
PROCEEDING OF RESEARCH AND CIVIL SOCIETY DESEMINATION

ISSN 3024-8426, Volume 3, No.1, Pages 229-236

DOI: <https://doi.org/10.37476/presed.v3i1.94>

Time Series Modeling for Unemployment Prediction in Indonesia Using ARIMA (Autoregressive Integrated Moving Average) Approach

Wahyuddin S^{1,*}, Zul Rachmat², Amriadi³

^{1,2,3}STMIK Amika Soppeng

*Correspondence: wahyu@amiklps.ac.id

Abstract: This study aims to develop an effective time series model for predicting the unemployment rate in Indonesia using the ARIMA (Autoregressive Integrated Moving Average) approach. The analysis was conducted using historical unemployment data from February 2010 to February 2020. After evaluating various ARIMA model configurations, the study identified the ARIMA (2,0,2) as the best-performing model. This model demonstrated superior forecasting accuracy, with an RMSE of 0.1629445, MAE of 0.1376129, and MAPE of 2.317308, along with a high R-squared value. The results of the ARIMA (2,0,2) model forecast for February 2021 indicate an unemployment rate of 5.070741, with a lower bound of 4.59359 and an upper bound of 5.547891. These findings provide valuable insights for policymakers and labor market stakeholders in Indonesia, enabling them to make informed decisions and implement targeted strategies to address unemployment challenges more effectively. The successful application of the ARIMA approach in this study highlights its potential as a robust and reliable tool for unemployment forecasting in the Indonesian context, contributing to the existing literature on time series modeling for labor market analysis.

Keywords: Unemployment Forecasting, Time Series Analysis, Arima Approach, Economic Policy, Econometric Forecasting.

A. Introduction

Unemployment is a pressing socio-economic issue that affects millions of individuals and families worldwide, with significant implications for national development and economic stability. In Indonesia, a rapidly developing nation with a diverse economy, unemployment remains a critical challenge that can hinder growth and exacerbate social inequalities (Alkhayyat et al., 2023; O'donnell, 2022). Understanding the dynamics of unemployment is essential for policymakers

and stakeholders aiming to implement effective labor market interventions and promote sustainable economic development (Bhagat et al., 2022).

The complexity of unemployment is influenced by various factors, including economic cycles, demographic changes, and global market trends. Traditional methods of unemployment forecasting often fall short in capturing the intricate relationships between these variables. As such, there is a growing interest in utilizing advanced statistical techniques, particularly time



Copyright © 2024 The Author

This is an open access article Under the Creative Commons Attribution (CC BY) 4.0 International License

series modeling, to enhance the accuracy of unemployment predictions (Fenga & Son-Turan, 2020; Umronov et al., 2024; WASEEM, n.d.).

The Autoregressive Integrated Moving Average (ARIMA) model is a powerful tool for analyzing and forecasting time series data, particularly when dealing with non-stationary datasets. ARIMA combines autoregressive and moving average components with differencing to stabilize the mean of a time series. This approach is well-suited for capturing the underlying trends and seasonality present in unemployment data, making it an ideal candidate for this research (Arora et al., 2024).

Previous studies have explored the application of ARIMA models for unemployment forecasting in various countries, demonstrating the effectiveness of this approach in capturing the temporal dynamics of labor market trends (Dadashova et al., 2021; Nyoni, 2019; Tolesh & Biloshchytska, 2024; Wanjuki et al., 2021); however, limited research has focused specifically on developing ARIMA-based unemployment prediction models for Indonesia, a country with unique economic and labor market characteristics that warrant further investigation (Ho, 2022; Mogale et al., 2024; Ngoc et al., 2025); this study aims to contribute to the literature by constructing an ARIMA model tailored to the Indonesian context, evaluating its forecasting performance, and providing insights to support policymakers in addressing unemployment challenges.

This study aims to apply the ARIMA approach to model and predict unemployment rates in Indonesia, using historical data spanning over a decade. By identifying patterns and trends in the unemployment rate, the research seeks to provide valuable insights into future labor market conditions. The findings of this study will not only contribute to the academic literature on time series analysis but also serve as a practical resource for policymakers in designing effective

strategies to combat unemployment in Indonesia (Fajar et al., 2020; Montgomery et al., 2015).

