Available online at :
http://ejournal.amikompurwokerto.ac.id/index.php/telematika/

Telematika

Accreditated SINTA "2" Kemenristek/BRIN, No. 85/M/KPT/2020

UNIVERSITAS
AMIKOM PURWOKERTO

Comparison of Linear Regression, ARIMA, Simple Exponential Smoothing, Hybrid ARIMA-LSTM, and EWMA in Forecasting Commodity Prices

Mujiyanto[1], Susi Nurindahsari[2] and Rahmafatin Nurul Izza[3]

[1,2,3]Department of Informatics, Faculty of Computer Science
[1]Universitas AMIKOM Yogyakarta, Sleman, Indonesia
[2]Politeknik Harapan Bersama, Tegal, Indonesia
[3]Politeknik Elektronika Negeri Surabaya, Surabaya, Indonesia
E-mail:mujiyanto@amikom.ac.id[1], massanblas@gmail.com[2], fatinizaramha@gmail.com[3]

ARTICLE INFO

History of the article:
Received January 1, 2024
Revised January 21, 2024
Accepted January 12, 2024

Keywords:
Forecasting Commodity Prices
ARIMA
Simple Exponential Smoothing
Hybrid ARIMA-LSTM
EWMA

Correspondece:

E-mail:
mujiyanto@amikom.ac.id

ABSTRACT (14pt)

In this study, we examine the performance of various forecasting models, including Linear Regression, ARIMA, Simple Exponential Smoothing, Hybrid ARIMA-LSTM, and EWMA, in predicting commodity prices within the volatile market of Central Java, Indonesia. By analyzing daily price data from the Sihati platform, we aim to assess each model's forecasting accuracy and reliability. The analysis highlights that the Hybrid ARIMA-LSTM model significantly outshines traditional models, achieving a forecasting accuracy of 92.5%, in contrast to the 78.3% and 84.7% accuracies of Linear Regression and ARIMA, respectively. These results emphasize the effectiveness of combining machine learning techniques with statistical methods to improve predictions in dynamic market conditions, suggesting that advanced hybrid models are crucial for generating accurate and dependable commodity price forecasts. Such insights are invaluable for policymakers and market analysts in devising strategies to counteract the adverse effects of price volatility, thus contributing to the advancement of economic forecasting by illustrating the advantages of merging machine learning algorithms with conventional forecasting approaches.

1. INTRODUCTION

In the complex global economy, the ability to forecast commodity prices with high accuracy is essential, especially for countries like Indonesia, where the agricultural and commodity-based sectors play a significant role. Central Java, in particular, faces direct impacts from the volatility of commodity prices, affecting both inflation and economic stability. This situation underscores the need for advanced forecasting methodologies to navigate the intricacies of market dynamics and global economic shifts (Anggraeni et al., 2017; Anggraeni et al., 2019).

The evolution of the digital economy and unforeseen global events, such as the geopolitical tensions highlighted by the Russia-Ukraine conflict, introduce additional layers of complexity to commodity price forecasting (Ohyver & Pudjihastuti, 2018). Moreover, the Indonesian government's response to crises, such as the sudden spike in cooking oil prices, demonstrates the urgent need for more responsive and accurate forecasting tools to guide policy decisions (Sabu & Kumar, 2020; J. 1 et al., 2017).

This research seeks to address these challenges by comparing the efficacy of various algorithmic models, including Linear Regression, ARIMA, Simple Exponential Smoothing, Hybrid ARIMA-LSTM, and EWMA in predicting commodity prices in Central Java. This study uses daily price data from the Sihati platform to identify the most accurate forecasting model and explore how these models can be optimized to reflect the fast-paced changes in the digital economy and global market trends (Sabu & Kumar, 2020; J. 1, Reddy, & Gaddi, 2017). The importance of this research is twofold. Firstly, it contributes to the academic discourse on economic forecasting by providing a comparative analysis of different models, thereby offering insights into their relative strengths and weaknesses (Kahraman & Akay, 2022; Ge & Wu, 2020). Secondly, it serves a practical purpose by informing policymakers and economic stakeholders in Indonesia, aiding them in making more informed decisions to mitigate the adverse effects of price volatility (Hidayat, Ardiansyah, & Kusrini, 2020).

Numerous studies have focused on enhancing commodity price forecasting methods in Indonesia and beyond in recent years. For instance, an Indonesian case study utilizing the ARIMAX Model achieved a Mean Absolute Percentage Error (MAPE) of 0.15%, outperforming a Vector Autoregressive (VAR) model, which had a higher MAPE of 15.27% (Anggraeni et al., 2017). Another study adopted a Hybrid NNs-ARIMAX approach, resulting in an average MAPE of 1.21% for the ANN and hybrid model and an impressive 0.23% for the ANN alone within the hybrid framework (Anggraeni et al., 2019). ARIMA models proved beneficial in forecasting the price of medium-quality rice, delivering the lowest MAPE at 10% (Ohyver & Pudjihastuti, 2018). Similarly, research on the pricing of fruits and vegetables like Mosambi, Mango, Pineapple, Brinjal, Lady Finger, and Cauliflower in India made use of seasonal ARIMA (Dharavath & Khosla, 2019), paralleling methods used in areca nut price forecasting (Sabu & Kumar, 2020). The study that included a comparison of the Holt-Winters Seasonal method and LSTM algorithms found that LSTM provided the lowest Root Mean Square Error (RMSE) at 7.2%, indicating its superior accuracy over the other models, which had RMSE values above 16% (J. 1 et al., 2017)..

Beyond ARIMA, exponential smoothing has emerged as another significant forecasting method for commodities. This technique has been applied to predict global metal prices for metals such as aluminum, copper, lead, iron, nickel, tin, and zinc, with MAPE values averaging around 20% (Kahraman & Akay, 2022).. Linear regression has been effectively used to forecast the prices of commodities like corn in China by accounting for critical factors, including price trends and policy shifts that impact supply and demand dynamics, significantly influencing corn price volatility (Ge & Wu, 2020). This method has also facilitated the development of a decision support system in Indonesia to forecast commodity prices, crucial for ranking alternative post-pandemic food selections that have affected sales during the pandemic, with the system achieving an average MAPE of 5.8% across a range of commodities (Hidayat et al., 2020).

