

ARIMA and LSTM Comparison for Forecasting Healthcare Service Costs in Bogor

Rizal Ashari Nampira^{1*}, Jenal Mutakin Sambas², Ika Nur Laily Fitriana³, Liyu Adhi Kasari Sulung⁴

^{1*} BPJS Kesehatan (Bogor Branch Office) & Universitas Terbuka

²BPJS Kesehatan (Bogor Branch Office)

³Fakultas Sains dan Teknologi, Universitas Terbuka

⁴Fakultas Ekonomi dan Bisnis, Universitas Indonesia

*Corresponding author's e-mail: rizal.ashari@bpjs-kesehatan.go.id

Abstract—The National Health Insurance (JKN) program, managed by BPJS Kesehatan, has experienced a significant increase in healthcare service costs, particularly at Advanced Referral Healthcare Facilities (FKRTL). This study aims to compare the forecasting accuracy of ARIMA and Long Short-Term Memory (LSTM) methods in predicting healthcare service costs in FKRTL Bogor from January 2014 to October 2024. The data, sourced from BPJS Kesehatan Branch Bogor, were analyzed using time series approaches. Model evaluation was conducted using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Results show that for 80% of training data, LSTM produced a MAPE of 8.85% and RMSE of IDR 6.98 billion, slightly outperforming ARIMA (0,1,1) with MAPE of 10.28% and RMSE of IDR 6.67 billion. For the 20% testing data, LSTM demonstrated significantly better accuracy, with an MAPE of 12.97% and RMSE of IDR 15.52 billion, compared to ARIMA's MAPE of 24.22% and RMSE of IDR 30.76 billion. Therefore, LSTM is considered more effective for short- to medium-term forecasting of JKN healthcare costs, particularly when dealing with complex and non-linear patterns.

Keywords: JKN; ARIMA; LSTM; forecasting; RMSE

1. INTRODUCTION

Based on Law of the Republic of Indonesia Number 40 of 2004 concerning the National Social Security System, the Government must provide social protection in the form of social security to all Indonesian citizens. One form of social security is ensuring access to healthcare services without geographical or financial barriers through the National Health Insurance (JKN) program managed by BPJS Kesehatan. Globally, healthcare expenditure continues to rise, with WHO (2022) reporting expenditures reaching US\$8.8 trillion, driven by demographic transitions and the growing burden of non-communicable diseases. In Indonesia, increasing cases of diabetes, hypertension, and cancer contribute significantly to healthcare spending at both primary (FKTP) and advanced referral healthcare facilities (FKRTL). During 2016–2018, FKRTL accounted for 79.51% of total JKN financing (Arumsari & Meliala, 2019), highlighting the importance of accurately forecasting healthcare service costs (bipelkes) to support sustainable national health insurance financing.

In recent years, time-series forecasting has become essential in supporting planning across various sectors, including healthcare. Traditional statistical models such as the Autoregressive Integrated Moving Average (ARIMA) have long been applied for linear time-series prediction (Hyndman & Athanasopoulos, 2021). However, healthcare costs often exhibit nonlinear, volatile, and multi-factor-driven patterns. This has driven increased use of deep learning methods such as Long Short-Term Memory (LSTM), which are capable of modeling long-term dependencies and nonlinear behavior in time series data (Cheng & Kuo, 2020). A growing body of comparative research demonstrates that LSTM frequently outperforms ARIMA in complex forecasting environments. Mahmud et al. (2025) showed that hybrid ARIMA–LSTM models significantly improve accuracy in COVID-19 forecasting. Alizadegan et al. (2024) found that LSTM and Bi-LSTM outperform ARIMA for energy consumption prediction. Hossain et al. (2025) similarly reported that ARIMA and SARIMA struggle with nonlinear renewable energy patterns, while LSTM achieved an R^2 above 0.98. Zabrina (2025) emphasized the limitations of ARIMA in Indonesian economic series when structural shifts occur. Additional evidence from Mahmud et al. (2025) indicated that deep learning models outperform classical approaches across epidemiological time-series. Furthermore, Gerakoudi et al. (2025) demonstrated that LSTM consistently surpasses ARIMA, Random Forest, and Decision Tree models in forecasting nonlinear maritime stock prices, with the hybrid ARIMA–LSTM model providing the best performance across multiple volatility scenarios. Collectively, these studies highlight the growing consensus that LSTM-based models are more effective than

traditional statistical methods for datasets characterized by nonlinear patterns, volatility, and evolving structural dynamics—properties commonly observed in healthcare service cost data.

