

Time Series Forecasting for the Spread of Covid-19 in Indonesia Using Hybrid ARIMA and Holt-Winters Methods

Damas Angga Pramudya¹, Nola Marina²

¹Information Systems Management, University of Gunadarma, Margonda Raya No. 100, Depok 16424, West Java

²Faculty of Industrial Technology, University of Gunadarma, Akses UI, Tugu, Depok 16451, West Java

Email address: ¹damasangga@gmail.com, ²nolamarina@gmail.com

Abstract— Forecasting is an activity to predict the future using data from the past. COVID-19 cases in Indonesia have an increasing trend pattern and predictions are needed in the future to reduce the rate of spread. This research has the purpose to produce the best data series forecasting model and forecasting results using a Hybrid model that combines ARIMA and Holt-Winters models to positive confirmed cases recovered cases and total deaths. The data used in this research was obtained from kawalcovid19.com for the spread of COVID-19 in Indonesia from March to October 2020. ARIMA analysis used the smallest estimation of Akaike's Information Criterion and Schwarz Bayesian Criterion criteria, and the evaluation parameters Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Square Error (MSE). Holt-Winters Exponential Smoothing analysis uses parameters alpha (α), beta (β), Sum Squared of Residual (SSR) criteria, and RMSE evaluation. The analysis of the Hybrid ARIMA and Holt-Winters methods uses a combination of ARIMA criteria and RMSE evaluation. Based on the results of the research, it can be concluded that the Hybrid ARIMA and Holt-Winters method is the best method for predicting the number of COVID-19 spread rates in November and December 2020 because it produces the smallest RMSE value.

Keywords— ARIMA, COVID-19, Forecasting, Hybrid, Holt-Winters.

I. INTRODUCTION

One of the countries affected by COVID-19 is Indonesia (Worldometers, 2020) as of October 3, 2020, ranking 23 in the number of positive cases. The spread of positive cases infected with the COVID-19 virus shows an increase. The Indonesian government has implemented policies in various fields to deal with the spread of COVID-19 (Ministry of Foreign Affairs, 2020). Therefore, it is important to analyze data to predict future conditions so that it can be used by the government as a consideration and take appropriate policies to suppress the spread of COVID-19. By using the concept of forecasting the spread of COVID-19, it can be used as information to prevent an increase in the number of positive confirmed cases in the future.

Time series forecasting can be used to generate future information using historical data. The methods used to predict future events are Autoregressive Integrated Moving Average (ARIMA), Holt-Winters, and Hybrid ARIMA and Holt-Winters. Research (Papastefanopoulos, Linardatos, and Kotsiantis, 2020) to predict positive cases in several countries affected by the spread of COVID-19 using the time series method with several models: ARIMA, Holt-Winters Additive

Models (HWAAS), TBAT, Facebook's Prophet, Deep AR, and N-Beats. The results of this research conclude that the ARIMA and TBAT models have RMSE statistical rating values of 1.70000 and 2.90000 which are smaller than the Prophet, Deep AR, and N-Beats models.

Research (Chintalapudi, Battineni, and Amenta, 2020) to predict the spread of COVID-19 in Italy with the ARIMA model. This research resulted in two ARIMA models. For confirmed cases, the ARIMA(1,2,0) model with an accuracy of 93.75% and recovered cases resulted in the ARIMA(3,2,0) model with an accuracy of 84.4. The research (Bissing, Klein, Chinnathambi, Selvaraj, and Ranganathan, 2019) aims to identify several predictive variables such as wind generation demand, temperature, and wind speed that can affect hourly electricity prices. The research was conducted by combining the regression model with Holt-Winters and ARIMA. The evaluation parameter used is the smallest Mean Absolute Percentage Error (MAPE) of each time series model. This research concludes that the combination of ARIMA and Holt-Winters outperforms other methods with a forecasting accuracy of 70%.

In this research, time series forecasting was carried out using the ARIMA, Holt-Winters, and Hybrid ARIMA and Holt-Winters models to predict the spread of COVID-19 in Indonesia by using data on total positive confirmed cases, total death cases, and recovered from August to October 2020. The accuracy evaluation used in this research is the Root Mean Square Error (RMSE), the smallest of any time series model in predicting the rate of spread of COVID-19 in Indonesia.

