COMPARISON OF FORECASTING VIOLENCE CASES NUMBER AGAINST WOMEN AND CHILDREN USING DOUBLE EXPONENTIAL SMOOTHING (DES) AND AUTO REGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) METHODS

Siti Nur Fadilah[1*]**, Putroue Keumala Intan**[2]**, Wika Dianita Utami**[3]

[1,2,3] *Prodi Matematika, Fakultas Sains dan Teknologi, UIN Sunan Ampel Surabaya*
Ahmad Yani St., No.117, Jemur Wonosari, Wonocolo, Surabaya, 60237, Indonesia

Corresponding author's e-mail: [1] sitinurfadilah576@gmail.com*

Abstract. *Violence is something that is being widely discussed. It is due to the increasing number of victims of violence in a scope where victims should feel safe. Therefore, the researchers took this case intending to predict the number of violence cases against women and children in Jakarta so that the government can anticipate the spike in cases and evaluate the policies that will be issued in this case. The data used was from the Office for the Empowerment of Child Protection and Population Control (DPPAPP) of DKI Jakarta Province from January 2018 to October 2021 to predict the number of cases in 2022. Based on the analysis results, it is known that the number of cases of violence against women and children has decreased throughout 2022. In addition, the accuracy of the model using the Double Exponential Smoothing (DES) method is 44.91%, and the Auto-Regressive Integrated Moving Average (ARIMA) is 39.03%.*

Keywords: Auto Regressive Integrated Moving Average (ARIMA), Double Exponential Smoothing (DES), forecast, violence against women and children.

1. INTRODUCTION

Since the first cases of violence against women and children have been widely discussed by the community. A total of 1,179 cases occurred in Jakarta, with physical violence such as domestic violence (domestic violence) and sexual violence against women and children as the most frequent forms of violence [1]. Factors are opinion differences in the household or economic conditions that cause men to commit acts of violence against women and children. If the virus can be stopped by giving vaccines, there is no vaccine for violence against women and children [2]. In addition, if this is not handled immediately, more and more parties will be harmed, not just the victims. Therefore, it is necessary to do forecasting in order to know the development of future cases to anticipate continuous spikes. If forecasting results continue to increase, the government should try to emphasize the rate of increase in cases. However, if the forecast results show a decline in cases, the government must remain vigilant if it experiences another spike one day. From these matters, it is known that forecasting the development of violence cases against women and children is very necessary.

Previous researchers have not widely studied forecasting cases of violence against women and children. Mariyana conducts a study under the title Spline Regression Modeling using the Penalized Spline method on the data on the number of violence cases against children in Indonesia with the MAPE error rate value of 12.81% [3]. The research results conducted by Al Farikhi explained that the DES method is the best method for forecasting with a MAPE value of 149.5 [4], and research by Ariska using the ARIMA method obtained a MAPE value of 4.83821 [5]. Therefore, researchers are interested in conducting research using the DES and ARIMA methods to predict violence cases against women and children to determine which one is more effective.

Exponential Smoothing is a moving average forecasting method with exponentially decreasing weighting. There are various types of Exponential Smoothing, namely Single Exponential Smoothing, Double Exponential Smoothing, and Triple Exponential Smoothing. Therefore, DES is also known as a forecasting method. DES is a time series quantitative forecasting method with trend data patterns. This method gives a decreasing weight to past observations and uses two smoothing parameters, namely alpha (α) and gamma (γ) [6]. The algorithm of Exponential Smoothing itself is exponential Smoothing, where the results of the Smoothing can also be used on data that is still unstable [7]. This method has one advantage: the relatively small amount of data and parameters used [8].

ARIMA method is a method that has a flexible nature. It is because this method can follow patterns from existing data. The model used from this method also uses the dependent variable, and the data used must also be stationary. If the data used is still not stationary, it is necessary to make improvements so that the data becomes stationary to produce accurate forecasts. This method, also known as the Box-Jenkins method, is very suitable for forecasting, especially in the short term. This method produces forecasts based on historical data patterns and uses a value approach [9]. In addition, the ARIMA method has an advantage, namely a reasonably high accuracy value which makes this method often used in time series forecasting [10].

This study aimed to predict the development of cases of violence against women and children in Jakarta using the DES and ARIMA methods. With the prediction of this case, it is hoped that the local government can increase the anticipation of the dangers of this violent case so that the case value does not increase. The prediction results are expected to be used as an alternative model for carrying out violence cases against women and children in Jakarta. It is also used as an optimization for the government in taking and evaluating a policy related to the case.

