



# Forecasting with ARIMA and LSTM in Bali Retail Industry

Wayan Gede Suka Parwita<sup>1</sup>, Putu Eka Suryadana<sup>2</sup>,  
and Gede Sugita Aryandana<sup>3</sup>

<sup>1,2,3</sup> Administration Business Department, Politeknik Negeri Bali, Bali, Indonesia  
gede.suka@pnb.ac.id

**Abstract.** Food waste significantly contributes to global greenhouse gas emissions, water and air pollution, and socio-economic issues, including in Bali Province. In Bali, increased tourism correlates with population growth and a rise in organic waste, reaching up to 59.10% of total waste. This is largely dominated by imported fruits such as apples, oranges, pears, and grapes, which are also used for religious ceremonies. To mitigate imported fruit waste, limiting imports based on accurate demand forecasting is crucial. Traditional methods like Autoregressive Integrated Moving Average (ARIMA) exhibit limitations in modeling complex non-linear and seasonal patterns, particularly given the fluctuating apple consumption in Bali. Conversely, the Long Short-Term Memory (LSTM) algorithm, a deep learning approach, demonstrates superior capability in capturing linear, non-linear, and dynamic data patterns. This research compares the performance of ARIMA and LSTM models for predicting apple transaction data, evaluating them using the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The aim is to facilitate more precise apple import projections, thereby reducing waste. The results of the research confirmed that the ARIMA method was found to be superior in computational efficiency compared to LSTM, as it required no repeated training iterations. In contrast, LSTM required 25 iterations, making the computation time longer. However, LSTM showed higher accuracy with a MAPE value of 56.58% and RMSE of 28.93, though the difference in RMSE was insignificant compared to ARIMA (28.84), as both algorithms failed to capture the trend of sales spikes during holidays.

**Keywords:** ARIMA, Bali Retail Industry, Forecasting, LSTM

## 1 Introduction

Food waste has globally brought a significant impact on greenhouse emissions (Urugo et al., 2024). As released by the Food and Agriculture Organization (FAO), food waste accounts for about one-third of total food production worldwide, or equivalent to 1.3 billion tonnes of food. In addition to increasing greenhouse emissions, food waste also has some impacts on water cleanliness and air pollution. Moreover, it takes up a large amount of land. Food waste also impacts local economies and social issues. The cost of producing, transporting, and disposing of wasted food is an additional burden for

© The Author(s) 2025

A. A. N. G. Saptika et al. (eds.), *Proceedings of the International Conference on Sustainable Green Tourism Applied Science - Engineering Applied Science 2025 (ICOSTAS-EAS 2025)*, Advances in Engineering Research 280,

[https://doi.org/10.2991/978-94-6463-878-3\\_22](https://doi.org/10.2991/978-94-6463-878-3_22)

both the government and society in Bali Province, without exception (Lipinski et al., 2013).

Bali Province is one of the provinces with the largest regional income from tourism. Increasing tourism growth rate in one of Bali's regencies is associated with higher population growth, thereby producing a greater waste level (Ratnawati et al., 2022). This can also not be apart from the increase in the amount of organic waste to 59.10% of the total waste produced (Widyarsana I Made Wahyu et al., 2020). Fruits dominate organic waste in Bali Province. The study by Maheswari et al. (2022) showed that the widely circulated fruits in Bali's traditional markets include Fuji apple, orange, pear, and grape, which are imported from abroad. In addition to consumption, these fruits are commonly used in religious activities.