II. THEORITICAL BASIC

A. Forecasting

Forecasting is predicting some event or activity in the future. Forecasting is essential in a wide variety of fields including business and industry, government, economics, environmental science, medicine, social sciences, politics, and finance. Forecasting is classified into the short, medium, and long term. Short-term forecasting is used to predict events only for a while (days, weeks, and months) into the future. Medium-term forecasts are used for the next 1 to 2 year time period. Long-term forecasts are used for the next several years (Montgomery, Jennings, Cheryl L., and Kulahci, 2015).

B. Time Series

A Time series model is a data with orientation based on time or sequential chronological observations on a variable. Most of the problems in forecasting use time-series data. Time series uses statistical properties of historical data to determine formal models and then estimates the unknown parameters based on least squares. (Montgomery, Jennings, Cheryl L. and Kulahci, 2015).

Time series analysis aims to learn patterns by sequence by identifying important parameters and estimating future values. There are several classical forecasting methods such as exponential smoothing, regression, and others that can be used to predict from a time series data. The method applied is following the data features, the amount of data, and the presence of autocorrelation (Abdallah, 2019).

C. Hybrid ARIMA and Holt-Winters

The Autoregressive Integrated Moving Average (ARIMA) model is based on time series assumptions with stationary and linear data connecting the input and output linear series (Waeto, Chuarkham, and Intarasit, 2017). Holt-Winters is a data series forecasting method with a fixed model. This method is a prediction scheme that uses the weighting of the values of the previous sequence observations to calculate future values. The goal of the Holt-Winters method with exponential smoothing is to smooth the original sequence and then use it to predict the future value of the main variable. This method can be used for forecasting with trend data, seasonal, or both (Shastri et al., 2018).

Hybrid ARIMA and Holt-Winters combined the two processes into one forecasting cycle. The initial stage is the ARIMA process in the form of model identification, model parameter estimation, diagnostic checks, and then entering the Holt-Winters process in the form of holt-winters model parameter estimation, error checking and the results of the ARIMA and Holt-Winters hybrid models.

Figure 1 display research methodology with The ARIMA and Holt-Winters Hybrid models. The hybrid method is a combination of the ARIMA and Holt-Winters methods. The ARIMA process is carried out at the time of initial data input, namely the model identification process through a transformation process (stationary and differentiation test), autocorrelation function (ACF) and partial autocorrelation function (PACF). The second stage is the estimation of the ARIMA model parameters using Akaike's Information Criterion (AIC) and Schwarz Bayesian Criterion (SBC). The third stage is a diagnostic examination using the Residual Correlogram. The fourth stage is the process carried out on the Holt-Winters model. The data generated from the ARIMA process is processed into the Holt-Winters model for the smoothing process. This stage estimates the parameters of the Holt-Winters model by using the alpha value (α) which indicates the smoothing control value, beta (β) displays the value containing the trend element and the Sum of Squares Residual. The fifth stage is error checking using RMSE. The sixth stage is to produce the best Hybrid ARIMA and Holt-Winters models. the seventh stage is to compare the Hybrid ARIMA and Holt-Winters models with the ARIMA and Holt

Winters models themselves. The eighth stage is forecasting and the last stage is displaying the forecasting results.

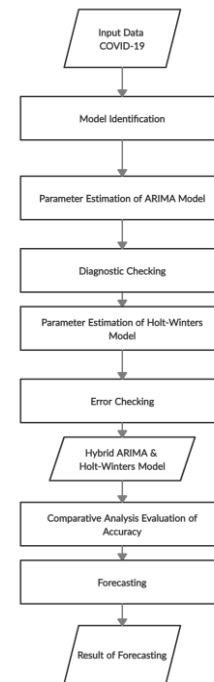


Fig. 1. Research Methodology

III. RESEARCH METHOD

This stage explains the research process systematically so that the research process runs well.

A. Data COVID-19

The data used in this research is weekly data of positive confirmed cases, recovered cases, and COVID-19 deaths in Indonesia from March 2, 2020, to October 26, 2020, with a total of 35 weeks. This data is public data sourced from the kawalcovid19 website with a total of 35 data (Kawal COVID19, 2020).

This research uses the EViews application to forecast the spread of COVID-19 for positive cases, recovered cases, and death cases in Indonesia.

