

Forecasting the Illiteracy Rate in Indonesia Using Time Series Methods With ARIMA and Holt-Winters Models

Muhammad Fahmy Adhiputra^{1✉}, Ana Kurniawati²

^{1,2} Business Information System, Computer Science Faculty, Universitas Gunadarma, Indonesia

✉ **Corresponding Author** : Muhammad Fahmy Adhiputra (e-mail: fahmyadhiputra@gmail.com)

Article Information	ABSTRACT
Article History Received : April 23, 2026 Revised : June 04, 2026 Published : July 04, 2026 Keywords: <i>Illiteracy Rate;</i> <i>Time Series Forecasting;</i> <i>ARIMA;</i> <i>Holt-Winters;</i> <i>Indonesia</i>	Illiteracy remains an important educational issue in Indonesia despite various government literacy programs. However, limited studies have compared time series forecasting methods for predicting illiteracy rates in Indonesia. Therefore, this study aims to identify the most accurate forecasting model using the Autoregressive Integrated Moving Average (ARIMA) and Holt-Winters methods. Annual illiteracy rate data from 2011 to 2021 were analyzed using the Box-Jenkins and Exponential Smoothing approaches. The novelty of this study lies in the comparative evaluation of ARIMA and Holt-Winters models in forecasting national illiteracy rates. Model performance was assessed using Root Mean Square Error (RMSE) and Mean Square Error (MSE). The results indicate that ARIMA(0,0,1) is the best ARIMA model, while Holt-Winters with $\alpha = 0.8$ and $\beta = 0.9$ provides the highest forecasting accuracy. Holt-Winters produced lower error values (RMSE = 0.33; MSE = 0.11) than ARIMA(0,0,1) (RMSE = 0.73; MSE = 0.54). This study contributes empirical evidence regarding the effectiveness of Holt-Winters for forecasting educational indicators and supports data-driven educational policy planning in Indonesia.

INTRODUCTION

Illiteracy remains a major educational challenge that limits individuals' ability to access information, acquire knowledge, develop skills, and actively participate in social and economic activities. Individuals with low literacy levels often face difficulties in adapting to technological developments and labor market demands, which may contribute to persistent poverty and social inequality. In Indonesia, education is recognized as a fundamental national objective, as stated in the 1945 Constitution and further reinforced through Law No. 20 of 2003 concerning the National Education System. Consequently, reducing illiteracy rates has become one of the government's priorities in improving the quality of human resources.

Various literacy programs have been implemented through formal, non-formal, and informal education pathways, including the National Movement for the Acceleration of Nine-Year Basic Education and Illiteracy Eradication. Despite these efforts, fluctuations in illiteracy rates continue to occur. High implementation costs, lengthy learning periods, and the tendency of some participants to relapse into illiteracy indicate that existing programs have not yet achieved optimal effectiveness. According to data from the Central Bureau of Statistics (BPS), the percentage of illiterate populations in Indonesia still varies across years, highlighting the importance of evidence-based planning and evaluation.

Forecasting techniques can support policymakers in anticipating future trends and allocating resources more effectively. Among the most widely used forecasting approaches are the

Autoregressive Integrated Moving Average (ARIMA) model and the Holt-Winters Exponential Smoothing model. ARIMA has been successfully applied to forecast inflation (Yahia & Abdellah, 2018), unemployment (Dritsaki, 2016), automobile sales (Shakti et al., 2017), and demand patterns (Fattah et al., 2018). Similarly, Holt-Winters has demonstrated reliable forecasting performance in unemployment studies (Dumi et al., 2015), natural gas production forecasting (Chase, 2019), and COVID-19 case prediction (Al et al., 2020). These findings suggest that both methods are capable of modeling time-dependent data and generating accurate forecasts across various domains.