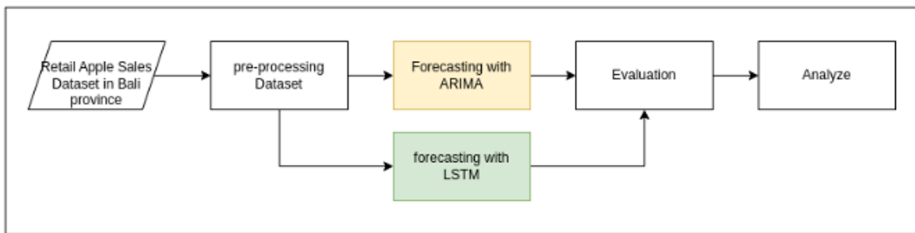
How to manage food waste in the retail industry. One way to manage organic waste from imported fruits is to limit their import in consideration of their needs. Forecasting in waste management is critical in predicting the demand for fruits, especially apples, purposely to determine the number of imported fruits to reduce organic waste (Muleta, 2021). A time series model is one common forecasting method using a statistical approach, such as Autoregressive Integrated Moving Average (ARIMA), which is proven to be accurate in projecting the number of forecasts for a trend or pattern in the future (Muleta, 2021). The Autoregressive Integrated Moving Average (ARIMA) model is a classic method for forecasting based upon a statistical approach. However, it has some limitations in modelling non-linear relationships and complex seasonal patterns.

In Bali Province, the apple consumption level fluctuates significantly, making it challenging to forecast seasonal consumption patterns. In this province, apple fruit is not only used for consumption but also as a means of (Wiwin, 2021). This then becomes a weakness of the ARIMA model in forecasting apple consumption data, which exhibits complex seasonal patterns. In contrast, one of the deep learning algorithms is capable of capturing the linear patterns and complex seasonal patterns (Youness & Driss, 2022). The Long Short-Term Memory (LSTM) algorithm can learn both non-linear data patterns and dynamic period data patterns (Chen et al., 2023). The Long Short-Term Memory (LSTM) algorithm is also able to capture the long-term dynamics and time dependence of non-stationary data. Therefore, theoretically, the LSTM algorithm has more advantages compared to ARIMA in forecasting time series-based data (Li et al., 2023).

This study aims to compare two forecasting algorithms, ARIMA and LSTM, for forecasting Apple transaction data in the form of a time series. The selection of these two algorithms was based on their respective advantages in modelling data with linear and non-linear characteristics. Once the forecasting data were obtained, they forecasting data were then evaluated using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The forecasting results are expected to help project a more targeted number of apple imports. Thus, the number of apples entering Bali Province can be adjusted to the actual needs of the community. This effort is expected to contribute to reducing the production of organic waste in view of the disproportionate stocks of unused apples.

## 2 Methodology

This research, as illustrated in Figure 1, consisted of five stages in comparing two algorithms to obtain forecasting data. The first stage began by obtaining the dataset of the Apple retail sales in Bali Province. The dataset then underwent a pre-processing stage to confirm data quality, including handling missing data, normalization, and transformation into a time series format. Subsequently, the data were predicted using two methods: ARIMA as a classical statistical model and LSTM as a deep learning-based model. The prediction results of both models were then evaluated using metrics such as RMSE and MAPE, and analysed to determine the best model to project the optimal number of apple imports to reduce the level of organic waste in Bali.



**Figure 1.** The Graph of Actual Data and Forecasting Using the Algorithms of ARIMA and LSTM

### 2.1 Dataset

The dataset used in this research referred to the data of organic food sales at retail stores in Bali Province. The data contained a history of daily sales transactions for all products, recorded from November 2015 to December 2017. The information available in the dataset included transaction date, product name, number of units sold, unit price, product category, and total sales value. The dataset was then used as the basis for building a time series forecasting model to more accurately predict the demand for organic food imports.

### 2.2 Data Pre-processing

The data used in this study were obtained from Apple sales data from retail stores in Bali Province, with a time span from November 2015 to December 2017. The data pre-processing stage began by selecting the columns used for forecasting, which consisted of 3 columns: transaction time, item name, and quantity of goods sold. After selecting three columns and eliminating unnecessary ones, pre-processing continued by selecting products related only to apples. The keyword Apple was used to select transaction rows selling Apples and the quantity of Apples sold. As forecasting was done using time series, it then required a history of the number of daily Apple transactions. If there were no transactions on a particular day, then the quantity data column would be replaced with 0. The data pre-processing resulted in a dataset with 3 columns with a total of 791 rows.

