JIGE 6 (4) (2025) 3247-3261
JURNAL ILMIAH GLOBAL EDUCATION
ejournal.nusantaraglobal.ac.id/index.php/jige
DOI: https://doi.org/10.55681/jige.v6i2.4694

Performance Evaluation of ARIMA and LSTM Models to Handle Multi-Interventions in Automobile Production Forecasting

Firda Aulia Maghfiroh[1*], Indahwati[1], Asep Saefuddin[1]
[1] Study Program on Statistics and Data Science, IPB University, Indonesia
*Corresponding author email: firdauliamffirda@apps.ipb.ac.id

Article Info

Article history:

Received September 20, 2025
Approved November 21, 2025

Keywords:

Multi-Input Intervention, LSTM, ARIMA, Car Production

ABSTRACT

Intervention refers to disturbances caused by internal or external variables, such as market changes, international events, or policy shifts. The dataset used in this study contains three intervention events, referred to as a multi-input intervention. The data consist of car production figures from PT Astra Daihatsu Motor obtained from the official GAIKINDO website. The forecasting task focuses on predicting PT Astra Daihatsu Motor's production, which was influenced by three major interventions: policy changes in 2013, the impact of the COVID-19 pandemic in 2020, and the increase in SUV production in 2022. This study compares ARIMA and LSTM models for car production forecasting. The dataset covers monthly production data from January 2010 to June 2024, with a total of 174 observations. RMSE, MAPE, and SMAPE are employed as accuracy measures. Based on the testing data (May 2023–June 2024), the results show that the LSTM model outperforms ARIMA in capturing trend patterns, with lower error values of RMSE (4587.65), MAPE (10.37), and SMAPE (10.39), compared to ARIMA with RMSE (5059.48), MAPE (11.59), and SMAPE (10.50). Accordingly, LSTM represents a relevant and robust modeling alternative for production forecasting in operational decision-making, owing to its flexibility and strong capability in capturing complex data patterns.

How to cite: Maghfiroh, F. A., Indahwati, I., & Saefuddin, A. (2025). Performance Evaluation Of Arima And Lstm Models To Handle Multi-Interventions In Automobile Production Forecasting. *Jurnal Ilmiah Global Education*, *6*(4), 3247–3261. https://doi.org/10.55681/jige.v6i4.4694

INTRODUCTION

Time series data is a collection of data arranged in a time sequence to facilitate the analysis of changes in patterns or trends within a particular period (Sirisha et al., 2016). This trend can be either upward or downward (Dubey et al., 2021). One of the main advantages of time series analysis is forecasting, which aims to predict future values (Masrawanti et al., 2019). It creates predictive models based on existing patterns and trends. Patterns in the data can be influenced by certain events that cause changes, called interventions (Iswari A et al., 2022). These interventions can be caused by external or internal factors, such as natural disasters, political changes, or economic crises (Aziza et al., 2023). Multiple interventions occur when multiple unexpected events affect time series data (Wei 2006). To account for the effects of interventions, intervention analysis is used to understand and correct their effects (Sohibien, 2018; Schaffer et al., 2021;Lorensya et al., 2022) as well as to improve the accuracy of the model. An example of time series data with multiple interventions is the automobile production process.

Cars serve as both personal and public transportation. However, their production can contribute to air pollution and traffic congestion. Therefore, forecasting vehicle production is important for both governments and manufacturers. Manufacturers can assist governments in planning transportation infrastructure while adapting production strategies to new standards. This study uses vehicle production data from PT Astra Daihatsu Motor, a company that experienced multiple intervention events. In 2013, production decreased due to internal policy changes. In April 2020, production decreased significantly due to the impact of the COVID-19 pandemic, but in 2022, SUV production increased. These events suggest the presence of multiple interventions affecting production data. To achieve more accurate forecasts, time series modeling must take these into account (Wiladibrata & Rifai, 2022).

Classical methods such as ARIMA (Autoregressive Integrated Moving Average) have been used to handle multiple interventions. Lorensya et al. (2022) and Schaffer et al. (2021) applied ARIMA with pulse and step functions to forecasting multiple steps and demonstrated the model's ability to provide accurate forecasts and respond to changes in the data. Therefore, ARIMA is a good choice for automated production data analysis of multiple interventions.

Technological advancements have enhanced statistical methodologies for machine learning applications. The Long Short-Term Memory (LSTM) model, a type of artificial neural network (ANN), has been selected to predict automobile production due to its ability to retain information over extended periods. LSTM excels in time series challenges where the intervals between key events are unpredictable. The iterative nature of LSTM's hidden layers yields more precise outcomes. Research indicates that LSTM can significantly minimize forecast errors in comparison to other models (Sautomo & Pardede, 2021). In a study by Primawati et al. (2023), LSTM was contrasted with Prophet for forecasting milk production impacted by interventions, demonstrating LSTM's superior performance. Furthermore, Sautomo & Pardede (2021) revealed that with proper hyperparameter selection, LSTM outperformed ARIMA in projecting daily government expenditures. Given that LSTM can effectively manage data with long-term dependencies, this study investigates whether it can adeptly address multiple interventions in time series data.

