BAREKENG: Journal of Mathematics and Its Applications
March 2025 Volume 19 Issue 1 Page 0407–0418
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss1pp0407-0418

COMPARATIVE PERFORMANCE OF SARIMAX AND LSTM MODEL IN PREDICTING IMPORT QUANTITIES OF MILK, BUTTER, AND EGGS

Ghardapaty Ghaly Ghiffary[1*], Eka Dicky Darmawan Yanuari[2], Khairil Anwar Notodiputro[3], Yenni Angraini[4], Laily Nissa Atul Mualifah[5]

[1,2,3,4,5]*Department of Statistics and Data Science, Faculty of Mathematics and Natural Science, IPB University Jln. Meranti, Kampus IPB Dramaga. Bogor 16680, Indonesia*

*Corresponding author's e-mail: * **ghardapatyghiffary@apps.ipb.ac.id***

Article History:

Received: 19th June 2024
Revised: 29th November 2024
Accepted: 30th November 2024
Published: 13th January 2025

Keywords:

Forecast;
HS04;
Import;
LSTM;
SARIMAX.

ABSTRACT

This study evaluates how well SARIMAX and LSTM models predict monthly imports of HS-04 commodities (butter, eggs, and milk) in Indonesia. Data were provided by BPS Statistics Indonesia, Bank Indonesia, Ministry of Trade, Trade Map, and Indonesia National Single Window and used from January 2006 to February 2024. The SARIMAX model included exogenous variables such as inflation rates, USD/IDR exchange rates, and major Indonesian holidays (Eid al-Fitr, Eid al-Adha, Christmas, and Lunar New Year). The results show that the SARIMAX and LSTM models predict the import volumes of butter, eggs, and milk with good accuracy. However, the SARIMAX model demonstrated superior forecasting accuracy, achieving a lower RMSE of 7547.89 and a MAPE of 13.16 compared to the LSTM model, which yielded an RMSE of 8787.73 and a MAPE of 14.89. The SARIMAX model performed significantly better when the lunar new year was added as an exogenous variable. In order to support price stability and economic growth in Indonesia, this research provides policymakers and industry stakeholders with critical information to optimize import management strategies for butter, eggs, and milk commodities. Accurate forecasts can contribute to price stability, enhanced food security, and sustainable economic development in Indonesia by enabling informed decisions on import quotas, tariff adjustments, investment in domestic production, and strategic reserves.

How to cite this article:

G. G. Ghiffary, E. D. D. Yanuari, K. A. Notodiputro, Y. Angraini and L. N. A. Mualifah., "COMPARATIVE PERFORMANCE OF SARIMAX AND LSTM MODEL IN PREDICTING IMPORT QUANTITIES OF MILK, BUTTER, AND EGGS," *BAREKENG: J. Math. & App.*, vol. 19, iss. 1, pp. 0407-0418, March, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Forecasting constitutes a methodological discipline aimed at predicting future events. This process entails utilizing historical data projected into the future through a systematic model [1]. Time series data, which is an accumulation of observations made at regular intervals like hourly, daily, weekly, monthly, or annually, is the particular application for it. Time series data typically displays four main patterns: cyclic, seasonal, trend, and constant [2]. Forecasting methodologies are bifurcated into two principal categories based on their approach: conventional methods and machine learning methods. The machine learning approach uses an algorithm to identify patterns in data without being aware of the data's creation process [3].

The Autoregressive Integrated Moving Average (ARIMA) model represents a traditional forecasting method that analyzes time series data to discern patterns and trends via data visualization. This technique seeks to determine the stationarity of the data concerning mean and variance [4]. The drawback of Arima is that it is unable to identify seasonal trends in the data. Data with seasonal patterns and trends can be analysed and predicted using a statistical model called SARIMA, or Seasonal Autoregressive Integrated Moving Average. The Autoregressive Integrated Moving Average (ARIMA) model is expanded upon by the SARIMA model, which incorporates seasonal elements [5]. However, SARIMA's usefulness is limited to univariate time series, which reduces its ability to effectively capture influences from outside sources. The Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) model addresses this shortcoming and enhances SARIMA by incorporating exogenous variables [6]. Comparative research on the efficacy of SARIMA and SARIMAX models in forecasting monthly rainfall demonstrated that SARIMAX exhibits superior performance [7]. The intricacy of rainfall data, which is frequently influenced by factors other than historical rainfall itself, can be better captured by SARIMAX with the addition of these extra variables. According to this study, SARIMAX performs better because it can account for these outside factors, producing forecasts that are more accurate. The results indicate that SARIMAX provides a more robust and dependable solution than SARIMA for forecasting tasks when external influences are significant, as demonstrated by its increased accuracy in predicting monthly rainfall. As time goes on, a number of new forecasting techniques have surfaced, particularly machine learning techniques, which have numerous benefits over conventional approaches. such as being able to process large amounts of data well, data with long-term dependencies and the ability to capture unseen patterns.

Machine learning is a different, more widely used forecasting technique. In this field, the long-short-term memory (LSTM) network is one often used technique. An improvement over the Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) is preferred because it is good at modeling time series data and can consistently produce longer-term predictions than other approaches [8]. In contrast to conventional RNN, which have trouble learning long-term relationships due to the vanishing gradient problem, which occurs when gradients get incredibly small during backpropagation, LSTM networks use unique memory cells to store and update information selectively over time. The input, output, forget, and memory gates are standard components of an LSTM design. The input gate vector governs how much the input vector influences the memory vector. The forget gate vector regulates the retention or deletion of previous memory. Meanwhile, the output gate vector governs memory's contribution to the hidden state [9]. Previous studies have demonstrated how well LSTM performs when compared to other machine learning models. An analysis by [10] , for instance, contrasted the accuracy of LSTM with that of the Double Random Forest (DRF) model in processing economic indicator data, demonstrating that LSTM performed better than DRF. In particular, LSTM generated more accurate forecasts, as seen by lower mean absolute percentage error (MAPE) and root mean square error (RMSE) measurements. The findings indicate that LSTM is more accurate in identifying complicated, nonlinear patterns in economic data, in addition to being better at capturing long-term interdependence.

