IJSA (2022), Vol. 06, No. 1, 143-160
DOI: https://doi.org/10.29244/ijsa.v6i1p143-160
Indonesian Journal of Statistics and Its Applications
Available at https://ijsa.stats.id
e-ISSN: 2599-0802

Price Prediction Model for Red and Curly Red Chilies using Long Short Term Memory Method*

Rizky Abdullah Falah[1] and Meuthia Rachmaniah[2‡]

[1,2]Department of Computer Science, IPB University, Indonesia
[‡]corresponding author: meuthiara@apps.ipb.ac.id

Copyright © 2022 Rizky Abdullah Falah and Meuthia Rachmaniah. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The price data of the Strategic Food Price Information Center from May 2018 to May 2021 in 34 provinces show a fluctuated trend. Our study aimed to build predictive modeling of red chili and curly chili prices in West Java province using the Long Short Term Memory method. The red chili and curly chili prices prediction model in our study was successfully constructed and is considered very representative of predicting prices in traditional and modern markets in West Java Province. The best parameter model for red chili in the traditional market is a neuron value of 64 and a learning rate of 0.0005, and in the modern market, there are neuron values of 48 and a learning rate of 0,005. For curly chili, the best parameter model in traditional markets is a neuron value of 48 and a learning rate of 0.00075, and in the modern market, there are neuron values of 32 and a learning rate of 0,001. All models use the number of the epoch 100. The best prediction model for the price of red chili and curly red chili in traditional markets obtained the smallest root mean square error values on the test data of 2.57% and 2.07%, respectively. Meanwhile, the best price prediction model in the modern market obtained the smallest root mean square error values on the test data of 2.11% and 2.17%, respectively. Based on the root mean square error value obtained, the model is better than the other research method and shows that the variation in the value produced by a model is close to the variation in the actual value.

Keywords: chili, demand and supply, long short term memory, price prediction, root mean square error.

* Received: Mar 2022; Reviewed: May 2022; Published: May 2022

1. Introduction

Data from the Central Statistics Agency (abbreviated BPS) shows high levels of red chili production and consumption per capita in several provinces in Indonesia. The highest and center of production and consumption levels of red chili occur in West Java, Central Java, East Java, West Sumatra, North Sumatra, and Aceh provinces. Based on (BPS, 2020), West Java ranks the highest chili production level, with 263,949 tons in 2019. The level of red chili consumption in West Java in 2019 was 52,945 tons. According to data from (BI, 2020), red chili in December 2020 was one of the contributors to inflation in the volatile food group, which was 31.50% month-to-month (mtm). Annually, red chilies experienced inflation of 45.68% year-on-year (yoy), so the average selling price of red chilies during 2020 reached Rp. 38,049/kg. (Helbawanti et al., 2021) stated that food prices for chicken meat, shallots, garlic, red chilies, and cayenne pepper fluctuated. These foodstuffs experienced changes from 2017 to 2021 and tended not to return to their lowest prices during 2017-2021.

At a certain point, red chili prices fluctuated somewhat elevated in many provinces of Indonesia. Based on researcher monitoring of the Strategic Food Price Information Center (abbreviated PIHPS) from May 2018 to May 2021, the price of red chili fluctuated in 34 provinces, such as in West Java Province. Red chili prices fluctuate every week. It can be seen in Figure 1 that the price of red chili has increased and decreased every week, and high supply and demand affected price fluctuation. Bad weather and the shift of chili farmers to other crops also affect chili price fluctuation (Jawal Anwarudin et al., 2015) . In chili, although there is an annual production surplus, daily price fluctuations increase the risk that affects the motivation of farmers to cultivate chili (Rachmaniah et al., 2021).

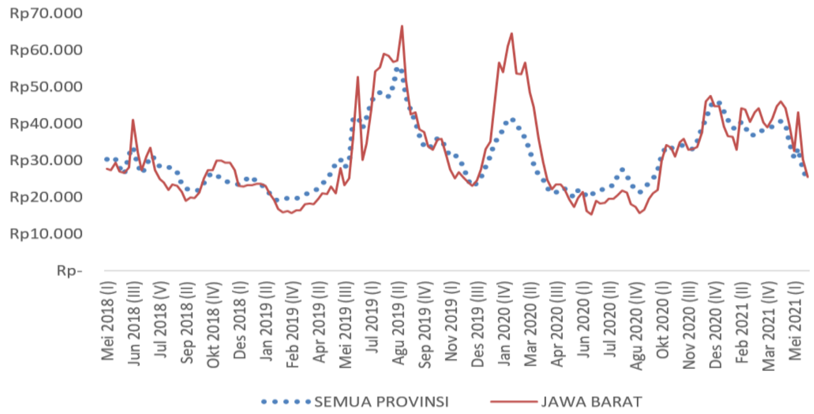


Figure 1: Red chili prices fluctuation in West Java Province from May 2018 to May 2021 (processed from the National PIHPS, 2021)

Irregular or fluctuating red chili prices can trigger economic inflation, especially during the harvest and non-harvest periods (Sepri et al., 2020). A price prediction analyzes the market's anticipation of the red chili and curly chili prices. This analysis provides information about the price increase and decrease in red chili and curly chili in West Java. This information will be helpful for the government, producers, business actors, and consumers to minimize the impact of price fluctuations during daily, weekly, or monthly periods. The price increases and decreases at a particular time prediction by building a model from time-series historical data of the red chili and curly chili prices. The price prediction model used a machine learning algorithm approach, namely the long short-term memory (LSTM) method, and statistical analysis for model development and evaluation.

Research by (Wulandari et al., 2017) discussed regression spline analysis to estimate Jakarta chili prices. In this study, the best mean absolute percentage error (MAPE) prediction was 9.57%, and the coefficient of determination was 86.41%. The regression spline utilized a three-knot approach in the third order. Another study by (Bagus Adiatmaja et al., 2019) discussed the price forecasting for red chilies in the East Java region using the extreme learning machine (ELM) method. In their study, the best mean absolute percentage error (MAPE) prediction value was 3.23%. This study utilized two features; the size of hidden neurons is three, with a range of weight values [-1.8, 1.8].

Our study proposed the LSTM method, devising a neural network to model chili price time-series data. We chose the LSTM method because, according to (Zahara et al., 2019), it is superior and more reliable in making predictions for long time-series data than other algorithms. In addition, the LSTM method is a comparative machine learning model that is more accurate in predicting food prices than the multilayer perceptron (MLP) and ARIMA methods (Sen et al., 2020). Hence, our study aims to build a price prediction model for red chili and curly red chili based on data on prices in traditional and modern West Java province markets. This study aims to anticipate the impact of fluctuations in commodity prices for red chili and curly red chili in the West Java Province. In particular, this study aims to model red chili and curly red chili price predictions using LSTM. Controlling fluxes in the stability of chili prices will result in better inflation control so that stakeholders do not experience difficulties in making decisions.