This analysis evaluates the effectiveness of ARIMA and LSTM in forecasting automobile production, with an emphasis on their capacity to manage multiple interventions. The models are assessed using RMSE and MAPE as metrics for accuracy. The findings of this study aim to offer valuable insights for future research and assist companies in making informed decisions about vehicle production.

METHODS

The analysis steps conducted with the car production data from GAIKINDO website are as follows:

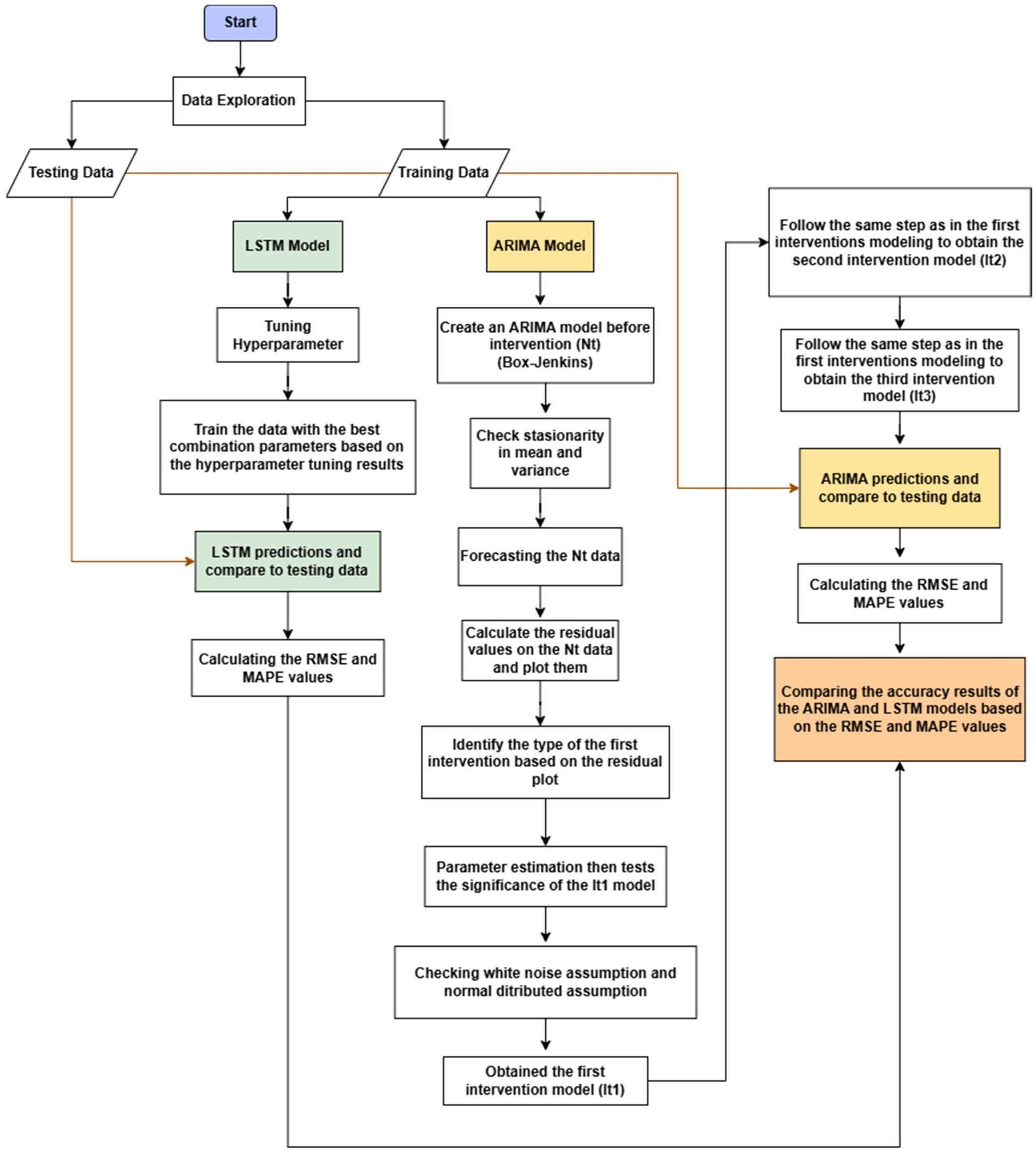

Fig. 1. Research Flowchart

RESULTS AND DISCUSSION

In the automobile production dataset from PT Astra Daihatsu Motor, a zero value was recorded at observation 124 (May 2020) due to the suspension of production activities during the COVID-19 pandemic, as reported by GAIKINDO. To prevent computational issues and ensure that the forecasting model could be properly estimated, a constant value of 1 was added to all observations in the dataset.

