JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 9, No. 3, July 2025, pp. 839-853

# Forecasting the Number of Dropout Student in Indonesia using ARIMA Model

**Aisyah Dhifa Az-Zahra[1*], Luthfia Azzahra Fajriati[1], Sherlyana Devita Sari[1]**
[1]Department of Mathematics, Maulana Malik Ibrahim State Islamic University, Indonesia
aisyahdhifa03@gmail.com

**Article History:**
Received : 20-05-2025
Revised : 20-06-2025
Accepted : 24-06-2025
Online : 01-07-2025

**Keywords:**
Forecasting;
Statistics;
ARIMA;
Dropout;
Government.

**ABSTRACT**

The high rate of dropout students in Indonesia remains a matter of considerable concern, as it erodes the quality of education and hinders the long-term development of human capital. The government of Indonesia has endeavored to address the issue of high dropout rates among students by implementing a range of initiatives. To demonstrate the effectiveness of this program, forecasting is necessary to measure and predict its outcomes. The purpose of this study is to utilize a time series approach, specifically the Autoregressive Integrated Moving Average (ARIMA) model, to predict the number of dropout students in the forthcoming years. This study employs a quantitative analysis using secondary data obtained from Statistics Indonesia (BPS) for the period 1970-2023. The ARIMA method is a statistical technique used to determine the most suitable forecasting model from historical data. This method has gained widespread popularity in the field of time series analysis due to its ability to manage non-stationary data effectively. The result shows that ARIMA (0,2,1) has the smallest AIC and meets the significant criteria model, also having the lowest MAPE value of 1.9%, indicating excellent forecasting accuracy. The plot of the result indicates a downward trend in the number of dropout students over the coming years. This downward trend aligns with the timeline of government interventions, suggesting a potential causal relationship between the implementation of educational support programs and the declining dropout rates. Thus, the prediction supports the effectiveness of these initiatives in mitigating dropout student in Indonesia.

Crossref
https://doi.org/10.31764/jtam.v9i3.31625


## A. INTRODUCTION

Education is a critical cornerstone for national development. Education serves an essential factor for sustainable economic growth because it provides individuals with the necessary skills and knowledge (Kopnina, 2020). However, in many developing countries, high dropout student rates pose a serious obstacle. The dropout rate significantly affects by hindering individual and societal progress (Mduma et al., 2019). Indonesia, one of developing country, also grappling this significant challenge. According to Statistics Indonesia, as of June 2023, 0,67% of primary school children are dropping out from the school indicating that 1 in every 1.000 student did not continue their education. In fact, primary school is the most crucial part of education that guides children toward higher level of learning. On the other hand, critical thinking begins to be develop at the primary school age (Florea & Hurjui, 2015). The percentage of dropout student in primary school is considerably lower compared with the percentage of dropout student in middle school and senior high school. The dropout rate is reached 6,93% and 21,61%.

The high number of dropout student is influenced by various factor. According to Yaneri et al. (2022), a multitude of factors have been identified as contributing to the phenomenon of school dropout, which can be categorized into two distinct categories: internal and external factors. Intelligence, motivation, cognitive abilities, and a lack of interest in schooling are all internal factors that must be taken into consideration (Singh & Alhamad, 2022). On the other hand, external factors encompass economic conditions, school-related issues, and social and cultural influences (Nurmalitasari et al., 2023). According to the research, The family's low economic status is the most dominant factor (Syafitri & Hadjam, 2017; Yaneri et al., 2022). Insufficient income often leads to a lack of attention to the child's education, thereby increasing the likelihood of school dropout. Mental health issues also take a crucial factor affecting the high rate of dropout student. A higher percentage of students with poor mental health dropped out of school, particularly at the vocational and higher education levels (Hjorth et al., 2016). In addition, socioeconomic such as low parental income are also associated with higher dropout rates, indicating that external factor alongside mental health contributes to the risk of dropping out.

To mitigate dropout rates, the Indonesian government has introduced several education programs as the Poor Student Assistance Program or "*Bantuan Siswa Miskin*", the Smart Indonesia Program or "Program Indonesia *Pintar*", and the School Operational Assistance Program or "*Bantuan Operasional Sekolah*". Several articles report that during the implementation of these program, many students in Indonesia had not obtained the programs (Marlini, 2016; Cahyaningtyas et al., 2022; Nurrokhmah, 2021). The reports suggest that the high number of students who failed to receive the programs was primarily caused by insufficient government communication and outreach to parents. As a response, the government began to intensify its outreach effort in the following years. To determine whether these efforts have been successful, forecasting can be used to predict future dropout student rate.

Forecasting is the process of predicting unreleased outcomes based on model estimates from realized data (Hegre et al., 2017). It involves making predictions about future events using past and present information, often through trend analysis (Moniz, 2022). Forecasting has been identified as a critical component in a multitude of disciplines, including education, politics, health, and economics. In education, forecasting is essential for planning student enrollment projections affecting institutional income, faculty needs, and budgets (Sinuany-Stern, 2021). Geopolitical forecasting aids decision-making for governments, organizations, and individual predicting event like elections, conflicts, and disease outbreaks (Juvina et al., 2020). The Centers for Disease Control and Prevention (CDC) has employed forecasting challenges to predict influenza-like illness trends, thereby demonstrating the application of forecasting in public health decision-making (Lutz et al., 2019). Forecasting also evolved in economics field using Bayesian methods gaining prominence due to their ability to incorporate uncertainty in models, parameters, and latent states (Martin et al., 2023).