2. Research Methods

2.1 Red Chili Price Fluctuation

Price fluctuation is a phenomenon of price volatility that repeatedly occurs throughout the year. Red chili often experiences high price oscillations. This problem is due to the seasonal nature of red chili production and is very susceptible to being influenced by climatic and weather conditions, so red chili commodities are perishable (Erviana et al., 2020). The highly erratic price movements, the amount of availability, and the quantity needed in the country for red chili commodities also contribute to the inflation rate in Indonesia. (Irawan, 2007) suggested that this tendency surfaces because price fluctuations open up business actors' opportunities to manipulate costs at the farm level in response to price movements at the consumer side. The supply-side leverage the price of an agricultural commodity and the demand side, so the rising and falling prices are caused by factors that affect the supply and demand sides (Sugiarto & Nangameka, 2013). Therefore, these problems affect the process of price transmission from producers to consumers.

Chili price fluxes often surface due to inequality between demand and supply. The central producers' limited chili supply causes the demand not to be met by domestic production. The extended distribution chain contributed to an increase in chili prices. On the other hand, oversupply of chili can cause chili prices to drop drastically. The drop in prices impacts decreasing farmers' interest in cultivating chilies. According to (Fitzsimmons & Fitzsimmons, 2011), supply chain management (SCM) is a service system for product delivery from suppliers with a flow process structure to the final consumer. In terms of processes, information technology coordinates all elements involved in the supply chain.

The chili marketplace is a chili commodity information system that creates a central market to bring together sellers, buyers, and other stakeholders to transact electronically (Rachmaniah et al., 2020). The chili marketplace application serves as a platform for web-based transactions for farmers' and consumers' chili transactions. By using digital platforms, farmers can reach more promising consumers, reducing supply chain length, generating broader sales, and increasing revenue. Tanihub and Sayurbox are samples of implementing agriculture-themed marketplaces in Indonesia. Both take advantage of web-based and mobile-based applications for buying and selling various kinds of daily fresh vegetable needs. This application is included in the business to customer (B2C) category because this application sells its products to customers.

According to (Achjari, 2000), e-commerce appealing the business world for two benefits, i.e., efficiency and effectiveness. Companies can gain efficiencies in marketing, labor, and overhead costs. Also, companies can open a virtual shop 24/7 continuously, and consumers can make transactions regardless of time. Through the chili marketplace, chili farmers get the certainty of a stable and competitive market price at a reasonable price. So that the welfare of farmers is maintained and the government is more accessible in making decisions on determining the standard price of national chili.

2.2 Long Short Term Memory

In 1997 (Hochreiter & Schmidhuber, 1997) introduced the LSTM algorithm to enhance an artificial neural network that can examine long-term data dependencies. However, there are differences between LSTM and recurrent neural networks (RNN). LSTM learns information to be discarded and stored within the memory cells. Each LSTM neuron carries a memory cell and a gate unit; memory cells store information, whereas gate units can manage stored information.

The LSTM consists of three vector gate units: forget, input, and output. The vector gates recognize the information stored in memory, how long it is stored, and when to recognize data patterns. The input gate vector monitors how many input vectors are allowed to affect the memory vector. The output gate vector monitors how much memory vector is stored in the hidden state (Kalchbrenner et al., 2016).

The LSTM method has several parameters determined during the training process, including learning rate and epoch. *Learning rate* is a parameter that monitors the magnitude of the weight transformation updated during the training process in response to the error value generated during the training process. While the epoch indicates the number of times, the algorithm runs for the overall data. The form of the LSTM equation (Olah, 2015) comprises equations (1), (2), (3), (4), (5), and (6). Figure 2 below shows an example of a typical LSTM architecture.

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

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

$$\tilde{c}_t = tanh(W_c \,.\, [h_{t-1}, x_t] + b_c) \tag{3}$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (4)$$
$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o) \quad (5)$$
$$h_t = o_t * tanh(c_t) \quad (6)$$

where x_t = time *t* input, h_{t-1}= time *t-1* output, c_{t-1}= time *t-1* cell state, f_t = forget gate, i_t = input gate, o_t= output gate, h_t = hidden state or the combined output vector, c_t= cell state, $\tilde{c}_t$ = new value candidate vector, W = weight matrix, *b* = intercept vector parameters, σ = the value of the sigmoid activation function, dan $tanh$ = the value of the activation function *tanh.*

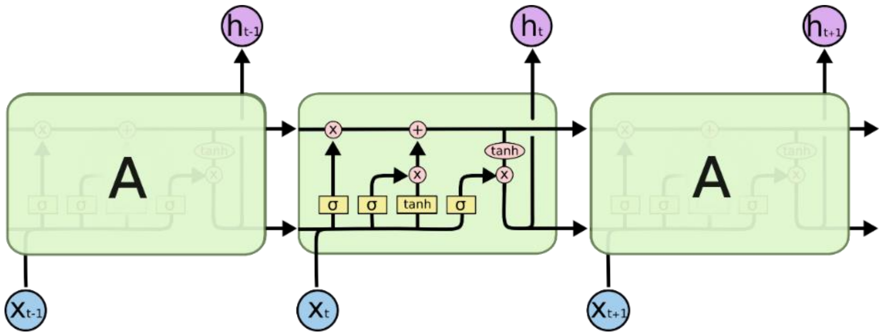


Figure 2: An example of an LSTM architecture (Olah 2015)

2.3 Source of Data

The secondary data were harvested from the Strategic Food Price Information Center (abbreviated PIHPS) website accessed at https://hargapangan.id. PIHPS is a food price information system established by Bank Indonesia to disseminate strategic food prices to the Indonesian community. This web contains various agricultural commodities is presented in daily, weekly, monthly, and graphical reports. The data taken through the https://hargapangan.id web page is then entered into a Microsoft Excel file and saved in CSV format for data modeling processing with the Python 3.7.8 programming language.

The data obtained is divided into four daily time series data in commodity prices for red chili and curly red chili in traditional and modern markets in West Java province. The data from 1 January 2018 to 31 October 2021 obtained was 5,600 observational data. Each dataset of commodity prices for large red chilies and curly red chilies in traditional and modern markets totals 1,400 observational data. This method of collecting daily price data comes from the recapitulation of the results of sampling several business actors based on final prices in several traditional markets and modern markets in West Java province. The dataset's attributes for model development are depicted in Table 1.