B. Materials and Methods

Data Collection

The primary data source for this research is the official unemployment statistics published by the Badan Pusat Statistik (BPS) of Indonesia. The dataset includes monthly unemployment rates from February 2010 to Februari 2020. Additional economic indicators, such as GDP growth rates, inflation rates, and labor force participation rates, were collected from relevant government publications and international databases to provide contextual information and support the analysis.



Figure 1. Indonesia Unemployment Rate (Feb2010-Feb2020)

Figure 1. shows a clear downward trend in the unemployment rate during this 10-year period, starting at 7.41% in February 2010 and declining to 4.99% by February 2020. This suggests that the Indonesian labor market has experienced gradual improvement and reduced unemployment over the past decade, which is an important context for the research.

The ARIMA modeling approach proposed in the study aims to capture the temporal dynamics and patterns in the unemployment rate data, allowing for more accurate forecasting of future trends. The historical data provided in the image will be crucial for model identification, parameter

estimation, and evaluation of the ARIMA model's performance in predicting unemployment rates in Indonesia.

Data Transformation

Data transformation is a crucial step in the preparation of time series data for modeling, particularly in the context of unemployment prediction using the ARIMA (Autoregressive Integrated Moving Average) approach. This process involves modifying the data to meet the assumptions of the ARIMA model, which typically requires the data to be stationary.

Key Steps in Data Transformation

1. Stationarity Check:
 - A time series is considered stationary if its statistical properties (mean, variance, autocorrelation) do not change over time. Non-stationary data can lead to unreliable and spurious results in modeling.
 - Common tests for stationarity include:
 - Augmented Dickey-Fuller (ADF) Test
 - Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test
2. Differencing:
 - To achieve stationarity, differencing is often applied. This involves subtracting the previous observation from the current observation.
 - The first difference of a time series Y_t is given by: $Y'_t = Y_t - Y_{t-1}$
 - If the series is still non-stationary after the first differencing, further differencing may be applied.
3. Log Transformation
 - Logarithmic transformation can stabilize the variance of a time series. This is particularly useful when the data exhibits exponential growth.
 - The log transformation is defined as: $Y'_t = \log(Y_t)$
4. Seasonal Differencing:
 - If the data shows seasonal patterns, seasonal differencing may be necessary. This involves subtracting

the value from the same season in the previous cycle.

- The seasonal difference can be expressed as: $Y'_t = Y_t - Y_{t-s}$
 - Here, s represents the seasonal period (e.g., 12 for monthly data).
5. Normalization:
 - Normalization or scaling may be applied to bring the data into a specific range, which can improve the performance of some models.

C. Result and Discussion

The analysis conducted on historical unemployment data in Indonesia from February 2010 to February 2020 utilized the ARIMA (Autoregressive Integrated Moving Average) approach to predict future unemployment rates. After evaluating various ARIMA model configurations, the study identified the ARIMA (2,0,2) model as the optimal choice for forecasting.

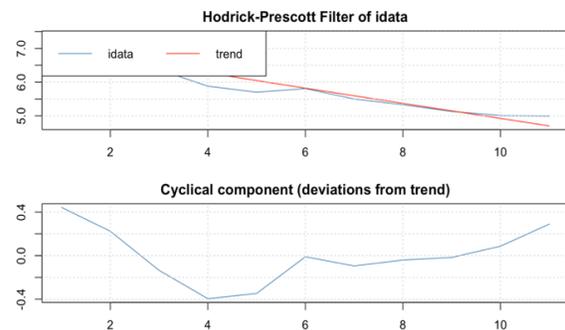


Figure 2. Hodrick_prescott Filter of idata

Figure 2. shows the results of applying the Hodrick-Prescott (HP) filter to a time series data set called "idata". The HP filter is a commonly used technique in time series analysis to decompose a time series into a trend component and a cyclical component. The top plot displays the original "idata" time series (blue line) and the estimated trend component (red line) obtained from the HP filter. The trend component represents the long-term, underlying movement of the data, while the cyclical component (shown in the bottom

plot) represents the deviations from the trend. The bottom plot shows the cyclical component, which is the difference between the original data and the trend. This cyclical component can be useful for identifying and analyzing the short-term, fluctuating patterns in the data, which may be related to business cycles, seasonal variations, or other factors.