This paper implements forecasting in education fields to predict the number of dropout student in Indonesia using Autoregressive Integrated Moving Average (ARIMA) models. It is a versatile time series analysis technique used for modeling and forecasting various phenomena. It has been applied to predict student enrollment in higher education and scholarship grant availability (Patulin, 2019). ARIMA comprises three components: Autoregressive (AR), Integrated (I), and Moving Average (MA), which can be used individually or in combination. The flexibility of these components is a key advantage of ARIMA, as it enables the model to accommodate a wide range of time series structures. This capacity renders ARIMA a highly effective tool for the analysis data (Le, 2024; Shivhare et al., 2021). The process of ARIMA modeling involves identifying the order of AR, I, and MA parts, estimating parameters, and diagnosing model adequacy. Proper model selection is crucial, as an incorrect order can lead to inaccurate forecast and misleading interpretations of trends. ARIMA models are particularly effective for short-term forecasting and can handle both stationary and nonstationary data through differentiation (Liu et al., 2020). By changing data from nonstationary data into a stationary form, ARIMA enhances the reliability of prediction and helps uncover hidden pattern.

The ARIMA models has been applied to predict college dropouts in Philippines, finding ARIMA (1,1,1) model performed best (Gambulao, 2023). The author posits that instructors and professors should strongly refer students who have accumulated absences, in accordance with the student and teacher handbook. This recommendation is based on findings that indicate a failure to report absences for intervention, as revealed by an initial interview with the Office of Student Affairs (OSA) director. Din investigated ARIMA's potential for estimating and forecasting higher education enrollment (Din, 2016). The effectiveness of the ARIMA model for forecasting student enrollment is confirmed by the finding of these study. These studies demonstrate the potential of ARIMA for dropout prediction, though their effectiveness may very depend on the specific context and data available.

According to previous studies and concerns about the high number of dropout student in Indonesia, this study aims to predict the number of students who will drop out in the fourth next year in Indonesia. By developing a reliable time series model, this research aims to provide accurate predictions that reflect ongoing trends and patterns in dropout behavior. The forecasted results are expected to serve as an early warning system for educational authorities and policymakers to take timely and targeted action. This study intends to offer insights that support the formulation of effective interventions to reduce school dropout rates. This study also can be used to assist the Indonesian government to developing strategies to push the rate of dropout student in Indonesia.

## B. METHODS

### 1. Data

The data used in this article is the primary data of dropout student in Indonesia. It can be accessed on BPS-Statistics Indonesia starting from 1970-2023. The data will be processed to forecast the number of dropout student in Indonesia. Data analysis, model selection, and forecasting process were conducted using Minitab software, which offers comprehensive tools for time series analysis.

**2. Data Stationarity Test**

a. Stationary in variance

A data stationarity test in variance assess whether the variance of a time series remains constant over time. The data is considered non-stationary in variance if the variance changes over time. In such cases, a Box-Cox transform can be employed for purpose of stabilizing the variance by transforming the data using parameter ($\lambda$) (Sharafi et al., 2017). According to Klawoon (2024) when the $\lambda$ is close to 1, its influence on the estimated of the reference interval is limited and its impact on other aspects of the process negligible. It is important to note that in cases where sample sizes are deemed to be adequate, the precision of the estimates of the population parameter known as lambda remains significant. Therefore, the data will be transformed iteratively until the estimated $\lambda$ approaches 1, ensuring that the transformation has minimal distortion on the reference interval while still improving variance stability.

b. Stationary in mean

In time series analysis, testing the mean stationarity constitutes a critical step, it is imperative that the data utilized in modelling is consistent and does not fluctuate over time. To ensure whether the data has stationary in mean we can use Augmented Dickey-Fuller (ADF) test. The ADF test can detect a unit root, which is an indication that the data is non- stationary (Guo, 2023). Several studies have been caried out to derive the general ADF formula such as the study by (Ajewole et al., 2020; Maitra & Politis, 2024; and Kim et al., 2024). To identify whether the data is stationary or not, we can make null hypotheses ($H_0$) and other hypotheses ($H_1$). The null hypotheses remains that the data was non stationary and the other hypotheses remains the data was stationary. The criteria of rejection is when the p-value under the significance level ($\alpha = 0{,}05$).

**3. Parameter Identification**

After stationarizing the time series data, the next step is to determine the order of AR and MA which is the parameter in ARIMA. Autocorrelation function (ACF) and partial autocorrelation function (PACF) are employed to identified the order of AR and MA by examining the plot (Imam, 2021). Autocorrelation refers to the correlation between current values and its past values (Riyadi et al., 2017). The ACF measures the overall strength of the relationship between current observations and their lagged values, making it useful for identifying the MA (q) component (Ratnesh & Kumar, 2019). Conversely, PACF measures the direct correlation between current values and specific lagged observations by eliminating the influence of intermediate lags, thus assisting in identifying the AR (p) component (Shumway & Stoffer, 2017).