B. Identification of Model

The identification stage is the initial stage in the Hybrid ARIMA and Holt-Winters method. This stage is similar to the ARIMA method by conducting a stationary test and producing the best model using autocorrelation (ACF), partial autocorrelation (PACF), and transformation (differencing and lagging). The data series for forecasting using the Hybrid ARIMA and Holt-Winters methods must be stationary concerning the mean and variance.

C. Parameter Estimation of ARIMA Model

The parameter estimation stage of the ARIMA model is to evaluate the model that has passed the data stationary testing stage and is continued by looking at the pattern of the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The estimated model that has been produced can be compared with several parameters, namely Akaike's

Information Criterion (AIC) and Schwartz's Bayesian Criterion (SBC). Determination of the best model estimation is seen from the Akaike Info Criterion and Schwarz Criterion with the smallest value (Abdallah, 2019).

D. Diagnostic Checking

The diagnostic check stage in the Hybrid ARIMA and Holt-Winters method is to analyze the probability value for each estimation model. This stage is carried out to determine whether the model used is good by looking at the residuals. Residual analysis can be said to be good if it has white noise by looking at the probability values in ACF and PACF which are not significant (p-value >= (0.05)).

E. Parameter Estimation of Holt-Winters Models

This stage uses the parameters of two main parameters for the estimation of the Hybrid ARIMA and Holt-Winters models, namely the alpha (α), beta (β) parameters with a comparison of the sum of squared residuals values. This research uses estimates of several constant values of alpha (α), beta (β) for each object of research for positive cases, recovered cases, and cases of death with 5 times forecasting (Silitonga, Himawan, and Damanik, 2020).

F. Error Checking

Based on the alpha (α), beta (β) values that have been determining in the previous stage for each data on the spread of COVID-19, namely positive confirmed cases, recovered cases, and death cases, one evaluation parameter is used for the error checking stage. The Hybrid ARIMA and Holt-Winters method in this research conducted an error examination by the Root Mean Squared Error parameter with the smallest value (Silitonga, Himawan, and Damanik, 2020). Refer to (1), displays the RMSE formula with the condition that is the actual value of the endogenous variable, at time t, is the change in the value of the endogenous variable at time t and T is the observation number in the simulation.

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{Y}_t - Y_t)^2} \tag{1}$$

G. Selection of Hybrid ARIMA and Holt-Winters Models

The selection stage for the Hybrid ARIMA and Holt-Winters models is the last process in predicting the spread of COVID-19 in Indonesia for positive confirmed cases, recovered cases, and death cases. The selection of the Hybrid ARIMA and Holt-Winters models is determined by the smallest value of the Root Mean Square Error parameter.

H. Comparative Analysis Evaluation of Accuracy

Holt-Winters Exponential Smoothing and Hybrid ARIMA Holt-Winters models on each research object have different accuracy evaluation values. The best models of ARIMA, Holt-Winters, and Hybrid ARIMA Holt-Winters will be compared with the parameter evaluation of the accuracy of Root Mean Square Error.

I. Forecasting

This stage describes the final process of evaluating the best model resulting from the comparison of ARIMA, Holt-Winters, and Hybrid ARIMA Holt-Winters. The best model will produce forecasting values in the period 2 November 2020 to 28 December 2020.

IV. RESULT AND DISCUSSION

The test in this research used 35 data for each positive confirmed case, recovered case, and death case of the spread of COVID-19 in Indonesia. Data was obtained from KawalCovid19 with initial observation data of 243 daily data and converted into weekly data with a total of 35 data (KawalCOVID19, 2020). This test was conducted to identify the best model for forecasting the spread of COVID-19 using a comparison of the ARIMA, Holt-Winters, and Hybrid ARIMA, and Holt-Winters methods. Tests are carried out by identifying the model of each method, model estimation, error checking, and in the final stage conducting a comparative analysis, evaluating the accuracy, and forecasting results using the best model.

A. Identification of Model

This stage is carried out to obtain stationary data and produce the best model using autocorrelation (ACF), partial autocorrelation (PACF), and transformation. Figure 2 displays data on the spread of COVID 19 in Indonesia by showing an increasing trend of growth.

This stationary test was carried out using Unit Root Tests with the Augmented Dickey-Fuller (ADF) test type. The parameter used in the ADF test is the Schwarz Info Criterion (SIC) with a probability value (p-value) less than 0.05 and a statistical value (t-statistic) less than 5%-1% of the critical value of the test (t-critical values). (Wei, 2016).