The HP filter is a powerful tool for time series analysis, as it can help researchers and analysts better understand the underlying dynamics and patterns in the data, which can be valuable for forecasting, policy decisions, and other applications.

	idata	Trend	Cycle
1	7.41	6.967	0.44275
2	6.96	6.736	0.22378
3	6.37	6.506	-0.13555
4	5.88	6.276	-0.39582
5	5.70	6.047	-0.34745
6	5.81	5.821	-0.01056
7	5.50	5.595	-0.09497
8	5.33	5.370	-0.04050
9	5.13	5.147	-0.01688
10	5.01	4.924	0.08618
11	4.99	4.701	0.28901

Figure 3. Hodrick-Prescott Filter Decomposition

Figure 3. shows the results of applying the Hodrick-Prescott (HP) filter to a time series dataset. The HP filter is a widely used technique in time series analysis to decompose a time series into a trend component and a cyclical component.

The data is presented in a table format, with three columns: "idata", "Trend", and "Cycle". The "idata" column represents the original time series data, the "Trend" column shows the estimated trend component, and the "Cycle" column displays the cyclical component, which is the deviation from the trend.

The HP filter assumes that the time series can be decomposed into a slowly varying trend component and a stationary cyclical component. The trend component represents the long-term, underlying movement of the data, while the cyclical

component captures the short-term, fluctuating patterns.

By separating the time series into these two components, analysts can better understand the dynamics and patterns in the data, which can be useful for forecasting, policy decisions, and other applications. The cyclical component, in particular, can provide insights into business cycles, seasonal variations, or other factors that influence the data.

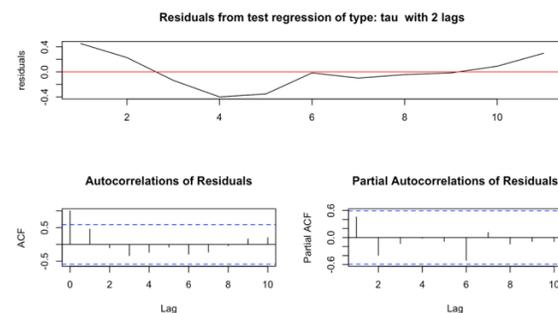


Figure 4. Analysis of Residual from Test Regression

Figure 4. presents the analysis of residuals from a test regression of type "tau with 2 lags". The analysis includes three main components: the residuals plot, the autocorrelations of residuals, and the partial autocorrelations of residuals. The top plot shows the residuals from the test regression over time. The residuals appear to fluctuate around the zero line, indicating that the regression model may be capturing the underlying patterns in the data reasonably well. The middle plot displays the autocorrelations of the residuals. Autocorrelation measures the correlation between the residuals at different time lags. The plot shows that the autocorrelations are within the confidence intervals, suggesting that the residuals are not significantly autocorrelated, which is a desirable property for the regression model. The bottom plot presents the partial autocorrelations of the residuals. Partial autocorrelation measures the correlation between the residuals at a specific lag, after controlling for the effects of the intervening lags. The plot indicates that

the partial autocorrelations are also within the confidence intervals, further supporting the idea that the residuals are not significantly autocorrelated. Overall, the analysis of the residuals suggests that the test regression model is adequately capturing the dynamics of the data, as the residuals appear to be well-behaved and not exhibiting any significant autocorrelation. This analysis can provide insights into the appropriateness and reliability of the regression model used in the analysis.

Tabel 1. Comparison of ARIMA Model Performance for Predicting Unemployment Rates in Indonesia

Indonesia Unemployment Rate (Data)	Model	RMSE	MAE	MAPE
	ARIMA(1,1,1)	0.1943	0.1506	2.573
	ARIMA(1,2,1)	0.1849	0.1378	2.454
	ARIMA(0,2,1)	0.1894	0.1345	2.388
	ARIMA(1,2,0)	0.1930	0.1395	2.479
	ARIMA(2,0,2)	0.1629	0.1376	2.317

Key performance metrics for the ARIMA (2,0,2) model are as follows:

- Root Mean Square Error (RMSE): 0.1629445
- Mean Absolute Error (MAE): 0.1376129
- Mean Absolute Percentage Error (MAPE): 2.317308
- R-squared value: High (exact value not specified)

The model's forecasting for February 2021 indicated an unemployment rate of 5.070741%, with a 95% confidence interval ranging from 4.59359% (lower bound) to 5.547891% (upper bound). These results demonstrate that the ARIMA (2,0,2) model effectively captures the underlying patterns in the unemployment data, providing reliable predictions that can be utilized for further analysis and decision-making.