Despite the extensive literature comparing ARIMA and LSTM in domains such as epidemiology, energy systems, and financial forecasting, research applying these methods to healthcare financing—particularly within Indonesia’s JKN system—remains limited. The unique characteristics of FKRTL cost data, which involve case complexity, demographic transitions, and medical inflation, present forecasting challenges that have not been addressed in previous studies. Thus, the novelty of this study lies in evaluating and comparing the performance of ARIMA and LSTM models for forecasting healthcare service costs (bipelkes) at advanced referral healthcare facilities using real-world operational data from BPJS Kesehatan Bogor Branch (2014–2024), providing a new empirical basis for model selection in national health insurance financial planning. This study assesses model accuracy using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) and aims to support BPJS Kesehatan in developing more responsive and data-driven strategies for sustainable health financing.

2. METHODS

The data used is secondary data from BPJS Kesehatan Bogor Branch Office, in the form of aggregated monthly total healthcare service costs at FKRTL (including public hospitals, private hospitals, and national referrals) from January 2014 to October 2024. The data were selected up to October 2024 because, according to Peraturan Presiden No. 82 Tahun 2018, FKRTL claims can be submitted a maximum of 6 months after the service is provided. By the time this paper was written, claims for October 2024 had entered a phase where submissions were no longer possible, ensuring data stability. The data represent bipelkes under the case-based group (CBG) system and exclude non-CBG services such as chemotherapy drugs, chronic medications, and special laboratory tests.

The forecasting process was conducted systematically through the following stages:

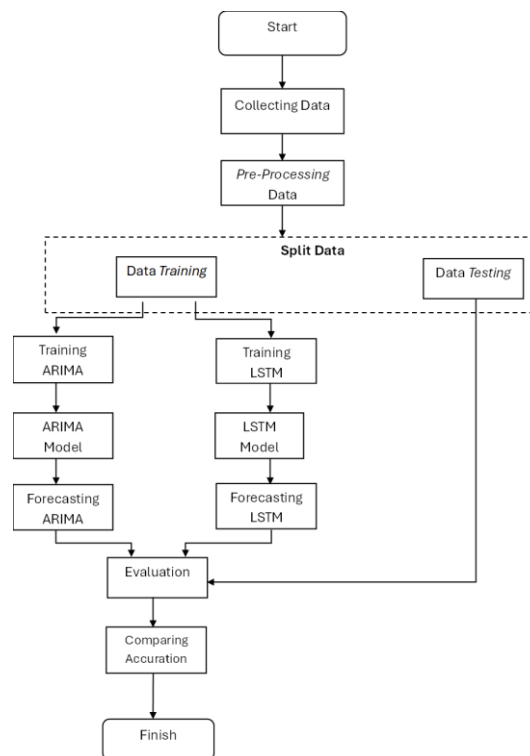


Figure 1. Research Flow

Figure 1 explains about at the initial stage, data collection was followed by data pre-processing, including format conversion and changing month names into English to be processed in Python. After pre-processing, the data were split into training and testing datasets. Subsequently, ARIMA and LSTM models were built, and forecasting was performed. Model testing began with developing ARIMA and LSTM models. For ARIMA, optimization was performed to find the optimal p , d , q parameters. The trained models were then used for forecasting, and accuracy was compared using MAPE and RMSE.