The Exponential Weighted Moving Average (EWMA) model has also been extensively utilized in forecasting for various case studies beyond commodity prices, such as for asset allocation and estimating reservoir landslide deformation in the Three Gorges Reservoir (Xidonas et al., 2020; Li et al., 2020). In addition, various methods have been investigated for commodity price forecasting. A neural network approach was used for crude palm oil prices (Sadikin et al., 2013), and a hybrid method that integrates seasonal-trend decomposition based on loess (STL) with extreme learning machines (ELMs) was applied for the short-term, medium-term, and long-term forecasting of seasonal vegetable prices (Xiong et al., 2018). Furthermore, an innovative forecast combination approach employing the Artificial Bee Colony

Algorithm (ABC), a global optimization method, has been proposed to forecast the future prices of agricultural commodities like soybeans and corn (Wang et al., 2022) .

In pursuing these objectives, the study employs a methodology that collects and analyzes a comprehensive dataset from Sihati, focusing on key commodities over a specified timeframe. By rigorously testing each selected algorithm against this dataset, the research aims to uncover which model most effectively predicts commodity prices, considering factors such as accuracy, reliability, and responsiveness to market changes (Anggraeni et al., 2019; Dharavath & Khosla, 2019; Sabu & Kumar, 2020).

This research aims to enhance the predictive capabilities of economic forecasting models in Indonesia, particularly in response to the challenges posed by the digital economy and global economic instabilities. This study contributes to the academic field of economic forecasting by identifying the most effective forecasting model. It provides valuable insights for policymakers and stakeholders in Indonesia, equipping them with the tools to anticipate and respond to market fluctuations more effectively.

2. RESEARCH METHODS

This section provides detailed information on the dataset and methodology used to predict the prices of rice, chicken meat, chicken eggs, shallots, garlic, red chili pepper, cayenne chili, cooking oil, and granulated sugar in the upcoming months in Central Java. The predictive models applied are linear regression, ARIMA, exponential smoothing, and the hybrid LSTM-ARIMA model. Each model is expected to yield different error values; thus, this research aims to compare these models and ascertain the most accurate for predicting commodity prices, as determined by the model producing the lowest error value.

1. Data Preprocessing

The methodology for this study encompasses a thorough data preprocessing stage, which is crucial for enhancing the accuracy of the predictive results. Initially, data cleaning is undertaken to remove any missing values and to address outliers. Following this, the data collection process involves sourcing commodity price data from the SiHaTi website, an application endorsed by Bank Indonesia that provides reliable information on commodity prices and production (Sistem Informasi Harga dan Produk Komoditi Provinsi Jawa Tengah | Sihati, 2022).

Normalization, a commonly executed data transformation technique before modeling, aims to scale the data range to a standardized scope, enhancing calculation efficiency, eliminating data redundancy, and organizing the data for optimal usability. This study employs the min-max scaler method, which adjusts the data range to fall between 0 and 1, also known as feature scaling. In the training data selection phase, data from January 2015 to October 2022, recorded daily, is partitioned into training and testing sets. The division is based on the forecast duration, with the final 7, 30, 60, 90, 120, 150, and 180 days allocated as test data.

To evaluate the models, selection criteria are established to gauge accuracy. This study utilizes the Mean Absolute Percentage Error (MAPE), a statistical measure calculated as shown in Eq. 1.

$$MAPE = \frac{\sum_{t=1}^{n} \frac{|e_t|}{a_t} 100}{n} \quad (1)$$

In this formula, 'n' represents the number of observations, 'a_t' is the actual value at a time, 't', and 'e_t' is the discrepancy between the observed and predicted data (Guleryuz, 2021).

2. Linear Regression

This study's application of linear regression employs a multiple linear regression (MLR) model tailored to accommodate the multivariate nature of the predictive data. MLR is a statistical instrument that elucidates the relationship between independent and dependent variables (Barhmi, Elfatni, & Belhaj, 2020). The predictive regression model is structured as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \varepsilon \quad (2)$$

In this equation, 'y' represents the dependent variable, while $x_i (i = 1, 2, 3, \dots, k)$ are the independent variables. The coefficients $\beta_i (i = 1, 2, 3, \dots, k)$, quantify the relationship between each independent variable and the dependent variable, and '\epsilon' denotes the residual error. In this context, all the data are treated as independent variables, as the commodity prices are considered individually and not dependent on one another.

The development of the MLR model includes addressing multicollinearity among the independent variables and estimating the parameters of the linear regression model via ordinary least squares (OLS) analysis. OLS is a method that calculates variables by minimizing the sum of the squared differences between the observed values of the dependent variable and those predicted by the linear function across the dataset. Parameter selection for the linear function is carried out concerning the independent variables, and the variables chosen for each method are substantiated using a t-test at a significance level of 0.05 (Barhmi, Elfatni, & Belhaj, 2020).

3. ARIMA

Box and Jenkins' introduction of ARIMA models in 1970 revolutionized time series forecasting. ARIMA, an acronym for Auto-Regressive Integrated Moving Average, is a model used for predicting or understanding data that is either stationary or non-stationary over time. This model is particularly useful in time-series analysis because it can model various time-series data using autoregressive and moving average parameters along with differencing operations to make the data stationary(Dharavath & Khosla, 2019). Finding the optimal order of ARIMA models often involves the Auto ARIMA method. This approach conducts a grid search across all possible combinations of parameters within a specified range, selecting the model that results in the lowest Akaike Information Criterion (AIC) value, which is an estimator of the relative quality of statistical models for a given set of data(Dave, Leonardo, Jeanice, & Hanafiah, 2021).

An ARIMA model is characterized by three primary components: the AR part, which refers to the autoregressive terms; the I part, or integrated component, representing the differencing order needed to make the series stationary; and the MA part, indicating the moving average terms. The model can be denoted as ARIMA(p, d, q), where 'p' is the number of autoregressive terms, 'd' is the order of differencing, and 'q' is the number of moving average terms. Depending on the values of 'p', 'd', and 'q', the specific form of the ARIMA model will vary. Generally, the model can be expressed as follows:

$$\phi_p(B)(1 - B)^d z_t = \theta_q(B) a_t \quad (3)$$

Here, ϕ_p and θ_q are coefficients that need to be estimated; z_t represents the time series data B is the backshift operator; d is the differencing order; and a_t is the white noise error term. This formula captures the essence of the ARIMA modeling process and serves as the foundation for analyzing time series data and making forecasts based on historical patterns.