Table 1. Original Production Data and Production Data Plus One

Period	Original Data	Data Plus One
01-01-2010	20336	20337
…	…	…
01-04-2020	1330	1331
01-05-2020	0	1
01-06-2020	333	334
01-07-2020	1553	1554
01-08-2020	6632	6633
…	…	…
01-06-2024	35038	35039

Automobile production data from PT Astra Daihatsu Motor, obtained from GAIKINDO, shows that between January 2010 and June 2024 the average production was 21361,96 units with a variance of 11438,94, indicating substantial fluctuations over time. The highest production was recorded in December 2022 (74,384 units), while the lowest occurred in May 2020 (0 units). Sharp declines were observed in January 2013 and May 2020, whereas a notable increase was recorded in January 2022. These significant variations, characterized by abrupt decreases and increases, are classified as interventions in statistical analysis.

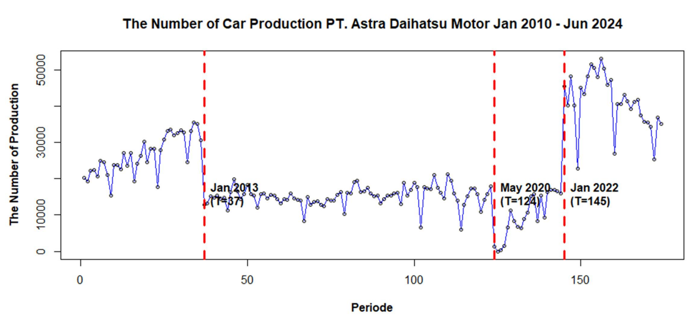

Fig. 2. Total Car Production of PT Astra Daihatsu Motor

3.1 ARIMA Model

Stationarity of the data was examined through both exploratory analysis and statistical testing. Exploratorily, stationarity in the mean was assessed using the ACF and PACF plots. The ACF plot in Figure 7 shows a slow decline, indicating that the data are non-stationary in the mean. To formally test stationarity in the mean, the Augmented Dickey-Fuller (ADF) test was applied. The ADF test result yielded a p-value greater than the 5% significance level (0,08336), suggesting that the data are non-stationary in the mean and therefore require differencing. Furthermore, the linear trend plot of production also indicates that the data are not linearly distributed and exhibit variability, necessitating an evaluation of the model's lambda value after differencing.

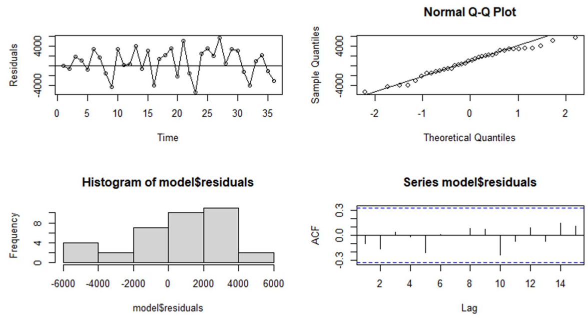

Fig. 3. ACF and PACF Plot of Car Production of PT Astra Daihatsu Motor

The plot of the differenced data indicates that the series is stationary in the mean, as fluctuations appear around zero without any discernible trend (Figure 8). The stationarity test using the Augmented Dickey-Fuller (ADF) on the differenced data produced a p-value of 0,01648, which is less than the 5% significance level, indicating that the data are stationary in the mean after the first differencing (d = 1).

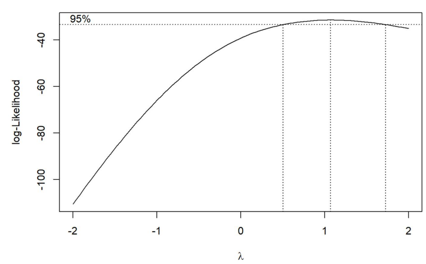

Fig. 4. Box-Cox Plot of Automobile Production Data from PT Astra Daihatsu Motor

Variance stationarity also needs to be examined once the data are confirmed to be stationary in the mean. The assessment of variance stationarity was conducted using the Box-Cox plot. The lambda (λ) value obtained from the Box-Cox plot in Figure 8 was 1.999924, which is greater than one, indicating that the data are stationary in variance; therefore, no transformation is required.