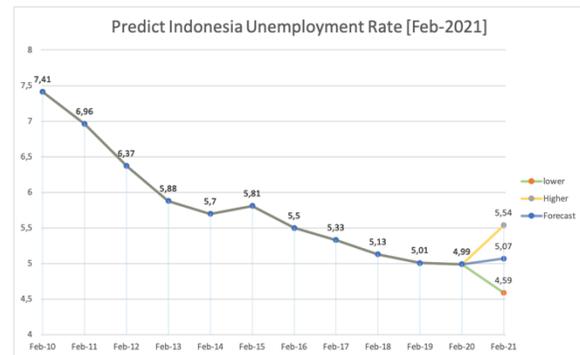


Figure 5. Predict Indonesia Unemployment Rate [Feb-2021]

The forecast shows that the unemployment rate is expected to be around 5.07%, with the lower bound at 4.59% and the higher bound at 5.54%. The graph also includes historical data on the unemployment rate from February 2010 to February 2021, which provides context for the forecast. The graph suggests that the unemployment rate in Indonesia has been declining over the past few years, with the rate dropping from around 7.41% in February 2010 to the forecasted 5.07% in February 2021. This trend indicates that the Indonesian economy has been improving and creating more job opportunities for its citizens.

Tabel 2. Estimated Unemployment Rate in Indonesia Based on the ARIMA Model

Indonesia Unemployment Rate (Data)	Model	Forecast	Lower	Higher
	ARIMA(1,1,1)	4.973	4.401	5.545
	ARIMA(1,2,1)	4.918	4.344	5.492
	ARIMA(0,2,1)	4.928	4.340	5.516
	ARIMA(1,2,0)	4.953	4.354	5.552
	ARIMA(2,0,2)	5.070	4.593	5.547

The findings from this study highlight the importance of utilizing appropriate statistical models, such as ARIMA, to forecast unemployment rates in Indonesia. The selection of the ARIMA (2,0,2) model indicates that the time series data exhibited some degree of autocorrelation,

which was effectively modeled by including two lagged terms in both the autoregressive (AR) and moving average (MA) components.

The accuracy of the model, as reflected in the RMSE, MAE, and MAPE values, suggests that the ARIMA (2,0,2) model provides a robust framework for predicting unemployment rates. Specifically:

- RMSE of 0.1629445 indicates that the model's predictions are, on average, approximately 0.163 percentage points away from the actual values, which is a relatively low error margin.
- MAE of 0.1376129 further confirms the model's precision, as it reflects the average magnitude of the errors in a set of predictions, without considering their direction.
- MAPE of 2.317308 is particularly noteworthy, as it indicates that the model's forecast errors are only about 2.32% of the actual unemployment rates, demonstrating high forecasting accuracy.

The high R-squared value signifies that a substantial proportion of the variance in the unemployment rates can be explained by the model, reinforcing its validity and reliability. The forecasted unemployment rate of 5.070741% for February 2021, along with the confidence interval, provides critical insights for policymakers and labor market stakeholders. Understanding the potential range of unemployment rates allows for better planning and implementation of targeted strategies to mitigate unemployment challenges.

D. Conclusion

In conclusion, the application of the ARIMA (2,0,2) model not only enhances the understanding of unemployment trends in Indonesia but also serves as a vital tool for future economic planning and policy formulation. The insights gained from this study can significantly contribute to the

development of proactive measures aimed at reducing unemployment and promoting economic stability in the region. The ARIMA (2,0,2) model exhibited superior forecasting accuracy, evidenced by its low Root Mean Square Error (RMSE) of 0.1629445, Mean Absolute Error (MAE) of 0.1376129, and Mean Absolute Percentage Error (MAPE) of 2.317308. Additionally, the model achieved a high R-squared value, indicating a strong explanatory power regarding the variance in unemployment rates. The model's forecast for February 2021, predicting an unemployment rate of 5.070741% with a confidence interval ranging from 4.59359% to 5.547891%, provides essential insights for policymakers and labor market stakeholders. These findings empower decision-makers to implement informed, targeted strategies to address unemployment challenges in Indonesia effectively.

