Data Science and Engineering (DSE) Record, Volume 3, issue 1.

Shallot Price Forecasting Model Using Hybrid ARIMA-LSTM Model

Kanokrot Phuruan[1] and Chompoonoot Kasemset [2]

[1] Graduate Program in Data Science, Chiang Mai University, Chiang Mai, Thailand
[2] Department of Industrial Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand

kanokrot_p@cmu.ac.th, chompoonoot.kasemset@cmu.ac.th

Abstract. Shallot is one of important agricultural products exported with high volume to many countries. Shallots are mainly cultivated at the northern region of Thailand. Price of shallot in different periods during a year is changed from many related parameters. This research aimed to develop the forecasting model of shallot's price using combination techniques from ARIMA and LSTM (ARIMA-LSTM). Considering independent parameters, ARIMA was applied for predicting effects from parameters with time-series and linear relationship, whereas LSTM was applied for predicting effects from parameters with non-linear relationship. Data collected 84 months during January 2014 to December 2020 were applied in this research. The accuracy of the proposed model was evaluated using three indicators including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The results presented that our ARIMA-LSTM model gave minimum values of RMSE, MAE and MAPE as 10.275 Baht, 8.512 Baht, and 13.618%, respectively. Moreover, the value of MAPE was in good forecasting level that can be implemented practically.

Keywords: Shallot price, ARIMA-LSTM, forecasting.

1 Introduction

Shallot is one of important agricultural products. Shallots can be grown year-round but are usually grown in the winter which will begin to mature within 70 to 110 days. If planted in the rainy season, harvesting can be done when shallots are 45 days [1]. Thailand has a high export volume of shallot. The buying market is Malaysia, Indonesia, Singapore, and the Middle East. There are countries in Europe also such as Netherlands, Germany, England, etc. The main areas where shallots are grown are in Thailand's Northern, Northeast, and West.

Even during January-March 2020, the average Agricultural Production Index was 139.0, down from 148.4 in the same period of 2019, or decreasing 6.3 per-cent, as shown in Fig. 1. While the Agricultural Price Index had an average of 137.3, increasing from 126.2 in the same period of 2019 or increasing 8.9 per-cent, as shown in Fig. 2. However, the shallot price situation in 2020 generates a lot of income for farmers when compared to 2019 [2]. It is expected that the farmers may increase the area of cultivation in the next production cycle. Due to higher prices and market demand.

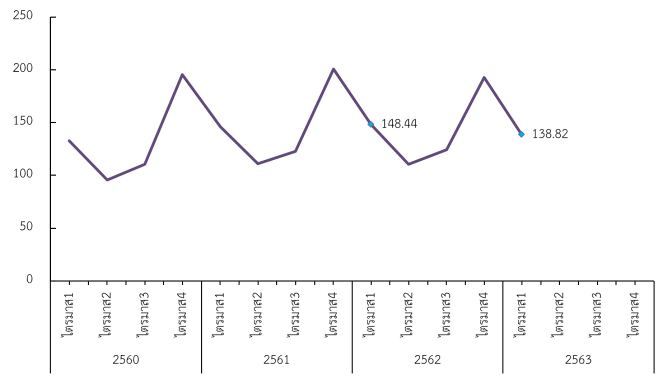

Fig 1. Quarterly Agricultural Production Index from 2017-2020.

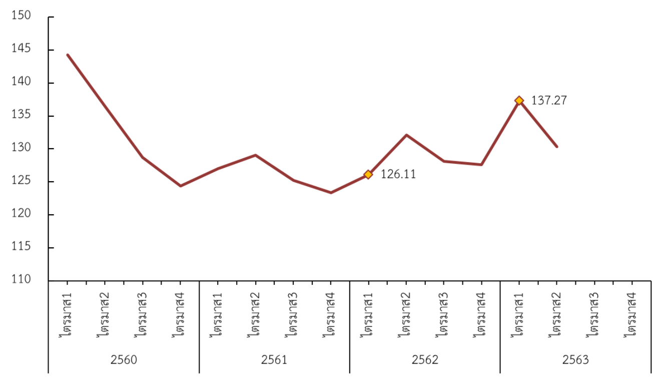

Fig 2. Quarterly Agricultural Price Index from 2017-2020.

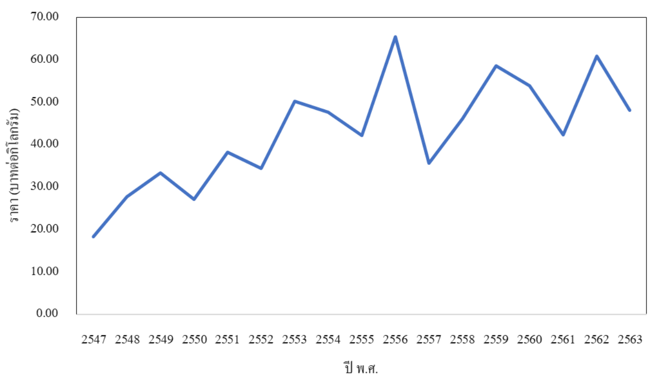

Fig 3. Monthly shallot price from 2017-2020.