**4. Autoregressive Integrated Moving Average (ARIMA)**

ARIMA (Autoregressive Integrated Moving Average) model, also known as the Box-Jenkins method, is widely employed in the field of time series analysis. This model is made up of three main components such as Autoregressive (AR), Integrated (I), and Moving Average (MA) are denoted by $(p, d, q)$ where $p$ represents the autoregressive, $d$ represent integrated or the number of differencing steps, and $q$ represents moving average (Utami et al., 2024). Parameters value of $p$ and $q$ can be determined using ACF and PACF (Grigonytė & Butkevičiūtė, 2016). ARIMA model is a generalization of ARMA model, both formulation have been presented in the past research such as (Box et al., 2016); (Gao et al., 2021);(Jain & Mallick, n.d.);(Murat et al., 2018). The time-dependent nature of a stationary time series denoted by $X_t$ can be modelled using an autoregressive (AR) component of order $p$. The autoregressive model of order $p$ is expressed by the following equation:

$$X_t = k + \sum_{i=1}^{p} \vartheta_i x_{t-i} + \xi_t \tag{1}$$

where $k$ is the constant, $\vartheta$ is AR model, $x_t$ is the observed value obtained at time $t$, and the value of $\xi_t$ is considered to be an error of a random value. In addition to AR terms, moving average terms of order $q$ are used, which incorporate the effects or previous error terms. The moving average model of order $q$ is given by the equation:

$$X_t = \xi_t + \sum_{j=1}^{q} \gamma_j \xi_{t-i} \tag{2}$$

where $\gamma_j$ is the MA model and $\xi_t$ is the error terms. The ARIMA model can be obtained by combining autoregressive and moving average model. The general formula of the ARIMA $(p, d, q)$ is expressed as

$$Y_t = k + \sum_{i=1}^{p} \vartheta_i y_{t-i} + \sum_{j=1}^{q} \gamma_j \xi_{t-i} \tag{3}$$

where $Y_t$ is the time series data that has been stationary, $k$ is the constant, $y_t$ is the observed value obtained at time $t$, $\vartheta_i$ is the autoregression expression, $\gamma_j$ is the moving average expression, and $\xi_t$ representing white noise errors.

## 5. Model Validation

a. White Noise Test

The white noise test is conducted to ensure that a time series process, particularly the residuals of a model, behaves randomly and exhibits no discernible pattern (Hassani et al., 2025). The process of white noise is characterized by a constant mean $E(a_t) = \mu_a$, which is typically assumed to be zero, constant in variance $var(a_t) = \sigma_a^2$, and zero covariance for all non-zero lags $\gamma_k = cov(a_t, a_{t+k}) = 0$ (Fahrin et al., 2019). The white noise test typically includes assessments of residual independence and residual normality test. To determine whether residuals from a model behave like white noise, analysts employ the white noise hypothesis test such as Ljung-Box test using the formula below

$$Q = \theta(\theta + 2) \sum_{i=1}^{j} \frac{\hat{\rho}_i^2}{\theta - j} \tag{4}$$

$\theta$ denotes the quantity of observations used, $j$ is the quantity of lags that are being subjected to testing, and $\hat{p}_i$ denotes the sample autocorrelation at lag $i$ (Luo, 2024). The hypothesis ($H_0$), which posits the absence of autocorrelation up to lag $j$. This hypothesis is considered to be rejected at the significance level $\alpha = 0{,}05$ if the $p - value$ is less than or equal to that level.

b. Akaike Information Criterion

AIC is tools for model selection in statistical modelling, including time series analysis (Cavanaugh & Neath, 2019). Introduces by Hirotugu Akaike in 1973, AIC provides a framework that integrates parameter estimation with the evaluation of a model's structure and complexity, allowing researches to balance model fit and parsimony effectively. This methods calculate a score based on how well the model fits the data using log-likelihood (Chakrabarti & Ghosh, 2011).

$$AIC = -2 \log L(\hat{\theta}) + 2k \tag{5}$$

$k$ is the parameter number and $\hat{\theta}$ is the maximum value of the likelihood function for the model (Shoko & Njuho, 2023). A lower AIC values are indicative of models that offer superior trade-offs between precision and complexity. It is particularly useful when comparing multiple candidate models fitted to the same dataset.

## 6. Accuracy Evaluation

a. Mean Absolute Percentage Error (MAPE)

MAPE is a widely used metric in time series analysis for evaluating the accuracy of forecasts. The objective of the present study is to ascertain the mean percentage discrepancy between predicted and acual values (Rahardja, 2024). Generally, lower MAPE values indicate that the forecasting model is more accurate, as it implies a smaller average prediction error relative to the actual values. According to Amini (2016), the relationship between the actual and forcasted values can be quantified using the following formula:

$$MAPE = \frac{1}{j} \sum_{i=1}^{j} \frac{|x_t - \hat{x}_t|}{x_t} \times 100\% \tag{6}$$

$n$ denotes the total number of observation, $x_t$ is the value of actual data that observed at time $t$, and $\hat{x}_t$ denotes the predicted or forecasted value. According to Chang et al. (2007), the MAPE accuracy criterion can be seen in Table 1.

**Table 1.** MAPE Accuracy Criterion

| MAPE | Signification |
|---|---|
| $< 10\%$ | Excellent Forecasting |
| $10\% - 20\%$ | Good Forecasting |
| $20\% - 50\%$ | Reasonable Forecasting |
| $> 50\%$ | Bad Forecasting |

## C. RESULT AND DISCUSSION

The time series plot of the number dropout student can be seen in figure 1.

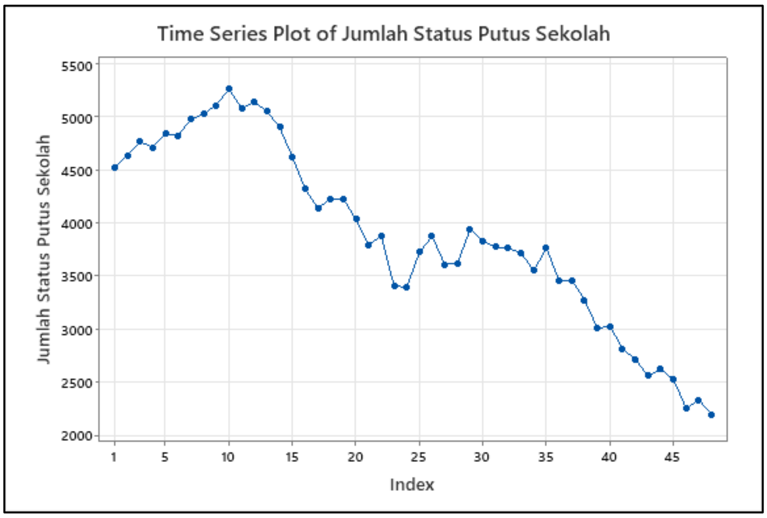


**Figure 1.** Time Series Plot of The Data

The plot presented that the data shows a downward trend over time which may indicate the positive impact of intervention policies in reducing school dropout rates.