2.4 Methods of Analysis

The stages of this research consist of the following stages: data collection, preprocessing and data exploration, data sharing, prediction model development using LSTM, and model evaluation (Figure 3). Remind that the data separation stage make a difference between training and testing data for modeling purposes.

Data Collection. The initial step of the research is to collect data from the PIHPS website into a Microsoft Excel file and save it in CSV format. Further, Python version 3.7.8 is used as a programming language to create the LSTM model.

Table 1: List of commodity price dataset attributes for red chili (CMB) and curly red chili (CMK) in West Java province for traditional (Pt) and modern (Pm) markets

Commodity	Attribute Name	Type of Data	Description
CMB	DateCMB	Datetime, Numeric	The date of the price growth
CMK	DateCMK	Datetime, Numeric	The date of the price growth
CMB	PriceCMBPt	Time series, Numeric	Traditional market red chili price
CMK	PriceCMKPt	Time series, Numeric	Traditional market curly red chili price
CMB	PriceCMBPm	Time series, Numeric	Modern market red chili price
CMK	PriceCMKPm	Time series, Numeric	Modern market curly red chili price

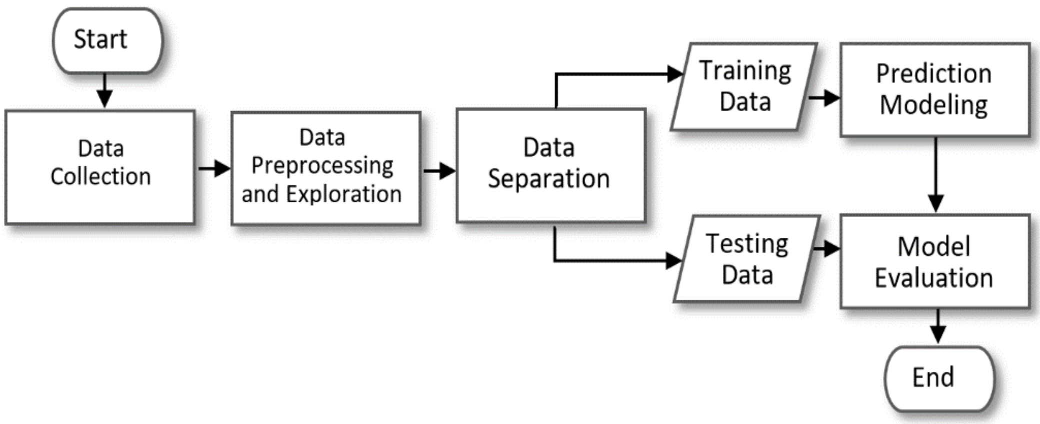


Figure 3: Research steps

Data Preprocessing and Exploration. At this stage, inspecting we examined missing data values, attribute selection, data normalization, data exploration to find out the description of data characteristics, and identifying data patterns by looking at data visualization plots. The LSTM is sensitive to data scale; hence we used normalization steps to avoid data with large values dominating data with small data ranges. In this study, min-max normalization used a 0-1 range of values. Equation (7) depicted the min-max normalization formula (Yu et al., 2009).

$$y_{i,j} = \frac{x_{i,j} - min(x_j)}{max(x_j) - min(x_j)} \quad (7)$$

where $y_{i,j}$ = normalization, $x_{i,j}$ = data to be normalized, $min(x_j)$ = the minimum data value, and $max(x_j)$ = the maximum data value.

Data Separation. The researcher divided the data that passed the preprocessing and data exploration stages into training and test data. The data sharing undergoes experimenting with a combination of percentage values for training and test data. In this study, the selection of the percentage value of the training data is 80%, and the test data is 20%. (Jauhari et al., 2016) stated that the test results show that the more training data, the more accurate the accuracy. Here, the diversification of data variations so that it can recognize various types of data at the time of testing. The dataset of the training data was from January 1, 2018, to January 23, 2021. Meanwhile, the test data dataset used is from January 24, 2021, to October 31, 2021. The LSTM algorithm applied to both the dataset and the training data model was further evaluated against the test data model.

Prediction Modeling. The modeling applied the LSTM algorithm, and the result of this learning process is a model that can predict the red chili and curly red chili prices in traditional and modern markets. LSTM will use equations (1), (2), (3), (4), (5), and (6) to calculate the vector for each gate unit. The value of x_t is input data on the price of red chili and curly red chili on the previous day, whose value converts into a vector.

The initial stage of LSTM modeling is parameter initialization. Here, the parameters required are the number of neuron units in the input layer, hidden layer, an output layer, batch size, number of epochs, and the optimizer in the form of learning rate. The optimizer is one of the parameters required to build the model using the Keras package, and the optimizer has an essential role in improving the model accuracy. The determination of the initialization in our study is random.

LSTM uses two activation functions, i.e., the $tanh$ activation function and the sigmoid activation function. The activation function is for generating the output of each layer. The activation function $tanh$ will convert a value of x into a range of values from -1 to 1. Meanwhile, the sigmoid activation function will convert a value of x into a value with a range of 0 to 1. Equations (8) and (9) depicted the formula for the activation function of sigmoid and $tanh$ (Karlic & Olgac, 2016).

$$\sigma = \frac{1}{1+ e^{-x}} \tag{8}$$

$$tanh(x) = 2\sigma(2x) - 1 \tag{9}$$

where x = input data, σ = the value of the sigmoid activation function, $tanh$ = the value of the activation function $tanh$.

Model Evaluation. One of the essential steps in building a predictive model is the model evaluation process. The evaluation of the model is to check the level of goodness of the model formed. In this study, the root mean squared error (RMSE) calculation becomes the mechanism for evaluating the model. According to (Wang et al., 2011), the RMSE calculation intends to catch how much the generated error from a prediction model. The closer the error value is to 0, the better the model made will be. Here we used equation (10) to calculate the RMSE value.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2}{n}} \tag{10}$$

where n = the amount of data, X = actual value, $\hat{X}$ = predictive value.

3. Result and Discussion

3.1 Data Preprocessing

Data preprocessing began with checking the missing data values in the dataset using the *isnull.sum* function in Python programming. The *isnull.sum* function will check the data values in all attribute columns in the dataset starting from the date attribute, commodity, province, traditional market price (IDR/kg), and modern market price (IDR/kg). The dataset of red chili and curly red chili prices in West Java Province used in this study contains missing values for two types of attributes. These attributes include traditional market prices and modern market price attributes. We interpreted the missing value *NaN* (not a number) in the dataset as not documented data of the object observed. Table 2 shows the total number of missing values in two types of attributes in the commodity price dataset of red chili and curly red chili out of 1.400 data each.