### 2.3 LSTM (Long Short Term Memory)

LSTM is a deep learning-based algorithm that can be used for data forecasting. The Long Short-Term Memory (LSTM) method was selected in this study due to its ability to selectively retain both long-term and short-term information, enabling it to capture complex patterns in apple sales data in the Province of Bali. The gate structure in LSTM allows the model to determine which information is relevant to store or discard, improving prediction accuracy for demand fluctuations influenced by seasonal factors and trends (Ilu & Prasad, 2023). As demonstrated in previous studies, LSTM can be combined with preprocessing techniques such as wavelet transforms to enhance forecasting performance on complex time-series data.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

The state update process in LSTM is carried out through two critical stages to determine which information needs to be retained and updated. The first stage involves using the sigmoid function to illuminate and select relevant information from the current input() and the previous hidden state. The sigmoid() function produces a value between 0 and 1, which acts as a gate to control how much new information will be initiated into the state.

$$C_t = f_t \odot C_{t-1} + i_t \odot C_t \quad (2)$$

The Output Gate in LSTM in the research on Apple sales forecasting in Bali Province functions to control what information from the cell state will be output as a hidden state (Li & Zhang, 2023), so that only relevant patterns such as seasonal trends or holiday spikes are used for prediction. With this mechanism, the model can selectively utilize long-term memory (e.g., the effect of the harvest season) while ignoring noise, thereby increasing the accuracy of monthly sales forecasting (Kenaan et al., 2024). The Output Gate in LSTM in the research on apple sales forecasting in Bali Province functions to control what information from the cell state will be output as a hidden state (Li & Zhang, 2023), so that only relevant patterns such as seasonal trends or holiday spikes are used for prediction. With this mechanism, the model can selectively utilize long-term memory (e.g., the effect of the harvest season) while ignoring noise, thereby increasing the accuracy of monthly sales forecasting.

### 2.4 ARIMA (Autoregressive Integrated Moving Averages)

ARIMA (Autoregressive Integrated Moving Averages) is a statistical-based algorithm that can be used for forecasting time series data. The ARIMA (Autoregressive Integrated Moving Average) model was chosen in this study as a statistical approach to forecast apple sales in the retail industry in Bali Province because of its ability to capture linear patterns in time series data. This model combines three main components: autoregressive (AR) to utilize historical data relationships, differentiation (I) to achieve data stationarity, and moving average (MA) to model random errors (Syahrini et al., 2023).

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right)(1 - L)^d X_t = \left(1 + \sum_{j=1}^q \theta_j L^j\right) \epsilon_t \quad (3)$$

This study conducts a forecasting analysis of apple sales trends in Bali Province, which exhibit seasonal patterns. The ARIMA algorithm is employed by modeling various combinations of the  $p$ ,  $d$ , and  $q$  parameters to determine the most suitable configuration. Accordingly, ARIMA provides a mathematical framework for predicting time series data characterized by historical patterns and errors, as demonstrated in Equation 3.

## 2.5 Evaluation

The results of apple fruit forecasting in the retail industry in Bali province were in the form of time series data. This research employed MAPE and RMSE to assess the accuracy of apple sales forecasting in the retail industry of Bali Province. MAPE was utilized to measure relative error in percentage, facilitating stakeholder interpretation. In contrast, RMSE was used to provide an absolute picture of the description of the deviation of the prediction from the actual value. In addition, evaluation using MAPE and RMSE also enabled to provide of complementary insights in assessing forecasting accuracy (Kramar & Alchakov, 2023).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \times 100\% \quad (4)$$

