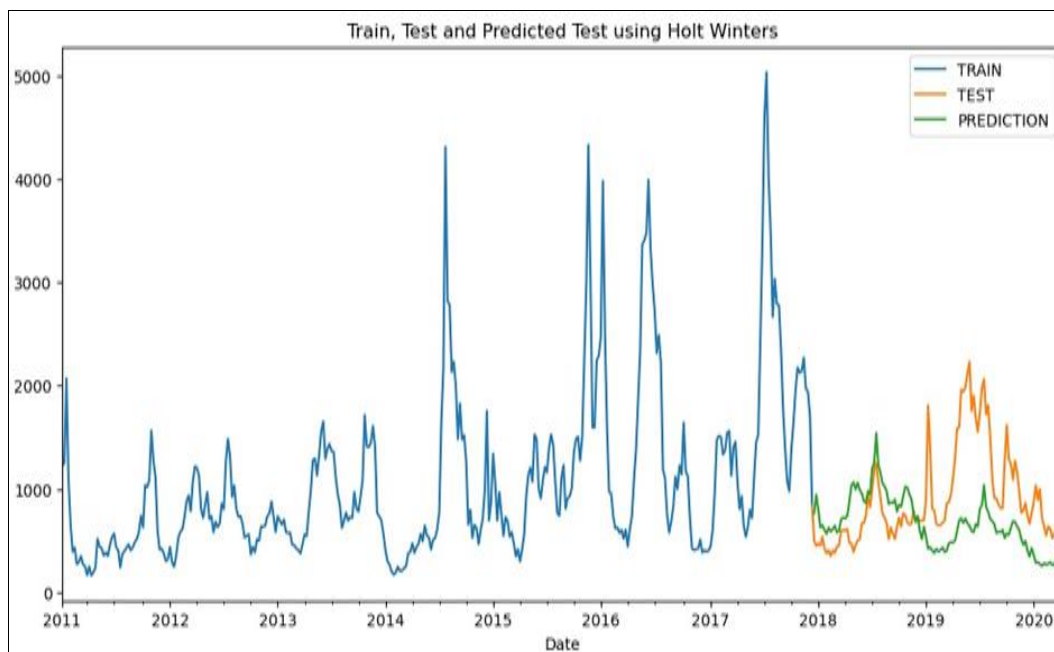


Fig. 3: Forecasting for weekly prices of tomato using Holt Winters method

In Figure 5, the forecast of tomato prices using the LSTM method demonstrates a remarkable level of accuracy, with minimal variation observed between the tested and predicted data. The green line representing the predicted data closely aligns with the orange line depicting the tested data, indicating that the LSTM model has successfully captured the underlying patterns and trends in the tomato price time

series. The minimal variation between the tested and predicted data points to the robustness of the LSTM method in handling the complexities of the tomato price fluctuations. LSTM's ability to learn long-term dependencies and capture context from sequential data enables it to make precise predictions, even in the presence of inherent volatility and fluctuations.