When considering the annual price of shallots from 2004 to 2020 as shown in Fig. 3. It was found that shallot price tends to increase, but still has high volatility. This may be caused by economic conditions, cultivation area, harvesting area, the number of households of shallot farmers, import and export volumes, the weather, and other factors. This research aims to develop a combination model between the time series model and machine learning model considering independent variables. The model's accuracy was evaluated by using three indicators including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). This prediction model can help farmers to plan for future cultivation

2 Literature Review

Ayse SOY TEMUR et al. [3] forecasted real estate sales in Turkey to determine how many houses will be sold in the next year. 124-month data set were used from January 2008 to April 2018 using ARIMA model and LSTM model. The results were evaluated using MAPE and MSE. The best model with the lowest error was the hybrid ARIMA-LSTM model.

S. Kiran and K. Manoj [4] forecasted monthly arecanut prices in Kerala, India from 2007 to 2017 using the SARIMA model, the Winter's Seasonal Forecasting model, and the Long Short-Term Memory model (LSTM) using the root mean square error (RMSE) to evaluate model's accuracy. The results showed that the SARIMA (1,0,0)(0,1,2) model had 16.5475 RMSE, the Winter's Forecasting Model had 18.0589 RMSE, and the LSTM model had 7.2780 RMSE, which can be concluded that the LSTM model is the most appropriate model for arecanut prices forecasting.

D. Jin et al. [5] forecasted the price of Chinese cabbages and radishes in Korean agricultural market using the Long short-term memory model (LSTM) and using the seasonal-trend-loess (STL) in the data preprocessing method and considering vegetable prices and meteorological data of each area in model. The results showed that the accuracies of the model in forecasting the prices of Chinese cabbages and radishes were 92.06% and 88.74%, respectively.
Data and Methodology

3 Data and Methodology

3.1 Data

This study collected data of monthly price of Northern shallot from the period from January 2014 to December 2020 from the Department of Internal Trade. In addition, related parameters, average monthly temperature, and rainfall of Chiang Mai, Lamphun and Phayao, were input in the LSTM model. Data splitting into 2 sets as shown in Fig. 4.

1) Training dataset: the data used to teach machine learning models to recognize pattern of time series data, from January 2014 to December 2019.

2) Testing dataset: the data used to evaluate model's accuracy, from January 2020 to December 2020.

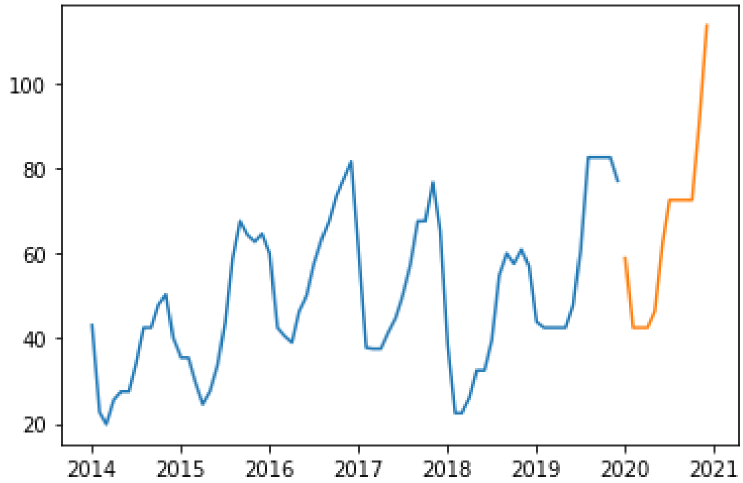

Fig 4. Monthly shallot price data split into training and testing dataset.

3.2 Methodology

1) ARIMA

ARIMA models is widely used approaches to time series forecasting to predict future value. ARIMA model is generally denoted as ARIMA(p, d, q) where:

p: the number of lag observations in the model (AR)
d: degree of differencing of times series data and previous value
q: the order of moving average model (MA) [6]

2) LSTM

LSTM model has been developed by using cell state and hidden state information to collect and process the data at the next interval. It uses different gates for computation consists of input gate, output gate, and forget gate where the cell remembered the value at a given interval and the three gates control the flow of data into and out of the cell based on the value of the weight. If there is a small value, it will not be able to pass. This prevents the occurrence of vanishing gradients. It is calculated according to the following equation:

$$i_t = \sigma(x_t W_i + h_{t-1} U_i + b_i) \quad (1)$$

$$f_t = \sigma(x_t W_f + h_{t-1} U_f + b_f) \quad (2)$$

$$o_t = \sigma(x_t W_{x^o} + h_{t-1} W_{h^o} + b_o) \quad (3)$$

$$c_t = (c_{t-1} \times f_t) + (i_t \times f_h(x_t W_c + h_{t-1} U_c + b_c)) \quad (4)$$

$$h_t = \sigma(c_t) \times o_t \quad (5)$$

where t is the interval, i is the input gate, f_t is the forget gate, o_t is the output gate, σ is the activation function, h_t is the hidden state, c_t is the cell state and W, U is the weight matrix for calculating previous input and hidden state and current hidden state respectively [7]. LSTM structure shown in Fig. 5.