Table 1. Summary of Previous Studies

Authors	Research Object	Method	Main Findings
Dritsaki (2016)	Unemployment Rate	SARIMA	Produced accurate unemployment forecasts
Shakti et al. (2017)	Automobile Sales	ARIMA	ARIMA effectively predicted sales trends
Fattah et al. (2018)	Demand Forecasting	ARIMA	ARIMA showed satisfactory forecasting accuracy
Chase (2019)	Natural Gas Production	Holt-Winters	Holt-Winters captured trend patterns effectively
Al et al. (2020)	COVID-19 Cases	ARIMA and Holt-Winters	Holt-Winters outperformed ARIMA in several forecasting scenarios
This Study	Illiteracy Rate in Indonesia	ARIMA & Holt-Winters	Comparison of forecasting performance on educational indicators

Although previous studies have extensively examined ARIMA and Holt-Winters models in economic, industrial, and public health contexts, limited research has compared these methods for forecasting illiteracy rates in Indonesia. Existing studies mainly focus on inflation, unemployment, sales, demand, and disease prediction, leaving a gap in the application of time series forecasting techniques to educational indicators, particularly illiteracy rates. Therefore, empirical evidence regarding the most suitable forecasting model for Indonesian illiteracy data remains limited.

The novelty of this study lies in the comparative evaluation of ARIMA and Holt-Winters models for forecasting illiteracy rates in Indonesia using annual data from 2011–2021. Unlike previous studies that focused primarily on economic and health indicators, this research applies both forecasting approaches to an educational indicator and systematically compares their predictive performance using Root Mean Square Error (RMSE) and Mean Square Error (MSE).

This study contributes to the forecasting literature by providing empirical evidence regarding the suitability of ARIMA and Holt-Winters models for educational time series data. In addition, the findings offer practical insights for policymakers by identifying a forecasting approach that can support data-driven literacy planning and evaluation.

Therefore, the objective of this study is to determine the most accurate time series forecasting model between ARIMA and Holt-Winters for predicting illiteracy rates in Indonesia.

RESEARCH METHODS

This study employed a quantitative approach using time series forecasting techniques to predict illiteracy rates in Indonesia. The data used were secondary data obtained from the Central Bureau of Statistics (BPS), consisting of annual illiteracy rates from 34 provinces during the period 2011–2021. The dataset contained 374 observations, representing the percentage of illiterate population in each province for each year. Data processing and forecasting analysis were conducted using EViews 12 software.

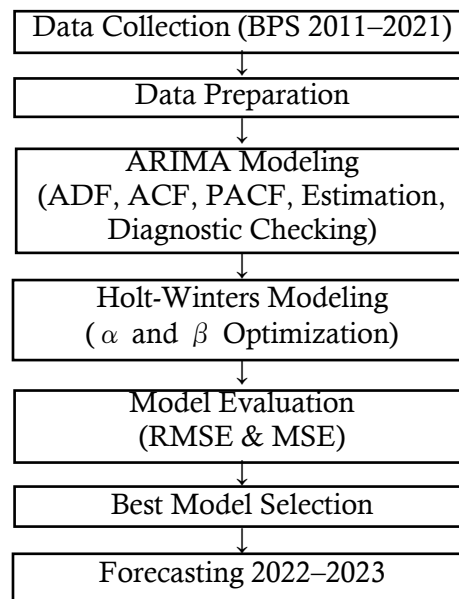


Figure 1. Research Framework

The research procedure consisted of four main stages: data preparation, model development, model evaluation, and forecasting. Two forecasting approaches were applied, namely the Box-Jenkins method using the Autoregressive Integrated Moving Average (ARIMA) model and the Exponential Smoothing method using the Holt-Winters model. Figure 1 illustrates the overall research framework used in this study.

ARIMA Model

The ARIMA model is generally expressed as:

$$\text{ARIMA}(p,d,q)$$

where p represents the autoregressive order, d represents the differencing order, and q represents the moving average order.

The ARIMA modeling procedure followed the Box-Jenkins methodology, consisting of model identification, parameter estimation, diagnostic checking, and forecasting. Data stationarity was examined using the Augmented Dickey-Fuller (ADF) test. If non-stationarity was detected, differencing was performed until stationarity was achieved. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) were then used to identify candidate models. Parameter estimation was evaluated using Akaike Information Criterion (AIC), Schwarz Bayesian Criterion (SBC), and Adjusted R-Squared values. Diagnostic checking was conducted to ensure that the residuals satisfied the white-noise assumption.