2. RESEACH METHODS

In this study, researchers used quantitative data. The data used was historical data on violence against women and children in Jakarta. The data was obtained from https://dppapp.jakarta.go.id/, which contained historical data on violence against women and children from January 2018 to October 2021. The following is the data obtained, presented in Table 1.

Table 1. Data on Violence Against Women and Children

Months	Number of Violence Cases Against Women and Children			
	2021	2020	2019	2018
January	41	121	40	164
February	55	50	109	169
…	…	…	…	…
December		59	40	48

The steps in this research were:
1. Input Data, which was to prepare the data in **Table 1** into the RStudio Software.
2. Preprocessing Data, which is to check whether there is empty data.
3. The formation of the DES Model is to form an equation model using the DES method.

Generally, this method uses two smoothing parameters with a value range of $0 - 1$ and is used for forecasting trend data patterns. Therefore, it requires trend values using different parameters from the parameters in the original series. This equation model is presented in the equation (4).

$$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} - b_{t-1}) \tag{1}$$
$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1} \tag{2}$$
$$F_t = S_{t-1} + b_t \tag{3}$$
$$F_{t+m} = S_t + b_t m \tag{4}$$

Where:
S_t = smoothing data in period t
b_t = smoothing trend in period t
F_{t+m} = forecasting in period t
m = number of forecasting periods

4. Formation of the ARIMA Model, which is to form an equation model using the ARIMA method.

ARIMA is an approach method used for forecasting the probability of an event in the future. This method requires stationary data. If the data to be used is still not stationary, it is necessary to do differencing as many as *d* times so that the data becomes stationary. Then, it is applied to the ARIMA model (p, d, q), where the p and q values are obtained from several models in ARIMA. There are three equation models used in the ARIMA method, namely:
a. *Auto Regressive* Model (AR)

$$Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \cdots + \theta_p Y_{t-p} + e_t \tag{5}$$

Where:
Y_t = forecasting variable
$Y_{t-1}, \cdots, Y_{t-p}$ = independent variable (lag of forecasting variable)
$\theta_1, \cdots, \theta_p$ = auto regressive parameters
e_t = *error* value

b. *Moving Average* Model (MA)

$$Y_t = e_t - \emptyset_1 e_{t-1} - \emptyset_2 e_{t-2} - \cdots - \emptyset_q e_{t-q} \tag{6}$$

Where:
$e_{t-1}, \cdots, e_{t-q}$ = error value at time t
$\emptyset_1, \cdots, \emptyset_q$ = moving average parameter

c. *Auto Regressive Moving Average* Model (ARMA)

$$Y_t = \theta_1 Y_{t-1} + \cdots + \theta_p Y_{t-p} + e_t - \emptyset_1 e_{t-1} - \cdots - \emptyset_q e_{t-q} \tag{7}$$

Then, in this method, the Kolmogorov Smirnov test was carried out, namely the normality test, where this test aimed to determine whether a variable has data that is normally distributed or not. A data is declared normally distributed if the value of Asymp. Sig. (2-tailed) is more significant than α (0,1). In addition, the Kolmogorov Smirnov test was used to determine the level of correspondence between the distribution function of the observations and a specific theoretical distribution function where which establishes a point that explains the maximum difference between the two [11]. This test can be used to compare the data in the sample to a normal distribution with the same mean and standard deviation.

5. Model Evaluation, an evaluation was carried out to obtain the level of accuracy after obtaining the equation model.

The evaluation of this model used the Mean Absolute Percentage Error (MAPE). After obtaining the model, it is necessary to evaluate the forecasting results [12]. It can be calculated using the equation to test the model's accuracy. (8)

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{Z_t - F_t}{Z_t} x100 \right|}{n} \tag{8}$$

The level of model accuracy using MAPE can be seen in Table 2.

Table 2. MAPE Value

| *Range* MAPE	Categories
< 10%	Very Good
10 – 20%	Good
20 – 50%	Sufficiently Good
>50%	Bad

6. Forecasting 2022, cases of violence against women and children throughout 2022 were forecasted.

After obtaining the models from the two methods, it was continued by predicting the number of violence cases against women and children throughout 2022. In addition, an analysis of the forecast results was carried out. After that, a comparison of the two methods was carried out based on the forecasting results.