2.1 Arima

Autoregressive Integrated Moving Average (ARIMA) is often referred to as the Box–Jenkins time series approach. ARIMA is a general Autoregressive Moving Average (ARMA) model that combines the Autoregressive (AR) process with the Moving Average (MA) process to process and construct composite time series models (Pandji et al., 2019). For short-term forecasting, ARIMA offers very high accuracy; however, for long-term forecasting, its accuracy tends to be lower. When generating forecasts, the ARIMA model completely ignores independent variables. To produce accurate short-term forecasts, ARIMA relies on the historical and current values of the dependent variable (Setiawan et al., 2020).

The following is the simple form of an Autoregressive model of order p , $AR(p)$, as shown in Equation (1):

$$x_t = c + \sum_{i=1}^p \phi_i x_{t-i} + \epsilon_t \quad (1)$$

Description:

x_t = stationary variable

C = constant

ϕ_i = autocorrelation coefficient at lags 1, 2, ..., p

ϵ_t = residuals, a Gaussian white noise series with zero mean and variance σ^2 .

Meanwhile, Equation (2) below represents the form of the Moving Average model.

$$x_t = \mu + \sum_{i=1}^q \theta_i \epsilon_t \quad (2)$$

Description:

x_t = observed value of the time series at time t

μ = mean of the stationary process (often simplified to zero)

θ_i = moving average coefficient from past residuals

ϵ_t = residuals, a Gaussian white noise series with zero mean and variance σ^2

We can obtain a combined model from the two models described above, resulting in an ARIMA model of order (p, q) . The formula in Equation (3) below represents the ARIMA model.

$$x_t = c + \sum_{i=1}^p \phi_i x_{t-i} + \sum_{i=1}^q \theta_i \epsilon_t \quad (3)$$

Where ϵ_t is Gaussian white noise with zero mean and variance σ^2 . The values of ϕ_i and θ_i are the coefficients of the AR and MA models, respectively, and are determined through an estimation process. The parameters p and q indicate the orders of the AR and MA components. The notation $ARIMA(p, d, q)$ represents the general form of the ARIMA model (Rachmawati, 2020).

2.2 LSTM

Long Short-Term Memory (LSTM) is a type of artificial neural network developed to overcome the limitations of Recurrent Neural Networks (RNNs) in handling long-term sequential data. With its memory cell structure and three main gates—input, forget, and output—LSTM is capable of retaining important information for longer periods, making it effective for processing complex and non-linear time series data. This capability has made LSTM excel in various applications, such as price prediction, sentiment analysis, and natural language processing (Cahyani et al., 2023).

In Indonesia, LSTM has been applied in various studies to address local issues. For example, Cahyani et al. (2023) used LSTM to predict national staple food prices such as rice and cooking oil. Their study demonstrated that LSTM, optimized

using the ADAM and RMSProp algorithms, was able to produce highly accurate predictions, as indicated by significant RMSE and R^2 values.

In addition, LSTM has also been used in the development of chatbots to support university students' mental health. Afrisia et al. (2024) implemented LSTM in a chatbot designed to provide emotional support for students. The model achieved 98% accuracy with a loss of 0.0614 after 50 training epochs, demonstrating LSTM's effectiveness in understanding conversational context and generating relevant responses.

In the economic field, Hartini et al. (2022) applied LSTM to predict the Wholesale Price Index (WPI) in Indonesia. The developed LSTM model achieved 98.85% accuracy with a MAPE of 1.14%, MSE of 0.0002, and RMSE of 0.0135, indicating excellent performance in modeling complex economic data.

Internationally, the LSTM method combined with ARIMA and CNN has been used for air quality forecasting in four cities in China (Duan et al., 2023). Meanwhile, in Brazil, Souza & Dantas (2024) designed a CNN–LSTM model to classify arrhythmia signs from ECG (electrocardiogram) signals using the MIT-BIH dataset.

Overall, the application of LSTM both in Indonesia and abroad demonstrates great potential in various fields, ranging from economics and environmental studies to healthcare. LSTM's ability to handle sequential data and retain long-term information makes it a highly valuable tool for analysis and prediction in local contexts.