Forecasting plays a pivotal role in guiding strategic decisions regarding import and investment policies for governments and industry stakeholders. One pertinent example of time series data in this context is import commodities. A goods classification system employing Harmonized System (HS) codes is outlined in [11], which addresses the establishment of a classification structure and the implementation of import duty tariffs on imported goods. Commodities under HS Code 04, encompassing butter, eggs, and milk, are particularly significant. These items are essential to Indonesia’s nutritional needs as they provide animal protein vital for bodily growth and maintenance. Because of local production cannot keep up with Indonesia's strong demand for these products, the country must rely on imports to make up the difference.

The significant imports of eggs and dairy products into Indonesia in 2022 demonstrate how crucial these goods are to supplying domestic demand. A substantial portion of Indonesia's imports, which are expected to total about USD 237.4 billion in 2022, will be agricultural goods including milk, butter, and eggs, according to the World Bank [12]. Importing these goods is essential for keeping inflation under control and prices stable. Imports contribute to a stable supply by enhancing domestic production, which reduces price volatility. According to the International Trade Administration, Indonesia's strategic import policies are intended to maintain price stability by balancing domestic supply and demand [13]. The importation of eggs and dairy products also helps the economy by sustaining sectors like retail and food processing, which in turn generate jobs. The goal of the Indonesian government's trade policies is to import necessities in order to increase national GDP and create jobs [14]. Consequently, forecasting future imports of commodities with HS Code 04 is of paramount importance. [15] carried out one of the research projects that forecasted milk imports. The seasonal trends observed in HS-04 commodity import data, coupled with its responsiveness to various factors such as inflation, exchange rates, and calendar variation effects, render it suitable for analysis using the SARIMAX model. A study by [16] utilized the effect of calendar variations to assess how Eid al-Fitr influences Indonesia's demand for Muslim boys' clothing. The findings indicated that the ARIMAX model surpasses the ARIMA model in terms of predictive accuracy. Similarly, research by [17] employed exchange rates as an exogenous variable in the ARIMAX model to forecast Nigeria's non-oil exports, while [18] elucidated the influence of both inflation and exchange rates on imports.

This research aims to achieve high forecasting precision by comparing the efficacy of SARIMAX and LSTM models in predicting imports of commodities with HS Code 04. It uniquely integrates exogenous variables such as significant calendar events (Eid al-Fitr, Eid al-Adha, Christmas, Lunar New Year), exchange rates (rupiah vs. US dollar), and inflation rates to enhance model robustness. The study employs RMSE and MAPE to evaluate performance. The SARIMA approach is appropriate for use when importing data of the HS 04 code that contains trends and seasonality. The model's accuracy should increase with the inclusion of exogenous variables. The merits of the LSTM technique, which can capture complicated and long-term patterns as well as non-stationary data, are compared. This dual-model comparison not only identifies the superior model for import forecasting but also provides insights into optimal conditions for each model, advancing the scholarly conversation and aiding commercial and governmental forecasting efforts.

2. RESEARCH METHODS

This research was conducted with 4 main stages which include: 1) Data pre-processing includes splitting the data into 85:15 proportions and normalizing the data; 2) Building the SARIMAX model; 3) Building the LSTM model and; 4) Model prediction and evaluation. Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are the metrics used to evaluate the models. The prediction error is expressed as a percentage of the actual value by MAPE, whereas the average size of the error is measured by RMSE. With smaller values denoting greater accuracy, these metrics are utilized to compare the SARIMAX and LSTM models' performances and identify the most accurate forecasting strategy.

2.1 Data

The dataset used in this study was obtained from Trade Map, Ministry of Trade, Bank Indonesia and Indonesia National Single Window and Central Bureau of Statistics. The dataset contains information on the import quantity of HS04 commodities (Milk, Butter and Eggs) per month measured in tons, from January 2006 to February 2024, along with exogenous variables. Research by [19] demonstrates that while price changes have no direct impact on the volume of imports reported, researchers can more accurately estimate trade benefits when utilizing quantity data as opposed to value data.

Table 1. Variables in The Data

Variable	Description
Import_Quantity_HS04	Import quantity variable of HS04 goods per month in tons.
Date	Variable that indicates monthly time.
Eid al-Fitr	Dummy variable (0,1) indicating the time of Eid al-Fitr in a particular month each year,
Eid al-Adha	Dummy variable (0,1) indicating the time of Eid al-Adha in a particular month each year,

Variable	Description
Lunar New Year	Dummy variable (0,1) indicating the time of Lunar New Year in a particular month each year,
Christmas Day	Dummy variable (0,1) indicating the time of Christmas Day in a particular month each year,
Inflation (%)	Variable of inflation rate per month,
USD to RP	Variable of monthly exchange rate of USD against Rupiah.

2.2 Research Methods

The software used for data analysis in this research is Python 3 and Microsoft Excel 2021. TensorFlow and Keras are essential packages for developing and refining the LSTM model, while statsmodels are used to implement the SARIMAX model. Important resources for scaling and assessing metrics like MSE and MAPE were made available by Sklearn. Numpy, Matplotlib.pyplot, and Pandas were used to handle and visualize the data. Moreover, Boxcox from Scipy and Adfuller from StatsModels. To ensure reliable model performance, statistics were employed for transformation and stationarity testing. In order to predict HS04 commodity imports, these packages made it possible to develop and compare SARIMAX and LSTM models effectively. The data analysis procedures carried out in this study are as follows:

a. Data Pre-processing

This research will compare the Autoregressive Integrated Moving Average with Exogenous input (SARIMAX) and Long Short Term-Memory models to predict the import quantity of HS04 commodities. At this point, the dataset has been progressively divided into training and test data, with a ratio of 85:15 based on time series data.