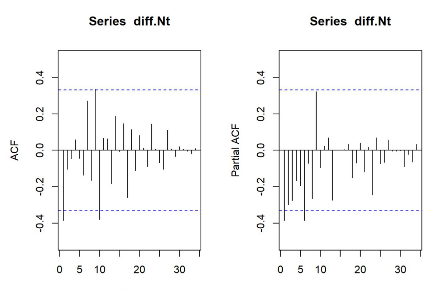

Fig. 5. ACF and PACF Plots of Differenced Automobile Production Data from PT Astra Daihatsu Motor

The identification results of the PACF and ACF plots in Figure 9 for the stationary data yielded a tentative ARIMA model, indicating that the ARIMA(3,1,5) model is significant for each parameter estimator and has the smallest AIC value of 684, as evidenced by p-values lower than the 5% significance level. Meanwhile, the ARIMA(4,1,5) and ARIMA(2,1,9) models also had parameter estimators that were significant at the 5% level but produced larger AIC values compared to the ARIMA(3,1,5) model.

Table 2. Estimated Parameter Values ARIMA Model Before Intervention (Nt)

Model	AIC
ARIMA (3,1,5)	684
ARIMA (4,1,5)	685,98
ARIMA (2,1,9)	685,68

All parameter estimation models were subsequently tested for model diagnostics. The results of the residual independence test using the Ljung-Box test on the ARIMA (3,1,5) model in Table 4.3 show that each residual lag tested has a p-value greater than the 5% significance level, indicating that no autocorrelation is present in the model residuals.

Table 3. Ljung-Box Test Values of the ARIMA Model Before Intervention (Nt)

Model	Lag	P-Value
ARIMA	1	0,554
(3,1,5)	15	0,8345

Residual normality was assessed using the Kolmogorov-Smirnov test, which yielded a p-value of 0,9028, exceeding the 5% significance level and indicating normally distributed residuals. This finding was further supported by the Q-Q plot and histogram (Figure 10), where the residual points closely followed the diagonal line, the histogram showed no sharp peaks, and the residuals demonstrated statistical independence as no autocorrelation lags exceeded the dashed blue boundaries.

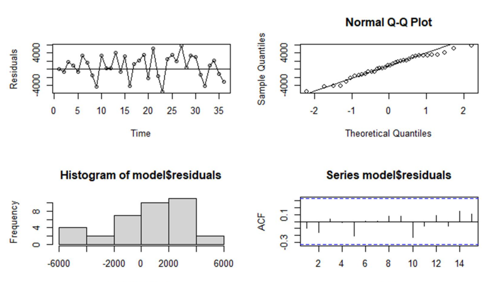

Fig. 6. Normal Distribution Plot of the ARIMA Model for Automobile Production Data of PT Astra Daihatsu Motor

The residual pattern plot with intervention in Figure 10 shows that prior to the intervention (T–36 to T–1), the residuals appeared stable with small fluctuations around zero, indicating that the model performed reasonably well on the pre-intervention data. However, after the intervention (T+1 onward), the residuals became substantially larger. These large fluctuations demonstrate that the model was unable to adequately capture the pattern of changes following the intervention, thus necessitating the use of an intervention function, namely the step function. The policy change within the company, specifically the transfer of Toyota brand assembly to PT

Toyota Manufacturing, caused a sudden and permanent shift in the time series data, indicating that the intervention followed a step function pattern.

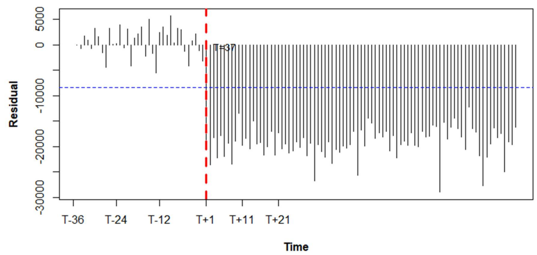

Fig. 7. Residual Plot of the First Intervention Model for Automobile Production Data of PT Astra Daihatsu Motor

The parameter estimation results of the intervention model are presented in Table 4. The intervention model, in which all parameter estimates are significant at the 5% significance level, was identified as the appropriate model. This is indicated by the p-values of each parameter estimator being smaller than the 5% significance threshold.

Table 4. Parameter Estimates of the First Intervention ARIMA Model (It1)

Model	Ordo	Coeff	P-Value
	AR(1)	2.8376e-01	
	AR(2)	2.4205e-01	
	AR(3)	-9.0319e-01	
	MA(1)	-9.6921e-01	
ARIMA (3,1,5)	MA(2)	-2.0040e-01	<0.0001
	MA(3)	1.1384e+00	
	MA(4)	-5.6401e-01	
	MA(5)	-3.6148e-02	
	ω_{0_1}	-1.7415e+04	

The Ljung-Box test results in Table 4 indicate that the first intervention model has no autocorrelation in the residuals, as the p-values of the tested residual lags are greater than the 5% significance level. Furthermore, the Kolmogorov-Smirnov test shows $p > 0.05$, confirming that the model residuals are normally distributed.