2.3 Evaluation

In forecasting model evaluations such as ARIMA and LSTM, two commonly used metrics are Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) (Umam, 2023). RMSE measures the square root of the average squared differences between the actual and predicted values, placing greater emphasis on larger errors. Mathematically, RMSE is formulated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (4)$$

Description:

\hat{y}_i = Forecasted Value

y_i = Actual Value

n = Numbers of Data

Conversely, MAPE provides a measure of error in percentage form and is calculated by taking the average of the absolute values of the relative errors. The formula is as follows:

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (5)$$

Description:

\hat{y}_i = Forecasted Value

y_i = Actual Value

n = Numbers of Data

Several recent studies have shown that the choice of evaluation metrics greatly influences the interpretation of a forecasting model's effectiveness. For example, in a study by Taslim & Murwantara (2024), LSTM demonstrated better performance than ARIMA in predicting volatile time series data, with lower RMSE and MAPE values. However, in larger datasets, ARIMA showed higher accuracy compared to LSTM, highlighting the importance of considering data characteristics when selecting models and evaluation metrics.

3. RESULTS AND DISCUSSION

Descriptive Statistics

The healthcare service cost utilization data consist of 130 data points, with no missing or unavailable entries. The average expenditure is IDR 66.12 billion, with a Standard Error (SE) of IDR 2.14 billion and a standard deviation of IDR 24.43 billion. The median value is IDR 64.91 billion, with no mode. The data range is IDR 122.53 billion, with a minimum value of IDR 10.44 billion (January 2014 claim) and a maximum value of IDR 122.53 billion (July 2024 claim).

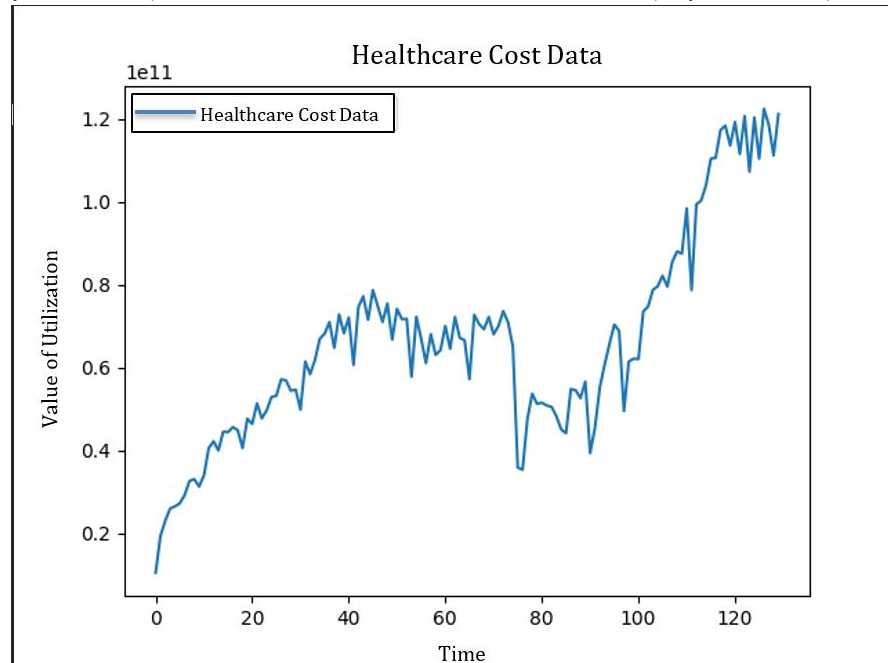


Figure 2. Trend of FKRTL Utilization in Bogor City, January 2014 – October 2024

Based on the graph in Figure 2, from January 2014 to July 2017 (data sets 1–43) there was an increase in healthcare service costs (*bipelkes*) in Bogor City. This occurred because of BPJS Kesehatan's efforts to expand service access and strengthen cooperation with various FKRTLs. During the period from August 2017 to March 2020 (data sets 44–75), *bipelkes* remained relatively stable at an average of IDR 69.29 billion, ranging from IDR 57.26 billion to IDR 78.74 billion. Starting in April 2020, *bipelkes* decreased (reaching IDR 35.86 billion) due to the spread of COVID-19, which caused a shift of JKN *bipelkes* claims to COVID-19 claims (from being covered by BPJS Kesehatan to being covered by the Ministry of Health). Then, in October 2021 (data set 94), *bipelkes* claims reached IDR 60.76 billion and subsequently remained relatively stable. Starting from June 2023 (data set 114), *bipelkes* reached IDR 100.37 billion, and from July 2023 to October 2024, the average was IDR 114.045 billion, ranging from IDR 100.37 billion to IDR 122.53 billion during that period.