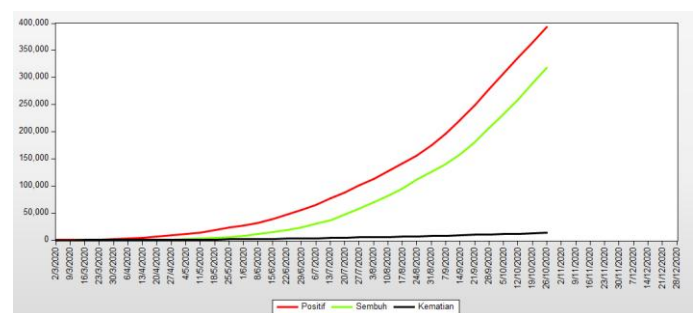


Fig. 2. Graph of the Spread of COVID-19 in Indonesia

Based on the results of the ADF test, the positive case data that has met the stationary test is differentiation level 2 (d = 2) with p-value 0.0015 < 0.05 and t-statistic -4.384051 < -2.957110. For recovered cases that have met the stationary test, it is differentiation level 2 (d = 2) with p-value 0.0001 < 0.05 and t-statistic -5.519146 < -2.957110 and for death cases that have met the stationary test is differentiation level 2 (d = 2) with p-value 0.0001 < 0.05 and t-statistic -5.519146 < -2.957110.

Based on the ACF and PACF plots, the researchers estimated that the ARIMA model of positive confirmed case data was autoregressive with an AR(5) value, namely ARIMA(5,2,0). The ARIMA model estimation for recovered case data is a moving average with an MA(8) value, namely ARIMA(0,2,8). Estimated ARIMA model data on AR(4) and MA(8) cases of death. So 3 models are suitable for mortality data, namely ARIMA(4,2,0), ARIMA(4,2,4), and ARIMA(0,2,4).

B. Parameter Estimation of ARIMA Models

The determination of the best model estimation is seen from the small Akaike Info Criterion and Schwarz Criterion values from each positive case data, recovered case, and death case as shown in Table I.

TABLE I. The Parameter Analysis Results of Positive Case

Model Estimation	Akaike Info Criterion	Schwarz Criterion
Positive ARIMA(5,2,0)	17.28753	17.42357
Recovered, ARIMA(0,2,8)	17.63040	17.67617
Death, ARIMA(4,2,0)	11.27157	11.40762

C. Diagnostic Checking

The diagnostic check stage is carried out to determine whether the model used is good by looking at the residuals. Residual analysis can be said to be good if it has white noise by looking at the probability values in ACF and PACF which are not significant ($p\text{-value} > = 0.05$).

ARIMA(5,2,0), ARIMA(0,2,8) and ARIMA (4,2,0) models produce lag 1 to lag 32, no significant lag ($p\text{-value} > 0.05$) shows no correlation between residuals, residuals are homogeneous and no pattern is shown. This shows that the residual ARIMA(5,2,0) model is white noise and well distributed.

D. Parameter Estimation of Holt-Winters Models

The parameter estimation stage of the Holt-Winters model in the EVIEWS application displays the parameter values of and the value of Sum of Squared Residuals. Table II is the result of the analysis of the Holt-Winters model for positive cases, recovered cases in Table III, and cases of death in Table IV below:

TABLE II. The Parameter Analysis Results of Positive Case

Forecasting	α	β	SSR
P1	0.2	0.3	8.09
P2	0.4	0.5	9.22
P3	0.5	0.7	3.73
P4	0.8	0.9	1.23
P5	0.9	0.9	1.04

TABLE III. The Parameter Analysis Results of Recovered Case

Forecasting	α	β	SSR
P1	0.2	0.3	9.79
P2	0.4	0.5	1.15
P3	0.5	0.7	4.71
P4	0.8	0.9	1.53
P5	0.9	0.9	1.30

TABLE IV. The Parameter Analysis Results of Death Case

Forecasting	α	β	SSR
P1	0.2	0.3	6728945
P2	0.4	0.5	784686.9
P3	0.5	0.7	320153.3
P4	0.8	0.9	103168.1
P5	0.9	0.9	87457.99

E. Error Checking

The error checking stage of the Hybrid ARIMA and Holt-Winters models uses the Root Mean Squared Error evaluation parameter with the smallest value. The analysis results of the Holt-Winters model for positive confirmed cases as shown in Tables V, VI and VII showed the values of, and the root mean square error. Based on Tables V, VI, and VII, the parameters and in the fifth forecast (P5) have the smallest RMSE value so that it can be determined for the positive case, recovered, and death forecasting process using $\alpha = 0.9$ which indicates the parameter to control the smoothing value and $\beta = 0.9$ with the smallest RMSE value.