### 1. Data Stationary Test

Stationarity test is conducted on both variance and mean. The Box-Cox plot was used to analyze whether the data has been stationary in variance. The initial trial revealed that the data exhibited non-stationary characteristics, as evidenced by the observed lambda value of 0,5. Therefore, the data transformation was required until the lambda approached 1. After first transformation the data was equal to 1, so the data can be said to be stationary in variance. The Box-Cox plot is displayed in Figure 2 and Figure 3.

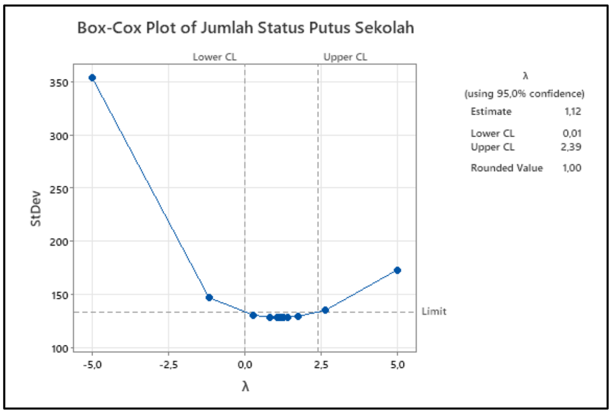


**Figure 2.** Box-Cox Plot When The $\boldsymbol{\lambda = 0,5}$

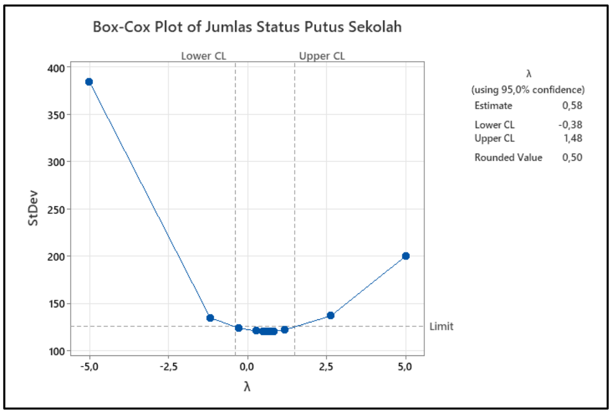


**Figure 3.** Box-Cox Plot When The $\boldsymbol{\lambda = 1}$

The stationarity in terms of mean will be tested using ADF test. The ADF test applied to the original time series produced $p$-value above the significance threshold ($\alpha = 0.05$), indicating that the null hypotesis of a unit root could not be rejected, and thus the data was non-stationary in mean. To address this, first differencing was applied. However, the result of $p$-value remained high (0.986), suggesting that first differencing was insufficent to achieve mean stationarity. Therefore, a second differencing was performed, which resulted $p$-value 0.000. This allowed to reject the null hypothesis and conlude that the data became stationary in mean. The detailed results of ADF tests are presented in Table 2.

**Table 2.** ADF Test Result

| | **$p$-value** | **Decision** |
|---|---|---|
| first differencing | 0,986 | Fail to reject $H_0$ |
| second differencing | 0,000 | Reject $H_0$ |

## 2. Model Identification

After the data has been made stationary, it can be used for model identification. The employment of ACF and PACF plots is instrumental in ascertaining the optimal sequence of autoregressive and moving average constituents within a time series model. The AR ($p$) parameter value was determined from the PACF plot and the MA ($q$) parameter value was determined from ACF plot lag that remains parameter value. The result of ACF and PACF can be seen in Figure 4.