3. RESULTS AND DISCUSSION

3.1. Prediction with DES method

Based on the data in **Table 1**, a total of 46 data were obtained, starting from January 2018 to October 2021. The representation of cases of violence against women and children in Jakarta can be seen in Figure 1.
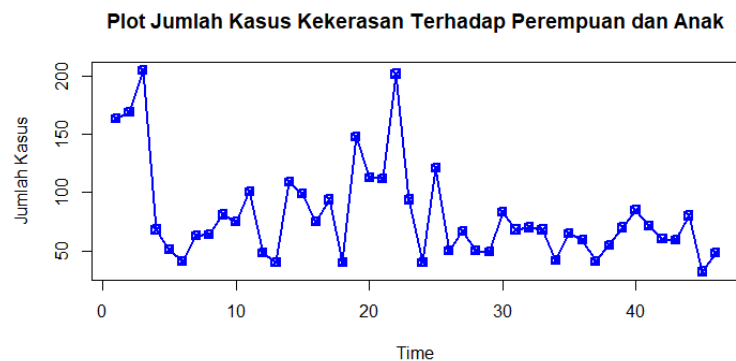
**Figure 1. The Number of Cases of Violence Against Women and Children**

Based on Figure 1, it can be seen that the data has increased over time. Therefore, it can be seen in the picture above that there is a trend that connects the points where the points connect the data every month. Furthermore, the parameter optimization was carried out by trial and error and the results of the parameter optimization are presented in Table 3.

Table 3. Parameter Optimization Results

Parameter Optimization	
Parameter α	Parameter β
0.5097743	0.03443752

After the parameter optimization results were obtained, the next step was to determine the single smoothing and trend of data on violence against women and children. Then, the forecasting using the DES

method was done. The results of forecasting data on violence against women and children from November 2021 to December 2022 can be seen in Figure 2.

Plot Hasil Peramalan
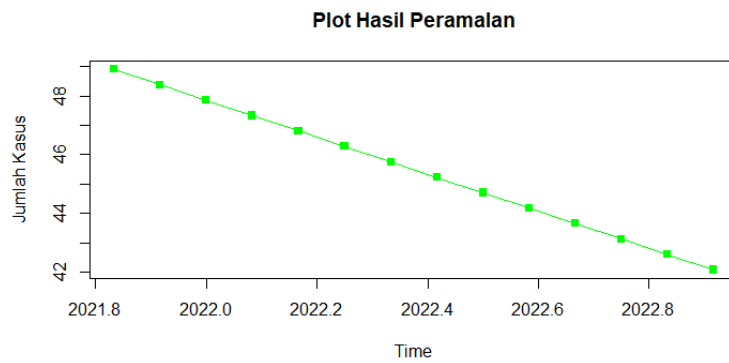
Figure 2. Forecasting Results

Based on Figure 2, forecasting results from November 2021 to December 2022 have decreased every month. Throughout 2021, starting from January to October, it can be seen that the number of cases has increased and decreased. In the forecasting results presented in Figure 3, it can be seen that the number of cases is at a reasonably stable value and is starting to decrease slowly, namely by one number every month.

3.2.   Prediction with ARIMA method

Based on Table 1. with a total of 46 data, it can be represented in Figure 3.
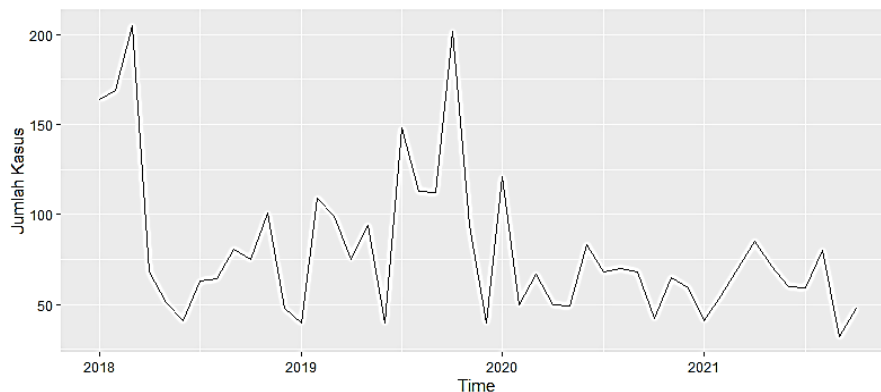