3.1 Forecasting with ARIMA

In the following data analysis, the ARIMA model approach was applied with optimization, resulting in the most optimal values for p , d , and q being (0,1,1) for 80% of the training data, with an AIC value of 4954.882.

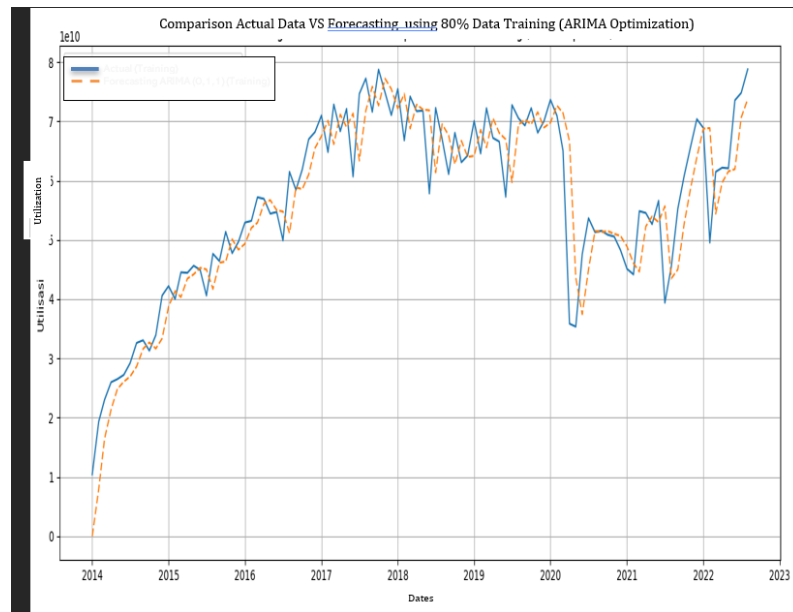


Figure 3. Comparison Chart of Actual Data vs. Forecast (80% Training Data) Using Optimized ARIMA (0,1,1)

In Figure 3, the forecasting model using 80% of the training data with ARIMA (0,1,1) reasonably represents the actual data. This is consistent with the MAPE value of 10.28% and the RMSE value of IDR 6,674,598,035 billion.

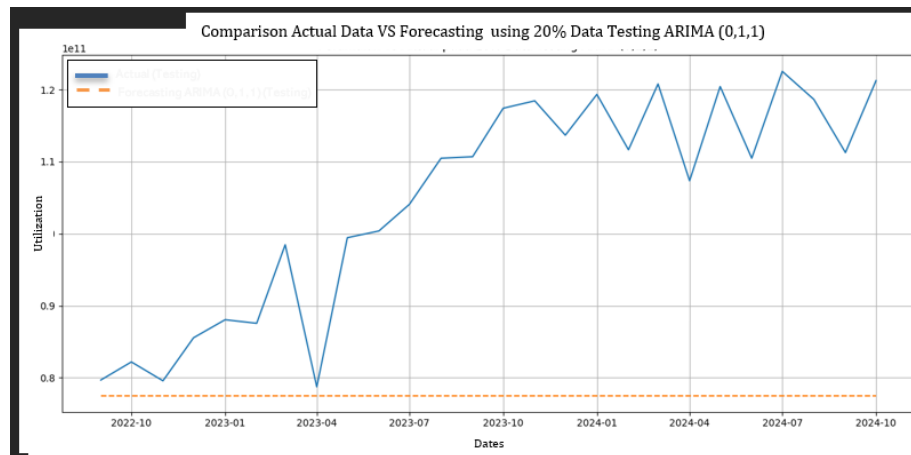


Figure 4. Comparison Chart of Actual Data vs. Forecast (20% Testing Data) Using ARIMA (0,1,1)

Based on Figure 4, a comparison using 20% testing data shows that the forecasted data does not reflect the actual data at all. The sharp increase observed between October 2022 and October 2023, combined with the relatively small number of data points in the testing set, caused ARIMA (0,1,1) to fail to capture the actual data pattern. This is consistent with the MAPE value of 24.22% and RMSE of IDR 30,759,267,895 billion. For MAPE, the value more than doubled compared to that using 80% training data, and for RMSE, the value increased more than fivefold.