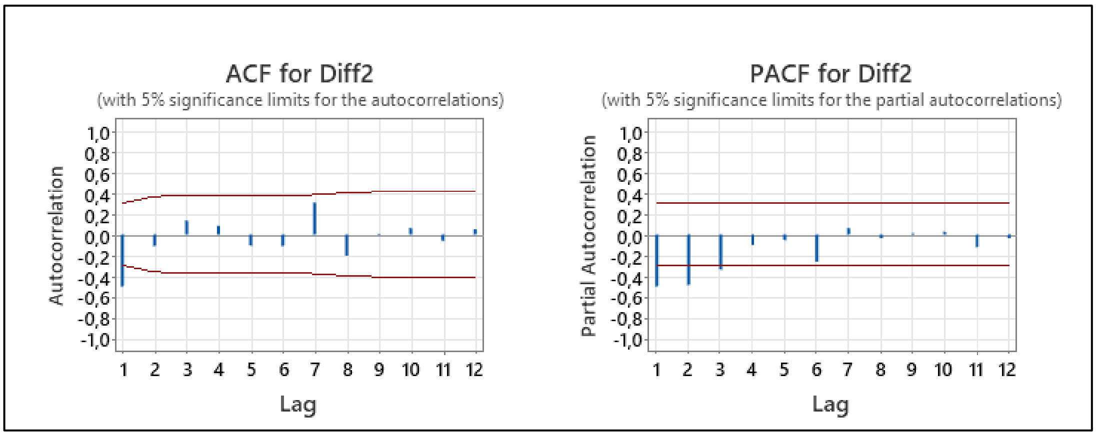


**Figure 4.** ACF and PACF Plot

Based on the ACF and PACF plots of the second differences, the ACF chart shows a significant spike at lag 1 followed by autocorrelation values that fall within the 5% significance bounds for subsequent lags. This indicates that the autocorrelation cuts off after lag 1, which is characteristic of a moving average process of order 1. Meanwhile, the PACF plot reveals a significant negative spike at lag 2 and a weaker spike at lag 3, while other lags remain within the significance bounds. This behavior suggests a possible autoregressive component at lag 2. It can be concluded that the potential ARIMA model is ARIMA(1,2,0), ARIMA(2,2,0), ARIMA(3,2,0), ARIMA(0,2,1), ARIMA(0,2,1), ARIMA(1,2,1), ARIMA(2,2,1), and ARIMA(3,2,1).

### 3. Significance and Residual White Noise Test

Significance test is used in the process of forecasting using ARIMA model to confirm the acceptability of certain parameters in model. The criteria $H_0$ that remains the model is non-significant will be rejected if $p - value < \alpha = 0{,}05$. The result of significance can be seen in the Table 3.

**Table 3.** Significance Test

| ARIMA | Parameter | P-value | Decision |
|---|---|---|---|
| (1,2,0) | $p = 1$ | 0,000 | Reject $H_0$ |
| (2,2,0) | $p = 1$ | 0,000 | Reject $H_0$ |
| | $p = 2$ | 0,001 | Reject $H_0$ |
| (3,2,0) | $p = 1$ | 0,000 | Reject $H_0$ |
| | $p = 2$ | 0,000 | Reject $H_0$ |
| | $p = 3$ | 0,024 | Reject $H_0$ |
| (0,2,1) | $q = 1$ | 0,000 | Reject $H_0$ |
| (1,2,1) | $p = 1$ | 0,620 | Fail to reject $H_0$ |
| | $q = 1$ | 0,000 | Reject $H_0$ |
| (2,2,1) | $p = 1$ | 0,000 | Reject $H_0$ |
| | $p = 2$ | 0,000 | Reject $H_0$ |
| | $q = 1$ | 0,000 | Reject $H_0$ |
| (3,2,1) | $p = 1$ | 0,000 | Reject $H_0$ |
| | $p = 2$ | 0,000 | Reject $H_0$ |
| | $p = 3$ | 0,005 | Reject $H_0$ |
| | $q = 1$ | 0,000 | Reject $H_0$ |

As presented in Table 3 the ARIMA (1,2,1) model is not meet the significance test requirements and therefore cannot be used in the subsequent test. In contrast, all parameters of ARIMA (1,2,0), ARIMA (2,2,0), ARIMA (3,2,0), ARIMA (0,2,1), ARIMA (2,2,1), and ARIMA (3,2,1) models meet significance test criteria, making them suitable for the next step. The residual test of white noise is shows. The residuals of the model were tested for white noise using the Ljung-Box test to assess autocorrelation at various lag. The result of white noise test is shows in Table 4.

**Table 4.** White Noise Test

| ARIMA | Lag | Chi square | DF | P-Value |
|---|---|---|---|---|
| (1,2,0) | 12 | 24,47 | 10 | 0,006 |
| | 24 | 43,35 | 22 | 0,004 |
| | 36 | 48,27 | 34 | 0,053 |
| (2,2,0) | 12 | 6,08 | 8 | 0,638 |
| | 24 | 17,43 | 20 | 0,625 |
| | 36 | 26,10 | 32 | 0,759 |
| (3,2,0) | 12 | 10,15 | 10 | 0,428 |
| | 24 | 24,50 | 22 | 0,321 |
| | 36 | 31,43 | 34 | 0,594 |
| (0,2,1) | 12 | 10,15 | 10 | 0,428 |
| | 24 | 24,50 | 22 | 0,321 |
| | 36 | 31,43 | 34 | 0,594 |
| (2,2,1) | 12 | 21,38 | 8 | 0,006 |
| | 24 | 41,70 | 20 | 0,003 |

| ARIMA | Lag | Chi square | DF | P-Value |
|---|---|---|---|---|
| | 36 | 44,56 | 32 | 0,069 |
| (3,2,1) | 12 | 9,01 | 7 | 0,252 |
| | 24 | 21,95 | 19 | 0,287 |
| | 36 | 27,29 | 31 | 0,658 |

The residuals of several ARIMA models were evaluated using the Ljung-Box test at lag 12,24, and 36. Models such as ARIMA(2,2,0), ARIMA(3,2,0), ARIMA(0,2,1), and ARIMA(3,2,1) produced $p$-values greater than 0.05 across all lags, indicating that their residuals do not exhibit significant autocorrelation. Therefore, these models are considered adequate in capturing the data's structure and the residuals behave like white noise. In contrast, the ARIMA(1,2,0) and ARIMA(2,2,1) models show significant autocorrelation at several lags, as reflected by $p$-values below 0.05 thus may not be suitable for forecasting. From acceptable models, the best one will be selected based on the lowest AIC value, which is summarized in the following Table 5.

