International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

Food Crop Farmer Index Price Prediction using LSTM during Covid-19 Pandemic

Irvan Setiawan[1], Tristyanti Yusnitasari[2], Teddy Oswari[3], Reni Diah Kusumawati[4]
[1]Faculty of Computer Science, Gunadarma University, Jakarta, Indonesia, Country-16421
[2]Faculty of Computer Science, Gunadarma University, Jakarta, Indonesia, 16424
[3]Faculty of Economic, Gunadarma University, Jakarta, Indonesia, 16424
[4]Faculty of Economic, Gunadarma University, Jakarta, Indonesia, 16424

Abstract**—** *The Indonesian economy is largely influenced by the agricultural sector because it has great potential in this agricultural country. The majority of the workforce in Indonesia work as farmers [4]. In addition, based on the main macroeconomic variables in the form of the composition of the labor force and the price index received by farmers (IT), the agricultural sector is considered to have a significant influence on the national economy. The price index variable received by farmers is a variable that indicates the level of development of farmers' production in Indonesia. During the Covid- 19 pandemic, which has existed since 2020, the agricultural sector still has an important role in national economic growth. However, this does not remove the impact experienced by the agricultural sector during the Covid-19 pandemic. One of the impacts received is the disruption of farmers' production in all agricultural areas. Therefore, to maintain the stability of national economic growth in the agricultural sector, predictions can be made on the IT variable. In this study, IT prediction or forecasting will be carried out using the Long Short Term Memory (LSTM) algorithm with parameters LSTM 200, dropout 0.2, and epoch 100. The results of the forecasting show that the LSTM model has an MSE value of 0.1173.*

Keywords**—** *Index price, food crop farmer, forecasting, time-series data, LSTM, MSE.*

I. INTRODUCTION

Agriculture is a sector that has great influence and potential to be exploited for the Indonesian economy. The agricultural sector in Indonesia has experienced a period of ups and downs [2]. The Indonesian agricultural sector is considered to be very influential on the national economy based on the main macroeconomic variables in the form of the composition of the workforce and the price index received by farmers (IT). Most of the workforce in Indonesia, especially for the small group, works in the agricultural sector [4]. The price index received by farmers (IT) is a value that shows the level of development of farmer production.

During the COVID-19 pandemic, Indonesia's agricultural sector is considered to still have a role in national economic growth. This is known based on data from the Central Statistics Agency (BPS) which shows that the agricultural sector continues to grow positively during the COVID-19 pandemic [6]. In the second and third quarters of 2020, the food crop subsector grew 9.23% and 7.14% respectively. Even so, there are still impacts from the COVID-19 pandemic on Indonesia's agricultural sector. One of them is the disruption of farmers' production in all regions (Commission IV DPR RI Press Release 23/4/2020) due to the impact on market & agricultural prices, customer supply chains slowing down and tend to decrease, decreasing the quality of health and labor of farmers, lack of work safety, and decreasing quality of food resources. Therefore, one solution to maintaining the stability of national economic growth in the agricultural sector is to predict the development of food crop farmers' production.

Prediction of the development of the production of food crop farmers can be done by implementing machine learning in a time series data. One of the time series data that can be used to predict the development of the production of food crop farmers is the price index data received by farmers during the COVID-19 (IT) pandemic. This data can be processed with machine learning so that it can be used to predict IT in the future. Prediction made by implementing machine learning in a time series is also known as forecasting. Prediction or forecasting will be done by analyzing the forecasting results obtained from Long Short Term Memory (LSTM) algorithm. The LSTM model is developed in this research have several limitations, such as:

1. The research will be conducted using LSTM algorithms only.
2. The dataset used is IT data received by food plants obtained from the Badan Pusat Statistik (BPS) from April 2020 to April 2021.
3. The model was developed using the python programming language.

The LSTM algorithm was chosen as the second algorithm of this research because the LSTM algorithm is a time series forecasting algorithm that can predict time series with a long period or time. In training the model, the input data will have a .csv format which contains 2 columns, namely the date and index column and 13 rows consisting of 13 pairs of months and index.