The MAPE method was selected for evaluating the forecasting of apple sales data in the retail industry due to its ability to compare time scales. Specifically, this study utilized a daily summary of apple sales as the time scale. The method to measure the error in percentage using the MAPE method can be seen in Equation 4. The MAPE method was also found useful to compare the performance of forecasting results on Apple sales summary datasets using the ARIMA and LSTM methods (Kenaan et al., 2024).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (A_i - F_i)^2} \quad (5)$$

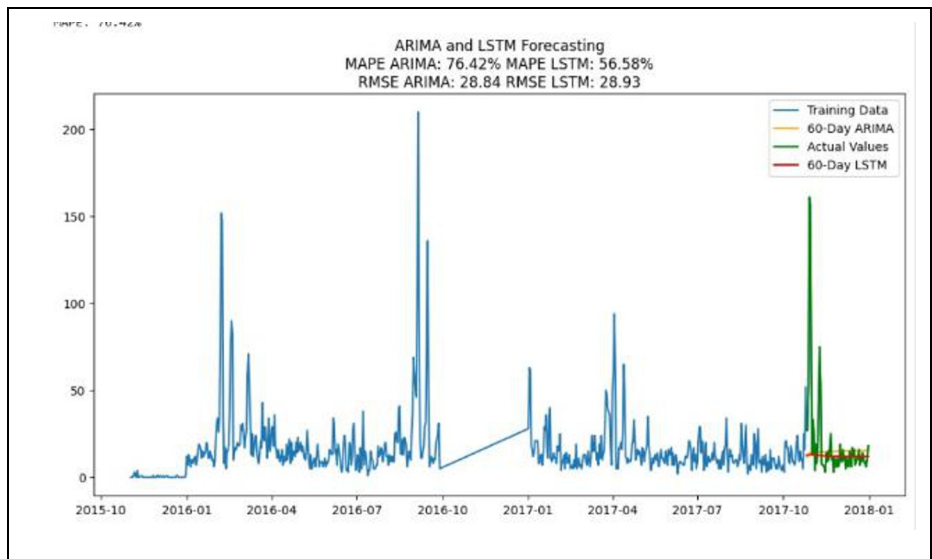
In addition to evaluation using the MAPE method, this research used the RMSE method to obtain insight from the results of forecasting apple fruit sales data in the retail industry in Bali. RMSE was used to measure the error level in the units of data for apple fruit forecasting results in Bali Province by giving greater weight to extreme errors (Ghareeb et al., 2023). It was also used to identify the average deviation of apple fruit forecasting results in each method. The formula of RMSE is presented in Equation (5). The larger error value indicated that the forecasting data showed a mismatch with the original data.

In addition to evaluation using the MAPE method, this research used the RMSE method to obtain insight from the results of forecasting apple fruit sales data in the retail industry in Bali. RMSE was used to measure the error level in the units of data for apple fruit forecasting results in Bali Province by giving greater weight to extreme errors. It was also used to identify the average deviation of apple fruit forecasting results in each method. The formula of RMSE is presented in Equation (5). The larger error value indicated that the forecasting data showed a mismatch with the original data.

### 3 Results and Discussion

#### 3.1 Result

In this experiment, the researcher compared two different algorithms for forecasting apple sales data in the retail industry in Bali Province. The computation for each method showed different results. In the ARIMA method, the computation took 6.2 seconds, while in the LSTM method it took 371.5 seconds.



**Figure 2.** The Graph of Actual Data and Forecasting Using the Algorithms of ARIMA and LSTM

As illustrated in Figure 2, blue lines refer to training data, green lines show the actual data for the last 60 days, orange lines refer to the prediction data, and red lines are the LSTM prediction data. The forecasting data consisted of Apple sales data from the last 60 days in Bali Province.

**Table 1.** Comparison of RMSE and MAPE Between the ARIMA Method and LSTM Method

Method	RMSE	MAPE
ARIMA	28.84	76.42%
LSTM	28.93	56.58%

Algorithm evaluation was carried out to predict the results of Apple sales forecasting data in Bali Province for the last 60 days. To evaluate the forecasting results in the ARIMA and LSTM methods, MAPE and RMSE were used. Table 1 presents the comparison of MAPE and RMSE between ARIMA and LSTM algorithms.