Table 2: Attributes containing missing values in the dataset.

Dataset	Attribute	Interval	Total Missing Value	Missing Value Percentage
Red chili price	Traditional market price	January 2018 - October 2021	464	33.1
	Modern market price	January 2018 - October 2021	456	32.5
Curly red chili price	Traditional market price	January 2018 - October 2021	465	33,2
	Modern market price	January 2018 - October 2021	458	32.7
Total			1,843	32.9

Relying upon the missing values in this research is done by replacing the missing data values with the average value calculation. The average monthly price value accompanies the number of missing data values from January 2018 to October 2021. The average monthly price used equation (11). Table 3 shows a sample of the results of handling missing values in the dataset in this study.

$$\bar{X} = \frac{\sum x_i}{n} \tag{11}$$

where $\bar{X}$ = average monthly price (IDR/kg), . $\sum x_i$.= sum of price data per month, n = number of price data per month.

3.2 Attribute Selection and Data Normalization

After examining the missing data, we conducted the attribute selection process and data normalization. At the attribute selection stage, we selected the required attributes to build a predictive model of all the attributes in the dataset. We omitted commodity and province attributes of the five attributes contained in the red chili and curly red chili commodity price dataset.

The data values for the price attributes of traditional and modern markets for red chili and curly red chili commodity datasets normalized into the data scale. The data rescaling overcomes the constraint of the sigmoid activation function in the LSTM cell, which is sensitive to significant numbers changes. It also avoids problems occurring in the learning process (using data training) required for forming predictive models (Zaytar & El Amrani, 2016). The normalization process employed the scikit-learn library and MinMaxScaler normalization. Min-max normalization changes the size of the data from the original range so that all values are in the value range 0 to 1. Normalizing the data will help speed up the learning process in machine learning (Ambarwari et al., 2020). The use of the proper data scaling method can optimize the performance of the LSTM algorithm. Table 4 depicts the normalization results.

Table 3: Sample of data preprocessing results for the commodity price dataset of large red chili (CMB) and curly red chili (CMK) in West Java

Attribute	Year	Month	PriceCMB results (IDR/kg)	PriceCMK results (IDR/kg)
Traditional market price	2018	January	40.335	37.750
	2018	February	44.142	40.460
	2018	March	44.473	43.588
	2018	April	47.723	38.759
	2018	May	37.902	32.375
	2018	June	44.652	34.735
Modern market price	2019	January	59.225	56.190
	2019	February	53.713	49.318
	2019	Mach	51.170	47.730
	2019	April	50.842	47.722
	2019	May	61.152	53.161
	2019	June	74.097	62.055

Table 4: Example of data before and after normalization on the attributes of traditional and modern markets prices

Dataset	Date	Traditional Market		Modern Market	
		Before	After	Before	After
Red chili Price	2018-01-01	40.335	0,29942997	52.911	0,19870915
	2018-01-02	41.400	0,31677524	53.750	0,21241830
	2018-01-03	41.850	0,32410423	53.750	0,21241830
	2018-01-04	40.200	0,29723127	53.750	0,21241830
	2018-01-05	39.200	0,28094463	52.911	0,19870915
Curly red chili Price	2018-01-01	37.750	0,29621036	59.600	0,37424893
	2018-01-02	36.450	0,27610209	59.600	0,37424893
	2018-01-03	36.250	0,27300851	58.650	0,35793991
	2018-01-04	35.900	0,26759474	58.650	0,35793991
	2018-01-05	35.650	0,26372776	58.650	0,35793991

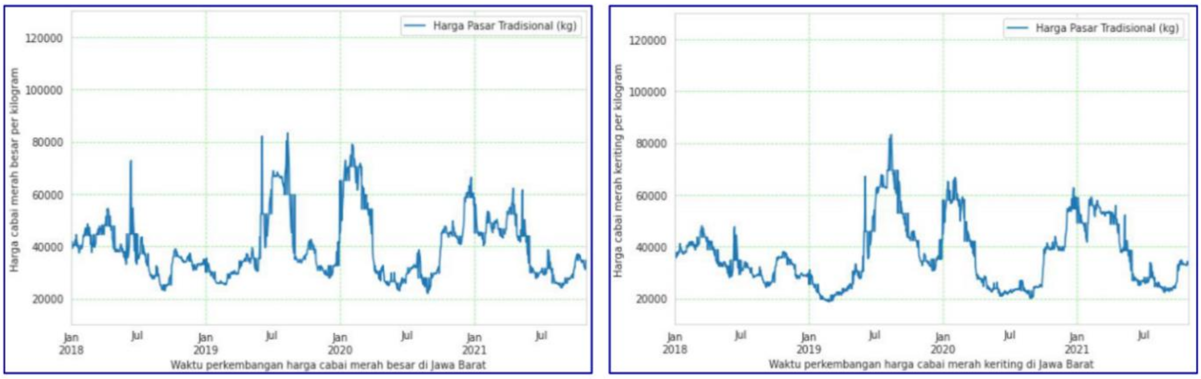
3.3 Data Exploration

Data exploration aimed to detect the description and identify the characteristics of the data presented in a daily period of 5.600 observational data—the description of the data characteristics applied via descriptive statistics of each attribute used in the dataset. Table 5 shows the descriptive statistical values of each attribute, starting from the average value, standard deviation, minimum and maximum values. In addition, we conducted data exploration using graphics of the time series data of red chili and curly red chili prices in traditional and modern markets to see data fluctuations.

Standard deviation is an index of variability of a distribution whose value will increase if the distribution is more diverse and varied. If the data spread is extensive to the average value, the standard deviation value will be significant, but if the data spread is minimal to the average value, the standard deviation value will be small (Yusniyanti & Kurniati, 2017). n addition, to see fluctuations in the data, we used daily time-series data graphics on red chili and curly red chili prices in traditional and modern markets in West Java province. The x-axis shows the data period, and the y-axis shows the price of red chili in kilograms. Figure 4(a) portrays fluctuations in red chili prices in the West Java traditional market. Data from January 2018 to October 2021 show a very fluctuating price pattern, with the standard deviation value for the attribute price of red chili in traditional markets being 12.3170. Figure 4(b) shows daily fluctuations in curly red chili prices in the West Java traditional market from January 2018 to October 2021 also has a wildly fluctuating pattern, with the standard deviation value for curly red chili price in the traditional market at 12.7354.