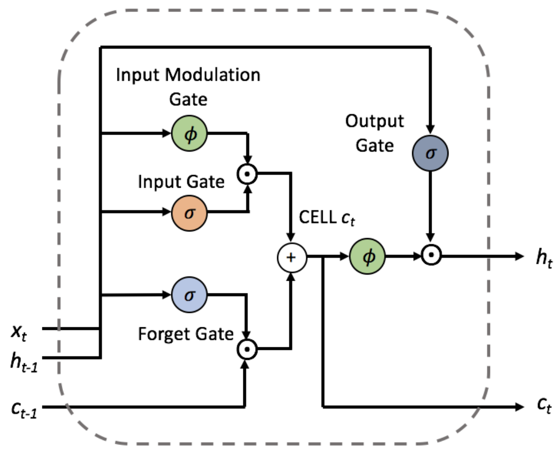

Fig. 5 LSTM structure.

3) ARIMA-LSTM

Many time series models are consisting of linear relationships and non-linear relationships. ARIMA is suitable for predicting linear relationships. LSTM is suitable for both of linear and non-linear relationships. In order to get better prediction results, hybrid models based on principle of separate modeling of linear and non-linear components of time series were employed. The hybrid model concept aims to make the model more diverse with better predictive results. The results obtained from the hybrid model and the results obtained from individual model, even though they are not related to each other, but it can reduce error or general variance. For this reason, hybrid models are the most successful model in forecasting [8].

To predict shallot price by ARIMA-LSTM hybrid model shown in Fig. 6.

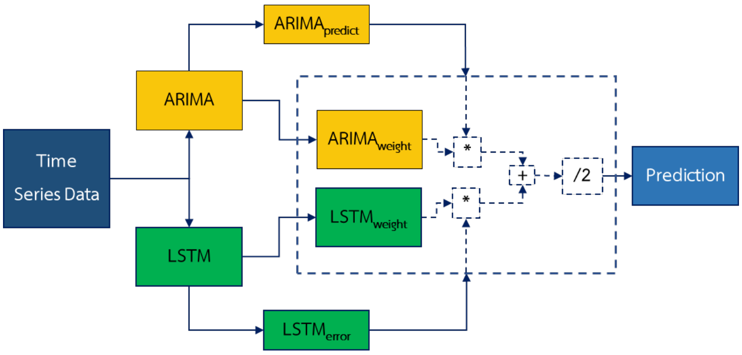

Fig. 6 Shallot price prediction by ARIMA-LSTM.

The time series forecasting model usually expressed as the sum of the linear and non-linear components as shown in equation 6.

$$y_t = L_t + N_t \qquad (6)$$

Where L_t is the linear component in time series whereas N_t shows non-linear component. In the hybrid model, L_t is predicted using the ARIMA model, then N_t is predicted using the LSTM model. The error values of are calculated according to equations 7 and 8.

$$LSTM_{error} = LSTM\_MEAN[error] \qquad (7)$$
$$ARIMA_{error} = ARIMA\_MEAN[error] \qquad (8)$$

The weights of each model were calculated according to equation 9 and 10.

$$LSTM_{weight} = \left(1 - \left(\frac{LSTM_{error}}{LSTM_{error}+ARIMA_{error}}\right)\right) \times 2 \qquad (9)$$
$$ARIMA_{weight} = 2 - LSTM_{weight} \qquad (10)$$

The result of the hybrid model can be calculated as equation 11 [3].

$$Hybrid_{predict}[i] = \left(\frac{\left(ARIMA_{weight} \times ARIMA_{pred}[i]\right) + \left(LSTM_{weight} \times LSTM_{error}\,[i]\right)}{2}\right) \qquad (11)$$

4) Evaluation

The model's accuracy was evaluated using three indicators including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) as shown in Table 1.

Table 1. Model evaluation formulas.