b. Building the SARIMAX Model

The divided dataset will next undergo an Augmented Dickey-Fuller (ADF) test to check for stationarity in the training data. The Dickey-Fuller test was introduced by [20] to detect the presence of a unit root in an autoregressive model, which became the basis for the development of the ADF test. The data is considered steady in the variance or the mean if the p-value is less than 0.05. The data will be converted if it is not stationary in the variance and will be differenced n times until it is stationary if it is not yet stationary in the mean. Data that has been stationary will be analyzed using decomposition to determine the seasonal pattern that occurs. The SARIMA model consists of a non-seasonal model (p, q) and a seasonal model (P, Q) identified by the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to determine the order (p, d, q) $(P, D, Q)_s$ in SARIMA. The determination of order p and q can be seen in the following Table 2:

Table 2. Order Determination of p, d, q

ACF	PACF	Tentative Model
Cuts-off after lag q	Tails-off	MA (q)
Tails-off	Cuts-off after lag p	AR (p)
Cuts-off after lag q	Cuts-off after lag p	MA(q) or AR(p) Choose the best model
Tails-off	Tails-off	Check all combinations of p and q, and choose the best model.
Tails-off (slowly)		Non-Stationary Model

The next step is to check the significance of the model parameters by considering the p-value of each order. If all the orders are significant, then it can proceed to the diagnostic test of the assumptions of the residuals by conducting a normality test on the residuals and a residuals freedom test. The model is said to be good if it has fulfilled the parameter significance and residuals diagnostic tests. Following the acquisition of the SARIMA model, the process proceeds to construct the SARIMAX model through the integration of exogenous factors into the time series data.

c. Building the LSTM Model

HS04 import quantity dataset will be normalized to change the value range to (0,1) and directly converted into supervised learning data. The LSTM model built contains LSTM with 100 and 200-units and Dense

Layer, neurons, inputs, loss function in the form of mean squared error, optimizer in the form of ADAM and activation functions in the form of Sigmoid and Tanh. The model that has been defined will be hyperparameter tuning using epochs of 100 and 150 with batch sizes of 1 and 32, the learning rate used is 0.0001 and 0.05, GridSearchCV uses the grid search method with k-fold cross-validation to evaluate the model. As an example, [21]'s research used GridSearchCV as a tuning hyperparameter.

d. Model Prediction and Evaluation

The best models from SARIMAX and LSTM are predicted as much as n data on test data to see a comparison of the ability of each model to predict data. Each model will also be evaluated using the smallest Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) to compare the best methods and models in forecasting data.

The flowchart of this research method is illustrated in Figure 1 below.

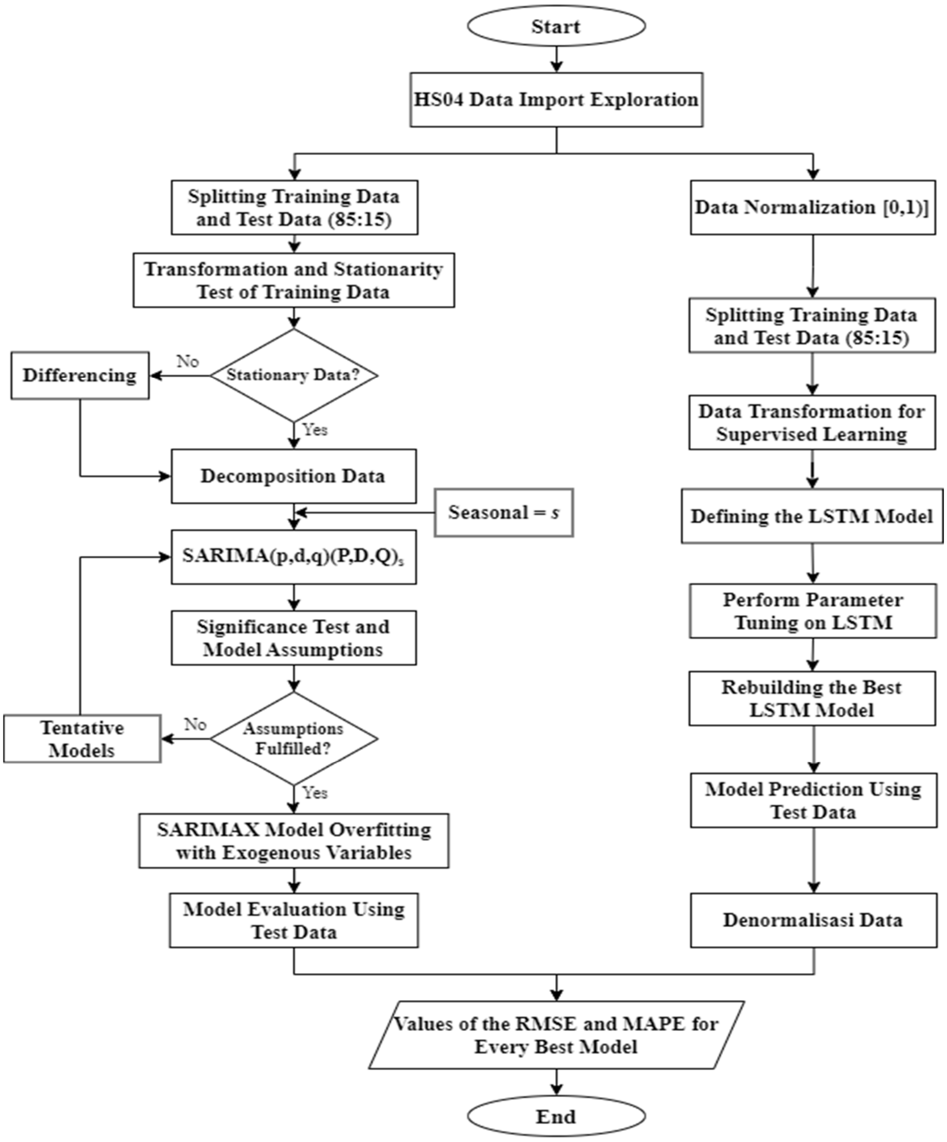


Figure 1. Flowchart of SARIMAX and LSTM Method

2.3 SARIMAX

The Seasonal Autoregressive Integrated Moving Average with Exogenous factors (SARIMAX) model is a development of the SARIMA model that involves exogenous variables that are considered to have an effect on endogenous variables. According to [6], the general form of the SARIMAX model can be written as follows:

$$\boldsymbol{\phi_p(B)\Phi_P(B^S)(1 - B)^d(1 - B^S)^D Z_t = \beta_1 X_{1,t} + \cdots + \beta_k X_{k,t} + \theta_q(B)\Theta_Q(B^S)\varepsilon_t} \tag{1}$$

with $X_{k,t}$ being k exogenous variables at t time with k = 1, 2, 3, ..., k.