Table 5. Ljung-Box and Kolmogorov-Smirnov Tests of the First Intervention ARIMA Model (It1)

Model	L-Jung	Kolmogorov-Smirnov
ARIMA (3,1,5)	0.4637	0.1538
ω_{0_1}	No Autocorrelation	Normal Distribution

The residual pattern plot with the second intervention in Figure 11 shows that prior to the intervention (T–123 to T+1), the residuals appeared relatively orderly despite some fluctuations, indicating that the model performed adequately on the first intervention data. However, for the second intervention (T+1 onward), the residuals exhibited a change in pattern, with several large residual values and an increased frequency of fluctuations. These substantial fluctuations demonstrate that the model was unable to properly capture the changes after the first intervention, thereby requiring the use of another intervention function, namely the pulse function. The occurrence of COVID-19 caused a sudden and temporary change in the time series data, indicating that the intervention followed a pulse function pattern.

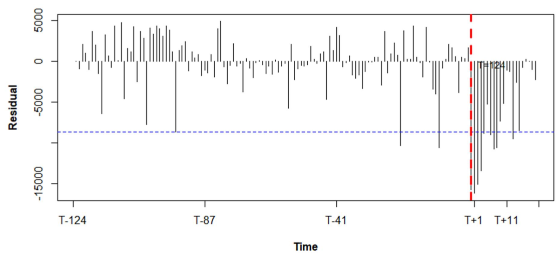

Fig. 8. Residual Plot of the Second Intervention Model for Automobile Production Data of PT Astra Daihatsu Motor

The estimation results of the second intervention model are presented in Table 6. This model, in which all parameter estimates are significant at the 5% level, is identified as the most appropriate. This conclusion is supported by the fact that the p-values of all parameter estimates are smaller than the 5% significance threshold.

Table 6. Parameter Estimates of the Second Intervention ARIMA Model (It2)

Model	Ordo	Coeff	P-Value
	AR(1)	-1.2682e-01	
	AR(2)	-7.9422e-02	
	AR(3)	-8.6253e-02	
	MA(1)	-1.9894e-01	
ARIMA (3,1,5)	MA(2)	-6.2809e-02	<0.0001
ω_{0_1}	MA(3)	-1.9347e-02	
	MA(4)	-1.8808e-01	
	MA(5)	1.2778e-01	
	ω_{0_2}	8.3932e-01	
	δ_{1_2}	-7.0173e+03	

The results of the Ljung-Box test in Table 7 indicate that the second intervention model does not exhibit autocorrelation in its residuals, as the p-values of the tested residual lags are greater than the 5% significance level. Furthermore, the results of the Kolmogorov-Smirnov test show $p > 0.05$, which means that the residuals of the model are normally distributed.

Table 7. Ljung-Box and Kolmogorov-Smirnov Tests of the Second Intervention ARIMA Model (It2)

Model	L-Jung	Kolmogorov-Smirnov
ARIMA (3,1,5)	0.6956	0.06494
$\omega_{0_2}, \delta_{1_2}$	No Autocorrelation	Normal Distribution

The residual pattern plot with the third intervention in Figure 12 shows that before the intervention, namely from T-145 to T+1, the residuals tended to remain stable and were mostly within tolerance limits, indicating that the model performed reasonably well on the second intervention data. However, the residuals of the third intervention, from T+1 onwards, indicate that the model could no longer adequately capture the data pattern. The large fluctuations demonstrate that the model was unable to capture the changes after the second intervention properly, thus necessitating the use of another intervention function, namely the step function. The introduction of new car types caused a sudden and assumed permanent change in the time series data, which indicates that the intervention occurred in the form of a step function.

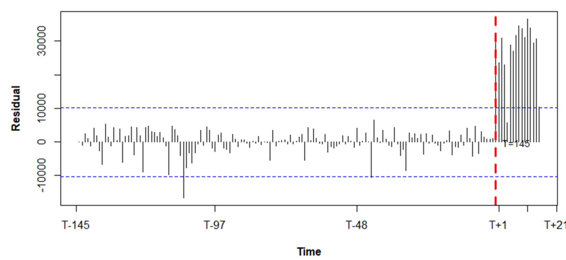

Fig. 9. Residual Plot of the Third Intervention Model for Automobile Production Data of PT Astra Daihatsu Motor

The parameter estimation results of the third intervention model are presented in Table 8. The intervention model in which all parameter estimates are significant at the 5% significance level is the third intervention model. This is evident from the p-values of each parameter estimator, which are smaller than the 5% significance level.