Table 5: Descriptive statistics of red chili and curly red chili price data

Dataset	Atribut	Minimum price (IDR/kg)	Maximum price (IDR/kg)	Average price (IDR/kg)	Standard Deviation
Red chili	Traditional market price	21.950	83.350	39.564	12,3170
	Modern market price	40.750	101.950	69.110	15,1091
Curly red chili	Traditional market price	18.600	83.250	36.786	12,7354
	Modern market price	37.800	96.050	63.246	14,9027

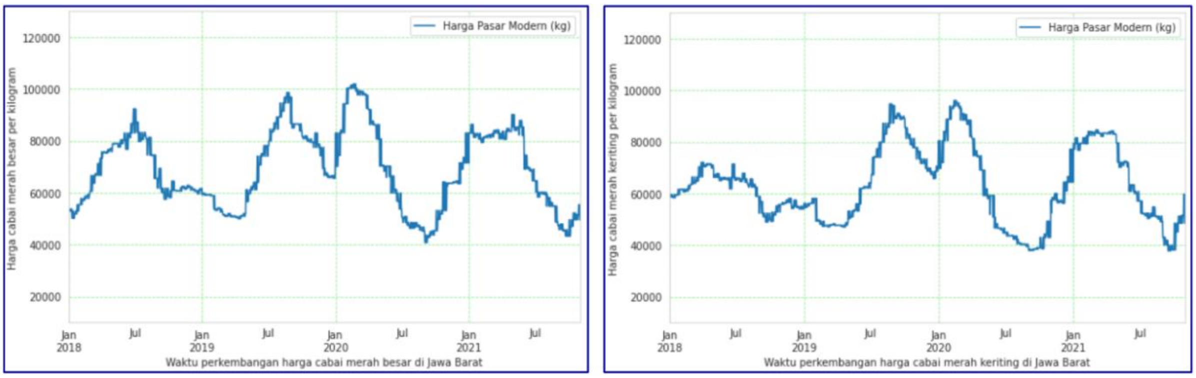


(a) Red Chili

(b) Curly Red Chili

Figure 4: The results of daily price exploration in traditional markets of (a) red chili and (b) curly red chili in West Java province for the period of January 2018 - October 2021

In Figure 5(a), daily red chili price fluctuations in the West Java modern market from January 2018 to October 2021 show volatile conditions. Following the data characteristics' description, the standard deviation value for the red chili price attribute in the modern market is 15.1091. Figure 5(b) shows a time series plot of the daily price growth of curly red chili in the West Java modern market from January 2018 to October 2021. The plot shows a somewhat volatile price pattern, as is the trend of red chili prices in the modern market. Based on the description of the data characteristics, the standard deviation value for the price attribute of curly red chili in the West Java modern market is 14.9027.



(a) Red Chili (b) CurlyRed Chili

Figure 5: The results of daily price exploration in modern markets of (a) red chili and (b) curly red chili in West Java province for the period of January 2018 - October 2021

3.4 Prediction Modeling with LSTM

We divided the normalized data into 80% of training and 20% of test data in this stage. Our model uses the LSTM method utilizing the Keras version 2.3 library package in Python programming. The initial stage of data training is the initialization of parameters used in the LSTM architecture, such as the size of neuron units within the input layer, hidden layer, output layer, learning rate, and epochs. Table 6 shows the LSTM network architecture for the prediction model for red chili and curly red chili prices in this study. This study uses one LSTM layer with 32, 48, and 64 neurons and one Dense layer. The activation function used in the LSTM layer is relu. The optimizer used in the LSTM architecture is adaptive moment estimation (Adam), and the batch size used is 16. In this study, we use different layers and number of neuron units aiming to see the difference in the predicted value generated. Our approach uses (Zhao et al., 2017) study to determine the most optimal architecture for the study. (Zhao et al., 2017) stated that the arrangement of the number of networks and layers determines optimal architecture for predictive analysis.

Table 6: LSTM architecture of red chili and curly red chili price model

Characteristic	Specification

Architecture	1 LSTM layer with 32, 48, 64 unit of neuron; 1 Dense layer
Activation function	Relu
Optimizer	Adam
Batch size	16

The learning rate value of the big and curly red chili price prediction model in traditional markets, the values tested are 0.0001, 0.00025, 0.0005, and 0.00075. Meanwhile, for the learning rate value of the big and curly red chili price prediction model in the modern market, the values tested are 0.001, 0.0025, 0.005, and 0.0075. The learning rate value chose based on the value that the parameters had tested by (Shiddiq et al., 2020) related to the temporal prediction model of CO and CO2 pollutants employing LSTM. We expected the values tested to affect the effectiveness and convergence of this study's LSTM algorithm training process.

3.5 LSTM Prediction Results

The data that has gone through the normalization stage is further separated into 80% training data and 20% test data. The dataset used to build the prediction model is red chili and curly red chili price data in the province of West Java. The model used the LSTM method in the Keras version 2.3 library package (Figure 6). Whereas Figure 7 shows the source code and libraries used in building prediction models using LSTM. We created the price prediction model for big red chili and curly red chili separately.

```
# Libraries and several packages used to build a big red chili price prediction model
import math
import pandas_datareader as web
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib.dates as mdates
sns.set_style('whitegrid')
%matplotlib inline
from datetime import datetime
from sklearn.preprocessing import MinMaxScaler
from keras.models import Sequential
from keras.layers import Dense, LSTM, Activation, Dropout
from sklearn.metrics import mean_squared_error
from sklearn.utils import shuffle
from tensorflow.keras import optimizers, Model
```

Figure 6: Segment of Phyton program for libraries and packages utilizing Keras version 2.3

The period of predictions from January 2018 to October 2021 pursues the training and test data percentage. Figures 8-11 show the comparison between the training data prediction value (red line), the red chili price data actual value (blue line), and the predicted value of the test data (green line) using the LSTM. In the graphic visualization, the prediction results using the said LSTM architecture can generate almost the same trendline pattern as the actual price data for red chili and curly red chili in 2021. It reveals that the prediction results almost aligned with the pattern constructed from the training dataset and the patterns of events over time.