**Table 5.** AIC Values

| Models | AIC |
|---|---|
| ARIMA (2,2,0) | 618,299 |
| ARIMA (3,2,0) | 615,311 |
| ARIMA (0,2,1) | 612,313 |
| ARIMA (3,2,1) | 622,232 |

## 4. Forecasting

Based on Table 6, the smallest AIC values are found in the ARIMA (0,2,1) model, indicating that this model will be used for forecasting with ARIMA. So, we will forecast the number of out of children using ARIMA (0,2,1). Prior to the presentation of the forecast, it is imperative to undertake a thorough examination of the model's MAPE. This procedure is instrumental in quantifying the precision of the data. The result of the MAPE given in Table 6.

**Table 6.** The Result of the MAPE ARIMA (0,2,1)

| Time Period | Forecast | SE Forecast | Lower | Upper | Actual | MAPE |
|---|---|---|---|---|---|---|
| 49 | 2132,59 | 176,595 | 1786,39 | 2478,78 | 2.164 | 1,4515% |
| 50 | 2067,17 | 255,876 | 1565,56 | 2568,79 | 1.970 | 4,9325% |
| 51 | 2001,76 | 320,949 | 1372,58 | 2630,95 | 1.980 | 1,0990% |
| 52 | 1936,35 | 379,392 | 1192,59 | 2680,11 | 1.914 | 1,1677% |
| 53 | 1870,94 | 434,064 | 1020 | 2721,87 | 1.890 | 1,0085% |
| Total Mape | | | | | | 1,9318% |

By using the ARIMA (0,2,1), the MAPE value of 1.931%, this shows the forecast results are very accurate. Then forecasting can be continued for the desired period. In this article, the researcher will forecast the number of school dropouts for 2024 - 2027. The results of the forecast can be seen in the following Table 7.

**Table 7.** Dropout Student Forecasting Result

| Year | Forecasting value |
|---|---|
| 2024 | 1740,11 |
| 2025 | 1674,70 |
| 2026 | 1609,28 |
| 2027 | 1543,87 |

The forecasting of the next four years indicates that the number of dropout student in Indonesia has the downward trend. This trend suggests that dropout rates are likely to continue decreasing if current conditions persist. This findings aligns with the conclusions of numerous studies that have demonstrated the efficacy of government programs in reducing the number of dropout student in Indonesia (Cahyaningtyas et al., 2022; Nurrokhmah, 2021).

## D. CONCLUSION AND SUGGESTIONS

Through the model identification, estimation, and validation, ARIMA(0,2,1) was selected as the most appropriate model for forecasting the number of dropout student in Indonesia. This selection was based on its lowest AIC value and MAPE of 1.931% or 1.9% indicating excellent forecasting ability. The forecast reveals a downward trend in dropout numbers from 2024 tp 2027. These finding provide a quantitative basis for policy planning and evaluation. The projected decline suppports the effectiveness of current government interventions and can guide the continuation or refinement of programs such as educational subsidies, outreach initiatives, and infrastructure improvements. Rather than assumptions, these forecasts serve as data driven evidence to inform strategic actions aimed at reducing dropout rates further.

## REFERENCES

Ajewole, K. P., Adejuwon, S. O., & Jemilohun, V. G. (2020). Test for Stationarity on Inflation Rates in Nigeria using Augmented Dickey Fuller Test and Phillips-Persons Test. *IOSR Journal of Mathematics*, *16*(3), 11–14. https://api.semanticscholar.org/CorpusID:220070542

Amini, M. H., Kargarian, A., & Karabasoglu, O. (2016). ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation. *Electric Power Systems Research*, *140(11)*, 378–390. https://doi.org/10.1016/j.epsr.2016.06.003

Box, G. E. P., Jenkins, G. M. J., Rainsel, G. C., & Ljung, G. M. (2016). *Time Series Analysis Forecasting and Control* (fifth). John Wiley & Sons. http://dx.doi.org/10.1111/jtsa.12194

Cahyaningtyas, A., Akbar, B. D. H., & Lestari, C. D. (2022). Dampak Bantuan Operasional Sekolah Terhadap Putus Sekolah di Indonesia. *HIMIE Economics Research and Olympiad (HERO)*, *1*(1), 116–122.

Cavanaugh, J. E., & Neath, A. A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Computational Statistics*, *11*(3), e1460. https://doi.org/10.1002/wics.1460

Chakrabarti, A., & Ghosh, J. K. (2011). AIC, BIC and Recent Advances in Model Selection. In handbook of *Philosophy of Science, Philosophy of Statistics, volume 7,* (pp. 583–605), ISSN 18789846, ISBN 9780444518620. https://doi.org/10.1016/B978-0-444-51862-0.50018-6

Chang, P.-C., Wang, Y.-W., & Liu, C.-H. (2007). The development of a weighted evolving fuzzy neural network for PCB sales forecasting. *Expert Systems with Applications*, *32*(1), 86–96. https://doi.org/10.1016/j.eswa.2005.11.021

Din, M. A. (2016). *ARIMA by Box Jenkins Methodology for Estimation and Forecasting Models in Higher Education*. Athens: *Atiner's Conference Paper Series*, No:EMS2015-1846. http://dx.doi.org/10.13140/RG.2.1.1259.6888

Fahrin, E., Hayati, M. N., & Siringoringo, M. (2019). *Penerapan Model Seasonal Autoregressive Fractionally Integrated Moving Average Pada Data Inflasi di Indonesia*. *10*(2), 113–118. https://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/568

Florea, N. M., & Hurjui, E. (2015). Critical Thinking in Elementary School Children. *Procedia - Social and Behavioral Sciences*, *180(16)*, 565–572. https://doi.org/10.1016/j.sbspro.2015.02.161

Gambulao, R. Jr. (2023). Forecasting Dropout Trend at King's College of The Philippines using ARI-MA Modeling. *International Journal of Multidisciplinary: Applied Business and Education Research*, *4*(6), 2094–2100. https://doi.org/10.11594/ijmaber.04.06.32