TABLE V. The Parameter Analysis Results of Positive Case Evaluation

Forecasting	α	β	RMSE
P1	0.2	0.3	14788.87
P2	0.4	0.5	4990.58
P3	0.5	0.7	3177.16
P4	0.8	0.9	1819.73
P5	0.9	0.9	1678.77

TABLE VI. The Parameter Analysis Results of Recovered Case Evaluation

Forecasting	α	β	RMSE
P1	0.2	0.3	15264.66
P2	0.4	0.5	5229.02
P3	0.5	0.7	3347.35
P4	0.8	0.9	1906.22
P5	0.9	0.9	1757.09

TABLE VII. The Parameter Analysis Results of Death Case Evaluation

Forecasting	α	β	RMSE
P1	0.2	0.3	420.81
P2	0.4	0.5	143.70
P3	0.5	0.7	91.79
P4	0.8	0.9	52.11
P5	0.9	0.9	47.97

F. Selection of Hybrid ARIMA and Holt-Winters Models

The selection stage for the Hybrid ARIMA and Holt-Winter models in Table VIII for positive confirmed cases has $\alpha = 0.9$ and $\beta = 0.9$ with an RMSE value of 1678.77. The recovered cases had $\alpha = 0.9$ and $\beta = 0.9$ with an RMSE value of 1757.09. The case of death had $\alpha = 0.9$ and $\beta = 0.9$ with an RMSE value of 47.97.

TABLE VIII. The Best Hybrid ARIMA and Holt-Winters Models to Positive Case, Recovered Case, and Case

Method	α & β	RMSE
Positive	0.9 & 0.9	1678.77
Recovered	0.9 & 0.9	1757.09
Death	0.9 & 0.9	47.97

G. Comparative Analysis Evaluation of Accuracy

The evaluation comparison analysis stage displays the best model of each method with the smallest Root Mean Square value. This stage compares the RMSE values between

ARIMA, Holt-Winters, and Hybrid ARIMA and Holt-Winters to get the best model and is used for forecasting the spread of COVID-19 in Indonesia. Table IX shows the comparison of the RMSE values for positive cases, Table X showed the comparison of RMSE values for positive cases, Table XI displays the comparison of RMSE values for positive cases.

TABLE IX. The Best Models to Positive Case

Method	RMSE
ARIMA(5,2,0)	47522.16
Holt-Winters (0.9, 0.9)	1821.49
Hybrid ARIMA dan Holt-Winters (0.9, 0.9)	1678.77

TABLE X. The Best Models to Recovered Case

Method	RMSE
ARIMA(0,2,8)	118200.40
Holt-Winters (0.9, 0.9)	1979.35
Hybrid ARIMA dan Holt-Winters (0.9, 0.9)	1757.09

TABLE XI. The Best Models to Death Case

Method	RMSE
ARIMA(4,2,0)	295.80
Holt-Winters (0.9, 0.9)	78.00
Hybrid ARIMA dan Holt-Winters (0.9, 0.9)	47.97

The best models for death cases are Hybrid ARIMA and Holt-Winters with parameters = 0.9 and = 0.9 indicating the smoothing control value and displaying data that has an element of trend.

H. Forecasting

Based on the analysis of the evaluation of the RMSE value on the ARIMA model, Holt-Winters and Hybrid ARIMA and Holt-Winters can be used for the forecasting process. Forecasting data is the value generated for each case of the spread of COVID-19. The following is Table XII which is the result of forecasting for the period 2 November 2020 to 28 December 2020.