Table 8. Parameter Estimates of the Third Intervention ARIMA Model (It3)

Model	Ordo	Coeff	P-Value
	AR(1)	-5.2327e-01	
	AR(2)	-4.4782e-01	
	AR(3)	-7.6822e-01	
	MA(1)	1.1623e-01	
ARIMA (3,1,5)	MA(2)	1.5313e-01	<0.0001
	MA(3)	5.4552e-01	
	MA(4)	-6.0844e-01	
	MA(5)	-1.3149e-01	
	ω_{0_3}	2.7784e+04	

The Ljung-Box test results in Table 9 show that the third intervention model has no autocorrelation in its residuals, as the p-value of the tested residual lags is greater than the 5% significance level. The Kolmogorov-Smirnov test results indicate a p-value < 0.05, meaning that the model residuals are not normally distributed. However, this can be tolerated based on the Central Limit Theorem. The Central Limit Theorem states that a distribution can approximate normality if the sample size is sufficiently large. Since the production data used in this study consists of 174 observations, the normality assumption of the residuals in the third intervention model is considered to be satisfied.

Table 9. Ljung-Box and Kolmogorov-Smirnov Tests of the Third Intervention ARIMA Model (It1)

Model	L-Jung	Kolmogorov-Smirnov
ARIMA (3,1,5)	0.6549	0.001316
$\omega_{0_2}, \delta_{1_2}$	No Autocorrelation	Normal Distribution

The ARIMA (3,1,5) multi-input intervention model will be applied to forecast the number of car production in future periods. The equation of the ARIMA (3,1,5) with the third intervention model can be expressed as follows:

$$Y_t = m_t + N_t \quad \text{...(1)}$$

With the multi-input intervention model :

$$m_t = -17415 S_{1,t} + 0{,}839 m_{2,t-1} - 7017{,}3 S_{2,t-1} + 27784\, S_{3,t} \quad \text{...(2)}$$

And ARIMA model :

$$N_t = -0{,}523 Y_{t-1} - 0{,}448 Y_{t-2} - 0{,}768 Y_{t-3} + 0{,}116 \varepsilon_{t-1} + 0{,}153 \varepsilon_{t-2} + 0{,}546 \varepsilon_{t-3} - 0{,}608 \varepsilon_{t-4} - 0{,}1315 \varepsilon_{t-5} \quad \text{....(3)}$$

3.2 LSTM

In practice, LSTM modeling does not provide definitive guidelines for determining parameter values such as the number of epochs, the number of units, learning rates, and batch sizes. Consequently, an experimental process is required to identify the optimal parameter combination that best fits the characteristics of the data. This process is conducted through hyperparameter tuning, which in this study is implemented using the Grid Search algorithm. The optimal parameters obtained from this tuning procedure are subsequently employed in the LSTM modeling stage, specifically during parameter initialization.

For the parameter initialization of the training data modeling, this study utilizes a single LSTM layer with configurations of 2 and 4 neurons, in combination with a single dense layer. The optimizer applied is Root Mean Square Propagation (RMSProp), with the number of epochs set to 50, 100, and 200, and batch sizes of 16 and 32. Furthermore, cross-validation is incorporated at this stage to ensure more reliable performance evaluation and to mitigate overfitting. The cross-validation technique adopted is TimeSeriesSplit, which is suitable for time series data as it preserves the temporal ordering of observations. A comprehensive summary of the parameter initialization process is provided in Table 10.

Table 10. Combination of Hyperparameters from LSTM Model Tuning

Parameters	Values
Neuron Units	2,4
LSTM Layer	1
Dense Layer	1
Activation Function	tanh,relu
Optimizer	RMSProp
Learning Rates	0,01, 0,001, 0,0001
Batch Size	16,32
Epoch	50,100,200

After identifying the optimal combination of hyperparameters, the model is reconstructed and trained using this best configuration on the training dataset. The training process is carried out to minimize prediction errors (loss) by adjusting and updating the model's weights and biases until an optimal state is achieved. Subsequently, the model is applied to the test dataset for prediction. The resulting model is then employed for further analysis on the test data and utilized for forecasting purposes.

In this study, Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) are employed as evaluation criteria to compare the predictive performance on the empirical data. These criteria assess prediction bias and error, with MAPE providing an average measure of relative error. The following presents the best-performing model obtained from the hyperparameter tuning process.