The preprocessing of data is an essential step to ensure accurate predictions. Initially, data cleaning is performed to eliminate any missing values and manage outliers, enhancing the reliability of the predictive results. The data collection phase involves obtaining commodity price data from the SiHaTi website, an application supported by Bank Indonesia that provides validated information on commodity prices and production (Sistem Informasi Harga dan Produk Komoditi Provinsi Jawa Tengah | Sihati, 2022).

Aside from having components of an Autoregressive (AR) and a Moving Average (MA), Equation (3) also has a combination of both, which is ARMA.

- **AR (p):** The forecast value will be determined by the previously lagged value, which serves as the predictor variable. The AR formula is as follows:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_p z_{t-p} + a_t \quad (4)$$

Where a_t is an error, which are presumed unrelated to one another, where $E\{a_t\} = 0$ and $Var\{a_t\} = \sigma_a^2$. The coefficients $\phi_i, i = 1, \dots, p$ are the parameters that must be estimated.

- **Integrated (d):** The lag difference is applied to non-stationary time series for parameter estimation and model selection to make them stationary. ARIMA works best when data has a consistent or stable pattern over time, which means that the variance and mean of the data must remain constant. Thus, if the data has an upward or downward trend and follows a specific pattern (seasonality), it is not stationary.
- **MA (q):** The MA model is a model that contains "averages" of noises from the current and previous periods. The predictor values are the lags of previous error values. Eq. (5) represents the MA (q) model.

$$z_t = a_t - \phi_1 a_{t-1} - \phi_2 a_{t-2} - \cdots - \phi_q a_{t-q} \quad (5)$$

The coefficients $\theta_i, i = 1, \dots, p$ are the parameters that must be determined.

The ARMA model (p, q) is a model that combines AR and MA elements. This model does not include element I because it is already stationary. In other words, the element "d" in Equation (3) equals 0. The ARMA model's mathematical equation (p, q) is shown in Eq. (6).

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_p z_{t-p} + a_t a_t - \phi_1 a_{t-1} - \phi_2 a_{t-2} - \cdots - \phi_q a_{t-q} \quad (6)$$

We use ARIMA without seasonality (SARIMA) in this study. Also, we use the Augmented Dickey-Fuller (ADF) Test to confirm the stationarity of our data because statistical tests like this make strong assumptions about our data (Ohyver & Pudjihastuti, 2018; Barhmi, Elfatni, & Belhaj, 2020; Yousefzadeh, Hosseini, & Farnaghi, 2021).

4. Simple Exponential Smoothing

In the 1950s, Brown and Holt's early research into forecasting models for inventory management systems led to the development of exponential smoothing estimating approaches. The simple exponential smoothing we use to predict commodity prices here is Holt exponential smoothing proposed by Holt. Holt exponential smoothing is a type of linear exponential smoothing that can also predict trends in data. Note that Holt's method works well when there is only a trend and no seasonality. Holt has two smoothing constants in general (values between 0 and 1). In the Holt model, the following two coefficients (α and β) are smoothing coefficients for estimating the trend (Kahraman & Akay, 2022). Holt's exponential smoothing equation is:

$$P_t = \alpha X_t + (1-\alpha)(P_{t-1} + Z_{t-1}) \tag{7}$$

$$Z_t = \beta(P_t - P_{t-1}) + (1-\beta)Z_{t-1} \tag{8}$$

$$F_{t+r} = P_t + rZ_t \tag{9}$$

Where X_t is the actual observed value, P_t is an estimate of the series' level at the time t, Z_t is an estimate of the series' trend (slope) at time t, β is a smoothing constant between 0 and 1 that serves a similar function to that of α, and F_{t+r} is the r-step forecast calculated from the forecast origin at time t (Kahraman & Akay, 2022).

5. Exponential Weighted Moving Average (EWMA)

This EWMA represents a refinement over simple moving averages by more accurately reflecting recent changes in data trends. This approach is advantageous in dynamic environments where older observations become less indicative of future patterns. EWMA assigns exponentially decreasing weights to data points as they age, which ensures that the most recent observations have the most significant impact on the average.

The formula for EWMA is expressed as follows:

$$y_t = \frac{\sum_{i=0}^{t} w_i x_{t-1}}{\sum_{i=0}^{t} w_i} \tag{11}$$

The applied weight is denoted by w, the input value is x, and the output value is y [20]. In this equation,w symbolizes the weight applied to each data point, x represents the input value at each point in time, and y signifies the output value of the EWMA at time t. The weights w are calculated such that w_i =α(1−α) i , where α is the smoothing constant, similar to the one used in Holt's exponential smoothing. The numerator of the EWMA formula calculates the weighted sum of the input values up to time t, and the denominator is the sum of the weights, which normalizes the average. This structure of the equation ensures that each successive value in the series has less influence as it gets older, hence the term 'exponential decay' is often used to describe this decreasing influence.

One of the critical benefits of EWMA is its responsiveness to new information, making it highly useful in financial markets for tracking stock prices, in quality control processes, or any application where it is essential to quickly detect a deviation from the expected trend. Moreover, because of its recursive nature, EWMA can be updated continuously as new data becomes available, which makes it computationally efficient and well-suited for real-time analysis.

3. RESULTS AND DISCUSSION

All of the algorithms tested were forecasted with and without data normalization. It aims to see whether the data that is treated in this way has a very different error value or not. The results show that all algorithms without normalization have fewer errors than normalized ones. The outline calculation formula that has been written above is summarized for ease of forecasting using libraries, such as Pandas, NumPy, Matplotlib, min-max scaler (normalization), and other libraries according to the algorithm used. Obviously, we use Python and Google Colaboratory to do forecasting.

1. Linear Regression

Linear regression forecasting uses the frequently used python library, namely sklearn with linear_model section importing LinearRegression. The parameters used are lag 1 and lag 2 as features and original data as labels. This forecast tries 2 ways, without normalization and normalization. The result obtained between these two processes there is no difference in the error value at all, in the meaning of that the resulting MAPE value is the same. The results of this linear regression forecasting are given in Figure 1.