II. RELATED WORK

There have been many time-series forecasting research using LSTM. These researches are been done to facilitate human work by making computers able to predict or forecast a value in a time-series data in the future. The researches that have been carried out are forecast on crop production such as rice and wheat from the National Food Security Mission in India. This research is being carried out in three different areas, which are the Area of Cultivation, Production, and Yielding of rice (Kiran Kumar Paidipati, 2019). The research

Irvan Setiawan, Tristyanti Yusnitasari, Teddy Oswari, and Reni Diah Kusumawati, "Food Crop Farmer Index Price Prediction using LSTM During Covid-19 Pandemic," *International Research Journal of Advanced Engineering and Science*, Volume 6, Issue 3, pp. 122-125, 2021.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

was carried out with creating a comparative analysis on the ARIMA and LSTM-NN models. Based on the analysis results, the LSTM-NN model is considered to be more capable in forecasting. The LSTM-NN model has a positive percentage error in forecasting value, while the ARIMA model has a negative value. In addition, there is also research that being conducted to predict quantitative amounts of rainfall and predictions of drought events to facilitate early warning of drought in Northeast China (Xianghua Wu et al. 2021). The forecasting in this research is conducted by comparative analysis between 3 models, which are Wavelete-ARIMA-LSTM (W-AL), Auto Regressive Integrated Moving Average (ARIMA), and Long Short-Term. Memory (LSTM). The results of the comparative analysis that have been carried out show that the W-AL hybrid model has a higher forecasting accuracy than the ARIMA and LSTM models.

There are two more researches that are related to data-series forecasting with ARIMA and LSTM. There is a research conducted for forecasting CPU usage (workload). Forecasting in this research was conducted by analyzing 2 models, which are the ARIMA model and the LSTM model on the Google cluster dataset dataset (29-day trace period). Based on the validated performance results based on the Root Mean Square Error (RMSE) value, it is known that the LSTM model has a better performance than the ARIMA model. ARIMA model with parameter (2,1,2) has a forecasting error value between 37,331% to 42,881% (overfitting). While the LSTM model which has Neurons: 5 and epoch: 3000 parameters has a forecasting error value between 17.566% to 23.65% (Deepak Janardhanan, 2017). There is also research that was being conducted to forecast the price of Bitcoin using the ARIMA and LSTM models. Forecasting is carried out on a dataset containing 10,000 data in the form of prices information on the Bitfinex website. Based on the results of forecasting the LSTM model with the Epochs: 100 parameter requires more time to train the model than the ARIMA model with parameter (1,1,0). But based on forecasting results it is known that the LSTM model can predict better than the ARIMA model. The LSTM model has an average error rate of 0.4765938 and a standard deviation of 2.092208 (Yiqing Hua, 2020).

III. Literature Review

A. Agriculture

Agriculture includes various activities such as farming, animal husbandry, fisheries, and also forestry. Farming is a job that dominates livelihoods in Indonesia because approximately 100 million people or nearly half of the total number of Indonesians work in the agricultural sector [4]. Agriculture is a major economic sector in developing countries. The role and contribution of the agricultural sector in economic development in a country has a priority position compared to other sectors. In developing countries, food production generally dominates the agricultural sector. If output increases due to increased productivity, farmers' income will tend to increase if demand can compensate for it.

The price index received by farmers (I_t) is a price index that shows the development of producer prices for farmer products [1]. This price index measures the average price change (fluctuation) in a period of a type of agricultural product at the producer price level. This price index is also used as supporting data in calculating agricultural sector income.

B. Price Index received by Farmers (I_t)

The price index received by farmers (I_t) is a price index that shows the development of producer prices for farmer products [1]. This price index measures the average price change (fluctuation) in a period of a type of agricultural product at the producer price level. This price index is also used as supporting data in calculating agricultural sector income. I_t is calculated based on the selling value of agricultural products produced by farmers including the food crop sector, horticulture, plantation crops, animal husbandry, fishermen and fish farming. In addition to the selling value of agricultural products, I_t is also calculated based on the Rural Producer Price Survey, Farmers Exchange Rate Weigh Chart Survey, Agricultural Census, and Farming Business Cost Structure Survey. The higher the I_t value, the higher the production value produced by farmers. Meanwhile, if it decreases, the income received by farmers will be lower.