Name	Formulas
Root Mean Squared Error	$RMSE = \sqrt{mean\big((Y_t - \hat{Y}_t)^2\big)}$
Mean Absolute Error	$MAE = mean(\lvert Y_t - \hat{Y}_t \rvert)$
Mean Absolute Percentage Error	$MAPE = \left(mean\left(\left\lvert \frac{Y_t - \hat{Y}_t}{Y_t} \right\rvert\right)\right) \times 100$

4 Results

The results obtained from the hybrid ARIMA-LSTM model are shown in Table 2.

Table 2. The hybrid ARIMA-LSTM model prediction results.

Month	Actual price (1)	Predicted value from hybrid ARIMA-LSTM model (2)	Error from hybrid (1) – (2)
Jan-20	58.86	66.88	-8.02
Feb-20	42.5	53.47	-10.97
Mar-20	42.5	51.4	-8.9
Apr-20	42.5	49.92	-7.42
May-20	46.39	55.24	-8.85
Jun-20	61.79	59.69	2.1
Jul-20	72.5	67.52	4.98
Aug-20	72.5	75.29	-2.79
Sep-20	72.5	79.71	-7.21
Oct-20	72.5	83.31	-10.81
Nov-20	91.58	86.83	4.75
Dec-20	113.5	88.18	25.32

Table 3. The hybrid ARIMA-LSTM model's error.

	MAE	RMSE	MAPE
ARIMA-LSTM	8.512	10.275	13.618

MAPE	Forecasting power
<10%	Highly accurate forecasting
10%~20%	Good forecasting
20%~50%	Reasonable forecasting
>50%	Weak and inaccurate forecasting
Source: Lewis (1982)	

Fig. 7 Criteria for evaluating predictions based on MAPE values.

Table 2 presented the predicted values from ARIMA-LSTM compared with actual price. Then, MAE, RMSE and MAPE were calculated and presented as Table 3 as 8.512 Baht, 10.275 Baht, and 13.618%, respectively. Based on Fig.7, interpretation of MAPE results for forecasting accuracy, MAPE within 10% to 20% presented good forecasting [9].

5 Discussion and Conclusion

This research using 84 months data of shallot price from January 2014 to December 2020. The hybrid ARIMA-LSTM model was predicted using ARIMA model to predict the linear relationship, and the LSTM model to predict the non-linear relationship. The prediction results were obtained from ARIMA prediction and LSTM prediction error. The results showed that the proposed hybrid ARIMA-LSTM produced values of MAE, RMSE and MAPE as 8.512 Baht, 10.275 Baht, and 13.618%, respectively. The value of MAPE was in good forecasting level that can be implemented practically. This is consistent with other studies in which the hybrid model was used to forecast time series data with the ARIMA-LSTM hybrid model that also gave good prediction results.

References

1. พลังเกษตร.com, การปลูกหอมแดง เจาะลึกขั้นตอน การปลูก สายพันธุ์และการเก็บเกี่ยว, 2019 [Online]. Available: https://www.palangkaset.com/ [Access August 30, 2020].

2. สำนักงานเศรษฐกิจการเกษตร, บทวิเคราะห์ ดัชนีเศรษฐกิจการเกษตร เดือนมิถุนายน 2563, 2020 [Online]. Available: http://www.oae.go.th/view/1/ดัชนีราคาและผลผลิต/TH-TH [Access July 19, 2020].
3. Ayse SOY TEMUR et al, "Predicting Housing Sales in Turkey Using ARIMA, LSTM And Hybrid Models", Journal of Business Economics and Management, Vol.20, pp. 927-928, 2019.
4. Kiran M. Sabu et al, "Predictive analytics in Agriculture: Forecasting prices of Arecanuts in Kerala", Procedia Computer Science, 171, pp. 699-708, 2020.
5. D. Jin, H. Yin, Y. Gu and S. J. Yoo, "Forecasting of Vegetable Prices using STL-LSTM Method", 2019 6th International Conference on Systems and Informatics (ICSAI), Shanghai, China, pp. 866-871, 2019.
6. อังคณา ตาเสนาและ ธีรศิลป์ กันธา, การพยากรณ์ผู้โดยสารสนามบินแม่สอด จังหวัดตาก. มหาวิทยาลัยราชภัฏกำแพงเพชร, 100-103, 2561.
7. เธียรศักดิ์ พลาดิศัยเลิศและ ธนิศา นุ่มนนท์, การเปรียบเทียบโมเดลการเรียนรู้ของเครื่องแบบต่างๆ สำหรับการทำนายราคาบิทคอยน์, วารสารเทคโนโลยีสารสนเทศลาดกระบัง, 6(1), 2561.
8. Shwet Ketu and Pramod Kumar Mishra, "A Hybrid Deep Learning Model for COVID-19 Prediction and Current Status of Clinical Trials Worldwide", Computers, Materials & Continua, Vol.66, pp. 1903-1904, 2021.
9. Colin David Lewis, "Industrial and Business Forecasting Methods", London: Butterworth Scientific, pp. 40, 1982.