2.4 Long Short-Term Memory (LSTM)

The idea of Long Short-Term Memory (LSTM), first proposed by Sepp Hochreiter and Jurgen Schmid Huber in 1997, is an extension of neural networks that controls access to memory cells through a unique gating mechanism. Basically, LSTM is capable of using memory units to store remote information and track various attributes of the text being analyzed [22].

An input gate, output gate, forget gate and a memory cell make up the basic components of the LSTM design. More specifically, LSTM utilizes three gate vectors, namely input gate, forget gate and output Input gate governs the extent to which the input vector can affect the memory vector. Forget gate controls how much of the previous memory vector needs to be forgotten. Meanwhile, the output gate controls how much of the memory vector is stored in the hidden state [9].

LSTM uses two types of activation functions, each with its own role. The sigmoid activation function is tasked with converting an x value into a value between 0 and 1. Meanwhile, the tanh activation function is tasked with converting an x value into a value between -1 and 1. The sigmoid and tanh activation function formulas [23] can be seen in the following equation:

$$\boldsymbol{\sigma = \frac{1}{1+e^{-x}}} \tag{2}$$

$$\boldsymbol{tanh(x) = 2\sigma(2X) - 1} \tag{3}$$

where x is the input data and is the value of the sigmoid activation function.

2.5 Model Evaluation

Performance evaluation of time series models is done to see which model performance is better. Performance evaluation can be seen from the accuracy of the forecast which can be calculated from the forecast residuals, including Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Deviation (MAD). All three can be calculated with the following formula [24].

$$\boldsymbol{RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(e_t)^2}} \tag{4}$$

Description:
$\mathrm{e_t}$ = Error or difference between the actual value and the predicted value at time t.
n = The total number of observations or data in a given period.

$$\boldsymbol{MAPE = \frac{1}{n}\sum_{t=1}^{n}|e_t|} \tag{5}$$

Description:
$\mathrm{e_t}$ = Error or absolute difference between the actual value and the predicted value at time t.
n = The total number of observations or data in a given period.

The model used yields excellent performance results when its MAPE value is less than 10%. The model utilized yields good performance results when the MAPE value is between 10% and 20%. Moreover, a MAPE value within the 20% to 50% range suggests that the model yields performance outcomes that remain practicable for application. When the MAPE value exceeds 50%, it signifies that the model being used performs poorly [25].

3. RESULTS AND DISCUSSION

3.1 Data Exploration

Commodities imported under HS Code 04 between January 2006 and February 2024 exhibit a range of trends and variations. The plot of data is displayed in **Figure 2** below.

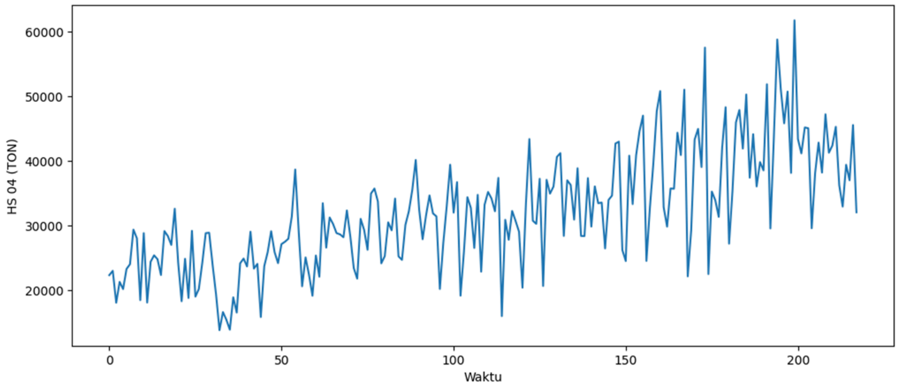


Figure 2. Time Series Plot of Import Commodities, HS Code 04

The time series plot **Figure 2** shows the volume of HS04 products imported from January 2006 to February 2024. Although there are fluctuations, the overall trend shows a general increase in import volumes. The increase over the Y-year period is approximately 43.57%. Significant import peaks are observed around the Lunar New Year, further suggesting the influence of calendar variations. The impact of calendar-related factors on import volumes is reinforced by the presence of a seasonal index, which is confirmed by the subsequent data decomposition **Figure 3**. Furthermore, the plot shows a non-constant variance and a distribution not centered around the mean, which is an indication of non-stationarity. This non-stationarity has important implications for forecasting, as it requires methods such as differencing or the use of models specifically designed for non-stationary data, such as the SARIMAX model used in this study.

