Kontinu: Jurnal Penelitian Didaktik Matematika
E-ISSN: 2656-5544
P-ISSN: 2715-7326
Vol. 8, No.1, Mei 2024
Hal. 1-19

Modeling Rainfall in Jakarta with Hybrid ARIMAX-ANN Model

Mahmudi[1)*], Issa Bella Safitri[2)], Dhea Urfina Zulkifli[3)]

Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia [1), 2), 3)]

*Corresponding email: mahmudi@uinjkt.ac.id

Abstrak. Perkiraan musim hujan dan musim kemarau di Indonesia semakin kompleks, memerlukan metode peramalan yang efektif. Penelitian ini difokuskan pada peramalan curah hujan di Jakarta yang meliputi Stasiun Kemayoran dan Stasiun Tanjung Priok dengan model Hybrid ARIMAX-ANN. Pada pemodelan ARIMAX, variabel eksogen yang dilibatkan adalah variabel kelembapan. Hasil akurasi model Hybrid ARIMAX-ANN akan dibandingkan dengan model ARIMAX. Hasil penelitian ini menunjukkan bahwa model Hybrid ARIMAX-ANN memberikan akurasi yang lebih baik. Pada Stasiun Kemayoran, model Hybrid ARIMAX-ANN (1,0,0) dengan 1 hidden layer menunjukkan MAPE yang lebih rendah, yakni sebesar 21.145% dibandingkan dengan model ARIMAX. Sementara itu, pada Stasiun Tanjung Priok, model Hybrid ARIMAX-ANN (1,0,0) dengan 2 hidden layer memiliki MAPE yang lebih rendah, yakni sebesar 37.416% dibandingkan dengan model ARIMAX. Hasil penelitian menunjukkan bahwa model Hybrid ARIMAX-ANN memberikan akurasi yang lebih baik dibandingkan dengan model ARIMAX pada pemodelan curah hujan di Jakarta. Penerapan model Hybrid ARIMAX-ANN pada Stasiun Kemayoran menghasilkan akurasi yang lebih baik dibandingkan dengan penerapan di Stasiun Tanjung Priok.

Kata kunci: ARIMAX, Hybrid ARIMAX-ANN, Curah Hujan

Abstract. *Forecasting Indonesia's rainy and dry seasons is increasingly complex, requiring effective forecasting methods. This research is focused on rainfall forecasting in Jakarta, which includes Kemayoran Station and Tanjung Priok Station, using the Hybrid ARIMAX-ANN model. In ARIMAX modeling, the exogenous variable involved is the humidity variable. The accuracy results of the Hybrid ARIMAX-ANN model will be compared with the ARIMAX model. The results of this study show that the Hybrid ARIMAX-ANN model provides better accuracy. At Kemayoran Station, the Hybrid ARIMAX-ANN (1,0,0) model with 1 hidden layer shows a lower MAPE, 21.145%, than the ARIMAX model. Meanwhile, at Tanjung Priok Station, the Hybrid ARIMAX-ANN (1,0,0) model with 2 hidden layers has a lower MAPE, which is 37.416% compared to the ARIMAX model. The results show that the Hybrid ARIMAX-ANN model provides better accuracy than the ARIMAX model in rainfall modeling in Jakarta. Applying the Hybrid ARIMAX-ANN model at Kemayoran Station produces better accuracy than the application at Tanjung Priok Station.*

Keywords: *ARIMAX, Hybrid ARIMAX-ANN, Rainfall*

Kontinu: Jurnal Penelitian Didaktik Matematika
E-ISSN: 2656-5544
P-ISSN: 2715-7326
Vol. 8, No.1, Mei 2024
Hal. 1-19

INTRODUCTION

It has been challenging to forecast Indonesia's rainy and dry seasons recently. Rainfall forecasting is essential in forecasting weather, managing water resources agriculture, and reducing the potential for natural disasters. According to (Lavers Villarini, 2013), accurate rainfall forecasting can help make better decisions in various sectors, improve water resource use efficiency, and reduce the risk of natural disasters. At this time, developments in meteorology and climatology are significant; rainfall needs to be analyzed in depth to face the challenges of climate change and its impact on daily life. Jakarta, one of Indonesia's most populous cities, has a growing population. It is located on the north coast of Java Island, including lowlands, with an altitude of 7 meters above sea level.

Due to complex natural conditions, Jakarta, one of Indonesia's metropolitan cities, often experiences difficulty predicting accurate rainfall (Smith & Katz, 2013). Jakarta is faced with climate change and extreme weather phenomena. Changes in extreme weather patterns, such as longer dry periods or heavy rains that increase in intensity can worsen the situation in Jakarta. Therefore, efforts to improve the ability to forecast rainfall are becoming increasingly important, not only to address current challenges but also to anticipate and mitigate future impacts of climate change.

The Autoregressive Integrated Moving Average (ARIMA) model, also known as the Box Jenkins method, is one of the most commonly used methods for time series modeling. The ARIMA model has been used in previous studies to forecast rainfall in Jambi (Badri et al., 2023), rainfall prediction in Padang City (Mailisa et al., 2023), rainfall forecasting at Class II Minangkabau Meteorological Station Padang Pariaman (Nur & Putri, 2024), and also for forecasting during the COVID-19 pandemic in Brazil (Ospina et al., 2023). The ARIMA model does not involve exogenous variables that affect the variables to be predicted. The development of the ARIMA model with the addition of exogenous variables forms a new model called the Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX) model. The ARIMAX model has been used in research related to rainfall forecasting in Gunungpati Semarang (Suryani et al., 2018), rainfall forecasting in Ketapang Regency (Wijayanti et al., 2021), rainfall forecasting in

Pangkalpinang (Amelia et al., 2021), and rainfall forecasting in Bali (Sumarjaya et al., 2021).