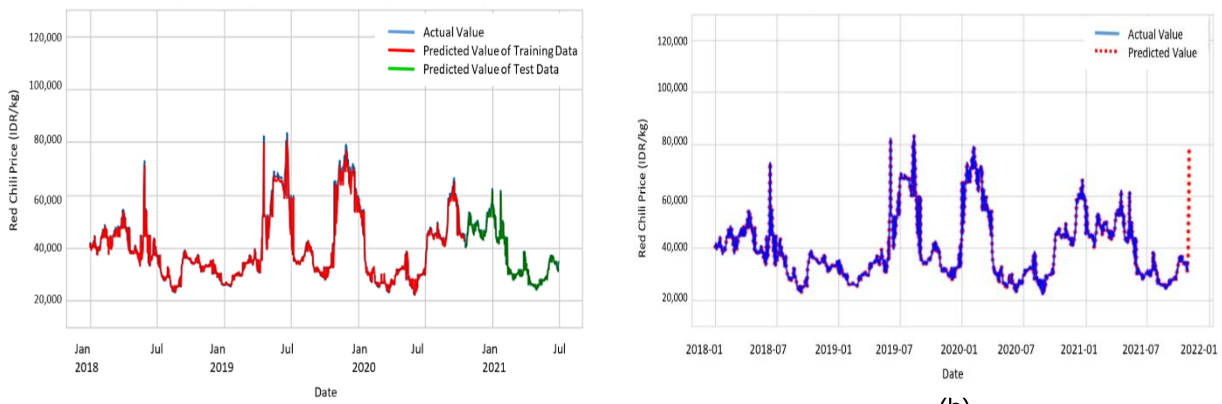
```
# Normalization of data using min-max normalization
dataset_1 = df_CabaiMerahBesar_PasTrad.values
scaler_1 = MinMaxScaler(feature_range = (0, 1))
dataset_1 = scaler_1.fit_transform(df_CabaiMerahBesar_PasTrad)

dataset_2 = df_CabaiMerahBesar_PasMod.values
scaler_2 = MinMaxScaler(feature_range = (0, 1))
dataset_2 = scaler_1.fit_transform(df_CabaiMerahBesar_PasMod)

# Divide training data and test data with percentage values of 80% and 20%
TRAIN_SIZE = 0.80
train_size = int(len(dataset_1) * TRAIN_SIZE)
test_size = len(dataset_1) - train_size
train, test = dataset_1[0:train_size, :], dataset_1[train_size:len(dataset_1), :]
print("Jumlah dari (training set, test set): " + str((len(train), len(test))))

train_size = int(len(dataset_2) * TRAIN_SIZE)
test_size = len(dataset_2) - train_size
train, test = dataset_2[0:train_size, :], dataset_2[train_size:len(dataset_2), :]
print("Jumlah dari (training set, test set): " + str((len(train), len(test))))
```

Figure 7: Segment program for source code and libraries used in building prediction models using LSTM

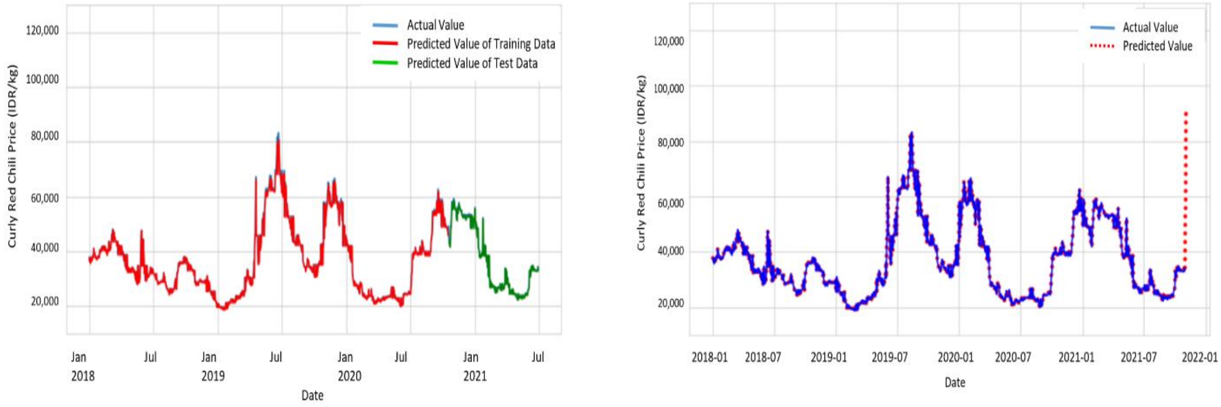


(a) (b)

Figure 8: Plot red chili prices in traditional markets (a) using a neuron value of 64, epoch 100, and a learning rate of 0.0005; (b) prediction after applying the resulted LSTM model for the period up to November 30, 2021

Figure 8(a) depicts a plot of the prediction of the red chili price in the West Java traditional market using the LSTM model employing several parameter experiments. These parameters include the number of neuron units, the number of epochs, and the learning rate value. It shows that the prediction results on the test data followed the movement of the actual data well. In addition, the prediction results on the training data also show an increase and decrease in prices in July 2019 and February 2020 (Figure 8(b)). Other prediction results on the training data can follow graphic patterns so that this model can follow trend changes in the data well. Based on the prediction results for the next 30 days from November 1, 2021, to November 30, 2021, the price of large red chilies will increase by IDR 35,000 – IDR 79,400 per kilogram.

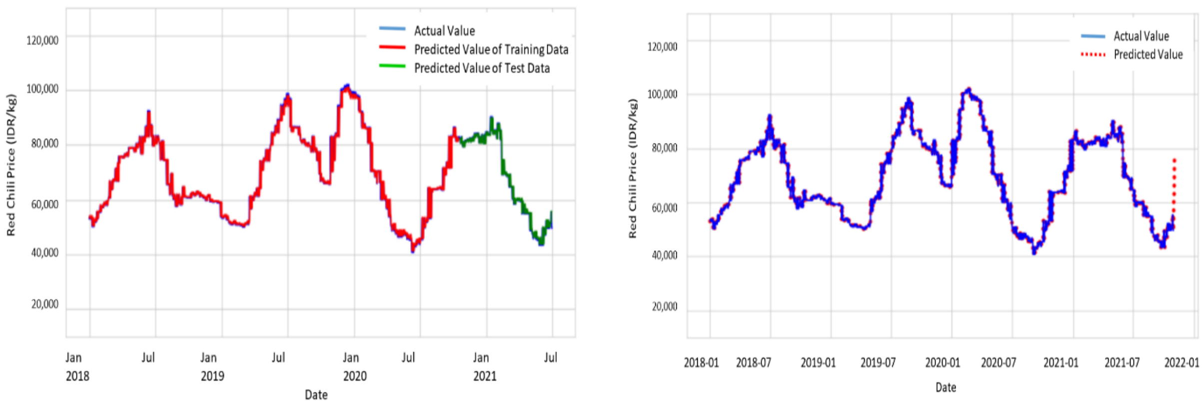
Figure 9(a) shows that the predictive value of curly red chili prices in traditional markets has decreased during the September 2019 period. Although it has not yet reached the actual data value, the prediction results for that period are good because they can predict the possibility of an increase or decrease in entirely different data. Based on the prediction results on the test data, this model can also follow the trend very well. The trend in the predicted data looks fluctuating and stable following the trend conditions in the actual data. Based on the prediction results for the next 30 days from November 1, 2021, to November 30, 2021, the price of curly red chili will increase by IDR 33,000 – IDR 91,300 per kilogram (Figure 9(b)).