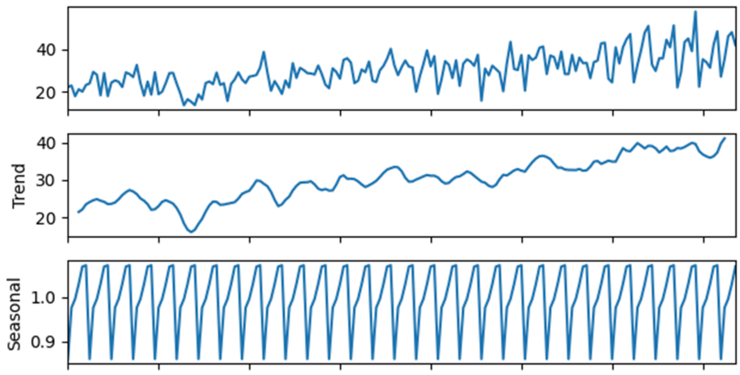


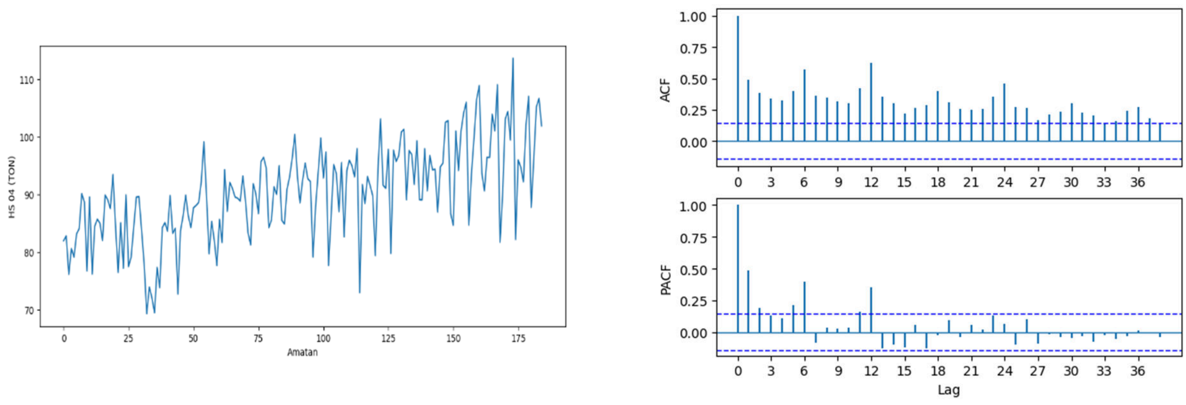
Figure 3. Decomposition of Import Commodity Data, HS Code 04

The decomposition of the HS04 commodity import time series in **Figure 3** illustrates the underlying structure of the data, comprising trend, seasonality, and residual components. Overall import growth is indicated by a generally rising trend, despite short-term fluctuations. A pronounced six-month cyclical pattern in the seasonal component is consistent with potential influences such as Indonesia's agricultural cycle and pre-Lunar New Year demand surges. This supports the inclusion of the Lunar New Year as an exogenous variable in subsequent SARIMAX modelling. Although not shown, analysis of the residual component—which represents unexplained variation—is critical to model validation. Any apparent patterns within the residuals require further examination, possibly by ACF and PACF plotting, to ensure model appropriateness and to inform refinement. This decomposition provides a robust basis for understanding the characteristics of the data and justifies the choice of modelling strategy.

3.2 SARIMAX Model

3.2.1 Data-Stationary

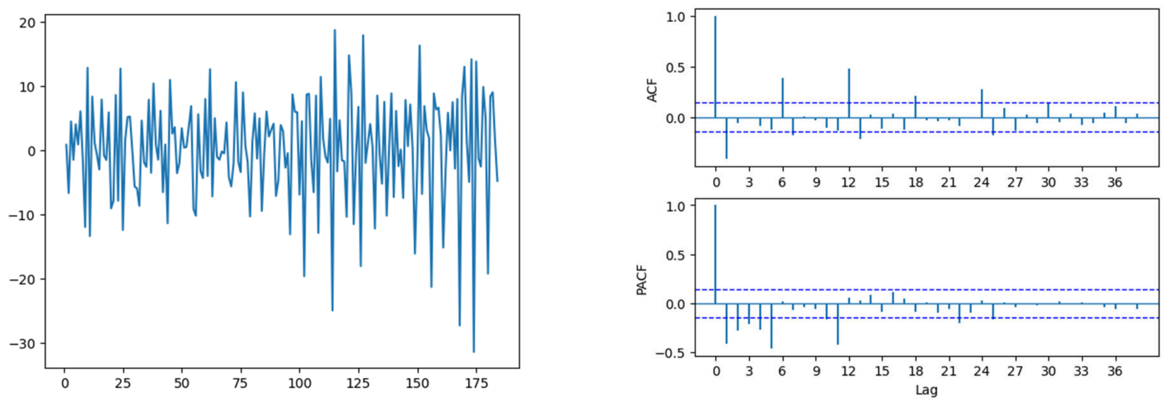
Plots of the partial autocorrelation function (PACF) and autocorrelation function (ACF) can be used to assess stationarity. Training data and test data were the two categories into which the data was previously separated. The test data is 15% from June 2021 to February 2024, and the training data is 85% from January 2006 to May 2021. A non-stationary variation treatment will subsequently be applied to the training data.



(a) **(b)**

Figure 4. (a) Plot of Transformed Training Data and (b) Plot of ACF PACF

As seen in **Figure 4** (a), transformation has been used to handle the non-stationary data in the variance in order to keep the window of values that occurs from being overly large. The import quantity of commodity HS04 is not stationary in the mean, as **Figure 4** (b) shows. The p-value of 0.9051, which is higher than the significance level ($\alpha = 0.05$), indicates that the Augmented Dickey-Fuller (ADF) test also demonstrates that this time series is not stationary at the mean. This suggests that there has been some variation in the mean and variance of the HS04 commodity import quantity over the period. To make a time series stationary, it is necessary to differentiate those that show signs of non-stationarity in terms of variety and mean value. There is only one differentiation ($d = 1$). **Figure 5** shows the plot of the stationary training data.



(a) **(b)**

Figure 5. (a) Plot of Training Data (d = 1) and (b) Plot of ACF PACF (d = 1)

Figure 5 illustrates the stationarity of the data, which is supported by the results of the ADF test, which show a p-value of 1.303e-08, which is less than the significance level ($\alpha = 0.05$).

3.2.2 Model Identification

The SARIMAX model can be used to model import quantity data because of the seasonal patterns in the data. The SARIMAX model is found using the ACF and PACF plots of the stationary data. In the non-seasonal model, all of the lags in the ACF and PACF displays are taken into account when calculating orders p and q. Orders P and Q are determined by taking into account the delays 6, 12, 18, and so on in the ACF and

PACF plots because the seasonal period in the seasonal model is every six months. The seasonal model on the ACF plot cut-off is at lag 4, the non-seasonal model's lag is at lag 1, and the PACF plot cut-off is at lag 5, as shown in **Figure 4**. Based on this, there are several SARIMAX models. It is discovered that the exogenous variable "Chinese New Year Day" has a significant impact on the suggested SARIMAX model through feature selection. A diagnostic test is performed on the tentative SARIMAX model to determine which model is appropriate for forecasting. To determine which SARIMAX model is best suited for forecasting, diagnostic tests are run on the tentative model.