Table 11. Best Combination of Hyperparameters from LSTM Model Tuning

Parameters	Values
Neuron Units	4
LSTM Layer	1
Dense Layer	1
Activation Function	relu
Optimizer	RMSProp
Learning Rates	0,01
Batch Size	16
Epoch	50

3.2 Evaluating ARIMA and LSTM Models

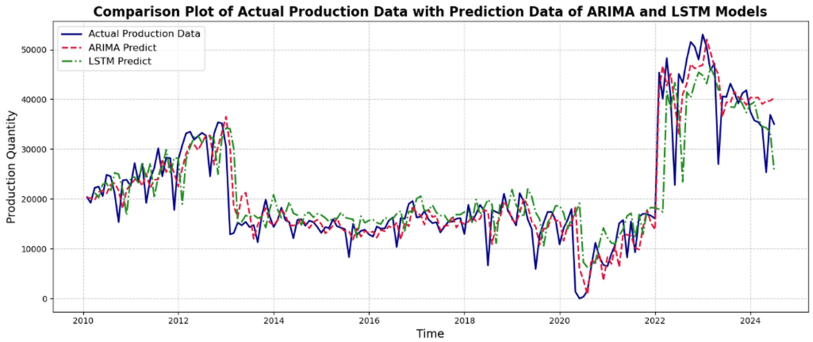

Fig. 10. Comparison Plot of Actual Production Data with Prediction Data of ARIMA and LSTM Models

The evaluation of model performance was conducted by comparing the forecasting results of ARIMA and LSTM with the actual car production data. The ARIMA model showed a relatively stable and consistent ability to capture the overall trend, but it was less responsive during periods of abrupt fluctuations, such as those observed in 2013 and 2022. In contrast, the LSTM model demonstrated greater adaptability in capturing sudden increases and declines, although it exhibited higher volatility in certain periods compared to the actual data.

The performance differences between the two models are primarily attributed to their distinct methodological frameworks. ARIMA was constructed by explicitly considering the structural characteristics of the data, including segmentation by intervention periods. The integration of step and pulse dummy variables enabled ARIMA to effectively represent structural changes in the historical series, thereby producing high training accuracy and coherent forecasting patterns.

Conversely, LSTM does not require preliminary statistical treatments, such as stationarity testing or the specification of intervention variables. Instead, it relies on the optimization of hyperparameter configurations without reference to classical statistical assumptions. Consequently, the LSTM model demonstrated relatively weaker performance on the training data but showed superior generalization on the testing data, particularly in capturing volatile patterns. In summary, ARIMA provides robust forecasts during the training phase owing to the incorporation of intervention dummies, whereas LSTM exhibits greater flexibility in adapting to new patterns during the testing phase.

Table 11. Best Combination of Hyperparameters from LSTM Model Tuning

Model	RMSE Training	MAPE Training	SMAPE Training	RMSE Testing	MAPE Testing	SMAPE Testing
ARIMA	4341.03	3750.11	18.65	5059.48	11.59	10.5
LSTM	6305.78	12075.023	23.42	4587.65	10.37	10.39

Based on the table 11, ARIMA demonstrated superior performance on the training data, consistent with its modeling approach that incorporates intervention treatments to capture structural changes. This allowed ARIMA to better represent historical patterns. In contrast, LSTM does not require stationarity assumptions or the construction of intervention variables;

thus, although its performance was less optimal on the training data, the model outperformed ARIMA on the testing data due to its stronger ability to generalize to new patterns.

The use of MAPE in this study presents limitations, as the presence of very small data values (close to zero) leads to disproportionately large error values, making the metric less representative. To address this, SMAPE was employed as a complementary measure to provide a more stable and fair evaluation of accuracy. The results indicate that, while ARIMA performs better in terms of consistency with historical data, LSTM delivers competitive results and even outperforms ARIMA across all three metrics (RMSE, MAPE, and SMAPE) in the testing phase

CONCLUSION

Based on the comparative analysis of the ARIMA and LSTM models, it can be concluded that each model exhibits distinct strengths. The ARIMA model demonstrates greater consistency in representing historical patterns by incorporating intervention factors, making it particularly suitable for academic analyses or research contexts that emphasize adherence to statistical assumptions. In contrast, the LSTM model shows superior performance in generalizing to testing data, especially in capturing sharp fluctuations in production levels.

With its independence from classical statistical assumptions, the LSTM model offers greater efficiency for practical applications, particularly for companies or industry stakeholders seeking reliable projections of future production. Although LSTM requires a relatively longer training time compared to ARIMA, its flexibility and ability to capture complex data patterns make it a highly relevant forecasting tool for supporting production planning and operational decision-making.

REFERENCES

Akaike, H. (1974). *A new look at the statistical model identification*. IEEE Transactions on Automatic Control, 19(6), 716–723.

Aziza VN, Hilali FM, Maghfiroh FA, Notodiputro KA, & Angraini Y. (2023). Comparison of SARIMA Intervention and PROPHET Models for Forecasting The Number of Airline Passenger At Soekarno-Hatta International Airport. *BAREKENG: J. Math. & App*, *Volume 17*(4), 2107–2120. https://doi.org/10.30598/barekengvol17iss4pp2107-2120

Bandara, K., Bergmeir, C., & Smyl, S. (2020). *Forecasting across time series databases using recurrent neural networks*. International Journal of Forecasting, 36(3), 855–873.