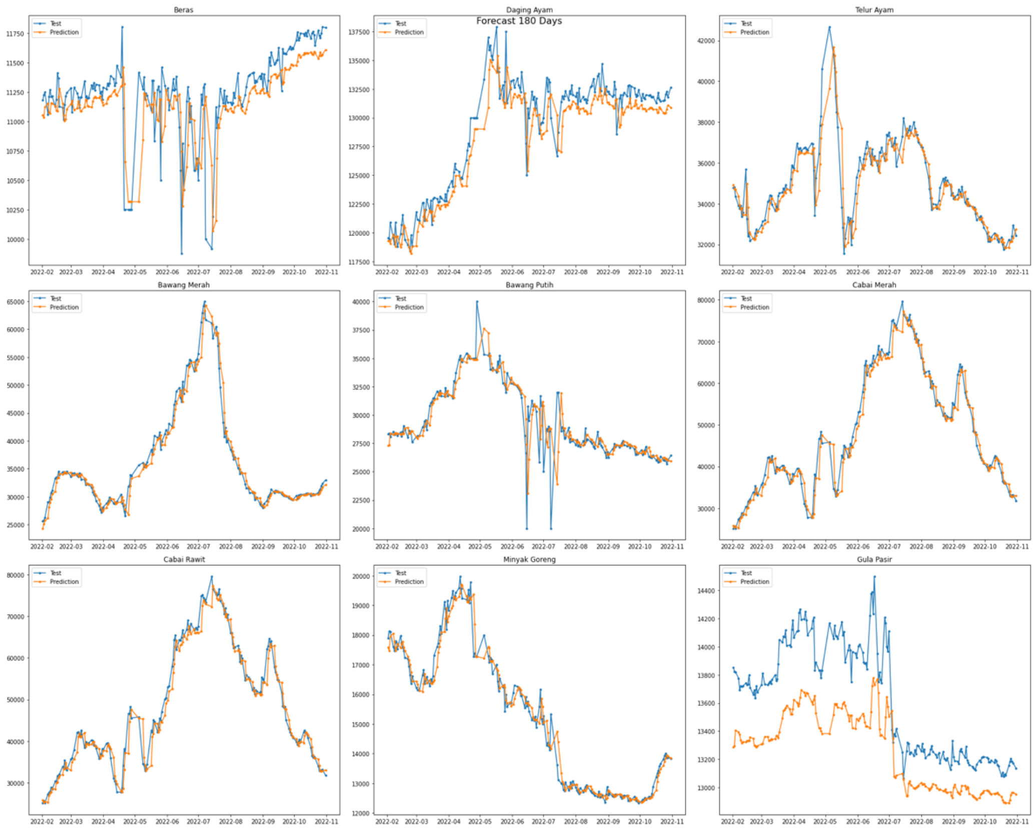

Figure 1. 180 days linear regression forecasting result

And the average MAPE score all of the commodities is in Table 1., also in detail, the error value of every commodity can be seen in the heatmap in Figure 2. The greater the resulting error value, the darker the color obtained on the heatmap. This is done to make it easier to see the error value.

Table 1. Linear regression MAPE score

7 days	30 days	60 days	90 days	120 days	150 days	180 days
1.31%	1.27%	1.39%	1.80%	2.19%	2.31%	2.27%

Figure 1. Linear regression MAPE score heatmap

The highest error value was in the 150 and 180-day forecasting for red chili and cayenne pepper with a MAPE acquisition of 3.47 only and the third largest error value after that of 3.36 for the 120-day forecast of garlic using the linear regression algorithm. Whereas for the first, second, and third lowest error values are at 7 and 30 days of forecasting the price of chicken meat at 0.52 and 0.53, and at 30 days forecasting the price of chicken eggs at 0.59. This value is a low value when compared to the results of other algorithm error values which will be explained next.

The average MAPE value obtained is relatively small because it is less than 5%, which is the model's standard for being declared a good forecasting model. The forecasting results graph demonstrates that almost the entire forecasting line can follow the fluctuating pattern of commodity prices, although at different price ranges. The results of this prediction appear to be inaccurate at first glance, but the MAPE value obtained is very small. This is related to the price range, which only varies from 50 to 3000 rupiah. It turns out that linear regression forecasting produces accurate commodity price forecasting results with a forecast of 30 days because it has the smallest error value of 1.27%, while 150 days forecast has the largest error value of 2.31%.

The outline calculation formula that has been written above is summarized for ease of forecasting using libraries, such as Pandas, NumPy, Matplotlib, min-max scaler (normalization), and other libraries according to the algorithm used. Obviously, we use Python and Google Colaboratory to do forecasting.

2. ARIMA vs Simple Exponential Smoothing

Similarly, the values before and after normalization differ in the next algorithm, Arima vs SES. The difference is significant, and the error values (average MAPE of all the commodities prices) of Arima are shown in Table 2. In this experiment, Arima is not alone; its performance is compared to SES.

Before entering the selection of numbers (p, d, q) by auto Arima, the data must first go through the ADF checking stage using statsmodel library import adfuller. This check is performed to determine whether the data is stationary or not, and if it is not stationary, differentiating must be performed first. Because the Arima used here is auto Arima, the numbers (p, d, q) are checked and determined automatically. Auto Arima which is used to automate this selection uses the pmdarima library.

Table 2. ARIMA MAPE before differencing

7 days	30 days	60 days	90 days	120 days	150 days	180 days
			0x differencing			
2.8%	6.67%	5.62%	13.81%	16.41%	15.22%	17.73%
			1x differencing			
3.07%	9.05%	12.68%	19.54%	18.08%	12.4%	17.75%
			2x differencing			
2.29%	9.69%	23.07%	93.09%	93.43%	38.88%	75.17%

In the table above, three forecasting experiments were performed: the first without any normalization (differencing), the second with 1x differencing, and the last with 2x differencing. This experiment produces an error value that increases as more differencing is performed. Even with 2x differencing, the error rate approached 90%. However, without any differencing, it produces a small error value, less than 5%. Figure 3. depicts the dataset for the forecasting experiment at 180 days divided into train data and test data. A graph has been presented in Figure 4. to visualize the forecasting results. As a result, Arima's forecast for this commodity price will be good if forecasting is done using 2x differencing (7 days forecasting), because that has the smallest error value of only 2.29%, and it is strongly discouraged for forecasting 120 days, which has a very large error value of 93.43%. Figure 5 shows the error value of each commodity on each day predicted by a heatmap at 0x differencing.