C. Long Short Term Memory (LSTM)

Long Short Term Memory (LSTM) is a development of the Recurrent Neural Network (RNN) by overcoming one of the shortcomings of RNN, which is the ability to manage information over a long period of time. Proposed by Sepp Hochreiter and Jurgen Schmidhuber in 1997, LSTM is widely chosen for time series based predictions because it is known to be superior and reliable in making predictions over a long period of time compared to other algorithms [9]. LSTM has a network structure as shown in Figure 1.

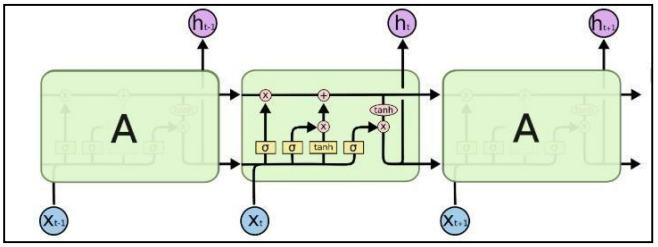

Fig. 1. LSTM

The LSTM model filters information through a gate structure to maintain and update the state of the memory cells. The door structure includes an input gate, forget gate, and output gate. Each memory cell has three sigmoid layers and one tanh layer. One LSTM cell has a path connecting the old memory cell ($Ct - 1$) to the new memory cell (Ct). Memory cell is a horizontal line that connects all the output layers on the LSTM. With this path, an old memory cell value can easily be passed to the new memory cell with minor modifications. The LSTM has the ability to add or delete previous information that entered the current cell. The sigmoid layer displays a number between zero and one, describing how much of each component should be allowed in. A zero value is interpreted as "may not enter" while a value of one is interpreted as "please enter".

Irvan Setiawan, Tristyanti Yusnitasari, Teddy Oswari, and Reni Diah Kusumawati, "Food Crop Farmer Index Price Prediction using LSTM During Covid-19 Pandemic," *International Research Journal of Advanced Engineering and Science*, Volume 6, Issue 3, pp. 122-125, 2021.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

IV. RESEARCH METHOD

The implementation of the LSTM methods in time series forecasting will be formulated in the flowchart in Figure 2.

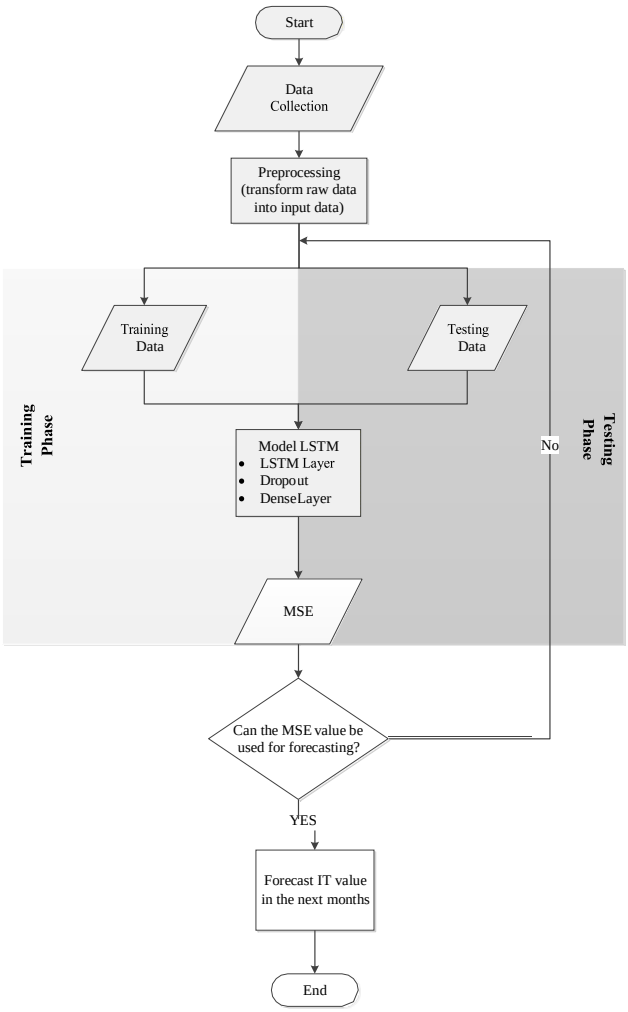

Fig. 2. System Flowchart

The following is an explanation of the flowchart in Figure 2:

1. Analysis of the time series forecasting model with the LSTM algorithms on the IT value of food crop farmers starting with data collection. The data collected is in the form of IT values received by food crop farmers every month from April 2020 to April 2021 which are obtained from the BPS Indonesia website.
2. Data preprocessing is done by transforming the raw data obtained from the BPS Indonesia website into structured data that can be processed as input data by the LSTM model. The data that is ready to use becomes input data, which will then be divided into training data and test data.
3. Training data is data that will be used as input data in the training phase. In the training phase, the training data will be trained using the LSTM model which consists of LSTM, dropout, and dense layer separately.
4. After LSTM models are trained, the model will be tested with test data in the testing phase separately. The result of this testing phase is the MSE value which is the benchmark for measuring the truth of the forecasting model results.

A. *Method of Collecting data*

The data used in this study were collected by downloading data in the form of the average price index value received by food crop farmers (IT) per month from the BPS Indonesia website. Because the research was conducted to analyze and predict the value of IT during the COVID-19 pandemic, the data collected came from data from 2020 to 2021 (to be precise from April 2020 to April 2021). The data, which consists of 13 months, will then be used as the time series dataset for the LSTM model after passing the preprocessing step. The preprocessing step taken on this dataset is to transform the raw data into data with a structure that can be accepted by LSTM model. This dataset will then be divided into training data and test data which will be used in the training phase and the testing phase of the LSTM models.

B. *LSTM Model*

The LSTM model is a sequential model that can run the process sequentially on the layers that have been arranged. The LSTM model used in this study is composed of 3 processes, namely LSTM, dropout, and dense layer. This dense layer functions to connect each neuron on a layer with neurons on another layer. Without a dense layer, the LSTM model will not be able to predict IT data for food crop farmers. In this study, the dense layer is useful for determining the loss value and MSE. Loss is a parameter containing a value that indicates poor forecasting results. Therefore, the lower the loss value, the better the prediction results. Meanwhile, the MSE value is determined based on the comparison between the prediction results in the form of the IT value of the food crop farmers with the original data. As with loss, the lower MSE value indicates a good prediction result.

V. DISCUSSION

A. *LSTM Parameter Determination*

The parameters used by the LSTM model in this study are:

- n_input: number of months to be forecast (forecast)
- LSTM: the value in the LSTM layer
- dropout: the value on the dropout
- epoch : number of model training

To train the LSTM model we need to determine the 4 most optimal parameters. We will test 4 different parameters by looking at the MSE value. To prevent underfit or conditions where the trained model is unable to forecast accurately, the selected parameter is a parameter with an MSE value that is not higher than 0.1 (MSE> 0.1). The following is a table of the results of the MSE test results to determine the parameters.

TABLE I. LSTM Model Parameter Testing

n_input	LSTM	dropout	epoch	MSE	loss
10	300	0.2	100	0.0893	0.0446
10	**200**	**0.2**	**100**	**0.1173**	**0.0587**
10	100	0.2	100	0.0793	0.0554

Based on the test data obtained from table I, it can be

Irvan Setiawan, Tristyanti Yusnitasari, Teddy Oswari, and Reni Diah Kusumawati, "Food Crop Farmer Index Price Prediction using LSTM During Covid-19 Pandemic," *International Research Journal of Advanced Engineering and Science*, Volume 6, Issue 3, pp. 122-125, 2021.

concluded that the LSTM model has the most optimal MSE value if it is trained with the parameters n_input, LSTM, dropout, and epoch of 10, 200, 0.2, 100.