### 3.2 Discussion

Based on the computational results, the ARIMA method had an advantage over the LSTM one. This was related to the LSTM algorithm requiring several iterations to train data based on the dataset used. The iteration of the LSTM algorithm performed in this study was 25 times. This, consequently, made the computation time of the LSTM algorithm longer.

In contrast to computation time, the evaluation results of the LSTM algorithm using the MAPE and RMSE methods showed better accuracy compared to ARIMA. The LSTM model achieved a MAPE of 56.58%, significantly lower than ARIMA's 76.42%, indicating its superior ability to minimize percentage errors in forecasting apple sales. However, while LSTM outperformed ARIMA in MAPE, the RMSE values were nearly identical, 28.93 for LSTM and 28.84 for ARIMA, suggesting that both models struggled similarly with large deviations in predictions. This similarity in RMSE highlights a key limitation of both methods: neither fully captures extreme fluctuations, such as sudden spikes in apple demand during Bali's holiday seasons. Despite LSTM's advantage in handling complex patterns, its computational cost and data requirements remain a drawback compared to ARIMA's simplicity and efficiency with smaller datasets.

Figure 2 illustrates that both ARIMA and LSTM produce nearly identical forecasting trends, which deviate significantly from the actual data, particularly during peak transaction periods. This limitation stems from ARIMA's inability to model non-linear trends and LSTM's potential underperformance when trained on insufficient or noisy data. While ARIMA's linear assumptions make it interpretable and easy to implement, it fails to adapt to sudden market changes. On the other hand, LSTM, despite its deep learning capabilities, may still miss abrupt demand surges if not properly tuned or if external factors (e.g., holidays) are not explicitly incorporated into the model. The similar RMSE scores further emphasize that both methods require enhancements, such as hybrid modeling or additional feature engineering, to better predict volatile trends in apple sales.

To improve forecasting accuracy, future research should focus on refining the LSTM algorithm by integrating external variables (e.g., holiday calendars, weather data) or adopting hybrid approaches that combine ARIMA's strengths in linear patterns with LSTM's ability to learn complex non-linear relationships. Additionally, leveraging advanced deep learning techniques, such as attention mechanisms or transformer-based

models, could help capture sudden demand spikes more effectively. While ARIMA remains a viable option for stable, linear trends, its rigidity limits its applicability in dynamic markets, such as Apple sales in Bali. Conversely, LSTM's flexibility comes at the cost of higher computational demands and a need for extensive training data. Balancing these trade-offs will be crucial for developing more reliable forecasting systems in agricultural supply chains.

## 4 Conclusion

The results of the research confirmed that the ARIMA method was found to be superior in computational efficiency compared to LSTM, as it required no repeated training iterations. In contrast, LSTM required 25 iterations, making the computation time longer. However, LSTM showed higher accuracy with a MAPE value of 56.58% and RMSE of 28.93, though the difference in RMSE was insignificant compared to ARIMA (28.84), as both algorithms failed to capture the trend of sales spikes during holidays. For future research, it is necessary to develop LSTM models that are better able to learn complex patterns, such as seasonal fluctuations, for more accurate predictions.