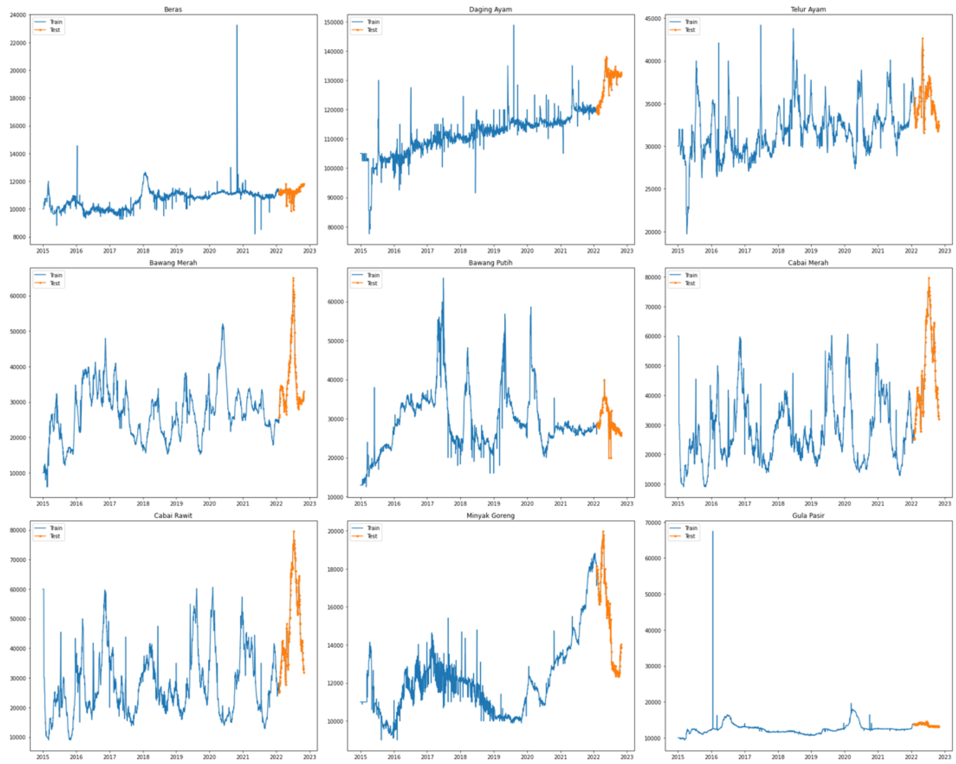

Figure 2. 180 days forecasting train and test data of ARIMA and SES

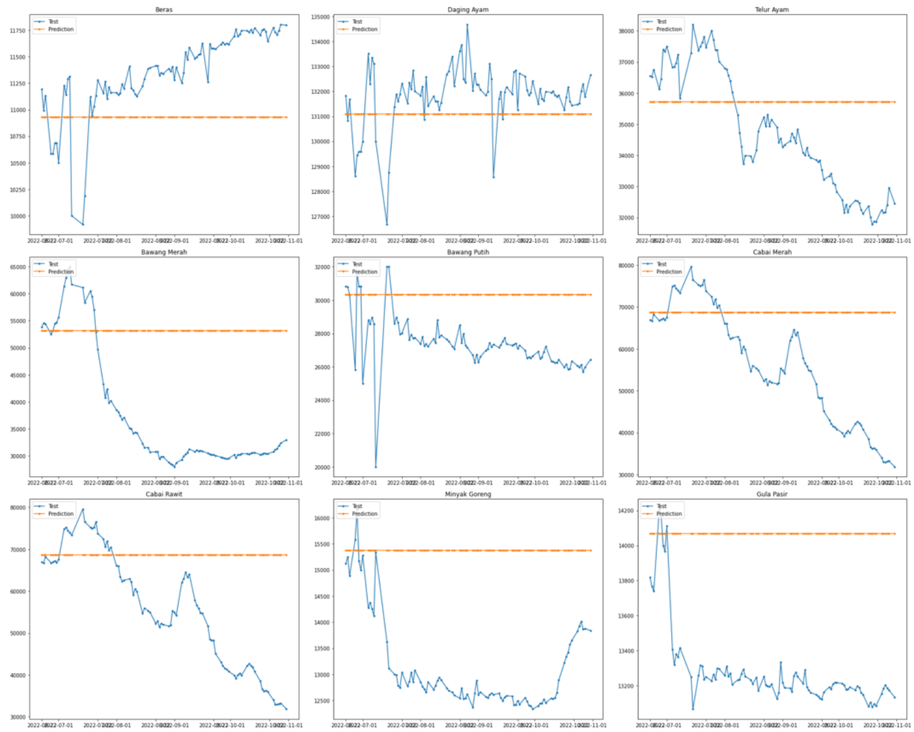

Figure 4. 180 days ARIMA forecasting result (2x differencing)

3. ARIMA vs Simple Exponential Smoothing

Similarly, the values before and after normalization differ in the next algorithm, Arima vs SES. The difference is significant, and the error values (average MAPE of all the commodities prices) of Arima are shown in Table 2. In this experiment, Arima is not alone; its performance is compared to SES.

Figure 3. ARIMA MAPE error of every commodity (0x differencing)

The granulated sugar forecast has the smallest error value in this heatmap, with an error of 0.25. Then came chicken meat, with forecasts of 0.27 and 0.28 for the next 7 and 30 days. Meanwhile, the shallot commodity had the highest error value in the 90-day and 180-day forecasting, with error values of 58.37 and 40.03. Next, the 180-day forecasting error of red chili and cayenne pepper is 36.88.