Holt-Winters Model

The Holt-Winters method was applied to capture trend patterns in the data through level and trend smoothing components. The model is formulated as follows:

Level:

$$L_t = \alpha Y_t + (1 - \alpha) (L_{t-1} + T_{t-1})$$

Trend:

$$T_t = \beta (L_t - L_{t-1}) + (1 - \beta) T_{t-1}$$

Forecast:

$$F_{t+m} = L_t + mT_t$$

where α denotes the level smoothing parameter, β denotes the trend smoothing parameter, L_t denotes the level component, T_t denotes the trend component, and F_{t+m} denotes the forecast value for period m . Several combinations of α and β values were tested to obtain the model with the lowest forecasting error.

Model Evaluation

The forecasting performance of the ARIMA and Holt-Winters models was evaluated using Root Mean Square Error (RMSE) and Mean Square Error (MSE). These measures are calculated as follows:

$$\begin{aligned} \text{RMSE} &= \sqrt{[\sum (Y_t - \hat{Y}_t)^2 / n]} \\ \text{MSE} &= [\sum (Y_t - \hat{Y}_t)^2 / n] \end{aligned}$$

where Y_t represents the actual value, $\hat{Y}_t$ represents the predicted value, and n represents the number of observations. The model with the lowest RMSE and MSE values was selected as the best forecasting model. Finally, the selected model was used to forecast Indonesia's illiteracy rate for the period 2022–2023.

RESULTS AND DISCUSSION

Stationarity Test

The Augmented Dickey-Fuller (ADF) test was conducted to examine data stationarity. The ADF statistic was -7.855012 with a probability value of 0.0011, which is below the 0.05 significance level. Therefore, the data were stationary at level and suitable for ARIMA modeling.

Table 1. ADF Test Results

Statistic	Value
ADF Statistic	-7.855012
Probability	0.0011
Critical Value (5%)	-4.107833

ARIMA and Holt-Winters Model Results

Based on the ACF and PACF analyses, three candidate ARIMA models were identified: ARIMA (1,0,0), ARIMA (1,0,1), and ARIMA (0,0,1). After parameter estimation and diagnostic checking, ARIMA (0,0,1) was selected as the best ARIMA model with RMSE = 0.73 and MSE = 0.54. For the Holt-Winters method, several combinations of smoothing parameters were evaluated. The optimal model was obtained with $\alpha = 0.8$ and $\beta = 0.9$, producing RMSE = 0.33 and MSE = 0.11.

Model Comparison

Table 2. Comparison of Forecasting Accuracy

Model	RMSE	MSE
ARIMA (0,0,1)	0.73	0.54
Holt-Winters	0.33	0.11

The Holt-Winters model outperformed ARIMA (0,0,1), as indicated by lower RMSE and MSE values. This result suggests that Holt-Winters is more effective in capturing the trend pattern of illiteracy rates in Indonesia. Similar findings were reported by Al et al. (2020), who found that Holt-Winters achieved higher forecasting accuracy than ARIMA for trend-dominated time series data.

Forecasting Results



Figure 1. Actual and Forecasted Illiteracy Rates in Indonesia (2011–2023)

Figure 1 shows that the Holt-Winters model closely follows the historical trend of illiteracy rates. The forecasting results indicate a continued decline in illiteracy rates during 2022–2023, suggesting that literacy improvement programs may continue to contribute to reducing illiteracy in Indonesia.

CONCLUSION

This study aimed to determine the most accurate time series model for forecasting the illiteracy rate in Indonesia by comparing the ARIMA and Holt-Winters methods. Based on the analysis, the Holt-Winters model outperforms the ARIMA model in terms of forecasting accuracy. The ARIMA(0,0,1) model, which is the best model in the Box-Jenkins approach, produces RMSE = 0.73 and MSE = 0.54. In contrast, the Holt-Winters model with parameters $\alpha = 0.8$ and $\beta = 0.9$ yields significantly lower error values, with RMSE = 0.33 and MSE = 0.11. These results indicate that the Holt-Winters model is more reliable for forecasting illiteracy rates in Indonesia.

The findings also show that the illiteracy rate tends to decline over time, although fluctuations still occur, indicating that literacy programs have not yet achieved consistent effectiveness. Therefore, the use of accurate forecasting models such as Holt-Winters can support policymakers in designing more effective literacy programs.