Figure 3. The Number of Violence Cases Against Women and Children

In Figure 3. It can be seen that the data is still not stationary for the average value. Therefore, it is necessary to do differencing until the data is stationary. Differencing was performed twice to make the data stationary. The results of differencing on the data are as shown in Figure 4.
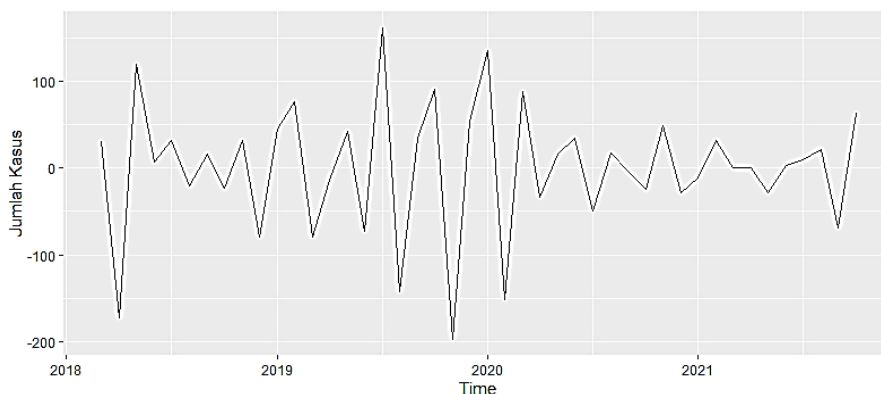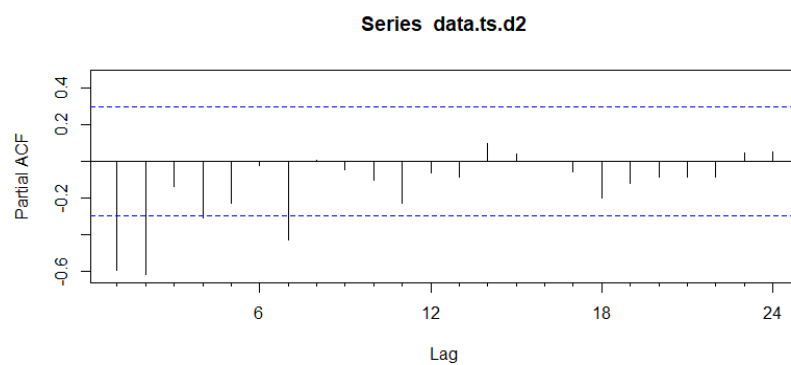

Figure 4. Number of Violence Cases After Differencing

Based on Figure 4, it can be seen that the data is stationary with respect to the average value. Then, identification is carried out based on the ACF and PACF plots which are represented in Figure 5 and Figure 6.

Series data.ts.d2
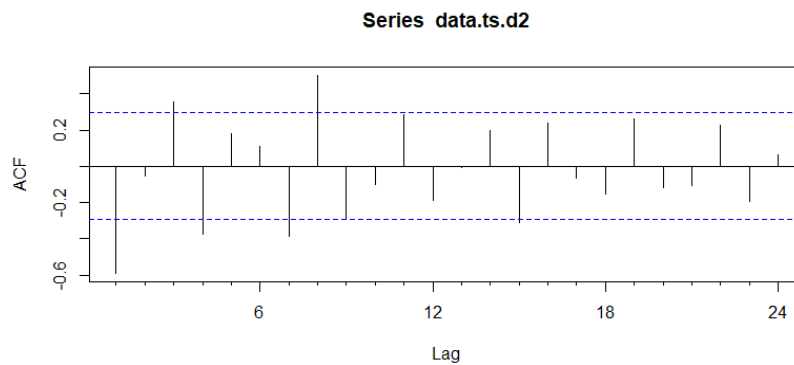
Figure 5. Stationary ACF Data Plot

Series data.ts.d2

Figure 6. Stationary Data PACF Plot

Based on Figure 5, it can be seen that there is a cut-off in the first Lag. Meanwhile, there is a cut-off in the first and second Lag in Figure 6. Then, the ARIMA model was formed with the best values presented in Table 4.

Table 4. ARIMA Model

Model	AIC
AR(2)	467.4813
MA(1)	466.9815
ARMA(2,1)	483.8236

Based on Table 4, the best ARIMA model is obtained, namely ARIMA (2,2,1) with parameter estimates presented in Table 5.