Benidis, K., Rangapuram, S. S., Flunkert, V., et al. (2020). *Neural forecasting: Introduction and literature overview*. IEEE Signal Processing Magazine, 37(6), 97–110.

Bergmeir, C., Hyndman, R. J., & Benítez, J. M. (2018). *Bagging exponential smoothing methods*. International Journal of Forecasting, 32(2), 303–312.

Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.

Cryer, J. D., & Chan, K. S. (2008). *Time Series Analysis: With Applications in R* (2nd ed.). Springer.

Dubey A K, Kumar A, Díaz V G, Sharma A K, & Kanhaiya K. (2021). Study And Analysis of SARIMA And LSTM in Forecasting Time Series Data. *Sustainable Energy Technologies and Assessments*, *Vol 47*. https://doi.org/10.1016/j.seta.2021.101474

Enders, W. (2015). *Applied Econometric Time Series* (4th ed.). Wiley.

Fildes, R., Ma, S., & Kolassa, S. (2019). *Retail forecasting: Research and practice*. International Journal of Forecasting, 35(1), 1–9.

GAIKINDO. (n.d.). *Indonesian Automobile Industry Data*. Gabungan Industri Kendaraan Bermotor Indonesia. Retrieved February 2, 2024, from https://www.gaikindo.or.id/indonesian-automobile-industry-data/

Hamilton, J. D. (2018). *Why you should never use the Hodrick–Prescott filter*. Review of Economics and Statistics, 100(5), 831–843.

Iswari A, Angraini Y, & Masjkur M. (2022). Comparison of The SARIMA Model and Intervention in Forecasting The Number of Domestic Passengers at Soekarno-Hatta International Airport. *Indonesian Journal of Statistics and Its Applications*, *Vol 6 No 1*(1), 132–146. https://doi.org/10.29244/ijsa.v6i1p132-146

Kilian, L., & Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Lorensya, Ambarwati L, & Mahatma Y. (2022). Analisis Model Intervensi Multi Input Fungsi Pulse dan Fungsi Step dengan Deteksi Outlier terhadap Harga Saham. *Jurnal EurekaMatika*, *Volume 10*(Issue 1), 61–72.

Masrawanti, Wahyuningsih S, & Hayati M N. (2019). Analisis Model Intervensi Fungsi Step Ganda untuk Peramalan Inflasi Indonesia (Studi Kasus: Inflasi Indonesia Tahun Analysis of Intervention Models Double Step Function for Forecasting Indonesia's Inflation (Case Study: Indonesia's Inflation in 2009-2017). *Jurnal EKSPONENSIAL*, *10 Nomor 2*(2), 119–126.

Papacharalampous, G., & Tyralis, H. (2018). *Evaluation of machine learning algorithms for time series forecasting*. Water Resources Research, 54(8), 5597–5613.

Primawati A, Sitanggang I S, Annisa, & Astuti D A. (2023). Perbandingan Kinerja Lstm dan Prophet untuk Prediksi Deret Waktu (Studi Kasus Produksi Susu Sapi Harian). *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, *Vol. 9, No. 3*, 428–435.

Ribeiro, M. H. D. M., Coelho, L. S., & dos Santos, A. G. (2020). *Multi-step ahead forecasting with deep learning*. Neurocomputing, 396, 173–184.

Sautomo S, & Pardede HF. (2021). Prediksi Belanja Pemerintah Indonesia Menggunakan Long Short-Term Memory (LSTM). *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(1), 99–106. https://doi.org/10.29207/resti.v5i1.2815

Schaffer AL, Dobbins TA, & Pearson SA. (2021). Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. *BMC Medical Research Methodology*, *21*(1). https://doi.org/10.1186/s12874-021-01235-8

Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications* (4th ed.). Springer.

Sirisha U M, Belavagi M C, & Attigeri G. (2016). Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison. *IEE Access*, *Vol 4*, 1–13. https://doi.org/10.1109/ACCESS.2017.DOI

Sohibien GPD. (2018). Analysis of The Effect of Fuel Price Policy on Jakarta Inflation by Using Multi-Input Intervention Model. *AIP Conference Proceedings*, 1–11. https://doi.org/10.1063/1.5054529

Taylor, J. W., & Letham, B. (2018). *Forecasting at scale*. The American Statistician, 72(1), 37–45.

Wei W W S. (2006). *Time Series Analysis Univariate and Multivariate Methods* (Second).

Wiladibrata M I, & Rifai N A K. (2022). Peramalan Produksi Mobil Menggunakan Metode Double Exponential Smoothing Dengan Algoritma Golden Section. *Bandung Conference Series: Statistics*, *Vol 2 No 2*, 507–511. https://doi.org/10.29313/bcss.v2i2.4776

Zhang, Y., & Yang, Q. (2021). *A survey on multi-task learning*. IEEE Transactions on Knowledge and Data Engineering, 34(12), 5586–5609.