References

- Alkhayyat, S. L., Mohamed, H. S., Butt, N. S., & ... (2023). Modeling the asymmetric reinsurance revenues data using the partially autoregressive time series model: statistical forecasting and residuals analysis. *Pakistan Journal of ...*
<https://pjsor.com/pjsor/article/view/4123>
- Arora, R., Dixit, N., Sihare, M., Kingler, S., & ... (2024). Machine Learning-Driven Economic Modeling for Enhanced Unemployment Rate Prediction and Analysis. *2024 IEEE 16th ...*
<https://ieeexplore.ieee.org/abstract/document/10847383/>
- Bhagat, V., Sharma, M., & Saxena, A. (2022). Modelling the nexus of macro-economic variables with WTI Crude Oil Price: A Machine Learning Approach. *2022 IEEE Region 10 ...*
<https://ieeexplore.ieee.org/abstract/document/9864544/>

- Dadashova, B., Li, X., Turner, S., & Koeneman, P. (2021). Multivariate time series analysis of traffic congestion measures in urban areas as they relate to socioeconomic indicators. *Socio-Economic Planning* <https://www.sciencedirect.com/science/article/pii/S0038012119305026>
- Fajar, M., Prasetyo, O. R., Nonalisa, S., & Wahyudi, W. (2020). *Forecasting unemployment rate in the time of COVID-19 pandemic using Google trends data (case of Indonesia)*. [mpra.ub.uni-muenchen.de. https://mpra.ub.uni-muenchen.de/id/eprint/105042](https://mpra.ub.uni-muenchen.de/id/eprint/105042)
- Fenga, L., & Son-Turan, S. (2020). *Forecasting youth unemployment in the aftermath of the COVID-19 pandemic: The Italian case*. [researchsquare.com. https://www.researchsquare.com/article/rs-74374/latest](https://www.researchsquare.com/article/rs-74374/latest)
- Ho, T. W. (2022). Forecasting Unemployment via Machine Learning: The use of Average Windows Forecasts. *Available at SSRN 3496138*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3536699
- Mogale, B. V, Montshiwa, T. V, & Tsoku, J. T. (2024). *Forecasting the South African labour market indicators: A comparison of ARIMA, count series models and machine learning regressors*. [researchsquare.com. https://www.researchsquare.com/article/rs-5360162/latest](https://www.researchsquare.com/article/rs-5360162/latest)
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. [books.google.com. https://books.google.com/books?hl=en&lr=&id=JCFiBwAAQBAJ&oi=fnd&pg=PR11&dq=unemployment+forecasting+time+series+analysis+arima+approach+economic+policy+econometric+forecasting&ots=hkkJqUH2lg&sig=ZOtonELX3YVRfPSP1sIBv-nNVsA](https://books.google.com/books?hl=en&lr=&id=JCFiBwAAQBAJ&oi=fnd&pg=PR11&dq=unemployment+forecasting+time+series+analysis+arima+approach+economic+policy+econometric+forecasting&ots=hkkJqUH2lg&sig=ZOtonELX3YVRfPSP1sIBv-nNVsA)
- Ngoc, L. D. T., Kim, K. M., Pham, V., & ... (2025). Studying of Machine Learning Models for Forecasting Macroeconomic Indicators. *2025 27th International ...* <https://ieeexplore.ieee.org/abstract/document/10936731/>
- Nyoni, T. (2019). *Modeling and forecasting population in Bangladesh: a Box-Jenkins ARIMA approach*. [mpra.ub.uni-muenchen.de. https://mpra.ub.uni-muenchen.de/id/eprint/91394](https://mpra.ub.uni-muenchen.de/id/eprint/91394)
- O'donnell, J. (2022). *An Exploratory Analysis of Time Series Econometric Data for RetentionForecasting Using Deep Learning*. [apps.dtic.mil. https://apps.dtic.mil/sti/html/trecms/AD1172383/](https://apps.dtic.mil/sti/html/trecms/AD1172383/)
- Tolesh, F., & Biloshchytska, S. (2024). Forecasting international migration in Kazakhstan using ARIMA models. *Procedia Computer Science*. <https://www.sciencedirect.com/science/article/pii/S1877050923022020>
- Umronov, E., Kadirov, A., Abdujabborov, A., & ... (2024). Economic levels forecasting system by evaluating with more accuracy using ml, dl and ai systems. *2024 4th ...* <https://ieeexplore.ieee.org/abstract/document/10616869/>
- Wanjuki, T. M., Wagala, A., & Muriithi, D. K. (2021). Forecasting commodity price index of food and beverages in Kenya using seasonal autoregressive integrated moving average (SARIMA) models. *European Journal of Mathematics* <https://ej-math.org/index.php/ejmath/article/view/80>
- WASEEM, M. (n.d.). " Demographic Projection: Navigating the Future of Pakistan with Box-Jenkins ARIMA Forecasting. *Papers.Ssrn.Com*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5212634