Gao, Y., Mosalam, K. M., Chen, Y., Wang, W., & Chen, Y. (2021). Auto-Regressive Integrated Moving-Average Machine Learning for Damage Identification of Steel Frames. *Applied Sciences*, *11*(13), Article 13. https://doi.org/10.3390/app11136084

Grigonytė, E., & Butkevičiūtė, E. (2016). Short-term wind speed forecasting using ARIMA model. *Energetika*, *62*(1–2), Article 1–2. https://doi.org/10.6001/energetika.v62i1-2.3313

Guo, Z. (2023). Research on the Augmented Dickey-Fuller Test for Predicting Stock Prices and Returns. *Advances in Economics, Management and Political Sciences*, *44*(1), 101–106. https://doi.org/10.54254/2754-1169/44/20232198

Hassani, H., Mashhad, L. M., Royer-Carenzi, M., Yeganegi, M. R., & Komendantova, N. (2025). White Noise and Its Misapplications: Impacts on Time Series Model Adequacy and Forecasting. *Forecasting*, *7*(1), 8. https://doi.org/10.3390/forecast7010008

Hegre, H., Metternich, N. W., Nygård, H. M., & Wucherpfennig, J. (2017). Introduction: Forecasting in peace research. *Journal of Peace Research*, *54*(2), 113–124. https://doi.org/10.1177/0022343317691330

Hjorth, C. F., Bilgrav, L., Frandsen, L. S., Overgaard, C., Torp-Pedersen, C., Nielsen, B., & Bøggild, H. (2016). Mental health and school dropout across educational levels and genders: A 4.8-year follow-up study. *BMC Public Health*, *16*(1), 976. https://doi.org/10.1186/s12889-016-3622-8

Imam, A. (2021). *Investigation of Parameter Behaviors in Stationarity of Autoregressive and Moving Average Models through Simulations*. *4*(4). https://doi.org/10.22377/ajms.v4i4.295

Jain, G., & Mallick, B. (n.d.). *A Study of Time Series Models ARIMA and ETS*. *International Journal of Modern Education and Computer Science*, *9*(4), 57–63. https://doi.org/10.5815/ijmecs.2017.04.07

Juvina, I., Larue, O., Widmer, C., Ganapathy, S., Nadella, S., Minnery, B., Ramshaw, L., Servan-Schreiber, E., Balick, M., & Weischedel, R. (2020). *Computer-Supported Collaborative Information Search for Geopolitical Forecasting* (W. T. Fu & H. Van Oostendorp, Eds.; pp. 245–266). Springer International Publishing. https://doi.org/10.1007/978-3-030-38825-6_12

Kim, S. G., Park ,Cheolyong, Hwang ,Sun-Young, Ha ,Jeongcheol, Park ,Inho, & and Kim, T. Y. (2024). Unit root tests and their challenges. *Communications in Statistics - Theory and Methods*, *53*(8), 2840–2847. https://doi.org/10.1080/03610926.2022.2150050

Klawonn, F., Riekeberg, N., & Hoffmann, G. (2024). Importance and Uncertainty of λ-Estimation for Box–Cox Transformations to Compute and Verify Reference Intervals in Laboratory Medicine. *Stats*, *7*(1), 172–184. https://doi.org/10.3390/stats7010011

Kopnina, H. (2020). Education for the future? Critical evaluation of education for sustainable development goals. *The Journal of Environmental Education*, *51*(4), 280–291. https://doi.org/10.1080/00958964.2019.1710444

Le, L.-H. (2024). Time series analysis and applications in data analysis, forecasting and prediction. *HPU2 Journal of Science: Natural Sciences and Technology*, *3*(1), Article 1. https://doi.org/10.56764/hpu2.jos.2024.3.1.20-29

Liu, T., Liu, S., & Shi, L. (2020). ARIMA Modelling and Forecasting. In T. Liu, S. Liu, & L. Shi, *Time Series Analysis Using SAS Enterprise Guide* (pp. 61–85). Springer Singapore. https://doi.org/10.1007/978-981-15-0321-4_4

Luo, L. (2024). Statistical model validation through white noise hypothesis testing in regression analysis and ARIMA models. *Theoretical and Natural Science*, *42*(1), 227–232. https://doi.org/10.54254/2753-8818/42/20240672

Lutz, C. S., Huynh, M. P., Schroeder, M., Anyatonwu, S., Dahlgren, F. S., Danyluk, G., Fernandez, D., Greene, S. K., Kipshidze, N., Liu, L., Mgbere, O., McHugh, L. A., Myers, J. F., Siniscalchi, A., Sullivan, A. D., West, N., Johansson, M. A., & Biggerstaff, M. (2019). Applying infectious disease forecasting to public health: A path forward using influenza forecasting examples. *BMC Public Health*, *19*(1), 1659. https://doi.org/10.1186/s12889-019-7966-8

Maitra, S., & Politis, D. N. (2024). Prepivoted Augmented Dickey-Fuller Test with Bootstrap-Assisted Lag Length Selection. *Stats*, *7*(4), 1226–1244. https://doi.org/10.3390/stats7040072

Marlini, H. (2016). EVALUASI PROGRAM BANTUAN SISWA MISKIN (BSM) DI SMP NEGERI 4 SANGGAU. *PUBLIKA : Juenal Ilmu Administrasi Negara*, *5*(3), 1–18. http://dx.doi.org/10.26418%2Fpublika.v5i3.1120