Table 5. ARIMA Model (2,2,1)

Parameters	Coefficients	Standard *Error*
AR1	-1.1710	0.0976
AR2	-0.7686	0.0989
MA1	-1.0000	0.0649

Then, the Kolmogorov Smirnov test was carried out on the ARIMA (2,2,1) model, and it was found that the residual value was normally distributed. Therefore, it can be concluded that this model is the best and will be used in forecasting. By using the best model, the forecasting results are obtained in Figure 7.
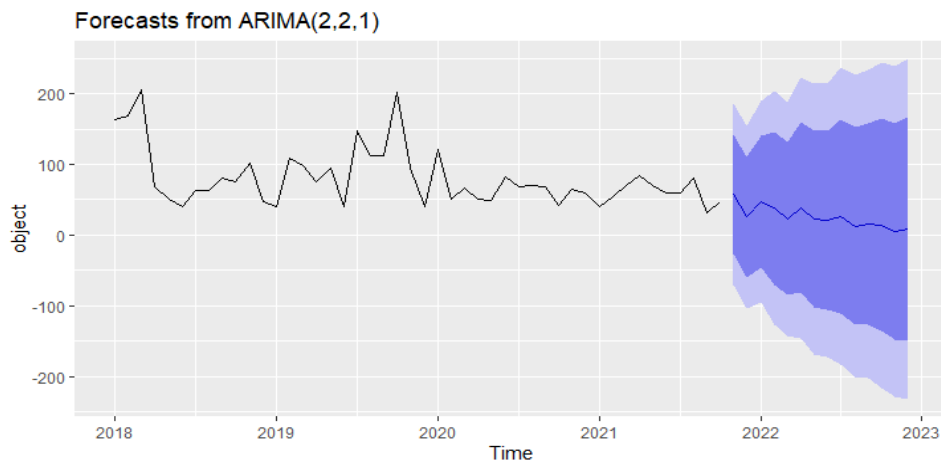
Figure 7. Forecasting Results

Based on Figure 7, forecasting results from November 2021 to December 2022 have decreased every month. It can be seen that the number of cases has decreased significantly every month. The decrease can be seen in the figure above. It decreased by almost half the value in November 2021 from November 2021 to December 2021.

3.3. Model Accuracy

The error rate measurement was carried out after obtaining the forecasting results using the two methods, namely the DES and ARIMA methods. To measure the accuracy of the model, the researcher used MAPE. If the MAPE value is more than 50%, the forecasting model is bad. On the other hand, if the MAPE value is less than 50%, the forecasting model is classified as good. The results of the forecasting error rates of the two methods are presented in Table 6.

Table 6. Model Accuracy

Model	Accuracy Value
DES Model	44.91953
ARIMA Model	39.0398

Based on Table 6, it can be seen that the accuracy of the model of the two methods is quite good because the value is still below 50%. However, the model's accuracy is very good in general, which is at a value below 10%.

3.4. Discussion

This study used the highest model accuracy in predicting violence cases against women and children. Zainal Arifin, et al who conducted research related to the number of unemployed using the DES method obtained a MAPE value of 9.71% [13]. There is also a study conducted by Lafnidita Farosanti, et al which predicts sales of medical and laboratory equipment using the ARIMA method and obtains an RMSE value of 4,129 [14]. In addition, in a study conducted by Yulinar, et al which compared the DES and ARIMA methods in the case study of imported values, the MAPE values for ARIMA and DES were 10.04 and 7.62, respectively [15]. From these studies, it is known that the DES and ARIMA methods have almost the same effectiveness. Thus, the researcher decided to compare the two methods for predicting cases of violence against women and children to determine which one is more effective.

4. CONCLUSION

Based on the research results on forecasting cases of violence against women and children in Jakarta using the DES and ARIMA methods, it can be concluded that cases of violence against women and children have decreased throughout 2022. Based on the value of the forecasting error rate, it is concluded that the ARIMA method is recommended for predicting cases of violence against women and children.

REFERENCES

[1]   M. G. Sopacua, "Konsep Ideal Pencegahan Kekerasan Dalam Rumah Tangga Terhadap Perempuan," vol. 4, 2022.