For future research, it is recommended to use a larger dataset to improve forecasting accuracy. In addition, other forecasting methods can be explored to provide comparative results. The use of more detailed data, such as monthly data, is also suggested to produce more representative and reliable forecasting outcomes.

ACKNOWLEDGEMENTS

The author would like to express sincere gratitude to Universitas Gunadarma, including the Rector, the Director of the Master Program in Technology and Engineering, and the Head of the Management Information Systems Study Program, for their institutional support. Special appreciation is extended to Dr. Ana Kurniawati for her invaluable guidance, support, and constructive feedback throughout the completion of this research. The author is also deeply

grateful to his family for their continuous encouragement, prayers, and unwavering support. Appreciation is further extended to fellow students of Batch 57 MMSI for their cooperation and shared experiences. Finally, the author acknowledges the Central Bureau of Statistics (Badan Pusat Statistik/BPS) for providing the data used in this study.

REFERENCES

- Ahmed, M., Hassan, S., & Ali, R. (2024). Forecasting socio-economic indicators using ARIMA and Holt-Winters models. *Sustainability*, 16(3), 1458.
- Al, F., Idress, R., & Sameeh, A. (2020). Prediction of the daily number of confirmed cases of COVID-19 in Sudan using ARIMA and Holt-Winters exponential smoothing. *International Journal of Development Research*. <https://doi.org/10.37118/ijdr.19811.08.2020>
- Badan Pusat Statistik. (2021). Persentase penduduk buta huruf. <https://www.bps.go.id/indicator/28/102/1/persentase-penduduk-buta-huruf.html>
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.
- Chase, O. (2019). Holt-Winters forecasting for Brazilian natural gas production. *International Journal of Innovation Education and Research*, 7(6), 119–129. <https://doi.org/10.31686/ijer.vol7.iss6.1559>
- Donate, J. P., Li, X., Sánchez, G. G., & de Miguel, A. S. (2019). Time series forecasting by evolving artificial neural networks with genetic algorithms, differential evolution and estimation of distribution algorithm. *Neural Computing and Applications*, 22(1), 11–20. <https://doi.org/10.1007/s00521-011-0741-0>
- Dritsaki, C. (2016). Forecast of SARIMA models: An application to unemployment rates of Greece. *American Journal of Applied Mathematics and Statistics*, 4(5), 136–148. <https://doi.org/10.12691/ajams-4-5-1>
- Dumi, K., Čeh Časni, A., & Žmuk, B. (2015). Forecasting unemployment rate in selected European countries using smoothing methods. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, 9(4), 1073–1078.
- Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, 10, 1–9. <https://doi.org/10.1177/1847979018808673>
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.
- Kourentzes, N., Fildes, R., & Petropoulos, F. (2023). Forecasting with exponential smoothing: A review. *International Journal of Forecasting*, 39(2), 567–590.
- Momin, B., & Chavan, G. (2018). Univariate time series models for forecasting stationary and non-stationary data: A brief review. In S. C. Satapathy & A. Joshi (Eds.), *Information and communication technology for intelligent systems (ICTIS 2017)* (pp. 219–226). Springer.
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. Wiley.
- Petropoulos, F., Makridakis, S., Assimakopoulos, V., & Nikolopoulos, K. (2022). Forecasting: Theory and practice. *International Journal of Forecasting*, 38(3), 705–871.
- Shakti, S. P., Hassan, M. K., Zhenning, Y., Caytiles, R. D., & N. Ch. S. N., I. (2017). Annual automobile sales prediction using ARIMA model. *International Journal of Hybrid Information Technology*, 10(6), 13–22. <https://doi.org/10.14257/ijhit.2017.10.6.02>

- Wang, X., Li, Y., & Zhang, H. (2023). Comparative analysis of ARIMA and exponential smoothing methods for time series forecasting. *Journal of Forecasting*, 42(4), 712–725.
- W. Gikungu, S. (2015). Forecasting inflation rate in Kenya using SARIMA model. *American Journal of Theoretical and Applied Statistics*, 4(1), 15.
<https://doi.org/10.11648/j.ajtas.20150401.13>
- Yahia, B. M. A., & Abdellah. (2018). Forecasting of Sudan inflation rates using ARIMA model. *International Journal of Economics and Financial Issues*, 8(3), 17–22.