B. LSTM Model Output

The LSTM model was trained for 100 times (epoch) with LSTM parameters 00, dropout 0.2. To train the LSTM model, 13 food crop farmer IT data were used during the COVID19 pandemic. After being trained, the LSTM model will be used to predict the IT value of food crop farmers in 10 months from July 2020. To see the prediction results (forecasting) in more detail, please see Figure 3.

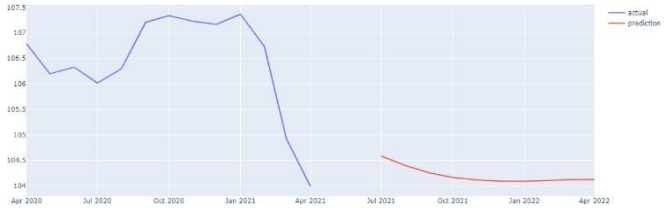

Fig. 3. LSTM Model Prediction Result Graph

In Figure 3, the blue line is the line flow that represents the real food crop farmer IT data flow, while the red line represents the predictive data in the form of the IT value of food crop farmers generated by the LSTM model starting from July 2021. The MSE value is obtained after training the LSTM model is 0.1173.

VI. CONCLUSION

The predictions (forecasting) obtained from the LSTM model have been able to carry out their functions properly. Based on the test results, this time series forecasting model can predict the IT data received by food crop farmers. The results produced by the LSTM model are MSE values and predictive data. The details of the results of this study can be seen in the following points:

- The dataset used as training data and test data in the LSTM model consists of IT data received by food crop farmers from July 2020 to July 2021 (13 months).
- LSTM model uses a dataset consisting of 13 months of training data.
- LSTM model has an MSE value of 0.1173 after being trained with the parameters n_input, LSTM, dropout, and epoch with values of 10, 200, 0.2, 100 respectively.
- The LSTM model is predicting IT values for the next 10 months starting from July 2021.
- The results of the forecasting of IT values received by food crop farmers during the COVID 19 pandemic using the LSTM model show that the LSTM model able to be used for predicting future IT values.

For future work, other researchers can improve the accuracy of the prediction results. The author recommends further research to use more data, and change the model architecture.

REFERENCES

[1] Badan Pusat Statistik, https://sirusa.bps.go.id/sirusa/index.php/indikator/65 and, https://daps.bps.go.id/file_artikel/77/arima.pdf , Accessed on: 2:46, 24/4/2021.
[2] Bustanul Arifin, "Analisis Ekonomi Pertanian Indonesia," 2004.
[3] Deepak Janardhanan, Enda Barrett, "CPU Workload forecasting of Machines in Data Centers using LSTM Recurrent Neural Networks and ARIMA Models," 2017.
[4] Kementerian Pertanian, https://www.pertanian.go.id/home/?show=news&act=vie w&id=2564. Accessed on : 17:42, 19/4/2021.
[5] Kiran Kumar Paidipati, Arjun Banik, "Forecasting of Rice Cultivation in India – A Comparative Analysis with ARIMA and LSTM-NN Models," 2019.
[6] KOMPAS, https://nasional.kompas.com/read/2020/11/18/12080261/sektor-pertanian-tumbuh-di-masa-pandemi-dinilai- sumbang-pertumbuhan-ekonomi?page=all , Accessed on: 18:48, 19/4/2021.
[7] Xianghua Wu, Jieqin Zhou, Huaying Yu, Duanyang Liu, Kang Xie, Yiqi Chen, Jingbiao Hu, Haiyan Sun and Fengjuan Xing, "The Development of a Hybrid Wavelet- ARIMA-LSTM Model for Precipitation Amounts and Drought Analysis," 2021.
[8] Yiqing Hua, "Bitcoin Price Prediction using ARIMA and LSTM," 2020.
[9] Zahara, S., Sugianto, & Ilmiddafiq, M. B., "Prediksi Indeks Harga Konsumen Menggunakan Metode Long Short Term Memory (LSTM) Berbasis Cloud Computing," 2019.