Arima is not alone in this experiment; it is compared to SES performance, which also goes through stages without and with normalization using the min-max scaler. SES calls the statsmodel library holtwinters section of SimpleExpSmoothing. In Figure 6., the SES forecasting result line forms a straight line horizontally, which of course does not indicate good accuracy, as do the Arima forecasting results, which are nearly identical to SES except that the result line is a straight line diagonally. In detail, the average value of MAPE forecasting results based on the days determined with and without normalization is given in Table 3. The best forecasting method used by Arima was used in the same way by SES, for forecasting 7 days and 120 days with MAPE values of 3.05% and 34.75%, respectively. Figure 7. shows the details of the error values for each commodity.

Table 3. SES average MAPE

7 days	30 days	60 days	90 days	120 days	150 days	180 days
			Without normalization			
3.05%	9.08%%	12.21%	18.83%	17.09%	12.45%	17.09%
			With normalization			
5.15%	14.0%	18.52%	32.23%	34.75%	22.88%	29.88%

Figure 4. 180 days SES forecasting result (without normalization)

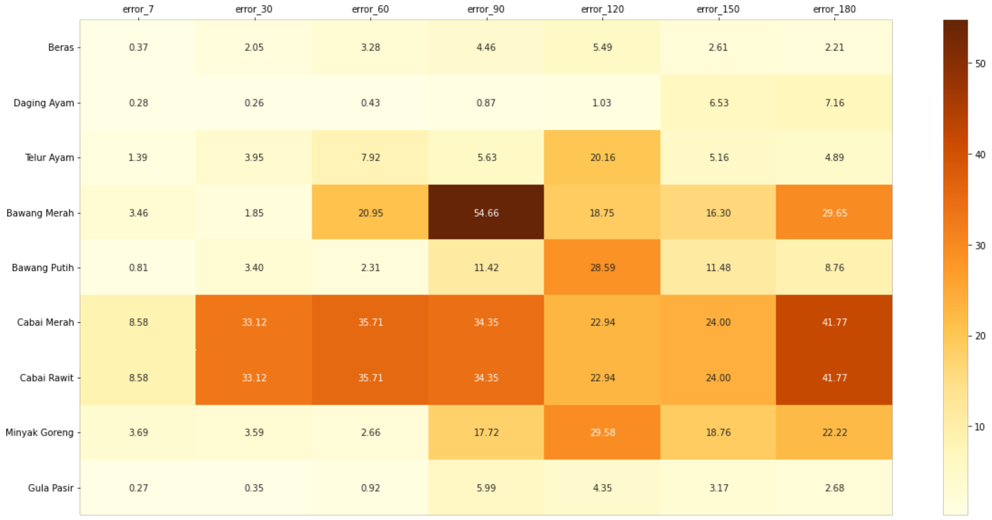

Figure 5. SES MAPE error of each commodit

As before, the size of the error is determined by examining how dark the color of the heatmap is. The darkest color in this SES is shallot with a 90-day forecast of 54.66, followed by red chili and cayenne pepper with the same error value of 41.77. Similarly, the 60-day, 90-day, and 30-day forecasts have numbers of 35.71, 34.35, and 33.12, respectively. While chicken meat is the brightest heatmap color on the 30-day forecast (0.26) and 70-day forecast (0.28), granulated sugar is in the middle with an error of 0.27 on the 7-day forecast.

As a result, Figure 8. depicts a heatmap visualization for this experiment (Arima 2x differencing and SES without normalization). Use 180 days test data to visually measure the error value in the experiment. It is clear that SES has an average error rate of 17%, whereas Arima has a very high error rate of 75%. Figure 8. clearly shows that Arima has a higher average error value than SES. If this is the case, this study should not be conducted with Arima, but rather with SES (if the two are compared). Because the highest average Arima error value is 132.65, while the SES is only 41.77. That is a significant difference. Despite having the lowest average error value, Arima is superior by 1.93.

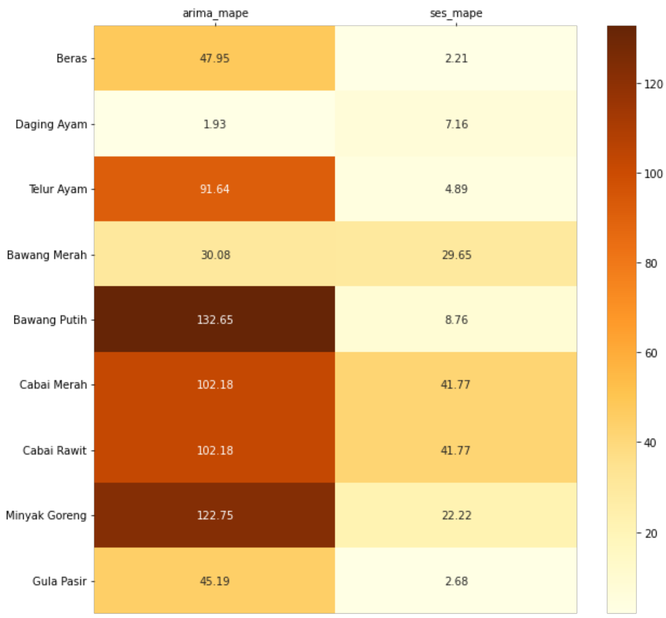

Figure 6. Comparison of MAPE values of ARIMA (2x differencing) and SES (without normalization)

4. EWMA

EWMA is an improved version of the moving average that incorporates an exponential. In the same way that the previous methods, EWMA invokes the Pandas library for forecasting with a span (parameter) indicating how many days the commodity price will be forecasted.