TABLE XII. Forecasting Results of Positive Cases, Recovered Cases, and Death Cases for the Period 2 November 2020 to 28 December 2020

Date	Positive Case	Recovered Case	Death Case
2-Nov-2021	471,570	527,696	14,547
9-Nov-2021	501,053	558,737	15,412
16-Nov-2021	531,429	590,665	16,303
23-Nov-2021	562,698	623,479	17,218
30-Nov-2021	594,860	657,181	18,158
7-Dec-2021	627,915	691,769	19,124
14-Dec-2021	661,861	727,245	20,114
21-Dec-2021	696,700	763,607	21,129
28-Dec-2021	732,430	800,856	22,170

V. CONCLUSION

A. Conclusion

Based on the results of the analysis and forecasting of the best models from ARIMA, Holt-Winters and Hybrid ARIMA and Holt-Winters using the RMSE accuracy evaluation parameter, it can be concluded as follows:

1. The spread of COVID-19 in Indonesia in November and December 2020 experienced an increase in positive confirmed cases, recovered cases, and death cases.

2. The best model for positive confirmed cases, recovered cases, and death cases are Hybrid ARIMA and Holt-Winters compared to ARIMA model and Holt-Winters model with the smallest RMSE evaluation value with = 0.9 and = 0.9 resulting in RMSE 1678.77, recovered cases yielding RMSE 1757.09 and cases of death with an RMSE value of 47.97.
3. The results of the research can be used by the central government to determine policies to prevent and control the spread of COVID-19 in Indonesia to reduce the rate of positive confirmed cases in each region.

B. Suggestion

Suggestions for the development of further research is testing the spread of COVID-19 forecasting in addition to the Hybrid ARIMA and Holt-Winters models, can use Long Short Term Memory (LSTM) to get more accurate forecasting results. From the research conducted, researchers can provide suggestions as follows:

1. For other researchers, these results can be used as a reference in forecasting data series with data that has a trend pattern such as the spread of the COVID-19 outbreak.
2. Tests for forecasting the spread of COVID-19 other than the Hybrid ARIMA and Holt-Winters models can use Long Short Term Memory (LSTM) to get more accurate forecasting results.

REFERENCES

- [1] Worldometers, "Covid-19 Coronavirus Pandemic," 2020. <https://www.worldometers.info/coronavirus/> (accessed Oct. 03, 2020).
- [2] Kemlu, "Kebijakan Pemerintah Republik Indonesia Terkait Wabah Covid-19," 2020. <https://kemlu.go.id/brussels/id/news/6349/kebijakan-pemerintah-republik-indonesia-terkait-wabah-covid-19> (accessed Oct. 03, 2020).
- [3] V. Papastefanopoulos, P. Linardatos, and S. Kotsiantis, "COVID-19: A comparison of time series methods to forecast percentage of active cases per population," *Applied Science*, vol. 10, no. 11, pp. 1–15, 2020.
- [4] N. Chintalapudi, G. Battineni, and F. Amenta, "COVID-19 virus outbreak forecasting of registered and recovered cases after a sixty-day lockdown in Italy: A data-driven model approach," *J. Microbiol. Immunol. Infect.*, vol. 53, no. 3, pp. 396–403, 2020, doi: 10.1016/j.jmii.2020.04.004.
- [5] D. Bissing, M. T. Klein, R. A. Chinnathambi, D. F. Selvaraj, and P. Ranganathan, "A Hybrid Regression Model for Day-Ahead Energy Price Forecasting," *IEEE Access*, vol. 7, pp. 36833–36842, 2019, doi: 10.1109/ACCESS.2019.2904432.
- [6] KawalCOVID19, "COVID-19 di Indonesia @kawalcovid19," 2020. <https://docs.google.com/spreadsheets/d/1ma1T9hWbec1pXlwZ89WakRk-OfVUQZsOCFl4FwZxzVw/edit#gid=2052139453> (accessed Oct. 26, 2020).
- [7] W. Wei, J. Jiang, H. Liang, L. Gao, B. Liang, J. Huang, N. Zang, Y. Liao, J. Yu, J. Lai, F. Qin, J. Su, L. Ye, H. Chen, "Application of a Combined Model with Autoregressive Integrated Moving Average (ARIMA) and Generalized Regression Neural Network (GRNN) in Forecasting Hepatitis Incidence in Heng County, China," *PLoS One*, vol. 11, no. 6, pp. 1–13, 2016.
- [8] F. D. M. Abdallah, "Role of Time Series Analysis in Forecasting Egg Production Depending on ARIMA Model," *Applied Mathematics*, vol. 9, no. 1, pp. 1–5, 2019.
- [9] P. D. P. Silitonga, H. Himawan, and R. Damanik, "Forecasting acceptance of new students using double exponential smoothing method," *Journal of Critical Review*, vol. 7, no. 1, pp. 300–305, 2020.