## References

- Lipinski, B., Hanson, C., James, L., Kitinoja, L., & Waite, R. (2013). *Installment 2 of "creating a sustainable food future" reducing food loss and waste*. <http://www.worldresourcesreport.org>.
- Chen, K.S., Lin, T.Y., Lin, K.P., Chang, P.T., & Wang, Y.C. (2023). Developing a novel long short-term memory network with seasonal wavelet transform for long-term wind power output forecasting. *International Journal of Computational Intelligence Systems*, 16(1), 191. <https://doi.org/10.1007/s44196-023-00371-x>.
- Ghareeb, S., Mahyoub, M., & Mustafina, J. (2023). A comparative time series analysis of the different categories of items based on holidays and other events. *2023 15th International Conference on Developments in ESystems Engineering (DeSE)*, 131–136. <https://doi.org/10.1109/DeSE58274.2023.10099814>.
- Ilu, S. Y., & Prasad, R. (2023). Improved autoregressive integrated moving average model for COVID-19 prediction by using statistical significance and clustering techniques. *Heliyon*, 9(2). <https://doi.org/10.1016/j.heliyon.2023.e13483>.
- Kenaar, A., Loudyi, D., Kissi, M., & El Bilali, A. (2024). Comparison of statistical time series and regression machine learning models to forecast the terraclimate dataset in the Bouregreg Chaouia Basin. *2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS)*, 1–7. <https://doi.org/10.1109/ICDS62089.2024.10756390>.
- Kramar, V., & Alchakov, V. (2023). Evaluation of the machine learning techniques for forecasting the seasonal time series. *2023 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT)*, 262–265. <https://doi.org/10.1109/USBREIT58508.2023.10158869>.
- Li, X., & Zhang, X. (2023). *A comparative study of statistical and machine learning models on carbon dioxide emissions prediction of China*. <https://doi.org/10.21203/rs.3.rs->

3070359/v1.

- Maheswari, A., Artini, N. I., & Dewi, L. (2022). Profil Pedagang Buah-Buahan di Pasar Adat Blahkiuh Kecamatan Abiansemai Kabupaten Badung. *Jurnal Agribisnis Dan Agrowisata (Journal of Agribusiness and Agritourism)*, 11, 370. <https://doi.org/10.24843/JAA.2022.v11.i01.p34>.
- Muleta, T. N. (2021). Climate change scenario analysis for Baro-Akobo river basin, Southwestern Ethiopia. *Environmental Systems Research*, 10(1), 24. <https://doi.org/10.1186/s40068-021-00225-5>.
- Ratnawati, D., Hendrawan, I., & Brasika, I. B. (2022). Potensi sampah masuk ke laut dari aktivitas darat di Kabupaten Badung, Provinsi Bali. *Journal of Marine Research and Technology*, 5(5). <https://doi.org/10.24843/JMRT.2022.v05.i01.p02>.
- Syahrini, I., Radhiah, R., & Damanik, W. F. (2023). Application of the autoregressive integrated moving average (ARIMA) Box-Jenkins Method in forecasting inflation rate in Aceh. *Transcendent Journal of Mathematics and Application*, 2(1).
- Urugo, M. M., Teka, T. A., Gemed, H. F., Mersha, S., Tessema, A., Woldemariam, H. W., & Admassu, H. (2024). A comprehensive review of current approaches on food waste reduction strategies. *Comprehensive Reviews in Food Science and Food Safety*, 23(5), e70011. <https://doi.org/https://doi.org/10.1111/1541-4337.70011>.
- Widyarsana I Made Wahyu, Damanhuri Enri, Ulhusna Nida, & Agustina Elprida. (2020). A preliminary study: Identification of stream waste quantity and composition in Bali Province, Indonesia. *E3S Web Conf.*, 148, 5005. <https://doi.org/10.1051/e3sconf/202014805005>.
- Wiwin, I. (2021). Implementasi Tri Hita Karana dalam pengembangan ekowisata menuju pariwisata berkelanjutan di Bukit Cemeng Kabupaten Bangli. *Jurnal Kajian Bali (Journal of Bali Studies)*, 11, 353. <https://doi.org/10.24843/JKB.2021.v11.i02.p06>.
- Youness, J., & Driss, M. (2022). LSTM deep learning vs ARIMA algorithms for univariate time series forecasting: A case study. *2022 8th International Conference on Optimization and Applications (ICOA)*, 1–4. <https://doi.org/10.1109/ICOA55659.2022.9934119>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