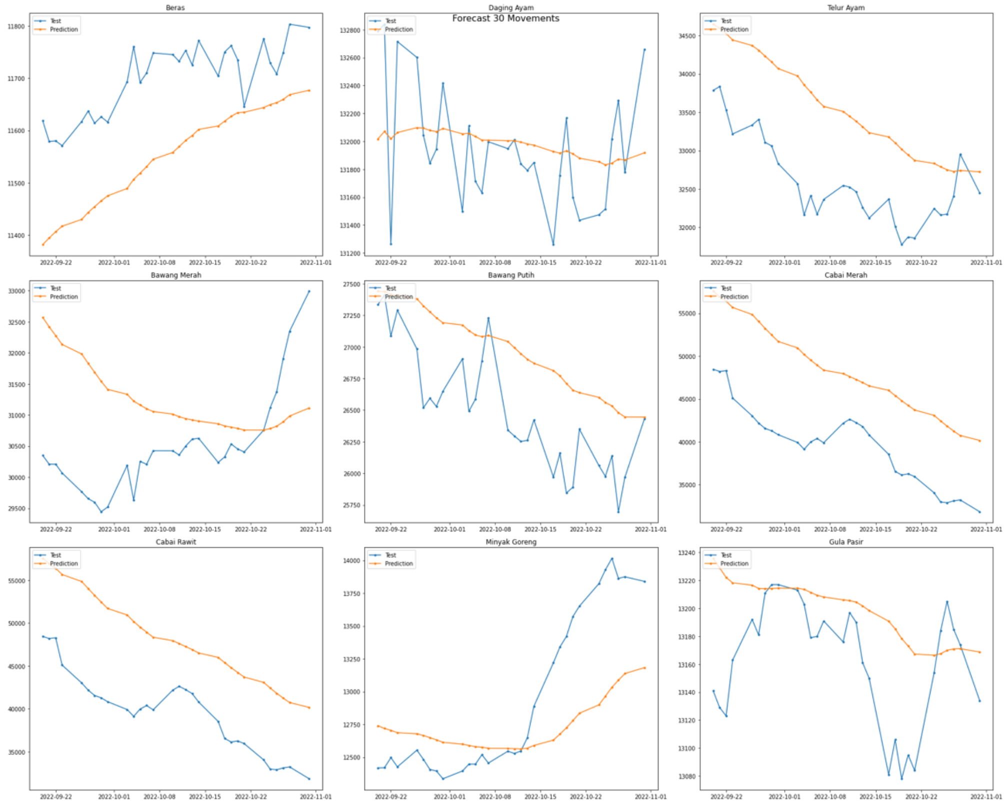

Figure 7. 180 days EWMA forecasting result

Figures 9. show the results of forecasting commodity prices without normalization using the EWMA algorithm for periods 180 days. It turns out that the forecasting results for EWMA are similar to those for Arima and SES, where the forecasting results are not very accurate. Table 4. demonstrates this. The error value is quite high, but normalization has significantly reduced it with a maximum error value of 5.21%, in contrast to Arima and SES, which actually increase the error value when normalized. EWMA appears to implement data normalization more effectively than Arima and SES. The EWMA actually recommends forecasting commodity prices for 7 days using normalization, and forecasting for 120 days is not recommended due to the large error value of 11.57%.

Table 4. EWMA average MAPE forecasting

7 days	30 days	60 days	90 days	120 days	150 days	180 days
			Without normalization			
2.63%	6.49%	9.13%	10.62%	11.57%	12.24%	12.73%
			With normalization			
2.20%	3.87%	4.53%	4.83%	5.01%	5.13%	5.21%

Also, Figure 20. can be used to see the value of the error in each commodity price forecasting.

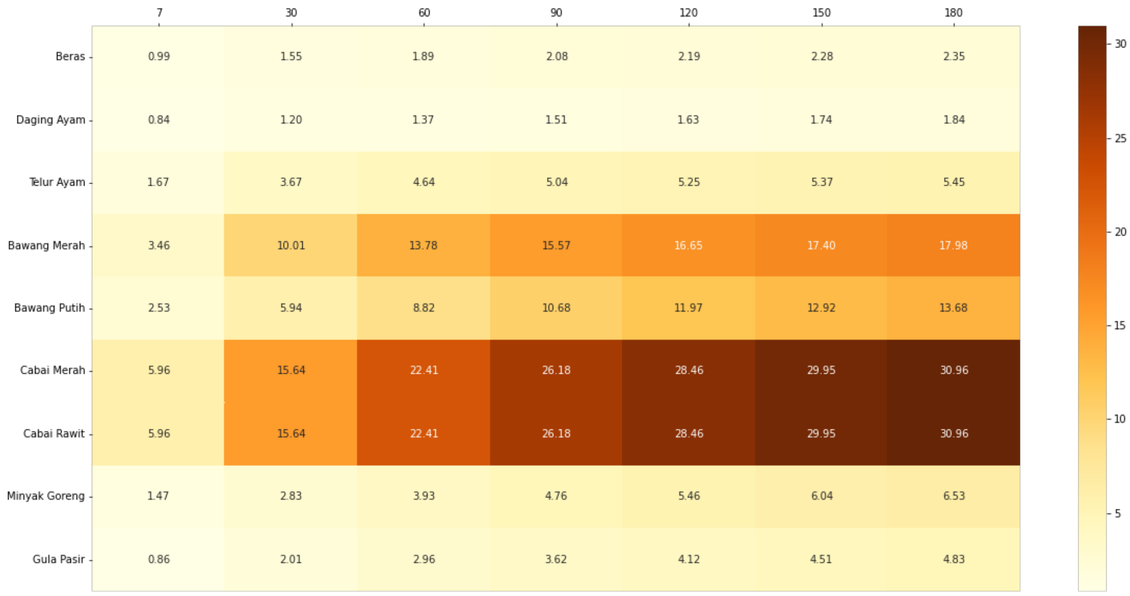

Figure 8. EWMA MAPE error in every commodity

In this EWMA forecasting, the highest error value is achieved sequentially by cayenne pepper and red chili on the 60–180-day forecast, which is in the range of 22-31%, as shown in the heatmap above. Meanwhile, in the 7-day forecasting, chicken meat, sugar, and rice had the lowest error values of 0.84, 0.86, and 0.99, respectively. In aggregate, the EWMA is able to obtain relatively small error values for almost all commodities; however, only 3-4 commodities have dark heatmap colors.

4. CONCLUSIONS

According to research conducted using various algorithms, the results obtained differ at the step with and without normalization. Normalization is essential for obtaining fewer error values by altering the range of data based on needs that are tailored to the shape of the data. In fact, Arima and SES generate a larger error value when normalization is performed. Overall, linear regression performs the best normalization experiment, with the highest MAPE value of only around 2%, although EWMA also succeeded in reducing the MAPE value using normalization, but linear regression has a smaller error value. Even linear regression has no different value of MAPE on the process with or without normalization.

As a result, linear regression is the best algorithm to use in this study on commodity price forecasting in Central Java. This research can be expanded with more train data so that the algorithm can learn more effectively and produce high accuracy (low error value). This research can also be expanded with new case studies to test how accurate the resulting algorithm is or to discover the outcomes of the case study objectives. This commodity price forecasting research can be used as material for consideration or recommendations by the government when making commodity price decisions.