3,2 Forecasting with LSTM

For the application of the LSTM method, the dataset was also divided into two parts: the first 80% of the data was used as the training set, and the remaining 20% was used as the testing set.

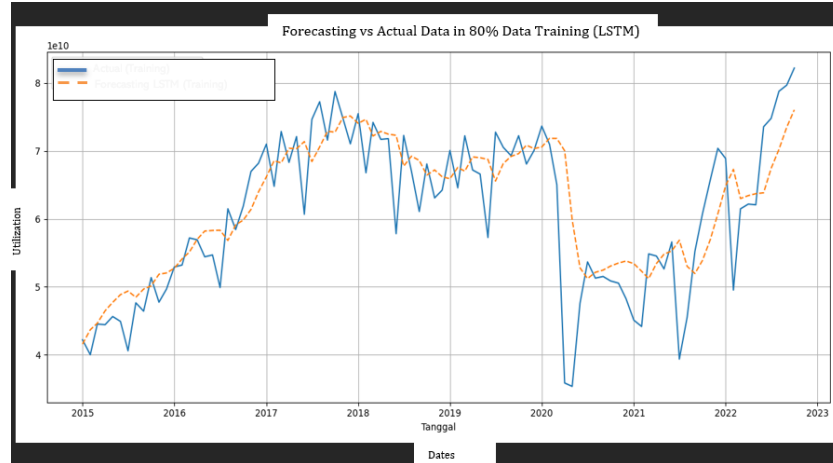


Figure 5. Comparison Chart of Actual Data vs. Forecast (80% Training Data) Using LSTM

Referring to Figure 5, the forecast data using LSTM on 80% of the training data reasonably represents the actual data; however, unlike ARIMA (0,1,1), the data appears to be overfitted. Based on the MAPE value, the result obtained is 8.85%, with an RMSE of IDR 6,984,222,955 billion.

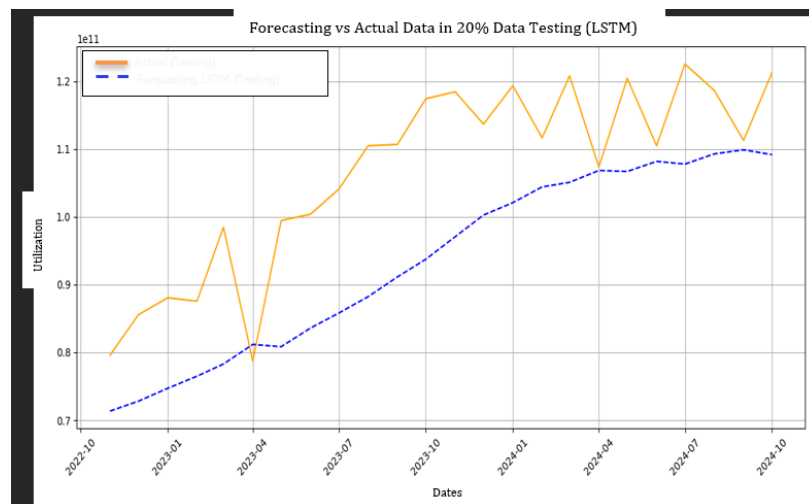


Figure 6. Comparison Chart of Actual Data vs. Forecast (20% Testing Data) Using LSTM

In Figure 6, the forecasting results using the LSTM method are able to follow the movement of the actual data in the 20% testing set. Compared to Figure 3, the use of the LSTM method is better at representing the actual data than ARIMA (0,1,1). This can be observed from the significantly smaller MAPE and RMSE values compared to ARIMA (0,1,1), namely 12.97% and IDR 15,521,547,180 billion.