Martin, G. M., Frazier, D. T., Maneesoonthorn, W., Loaiza-Maya, R., Huber, F., Koop, G., Maheu, J., Nibbering, D., & Panagiotelis, A. (2023). *Bayesian Forecasting in Economics and Finance: A Modern Review* (arXiv:2212.03471). arXiv. https://doi.org/10.48550/arXiv.2212.03471

Mduma, N., Kalegele, K., & Machuve, D. (2019). A Survey of Machine Learning Approaches and Techniques for Student Dropout Prediction. *Data Science Journal*, *18(1)*, 14. https://doi.org/10.5334/dsj-2019-014

Moniz, A. B. (2022). Forecasting and Responsible Innovation: A Book Review. *Frontiers in Sociology*, *7*, 835277. https://doi.org/10.3389/fsoc.2022.835277

Murat, M., Malinowska, I., Gos, M., & Krzyszczak, J. (2018). Forecasting daily meteorological time series using ARIMA and regression models. *International Agrophysics*, *32*(2), 253–264. https://doi.org/10.1515/intag-2017-0007

Nurmalitasari, Awang Long, Z., & Faizuddin Mohd Noor, M. (2023). Factors Influencing Dropout Students in Higher Education. *Education Research International*, *2023*(1), 7704142. https://doi.org/10.1155/2023/7704142

Nurrokhmah, M. (2021). Program Indonesia Pintar (PIP): Implementasi Kebijakan Kesejahteraan dalam Upaya Meningkatkan Angka Partisipasi Pendidikan. *Jurnal Paradigma : Jurnal Multidisipliner Mahasiswa Pascasarjana Indonesia*, *2*(1), 37–48. https://doi.org/10.22146/jpmmpi.v2i1.66256

Patulin, E. P. (2019). Scholarship Grants Prediction using Autoregressive Integrated Moving Average (ARIMA) Algorithm. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(3), 551–555. https://doi.org/10.30534/ijatcse/2019/33832019

Rahardja, D. (2024). Statistical time-series forecast error and bias assessment using hold-out samples and a single benchmark (data-reduction) scoring metric. *Journal of Statistics and Management Systems*, *27*(3), 569–578. https://doi.org/10.47974/JSMS-974

Ratnesh, G., & Kumar, S. A. (2019). Estimation of Autocorrelation function and partial Autocorrelation function of Evapotranspiration for Ranchi district, India, in context of time series Modelling. *International Journal of Scientific Research and Reviews*, *08*(04), 21–31. https://doi.org/10.37794/IJSRR.2019.8403

Riyadi, M. A. A., Pratiwi, D. S., Irawan, A. R., & Fithriasari, K. (2017). Clustering stationary and non-stationary time series based on autocorrelation distance of hierarchical and k-means algorithms. *International Journal of Advances in Intelligent Informatics*, *3*(3), 154. https://doi.org/10.26555/ijain.v3i3.98

Sharafi, M., Ghaem, H., Tabatabaee, H. R., & Faramarzi, H. (2017). Forecasting the number of zoonotic cutaneous leishmaniasis cases in south of Fars province, Iran using seasonal ARIMA time series method. *Asian Pacific Journal of Tropical Medicine*, *10*(1), 79–86. https://doi.org/10.1016/j.apjtm.2016.12.007

Shivhare, N., Rahul, A. K., Dwivedi, S. B., & Dikshit, P. K. S. (2021). ARIMA based daily weather forecasting tool: A case study for Varanasi. *MAUSAM*, *70*(1), 133–140. https://doi.org/10.54302/mausam.v70i1.179

Shoko, C., & Njuho, P. (2023). Arima Model In Predicting Of Covid-19 Epidemic For The Southern Africa Region. *African Journal of Infectious Diseases*, *17*(1), 1–9. https://doi.org/10.21010/Ajidv17i1.1

Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples*. Springer International Publishing. https://doi.org/10.1007/978-3-319-52452-8

Singh, H. P., & Alhamad, I. A. (2022). Influence of National Culture on Perspectives and Factors Affecting Student Dropout: A Comparative Study of Australia, Saudi Arabia, and Ethiopia. *Archives of Business Research*, *10*(11), 287–300. https://doi.org/10.14738/abr.1011.13508

Sinuany-Stern, Z. (2021). *Forecasting Methods in Higher Education: An Overview*. In: Sinuany-Stern, Z. (eds) Handbook of Operations Research and Management Science in Higher Education.

International Series in Operations Research & Management Science, vol 309, 131–157. Springer, Cham. https://doi.org/10.1007/978-3-030-74051-1_5

Syafitri, D. U., & Hadjam, M. N. R. (2017). Religiusitas: Faktor Protektif Pengasuhan Orangtua dengan Status Sosial Ekonomi Rendah. *Psikologika: Jurnal Pemikiran dan Penelitian Psikologi*, *22*(2), 1–14. https://doi.org/10.20885/psikologika.vol22.iss2.art1

Utami, D. S., Huda, N. M., & Imro'ah, N. (2024). ARIMA Time Series Modeling with the Addition of Intervention and Outlier Factors on Inflation Rate in Indonesia. *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, *8*(1), 256. https://doi.org/10.31764/jtam.v8i1.17487

Yaneri, A., Suviani, V., & Vonika, N. (2022). Analisis Penyebab Anak Putus Sekolah Bagi Keluarga Miskin (Studi Kasus Anak Usia Sekolah Pada Keluarga Miskin di Kampung Lio Kota Depok). *Jurnal Ilmiah Perlindungan dan Pemberdayaan Sosial (Lindayasos)*, *4*(1), 76-89. https://doi.org/10.31595/lindayasos.v4i1.554