(a) (b)

Figure 9: Plot of curly red chili prices in traditional markets (a) using a neuron value of 48, epoch 100, and a learning rate of 0.00075; (b) prediction after applying the resulted LSTM model for the period up to November 30, 2021

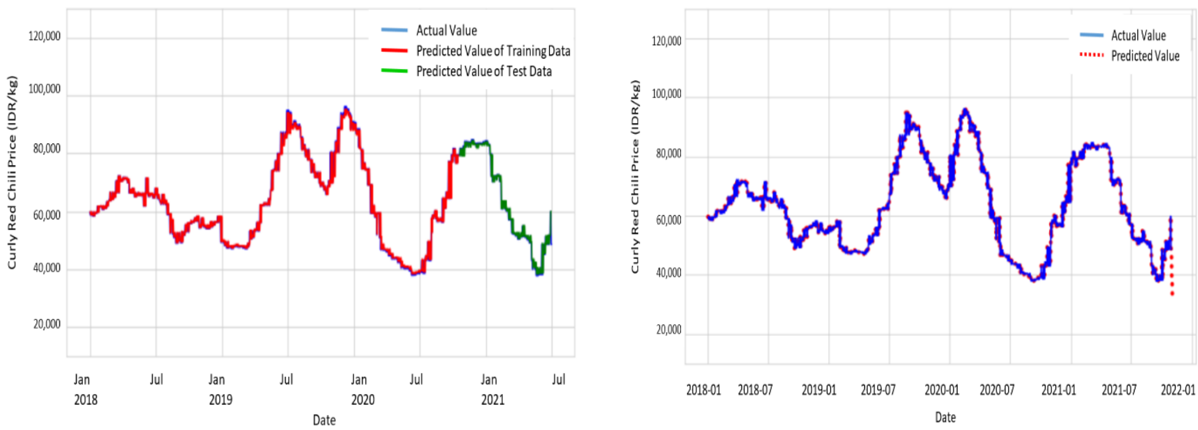
Figure 10(a) is a graphical visualization of the prediction results of red chili prices in the West Java modern market using the LSTM model with different parameter experiments. A *different value* is the learning rate value used in the price attribute in the modern market. The prediction results show a predictive value for the price of red chili in the modern market, which decreased in February 2020. Based on the prediction results obtained, the price of red chili during that period did not penetrate the IDR 100,000 per kilogram. For the prediction results of the test data, it appears that the model aligns with the actual data movement very well so that it produces a minimum error value in the learning rate value experiment of 0.005. Based on the prediction results for the next 30 days, from November 1, 2021, to November 30, 2021, the price of large red chilies will increase by IDR 49,565 – IDR 76,500 per kilogram (Figure 10(b)).



(a) (b)

Figure 10: Plot of red chili prices in modern markets (a) using a neuron value of 48, epoch 100, and a learning rate of 0,005; (b) prediction after applying the resulted LSTM model for the period up to November 30, 2021

Finally, Figure 11(a) shows that the prediction results of curly red chili prices in the modern market experienced a price decline between September 2019 and February 2020. The prediction results in price declines during this period showed that curly red chili price did not penetrate the maximum price per kilogram based on the descriptive statistical values obtained. The resulting prediction follows the pattern formed from the training data in the next period. Hence, the resulting prediction on the test data can also align with the actual data values. This LSTM model also goes through several parameters different from the parameters used for price attributes in traditional markets. The best parameter model obtained is in the experiment the learning rate value of 0.001. Based on the prediction results for the next 30 days, from November 1, 2021, to November 30, 2021, the price of curly red chili will experience a price decline of IDR 48,520 – IDR 30,690 per kilogram (Figure 11(b)).



(a) (b)

Figure 11: Plot of curly red chili prices in modern markets (a) using a neuron value of 32, epoch 100, and a learning rate 0,001; (b) prediction after applying the resulted LSTM model for the period up to November 30, 2021

3.6 Evaluation of Predicted Result Model

At this stage, the evaluation of the model aimed to see the level of goodness of the model formed. The evaluation calculated the root mean squared error (RMSE) value. The smallest RMSE values obtained in modeling predictions of red chili and curly red chili prices in traditional markets are 2.57% and 2.07%, respectively. Meanwhile, the smallest RMSE value obtained in modeling the prediction of red and curly red chili prices in the modern market is 2.11% and 2.17%, respectively. The essence of the RMSE value can define the best model advancement because it indicates the actual error (Chai & Draxler, 2014). Therefore the smaller the resulting RMSE value, it indicates that the variation in the value produced by a model is close to the variation in the actual value.

The percentage of error values generated in this study resulted in a smaller error value than in previous studies related to the prediction modeling of chili commodity prices in Indonesia. Research by (Wulandari et al., 2017) discusses regression spline analysis to estimate chili prices in Jakarta, producing the best error value of 9.57%. Another study by (Bagus Adiatmaja et al., 2019) examines the price forecasting for red chilies in the East Java region using the extreme learning machine (ELM) method. In this study, the best error value was 3.23%. Therefore, the LSTM method successfully and much more satisfactory forms a prediction model for chili prices in Indonesia, especially in the West Java region.

4. Conclusion and Recommendation

The prediction model for red chili and curly red chili using LSTM has been successfully established and considered representative to predict prices in traditional and modern markets in West Java province. The trend of increase and decrease in the red chili and curly red chili price with the prediction results formed using the LSTM method aligned with the trend of the actual value, resulting in a lower error value than the previous research method. This prediction model can provide information about the increase and decrease in the price of red chili and curly red chili in West Java Province. Therefore, the information might help the government, producers, business actors, and consumers to anticipate the impact of price fluctuations on daily, weekly, or monthly periods. Suggestions for further research are to continue researching the prediction of commodity prices for red chili and curly red chili using price data in producer centers and price data of wholesalers in West Java Province. We also recommended using data sharing through cross-validation for further research on time series data. In addition, we recommend a simultaneous equations system for red chili, curly red chili, and red cayenne pepper.