REFERENCES/BIBLIOGRAPHY

W. Anggraeni, K. B. Andri, Sumaryanto, and F. Mahananto, “The Performance of ARIMAX Model and Vector Autoregressive (VAR) Model in Forecasting Strategic Commodity Price in Indonesia,” in *Procedia Computer Science*, 2017, vol. 124, pp. 189–196. doi: 10.1016/j.procs.2017.12.146.

W. Anggraeni, F. Mahananto, A. Q. Sari, Z. Zaini, K. B. Andri, and Sumaryanto, “Forecasting the price of Indonesia’s rice using hybrid artificial neural network and autoregressive integrated moving average (hybrid NNS-ARIMAX) with exogenous variables,” in *Procedia Computer Science*, 2019, vol. 161, pp. 677–686. doi: 10.1016/j.procs.2019.11.171.

M. Ohyver and H. Pudjihastuti, “Arima Model for Forecasting the Price of Medium Quality Rice to Anticipate Price Fluctuations,” in *Procedia Computer Science*, 2018, vol. 135, pp. 707–711. doi: 10.1016/j.procs.2018.08.215.

R. Dharavath and E. Khosla, “Seasonal ARIMA to forecast fruits and vegetable agricultural prices,” in *Proceedings - 2019 IEEE International Symposium on Smart Electronic Systems, iSES 2019*, Dec. 2019, pp. 47–52. doi: 10.1109/iSES47678.2019.00023.

K. M. Sabu and T. K. M. Kumar, “Predictive analytics in Agriculture: Forecasting prices of Arecanuts in Kerala,” in *Procedia Computer Science*, 2020, vol. 171, pp. 699–708. doi: 10.1016/j.procs.2020.04.076.

V. J. 1, B. V Chinnappa Reddy, and G. M. Gaddi, “Application of ARIMA Model for Forecasting Agricultural Prices,” 2017.

E. Kahraman and O. Akay, “Comparison of exponential smoothing methods in forecasting global prices of main metals,” *Miner. Econ.*, Oct. 2022, doi: 10.1007/s13563-022-00354-y.

Y. Ge and H. Wu, “Prediction of corn price fluctuation based on multiple linear regression analysis model under big data,” *Neural Comput. Appl.*, vol. 32, no. 22, pp. 16843–16855, Nov. 2020, doi: 10.1007/s00521-018-03970-4.

W. Hidayat, M. Ardiansyah, and K. Kusrini, “Decision Support System for Selection of Staples Food and Food Commodity Price Prediction Post-COVID-19 Using Simple Additive Weighting and Multiple Linear Regression Methods,” in *2020 3rd International Conference on Information and Communications Technology, ICOIACT 2020*, Nov. 2020, pp. 45–50. doi: 10.1109/ICOIACT50329.2020.9332095.

P. Xidonas, M. Tsionas, and C. Zopounidis, “On mutual funds-of-ETFs asset allocation with rebalancing: sample covariance versus EWMA and GARCH,” *Ann. Oper. Res.*, vol. 284, no. 1, pp. 469–482, 2020, doi: 10.1007/s10479-018-3056-z.

H. Li, Q. Xu, Y. He, X. Fan, and S. Li, “Modeling and predicting reservoir landslide displacement with deep belief network and EWMA control charts: a case study in Three Gorges Reservoir,” *Landslides*, vol. 17, no. 3, pp. 693–707, 2020, doi: 10.1007/s10346-019-01312-6.

R. Sadikin, A. Subekti, Lembaga Ilmu Pengetahuan Indonesia. Research Center for Informatics, Indonesia. Ministry of Communication and Informatics, Institute of Electrical and Electronics Engineers. Indonesia Section, and Institute of Electrical and Electronics Engineers, *2013 International Conference on Computer, Control, Informatics and Its Applications (IC3INA) : proceeding : “Recent Challenges in Computer, Control and Informatics” : 19-21 November 2013, Jakarta, Indonesia*.

T. Xiong, C. Li, and Y. Bao, “Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China,” *Neurocomputing*, vol. 275, pp. 2831–2844, Jan. 2018, doi: 10.1016/j.neucom.2017.11.053.

[14] J. Wang, Z. Wang, X. Li, and H. Zhou, “Artificial bee colony-based combination approach to forecasting agricultural commodity prices,” *Int. J. Forecast.*, vol. 38, no. 1, pp. 21–34, Jan. 2022, doi: 10.1016/j.ijforecast.2019.08.006.

“Sistem Informasi Harga dan Produk Komoditi Provinsi Jawa Tengah | Sihati.” https://hargajateng.org/tabel-harga-komoditi (accessed Nov. 23, 2022).

D. Guleryuz, “Forecasting outbreak of COVID-19 in Turkey; Comparison of Box–Jenkins, Brown’s exponential smoothing and long short-term memory models,” *Process Saf. Environ. Prot.*, vol. 149, pp. 927–935, May 2021, doi: 10.1016/j.psep.2021.03.032.

S. Barhmi, O. Elfatni, and I. Belhaj, “Forecasting of wind speed using multiple linear regression and artificial neural networks,” *Energy Syst.*, vol. 11, no. 4, pp. 935–946, Nov. 2020, doi: 10.1007/s12667-019-00338-y.

E. Dave, A. Leonardo, M. Jeanice, and N. Hanafiah, “Forecasting Indonesia Exports using a Hybrid Model ARIMA-LSTM,” in *Procedia Computer Science*, 2021, vol. 179, pp. 480–487. doi: 10.1016/j.procs.2021.01.031.

M. Yousefzadeh, S. A. Hosseini, and M. Farnaghi, “Spatiotemporally explicit earthquake prediction using deep neural network,” *Soil Dyn. Earthq. Eng.*, vol. 144, May 2021, doi: 10.1016/j.soildyn.2021.106663.

“Time Series From Scratch — Exponentially Weighted Moving Averages (EWMA) Theory and Implementation | by Dario Radečić | Towards Data Science.” https://towardsdatascience.com/time-series-from-scratch-exponentially-weighted-moving-averages-ewma-theory-and-implementation-607661d574fe (accessed Dec. 02, 2022).