The following is a summary table comparing the MAPE and RMSE metric values for the ARIMA and LSTM methods in the 80% training data phase and the 20% testing data phase:

Table 1. Comparison of MAPE and RMSE Metrics for ARIMA and LSTM Methods Using 80% Training Data and 20% Testing Data

| | MAPE | | RMSE | |
|--------------|-------------------|------------------|-------------------|------------------|
| | 80% data training | 20% data testing | 80% data training | 20% data testing |
| ARIMA | 10,28% | 24,22% | 6.674.598.035 | 30.759.267.895 |
| LSTM | 8,85% | 12,97% | 6.984.222.955 | 15.521.547.180 |

In Table 1, for the MAPE metric, LSTM performs much better than ARIMA in both the testing and training data. For the RMSE metric, ARIMA is slightly better for the training data; however, for the testing data, LSTM performs significantly better than ARIMA.

4. CONCLUSION

For the comparison of forecasting method effectiveness using 80% training data, the ARIMA (0,1,1) method achieved a MAPE of 10.28% and an RMSE of IDR 6.67 billion. Meanwhile, the LSTM method, with 80% training data, obtained a MAPE of 8.85% and an RMSE of IDR 6.98 billion. It is concluded that for the 80% training data phase, LSTM performed better in terms of MAPE, with a difference of 1.43% lower than ARIMA (0,1,1), while for RMSE, ARIMA (0,1,1) was slightly better, with a value 0.31 billion lower than LSTM. For the 20% testing data phase, the LSTM method was far superior to ARIMA (0,1,1). LSTM achieved MAPE and RMSE values of 12.97% and IDR 15.52 billion, respectively, whereas ARIMA (0,1,1) recorded MAPE and RMSE values of 24.22% and IDR 30.76 billion. It is evident that for both MAPE and RMSE in the 20% testing phase, ARIMA's values were nearly double those of LSTM.

Reference

- [1] S. P. Afrisia, F. M. Hana, and W. C. Wahyudin, "Implementasi Metode Long Short Term Memory (LSTM) pada Chatbot Kesehatan Mental Mahasiswa," *Sainteks*, vol. 21, no. 2, p. 107, Oct. 2024, doi: 10.30595/sainteks.v21i2.23869.
- [2] H. Alizadegan, B. Rashidi Malki, A. Radmehr, H. Karimi, and M. A. Ilani, "Comparative study of long short-term memory (LSTM), bidirectional LSTM, and traditional machine learning approaches for energy consumption prediction," *Energy Exploration and Exploitation*, vol. 43, no. 1, pp. 281–301, Jan. 2025, doi: 10.1177/01445987241269496.
- [3] I. Arumsari and A. Meliala, "UTILIZATION REVIEW PADA FASILITAS KESEHATAN RUJUKAN TINGKAT LANJUTAN BADAN PENYELENGGARA JAMINAN SOSIAL (BPJS) KESEHATAN KANTOR CABANG SLEMAN," *Jurnal Kebijakan Kesehatan Indonesia*, vol. 08, no. 04, pp. 196–201, Dec. 2019.
- [4] Badan Pusat Statistik, "Profil Statistik Kesehatan 2023," vol. 7, 2023.
- [5] J. Cahyani, S. Mujahidin, and T. P. Fiqar, "Implementasi Metode Long Short Term Memory (LSTM) untuk Memprediksi Harga Bahan Pokok Nasional," *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 11, no. 2, p. 346, Jul. 2023, doi: 10.26418/justin.v11i2.57395.
- [6] N. Cheng and A. Kuo, "Using Long Short-Term Memory (LSTM) Neural Networks to Predict Emergency Department Wait Time," *Studies in health technology and informatics*, vol. 272, pp. 199–202, Jun. 2020, doi: 10.3233/SHTI200528.
- [7] J. Duan, Y. Gong, J. Luo, and Z. Zhao, "Air-quality prediction based on the ARIMA-CNN-LSTM combination model optimized by dung beetle optimizer," *Scientific Reports*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-36620-4.
- [8] K. Gerakoudi, D. Georgoulas, and P. J. Stavroulakis, "A novel hybrid ARIMA-LSTM model for maritime shipping stock forecasting: comparative evidence against statistical and machine learning benchmarks," *Maritime Business Review*, pp. 1–19, Aug. 2025, doi: 10.1108/mabr-04-2024-0035.
- [9] Hartini, F. Insani, Novriyanto, and S. Sanjaya, "Impelmentasi Long Short Term Memory Neural Network untuk Prediksi Indeks Harga Perdagangan Besar," Pekanbaru, Oct. 2022.
- [10] M. L. Hossain, S. M. N. Shams, and S. M. Ullah, "Time-series and deep learning approaches for renewable energy forecasting in Dhaka: a comparative study of ARIMA, SARIMA, and LSTM models," *Discover Sustainability*, vol. 6, no. 1, Dec. 2025, doi: 10.1007/s43621-025-01733-5.
- [11] R. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 03 ed. OText, 2021.