[2]   A. Kamal, P. Anggraini, and Renita Astri, "Web Untuk Pengaduan Bagi Korban Kekerasan Terhadap Perempuan dan Anak," *J. Sains dan Inform.*, vol. 4, no. 1, pp. 63–69, 2018.

[3]   M. Indriyani, *Pemodelan regresi spline menggunakan metode penalized spline pada data jumlah kasus kekerasan terhadap anak di indonesia*. 2020.

[4]   A. H. Al Farikhi and M. Y. Darsyah, "Perbandingan Autoregressive Integrated Moving Average (Arima) dan Double Exponential Smoothing pada Peramalan Curah Hujan di Provinsi Aceh," *Statistika*, vol. 4, no. 1, pp. 471–478, 2018.

[5]   A. K. Rachmawati, "Peramalan Penyebaran Jumlah Kasus Covid19 Provinsi Jawa Tengah dengan Metode ARIMA," *Zeta - Math J.*, vol. 6, no. 1, pp. 11–16, 2020, doi: 10.31102/zeta.2021.6.1.11-16.

[6]   S. Kinasih, A. Agoestanto, and Sugiman, "Optimasi Parameter pada Model Exponential Smoothing Menggunakan Metode Golden Section untuk Pemilihan Model Terbaik dan Peramalan Jumlah Wisatawan Provinsi Jawa Tengah," *Unnes J. Math.*, vol. 7, no. 1, pp. 38–46, 2018.

[7]   E. S. Sembiring and Z. Syahputra, "Prediksi Aktivitas Tanpa Masker Dengan Kombinasi Metode Single Exponential Smoothing Dan Fuzzy Time Series," vol. 2, no. 1, pp. 57–62, 2022.

[8]   M. H. I. Priambudi, E. L. Amalia, and A. N. Pramudhita, "Sistem Informasi Peramalan Jumlah Pengunjung Menggunakan Metode Double Exponential Smoothing (Studi Kasus Body Gym Kota Malang)," *J. Inform. Polinema*, vol. 7, no. 1, pp. 23–28, 2020, doi: 10.33795/jip.v7i1.431.

[9]   C. Chandra and S. Budi, "Analisis Komparatif ARIMA dan Prophet dengan Studi Kasus Dataset Pendaftaran Mahasiswa Baru," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 278–287, 2020, doi: 10.28932/jutisi.v6i2.2676.

[10]   F. Fejriani, M. Hendrawansyah, L. Muharni, S. F. Handayani, and Syaharuddin, "Forecasting Peningkatan Jumlah Penduduk Berdasarkan Jenis Kelamin menggunakan Metode Arima," *J. Kajian, Penelit. dan Pengemb. Pendidik.*, vol. 8, no. 1 April, pp. 27–36, 2020, [Online]. Available: http://journal.ummat.ac.id/index.php/geography/article/view/2261/pdf.

[11]   H. D. Saribu, "Pengaruh Brand Equity Terhadap Purchase Intention," *J. Mantik Penusa*, vol. 3, no. 1, pp. 21–26, 2019, [Online]. Available: http://e-jurnal.pelitanusantara.ac.id/index.php/mantik/article/view/574.

[12]   I. Nabillah and I. Ranggadara, "Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut," *JOINS (Journal Inf. Syst.*, vol. 5, no. 2, pp. 250–255, 2020, doi: 10.33633/joins.v5i2.3900.

[13]   Z. Arifin, J. Herliani, and Hamdani, "Peramalan Pengangguran Menggunakan Metode Double Exponential Smoothing Di Provinsi Kalimantan Timur," *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 1, pp. 24–29, 2019.

[14]   L. Farosanti and H. Mubarok, "Analisa Peramalan Penjualan Alat Kesehatan dan Laboratorium di PT . Tristania Global Indonesia Menggunakan Metode ARIMA," *J. Inform. Merdeka Pasuruan*, vol. 7, no. 1, pp. 14–18, 2022, doi: http://dx.doi.org/10.37438/jimp.v7i1.428.

[15]   Y. I. AJUNU, N. ACHMAD, and M. R. F. PAYU, "Perbandingan Metode Autoregressive Integrated Moving Average Dan Metode Double Exponential Smoothing Dari Holt Dalam Meramalkan Nilai Impor Di Indonesia," *Jambura J. Probab. Stat.*, vol. 1, no. 1, pp. 37–46, 2020, doi: 10.34312/jjps.v1i1.5393.