Table 3. Tentative Model Comparison

SARIMAX Model	Parameter Significance	Diagnostic Test			Model Evaluation	
		Normality	White-Noise	Heteroskedasticity	MAPE	RMSE
$(1,1,0)(1,0,0)_6$	Achieved	Not Achieved	Achieved	Achieved	15.62	7532.94
$(0,1,1)(1,0,0)_6$	Achieved	Achieved	Achieved	Achieved	13.16	7547.89
$(1,1,1)(1,0,0)_6$	Not Achieved	Achieved	Achieved	Achieved	12.93	7412.60
$(2,1,0)(1,0,0)_6$	Achieved	Not Achieved	Achieved	Achieved	18.47	8586.27
$(2,1,1)(1,0,0)_6$	Not Achieved	Achieved	Achieved	Achieved	12.88	6832.38
$(3,1,0)(1,0,0)_6$	Not Achieved	Achieved	Achieved	Achieved	21.62	9653.42
$(3,1,1)(1,0,0)_6$	Not Achieved	Achieved	Achieved	Achieved	12.97	6983.78
$(4,1,0)(1,0,0)_6$	Not Achieved	Achieved	Achieved	Achieved	19.52	8856.08
$(4,1,1)(1,0,0)_6$	Not Achieved	Achieved	Achieved	Achieved	12.87	6816.28
$(5,1,0)(1,0,0)_6$	Achieved	Achieved	Not Achieved	Achieved	12.10	6521.25
$(5,1,1)(1,0,0)_6$	Not Achieved	Achieved	Achieved	Achieved	12.86	6815.27

Table 3 shows that SARIMAX $(0,1,1)(1,0,0)_6$ is the training data model that satisfies parameter significance, diagnostic assumptions, and the best model evaluation value. Other tentative SARIMAX models, on the other hand, are still insufficient because they fail to meet certain diagnostic assumptions and are not real levels. The SARIMAX $(0,1,1)(1,0,0)_6$ equation for the training data model is as follows.

$$(1-0.2899B^6)(1-B)Y_t = (1-0.9102B)\varepsilon_t - 2.0208X_{imlek}$$
$$Y_t = Y_{t-1} - 0.9102\varepsilon_{t-1} + 0.2899Y_{t-6} - 0.2899Y_{t-7} - 2.0208X_{imlek} + \varepsilon_t$$

3.3 LSTM Model

HS04 commodity import quantity data must be normalized to a scale of [0, 1] before modelling using Long Short-Term Memory (LSTM). This normalization is necessary so that the range of the data is not too wide and there is no interference with the modelling process. In the model training, there are several parameters that affect the final outcome. **Table 4** shows the LSTM architecture for the estimation of the import volume of HS04 commodities.

Table 4. LSTM Model Specification

Characteristic	Specification
Architecture	3 *Layer* LSTM with *neuron* and 1 *Layer Dense*
Optimizer	ADAM (*Adaptive Momentum*)
Loss Function	*Mean Squared Error*
Batch size	1, 32
Epoch	100, 150
Neuron	100, 200
Units	100, 200
Learning Rate	0.05, 0.0001

The use of 3 LSTM layers, 1 dense layer, neurons [100, 200], and epochs [100, 150] has no specific reference in each LSTM model. The ideal number of neuron units needs to be determined to model the time series using training data. The LSTM model, or the optimal model obtained from the training data, is evaluated as the optimal model of each combination of neurons, epochs, batch size of 32, and learning rate of 0.0001 to predict the training data.

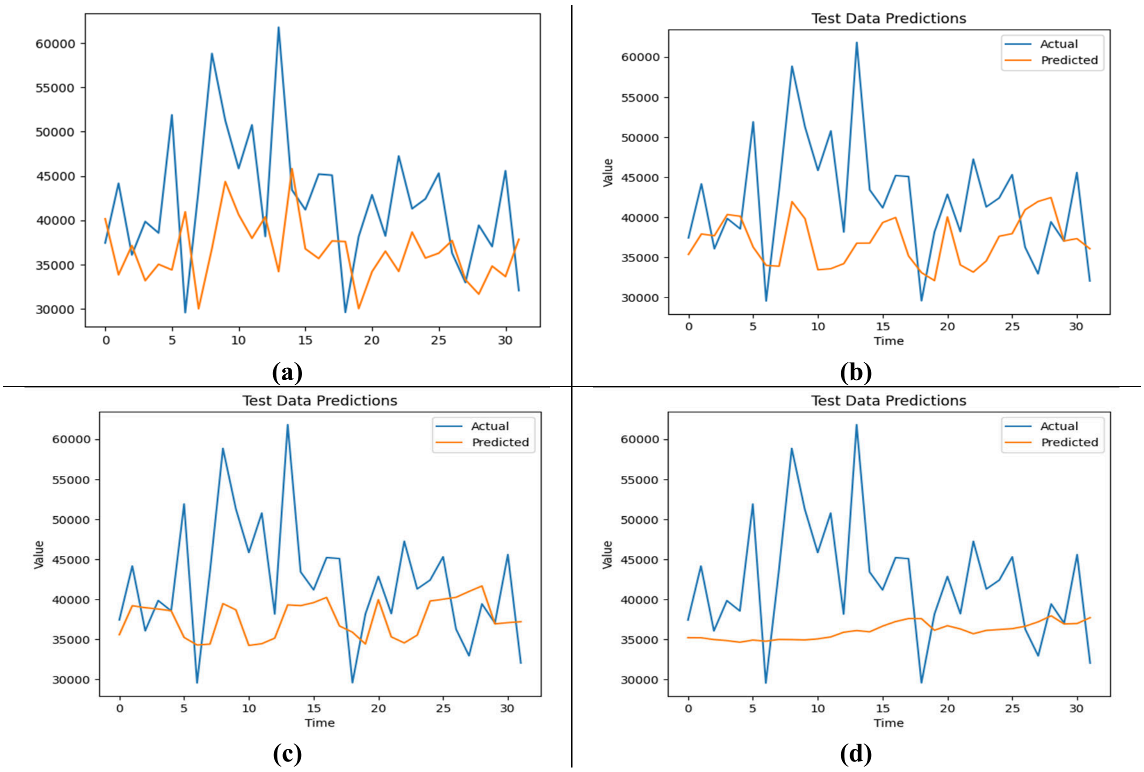


Figure 6. Model LSTM 1 (a); Model LSTM 2 (b); Model LSTM 3) (c); Model LSTM 4 (d)