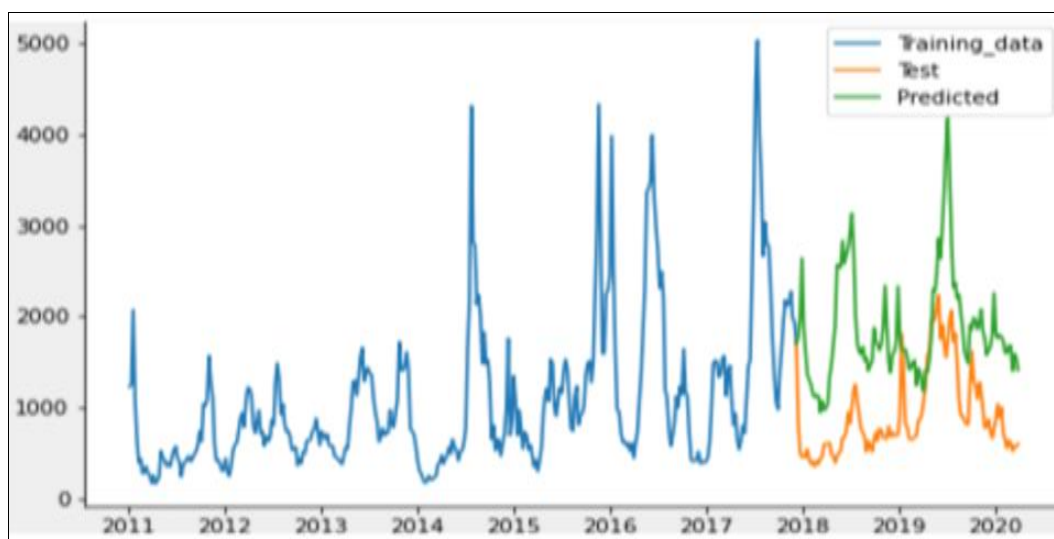


Fig.4: Forecasting for weekly prices of tomato using SARIMA model

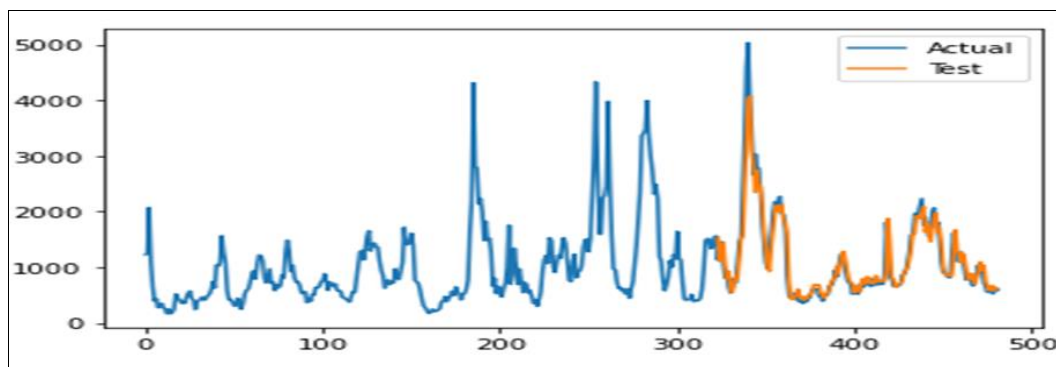


Fig. 5: Forecasting for weekly prices of tomato using LSTM

The table 1 provided shows the performance analysis of the three forecasting methods: SARIMA, Holt Winters, and LSTM. The evaluation is based on three error metrics: Residual Mean Square Error (RMSE), Mean Absolute Error

(MAE), and Mean Absolute Percentage Error (MAPE). These metrics are commonly used to assess the accuracy of forecasting models.

Table 1: Performance analysis

Model	Residual Mean Square Error	Mean Absolute Error	Mean Absolute Percentage Error
SARIMA (0, 1, 0)x(2, 1, 0, 52)	1074.89	947.57	0.63
Holt-Winter's Seasonal Method	568.55	449.42	0.48
LSTM	279.59	176.40	0.15

The SARIMA model exhibits comparatively higher error metrics, with RMSE and MAE values indicating significant deviations from actual tomato prices. Its MAPE of 0.63 reflects a substantial 63% average prediction error. Holt Winters performs better than SARIMA but is still surpassed by LSTM, showcasing lower error metrics with a MAPE of 0.48, suggesting a 48% average error. LSTM stands out as the most accurate method, boasting the lowest error metrics with a MAPE of 0.15, indicating a mere 15% average prediction error. Consequently, the study concludes LSTM as the superior method for forecasting tomato prices in the Kolar market, attributing its success to its capability to grasp long-term dependencies and contextual nuances within sequential data, ensuring more precise predictions compared to SARIMA and Holt Winters.

Conclusion

This study aimed to assess tomato price forecasting methods in the Kolar market, crucial for India's agricultural sector. SARIMA, Holt Winters, and LSTM were evaluated due to their significance for farmers, policymakers, and consumers. The analysis revealed the complexity and seasonality of tomato price data, posing challenges for accurate forecasting. LSTM emerged as the most accurate method, surpassing SARIMA and Holt Winters, as evidenced by lower RMSE, MAE, and MAPE values.

References

- Guo Y, Tang D, Tang W, Yang S, Tang Q, Feng Y, *et al.* Agricultural price prediction based on combined forecasting model under spatial-temporal influencing factors. *Sustainability*. 2022;14(17):10483. DOI: 10.3390/su141710483.
- Jadhav V, Chinnappa RB, Gaddi GM. Application of ARIMA model for forecasting agricultural prices. *J Agr Sci Tech*. 2017;19:981-992. Available from: <http://jast.modares.ac.ir/article-23-2638-en.html>.
- Li GQ, Xu SW, Li ZM. Short-term price forecasting for agro-products using artificial neural networks. *Agric Agric Sci Procedia*. 2010;1:278-287. DOI: 10.1016/j.aaspro.2010.09.035.
- Mahto AK, Alam MA, Biswas R, Ahmed J, Alam SI. Short-term forecasting of agriculture commodities in context of Indian market for sustainable agriculture by using the artificial neural network. *J Food Qual*. 2021;2021:1-13. DOI: 10.1155/2021/9939906.
- Paul RK, Yeasin M, Kumar P, Kumar P, Balasubramanian M, Roy HS, *et al.* Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India. *PLoS One*. 2022;17(7). DOI: 10.1371/journal.pone.0270553.
- Sabu KM, Kumar TM. Predictive analytics in agriculture: Forecasting prices of arecanuts in Kerala. *Procedia Comput Sci*. 2020;171:699-708. DOI: 10.1016/j.procs.2020.04.076.
- Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D Nonlinear Phenom*. 2020;404:132306. DOI: 10.1016/j.physd.2019.132306.
- Shin S, Lee M, Song SK. A prediction model for agricultural products price with LSTM network. *J Korea Contents Assoc*. 2018;18(11):416-29. DOI: 10.3390/agriculture12020256.
- Yin H, Jin D, Gu YH, Park CJ, Han SK, Yoo SJ. STL-ATTLSTM: vegetable price forecasting using STL and attention mechanism-based LSTM. *Agriculture*. 2020;10(12):612. DOI: 10.3390/agriculture101206.