- [12] Kementerian Kesehatan, “PROFIL KESEHATAN INDONESIA 2022,” 2022.
- [13] A. Mahmud *et al.*, “Hybrid ARIMA-LSTM for COVID-19 forecasting: A comparative AI modeling study,” *PeerJ Computer Science*, vol. 11, 2025, doi: 10.7717/peerj-cs.3195.
- [14] B. Pandji, Indwiarti, and A. Rohmawati, “PERBANDINGAN PREDIKSI HARGA SAHAM DENGAN MODEL ARIMA DAN ARTIFICIAL NEURAL NETWORK,” *Indonesia Journal of Computing*, vol. 04, no. 2, pp. 189–198, Sep. 2019, doi: 10.21108/indojc.2019.4.2.344.
- [15] A. K. Rachmawati, “Peramalan Penyebaran Jumlah Kasus Covid19 Provinsi Jawa Tengah dengan Metode ARIMA,” *Zeta - Math Journal*, vol. 6, no. 1, pp. 11–16, Dec. 2020, doi: 10.31102/zeta.2021.6.1.11-16.
- [16] Z. Raissa, Z. Raissa, and K. Kunci, “A COMPARATIVE ANALYSIS OF FINANCIAL PERFORMANCE FORECASTING MODELS: ARIMA, ARIMA-GARCH & LSTM IN INDONESIAN BANKING STOCKS”.
- [17] H. Setiawan, E. Utami, and H. al Fatta, “Penerapan Arima Dan Artificial Neural Network Untuk Prediksi Penderita DBD Di Kabupaten Sragen,” *Majalah Ilmiah Bahari Jogja*, vol. 18, no. 2, pp. 64–78, Aug. 2020, doi: 10.33489/mibj.v18i2.220.
- [18] I. L. Souza and D. O. Dantas, “Cardiac Arrhythmia Detection in Electrocardiogram Signals with CNN-LSTM,” in *International Conference on Pattern Recognition Applications and Methods*, Science and Technology Publications, Lda, 2024, pp. 304–310. doi: 10.5220/0012362600003654.
- [19] D. G. Taslim and I. M. Murwantara, “Comparative analysis of ARIMA and LSTM for predicting fluctuating time series data,” *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 3, pp. 1943–1951, Jun. 2024, doi: 10.11591/eei.v13i3.6034.
- [20] K. Umam, “MIND (Multimedia Artificial Intelligent Networking Database Perbandingan Metode ARIMA dan LSTM pada Prediksi Jumlah Pengunjung Perpustakaan,” *Journal MIND Journal | ISSN*, vol. 8, no. 2, pp. 119–129, 2023, doi: 10.26760/mindjournal.v8i2.119-129.
- [21] WHO, “Global spending on health Rising to the pandemic’s challenges,” 2022. Accessed: Jun. 13, 2025. [Online]. Available: <https://www.who.int/publications/i/item/9789240064911>