Figure 6 shows the performance of four different LSTM models. Each was trained with a different combination of hyperparameters. In order to select the optimal model for the prediction of the test data, a comparative analysis was performed using two evaluation metrics commonly used in predictive analysis, namely mean absolute percentage error (MAPE) and root mean square error (RMSE). MAPE measures the mean absolute percentage error between predicted and actual values, while RMSE measures how much the model prediction differs from actual by squaring the error. Both metrics provide a comprehensive view of how well the model is able to generalize to the unseen data (test data). The results of this comparative analysis can be seen in Table 5. It shows the MAPE and RMSE values of each model based on the training data used.

Table 5. LSTM Model Comparison

Neuron	Epoch	Units	Model Evaluation	
			MAPE	RMSE
100	100	100	17.89	10018.01
	150	150	15.95	9148.43
200	100	100	14.89	8787.73
	150	150	16.74	9813.82

The results of this analysis show that the LSTM model (c), trained with a batch size configuration of 32, 100 epochs, and 100 neuron units, provides the best performance in predicting the test data. MAPE and RMSE were 14.89% and 8787.73, respectively, reflecting very good predictive accuracy. This result confirms that the LSTM(c) model has superior capabilities compared to the other models. Furthermore, Figure 6 manages to visualize the performance of each model by comparing the evaluation metrics used (MAPE and RMSE).

3.4 Best Model Selection

Based on forecasting results on out-of-sample data (training data), namely test data from the HS04 import quantity variable during the period of June 2021 – February 2024, a comparison of the prediction performance of the SARIMAX and LSTM models can be observed.

Table 6. Best Model Comparison

Model	Model Evaluation	
	MAPE	RMSE
SARIMAX $(0,1,1)\ (1,0,0)_6$	13.16	7547.89

Model	Model Evaluation	
	MAPE	RMSE
LSTM (*epoch 100*, *neuron* 100, *batch size* 32)	14.89	8787.73

The SARIMAX (0,1,1) $(1,0,0)_6$ model yields the highest accurate model evaluation value for HS04 commodity import amount data prediction, as **Table 6.** shows. The results of this study also show that the SARIMAX model with calendar variation effect variables is suitable for predicting comparable events because other data may have calendar variation effects. SARIMAX statistically uses exogenous variables to adequately manage seasonality and periodic shifts, demonstrating resilience in datasets with significant temporal patterns. However, when working with data that displays non-linear dependencies, its linear nature limits its performance. The LSTM model, in contrast, statistically captures complex, non-linear relationships in the time series data by incorporating a more complex, multi-layered network structure. LSTM's adaptability in learning from non-linear patterns makes it appropriate for data with complex variable relationships, even though this raises the computational demand. Therefore, the statistical trade-off is between the depth of pattern recognition of LSTM and the efficiency and interpretability of SARIMAX, with the choice of model relying on the linearity and complexity of the time series data.

4. CONCLUSIONS

This study compared the performance of SARIMAX and LSTM models in predicting monthly import volumes of HS04 products (milk, eggs, and butter) in Indonesia. The SARIMAX (0,1,1) $(1,0,0)_6$ model, which includes exogenous variables such as inflation, the exchange rate between the USD and the IDR, and major Indonesian holidays, showed a superior forecasting accuracy. In particular, the inclusion of the lunar new year as an exogenous variable significantly improved the performance of the SARIMAX model. It effectively captured the observed import fluctuations around this holiday. SARIMAX achieved a lower RMSE of 7547.89 and MAPE of 13.16 compared to LSTM, which achieved a RMSE of 8787.73 and MAPE of 14.89 when evaluated on a test set of 15% of data (January 2006 to February 2024). While the LSTM model has the potential to capture non-linear relationships, its performance in this context, characterized by strong seasonality and exogenous influences, was outperformed by the SARIMAX model. This suggests that SARIMAX may be more suitable for forecasting HS04 commodity imports in Indonesia, given the specific data characteristics. However, it is important to note that this study focused only on Indonesian import data.

Future research could explore different ways of improving the forecast of HS04 commodity imports. The development of statistical techniques specifically tailored to capture the unique characteristics of import data, such as intermittent demand and supply chain disruptions, could yield significant improvements. The exploration of the potential of stacking models, which combine the strengths of machine learning models such as LSTM with the robust statistical foundation of SARIMAX, is a promising direction. Specifically, the exploration of ensemble methods that incorporate both linear (SARIMAX) and non-linear (LSTM) components could exploit the ability of LSTM to capture complex relationships while maintaining the interpretability and performance of SARIMAX in the presence of strong seasonality and exogenous factors. This could lead to more accurate and robust forecasts. In addition, future research will need to examine how other exogenous variables, including government trade policy, global commodity prices, and climate patterns, might be included to improve modelling.

ACKNOWLEDGMENT

We are grateful to the BPS Statistics Indonesia, Bank Indonesia, Ministry of Trade, Indonesia National Single Window, and the Department of Statistics and Data Science IPB University for making this study possible.

REFERENCES

[1] J. Heizer and B. Render, *Manajemen Operasi*, 9th ed., vol. 1. Jakarta: Salemba Empat, 2009.

[2] S. K. Dwi Arini, F. M. Afendi, and P. Silvianti, "Penanganan Pencilan pada Peramalan Data Deret Waktu Menggunakan Metode Pemulusan Holt dan Robust Holt," *Xplore: Journal of Statistics*, vol. 10, no. 2, pp. 112–128, May 2021, doi: 10.29244/xplore.v10i2.205.