Acknowledgment. This research was supported by the Computer Science Department of Faculty of Mathematics and Natural Sciences, IPB University. The authors thanks the Department for providing research tools and their useful suggestions.

Daftar Pustaka

Achjari, D. (2000). Potensi manfaat dan problem di e-commerce. *Jurnal Ekonomi Dan Bisnis Indonesia*, *15*(3): 388–395.

Adiatmaja, P. B., Setiawan, B. D., & Wihandika, R. C. (2019). Peramalan Harga Cabai Merah Besar Wilayah Jawa Timur Menggunakan Metode Extreme Learning Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(6): 5444–5449.

Ambarwari, A., Adrian, Q. J., & Herdiyeni, Y. (2020). Analisis pengaruh data scaling terhadap performa algoritme machine learning untuk identifikasi tanaman. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *4*(1): 117–122.

Anwarudin, M. J., Sayekti, A. L., Mahendra, A. K., & Hilman, Y. (2015). Dinamika produksi dan volatilitas harga cabai: Antisipasi strategi dan kebijakan pengembangan. *Pengembangan Inovasi Pertanian*, *8*(1): 33–42.

[BPS] Badan Pusat Statistik. (2020). *Distribusi Perdagangan Komoditas Cabai Merah Indonesia 2020* (M. Karmiati, L. I. Fadilah, & R. Suerlianto, Eds.). Jakarta (ID): Badan Pusat Statistik.

[BI] Bank Indonesia. (2020). *Analisis Inflasi Desember 2020.* Jakarta (ID): Bank Indonesia.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.*, *7*(3): 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Erviana, V., Syaukat, Y., & Fariyanti, A. (2020). Analisi transmisi harga cabai merah besar di provinsi Jawa Barat. *Jurnal Ekonomi Pertanian Dan Agribis (JEPA)*, *4*(1): 77–86.

Fitzsimmons, J. A., & Fitzsimmons, M. J. (2011). *Service Management: Operations, Strategy, Information Technology* (Seventh Ed). New York, USA: McGraw-Hill.

Helbawanti, O., Saputro, W. A., & Ulfa, A. N. (2021). Pengaruh harga bahan pangan terhadap inflasi di Indonesia. *Jurnal Ilmu-Ilmu Pertanian*, *5*(2): 107–116.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(1997): 1735–1780.

Irawan, B. (2007). Fluktuasi Harga, Transmisi Harga dan Marjin Pemasaran Sayuran dan Buah. *Analisis Kebijakan Pertanian*, *5*(4): 358–373.

Jauhari, D., Himawan, A., & Dewi, C. (2016). Prediksi Distribusi Air PDAM Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation Di PDAM Kota Malang. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *3*(2): 83–87. https://doi.org/10.25126/jtiik.201632155

Kalchbrenner, N., Danihelka, I., & Graves, A. (2016). Grid long short-term memory. *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, 1–15.

Karlic, B., & Olgac, A. V. (2016). Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks. *JInternational Journal of Artificial Intelligence and Expert Systems (IJAE)*, *1*(4): 111–122. https://doi.org/10.1088/1742-6596/1237/2/022030

Olah, C. (2015). Understanding LSTM Networks. Retrieved from Web Page website: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Rachmaniah, M., Krismanti, M. M., & Darissalam, M. I. (2020). Tokocabai marketplace application based on web using extreme programming method. *2020 International Conference on Computer Science and Its Application in Agriculture, ICOSICA 2020*.

Rachmaniah, M., Suroso, A. I., Syukur, M., & Hermadi, I. (2021). Strategic Food Risks – Chili'S Agrosystem Perspective. *Jurnal Manajemen Dan Agribisnis*, *18*(1): 19–31. https://doi.org/10.17358/jma.18.1.19

Sen, S., Sugiarto, D., & Rochman, A. (2020). Komparasi metode multilayer perceptron (MLP) dan long short term memory (LSTM) dalam peramalan harga beras. *Ultimatics : Jurnal Teknik Informatika*, *12*(1): 35–41. https://doi.org/10.31937/ti.v12i1.1572

Sepri, D., Fauzi, A., Wandira, R., Riza, O. S., Wahyuni, Y. F., & Hutagaol, H. (2020). Prediksi Harga Cabai Merah Menggunakan Support Vector Regression. *Computer Based Information System Journal*, *02*(October): 1–5.

Shiddiq, M. I., Sitanggang, I. S., & Agmalaro, M. A. (2020). *Model Prediksi Temporal Polutan Co Dan Co 2 Menggunakan Long Short Term Memory*. Institut Pertanian Bogor.

Sugiarto, Y. H., & Nangameka, Y. (2013). Faktor-faktor yang mempengaruhi naik – turunnya harga cabai merah menurut pendapat petani di kabupaten situbondo. *Agribios*, *11*(2): 1–16.

Wang, X., Guo, P., & Huang, X. (2011). A review of wind power forecasting models. *Energy Procedia*, *12*: 770–778. https://doi.org/10.1016/j.egypro.2011.10.103

Wulandari, H., Kurnia, A., Sumantri, B., Kusumaningrum, D., & Waryanto, B. (2017). Penerapan Analisis Regresi Spline Untuk Menduga Harga Cabai Di Jakarta. *Indonesian Journal of Statistics and Its Applications*, *1*(1): 1–12. https://doi.org/10.29244/ijsa.v1i1.47

Yu, L., Pan, Y., & Wu, Y. (2009). Research on data normalization methods in multi-attribute evaluation. *Proceedings - 2009 International Conference on Computational Intelligence and Software Engineering, CiSE 2009. IEEE*, 1–5. https://doi.org/10.1109/CISE.2009.5362721

Yusniyanti, E., & Kurniati, K. (2017). Analisa puncak banjir dengan metode MAF (Studi kasus sungai Krueng Keureuto). *Jurnal Einstein*, *5*(1): 7–12. https://doi.org/10.24114/einstein.v5i1.7224

Zahara, S., Sugianto, & M. Bahril Ilmiddafiq. (2019). Prediksi Indeks Harga Konsumen Menggunakan Metode Long Short Term Memory (LSTM) Berbasis Cloud Computing. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *3*(3): 357–363. https://doi.org/10.29207/resti.v3i3.1086

Zaytar, M. A., & El Amrani, C. (2016). Sequence to Sequence Weather Forecasting with Long Short-Term Memory Recurrent Neural Networks. *International Journal of Computer Applications*, *143*(11): 1–5.

Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y., & Liu, J. (2017). LSTM network: A deep learning approach for Short-term traffic forecast. *IET Intelligent Transport Systems*, *11*(2): 68–75. https://doi.org/10.1049/iet-its.2016.0208