[3] T. Hastie, R. Tibshirani, and J. Friedman, *Springer series in statistics the elements of statistical learning*. New York: Springer, 2009.

[4] G. T. Meilania, A. V. Septiani, E. Erianti, K. A. Notodiputro, and Y. Angraini, "Pemodelan ARIMA-GARCH dalam Peramalan Kurs Rupiah Terhadap Yen dengan Masalah Keheterogenan Ragam," *Ekonomis: Journal of Economics and Business*, vol. 8, no. 1, p. 165, Mar. 2024, doi: 10.33087/ekonomis.v8i1.1294.

[5] V. P. Ariyanti and Tristyanti Yusnitasari, "Comparison of ARIMA and SARIMA for Forecasting Crude Oil Prices," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 2, pp. 405–413, Mar. 2023, doi: 10.29207/resti.v7i2.4895.

[6] C. J. Pangestu, E. Piantari, and M. Munir, "Prediction of diarrhea sufferers in Bandung with Seasonal Autoregressive Integrated Moving Average (SARIMA)," *Journal of Computers for Society*, vol. 1, no. 1, pp. 61–79, Jun. 2020.

[7] A. Ibrahim and A. O. Musa, "ON THE PERFORMANCE OF SARIMA AND SARIMAX MODEL IN FORECASTING MONTHLY AVERAGE RAINFALL IN KOGI STATE, NIGERIA," *FUDMA JOURNAL OF SCIENCES*, vol. 7, no. 6, pp. 24–31, Dec. 2023, doi: 10.33003/fjs-2023-0706-2095.

[8] S. Zahara, Sugianto, and M. Bahril Ilmiddafiq, "Prediksi Indeks Harga Konsumen Menggunakan Metode Long Short Term Memory (LSTM) Berbasis Cloud Computing," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 3, no. 3, pp. 357–363, Dec. 2019, doi: 10.29207/resti.v3i3.1086.

[9] N. Kalchbrenner, I. Danihelka, and A. Graves, "Grid Long Short-Term Memory," Jul. 2015.

[10] A. P. Ratnasari, B. Susetyo, and K. A. Notodiputro, "COMPARISON OF DOUBLE RANDOM FOREST AND LONG SHORT-TERM MEMORY METHODS FOR ANALYZING ECONOMIC INDICATOR DATA," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 17, no. 2, pp. 0757–0766, Jun. 2023, doi: 10.30598/barekengvol17iss2pp0757-0766.

[11] *Penetapan Sistem Klasifikasi Barang dan Pembebanan Tarif Bea Masuk atas Barang Impor*. Indonesia: BN 2022 (316), 2022.

[12] World Bank, "Indonesia Trade Statistics," World Integrated Trade Solution (WITS).

[13] International Trade Administration, "Indonesia Country Commercial Guides," Country Commercial Guides - Trade.gov.

[14] Oosga, "Trade Policy Overview of Indonesia in 2023," Import & Export in Indonesia (2023): Trade Policies & Regulation.

[15] S. A. Rodia and M. A. Ridla, "Prediksi Nilai Impor Susu Segar Menggunakan Algoritma Backpropagation," *Ahsana: Jurnal Penelitian dan Pengabdian kepada Masyarakat*, vol. 2, no. 1, pp. 27–32, 2024.

[16] M. H. Lee and Suhartono, "Calendar variation model based on ARIMAX for forecasting sales data with Ramadhan effect," *Proceedings of the Regional Conference on Statistical Sciences*, vol. 10, no. 2010, pp. 349–361, Jun. 2010.

[17] U. A. Victor-Edema and I. D. Essi, "Autoregressive integrated moving average with exogenous variable (ARIMAX) model for Nigerian non-oil export," *European Journal of Business and Management*, vol. 8, no. 36, pp. 29–34, 2016.

[18] A. P. Sonia and N. D. Setiawina, "Pengaruh Kurs, Jub Dan Tingkat Inflasi Terhadap Ekspor, Impor Dan Cadangan Devisa Indonesia," *E-Jurnal Ekonomi Pembangunan Universitas Udayana*, vol. 5, no. 10, pp. 1077–1102, Oct. 2016.

[19] S. M. Ferguson and A. Smith, "Import demand elasticities based on quantity data: Theory and evidence," *Canadian Journal of Economics/Revue canadienne d'économique*, vol. 55, no. 2, pp. 1027–1056, May 2022, doi: 10.1111/caje.12596.

[20] D. A. Dickey and W. A. Fuller, "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *J Am Stat Assoc*, vol. 74, no. 366a, pp. 427–431, Jun. 1979, doi: 10.1080/01621459.1979.10482531.

[21] H. Darmawan, M. Yuliana, and Moch. Z. Samsono Hadi, "GRU and XGBoost Performance with Hyperparameter Tuning Using GridSearchCV and Bayesian Optimization on an IoT-Based Weather Prediction System," *Int J Adv Sci Eng Inf Technol*, vol. 13, no. 3, pp. 851–862, Jun. 2023, doi: 10.18517/ijaseit.13.3.18377.

[22] A. Karpathy, J. Johnson, and L. Fei-Fei, "Visualizing and Understanding Recurrent Networks," Jun. 2015.

[23] B. Karlik and A. V. Olgac, "Performance analysis of various activation functions on generalized MLP architectures of neural network," *International Journal of Artificial Intelligent and Expert System (IJAE)*, vol. 1, no. 4, pp. 111–122, Feb. 2011.

[24] W. Zha *et al.*, "Forecasting monthly gas field production based on the CNN-LSTM model," *Energy*, vol. 260, p. 124889, Dec. 2022, doi: 10.1016/j.energy.2022.124889.

[25] J. J. M. Moreno, A. P. Pol, A. S. Abad, and B. C. Blasco, "Using the R-MAPE index as a resistant measure of forecast accuracy," *Psicothema*, vol. 25, no. 4, pp. 500–506, 2013.