Artificial Neural Network (ANN) models are machine learning algorithms used to identify patterns in data with human intelligence. The ANN model consists of an input layer and an output layer. ANN models can help in predicting rainfall. Previous research has applied the ANN model to analyze runoff based on rainfall in the Upper Brantasan Subwatershed (Suhartanto & Cahya, 2019). In addition, the same model has also been used to predict rainfall at Kemayoran Meteorological Station (Putra & Rani, 2020), as well as to predict monthly rainfall in Wajo Regency, South Sulawesi (Fitriyanti, 2022).

The Hybrid ARIMAX-ANN model is a combination of the ARIMAX model and the ANN model. In the Hybrid ARIMAX-ANN model, the ANN model is used to predict errors from the ARIMAX model so that the Hybrid ARIMAX-ANN model can improve forecasting accuracy. Previous research related to the Hybrid ARIMAX-ANN model has applied it in forecasting non-cash payment transactions with calendar configuration (Putera, 2020), forecasting rice production damage caused by the Yellow Stem borer pest (Supriya, 2020), and forecasting rice production in the fall (Neog et al., 2022).

This study will use the Hybrid ARIMAX-ANN model to model rainfall in Jakarta. The Hybrid ARIMAX-ANN model is compared with the ARIMAX model to see its accuracy. The purpose of comparing the two models is to determine whether the Hybrid ARIMAX-ANN model significantly improves rainfall forecasting accuracy in Jakarta compared to the ARIMAX model.

RESEARCH METHODS

This research will compare the accuracy of ARIMAX and Hybrid ARIMAX-ANN models in forecasting rainfall in the Jakarta area. The data used are rainfall data and humidity and temperature data as exogenous variables obtained from two Meteorological Stations in Jakarta, namely Kemayoran Meteorological Station and Tanjung Priok Maritime Meteorological Station. The data used is monthly data from July 2018 to June 2023. The data is divided into training data from July 2018 to December 2022 and testing data from January 2023 to June 2023.

Kontinu: Jurnal Penelitian Didaktik Matematika
E-ISSN: 2656-5544
P-ISSN: 2715-7326
Vol. 8, No.1, Mei 2024
Hal. 1-19

This research focuses on forecasting rainfall in Jakarta using the ARIMAX model with exogenous variables, namely humidity and temperature. In addition, it is also modeled with Hybrid ARIMAX-ANN, which combines artificial neural networks (ANN) to improve forecasting accuracy. The stages of ARIMAX and Hybrid ARIMAX-ANN modeling are as follows.

1. Data Exploration.

   Data exploration was conducted to determine the maximum value, minimum value, mean, and standard deviation.

2. Data Stationarity Testing.

   Data stationarity in variance and mean are checked before ARIMAX modeling is performed. Checking data stationarity in variance uses the Box-Cox test. Because there is a zero value in the data, the data is first added with one so Box-Cox testing can be done. When the λ value in the Box-Cox test is one, the data is considered stationary in variance. Checking stationarity in the mean is done with the Augmented Dickey-Fuller (ADF) test. The hypothesis used in the ADF test is as follows (Shumway & Stoffer, 2016).

   $H_0: \phi = 1$ (data has unit root or not stationary).

   $H_1: \phi < 1$ (data does not have unit root or stationary).

3. ARIMA Modeling.

   ARIMA modeling uses Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to determine the ARIMA model parameters. After that, a significant ARIMA model is selected. The significant ARIMA model is then tested for normality using the Jarque-Bera test and tests for autocorrelation using the Ljung-Box test to ensure that the selected model meets the basic assumptions of data analysis (Juliana et al., 2023). The following are normality test criteria.

   H_0 : normally distributed data.

   H_1 : data is not normally distributed.

   The following autocorrelation test criteria are as follows :

   H_0 : data does not have autocorrelation.

   H_1 : data there is autocorrelation.

Kontinu: Jurnal Penelitian Didaktik Matematika
E-ISSN: 2656-5544
P-ISSN: 2715-7326
Vol. 8, No.1, Mei 2024
Hal. 1-19

After the normality and autocorrelation tests, the ARIMA model equation will be made based on the smallest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values to select the model that best fits the data.

4. ARIMA Model into ARIMAX Model.

   ARIMAX modeling is a continuation of the significant ARIMA model added with exogenous variables, namely humidity and temperature. Because the exogenous variable of temperature is not significant, an ARIMAX model is made with the exogenous variable of humidity. Selection of the best model in the ARIMAX model is based on the AIC and BIC value criteria. The best ARIMAX model is also tested for normality using the Jarque-Bera test and autocorrelation using the Ljung-Box test. After that, the ARIMAX model equation will be made.

5. Hybrid ARIMAX-ANN Model

   After modeling ARIMAX, the residue of the ARIMAX model is obtained. The residue of the ARIMAX model will be modeled in the ANN model. Then, the ARIMAX and ANN models are combined to improve accuracy in the Hybrid ARIMAX-ANN model.

6. Model Evaluation.

   A comparison of the accuracy results of the Hybrid ARIMAX-ANN model with the ARIMAX model using Mean Absolute Percentage Error (MAPE) is conducted. Therefore, it will be shown whether using the Hybrid ARIMAX-ANN model in forecasting data significantly increases accuracy compared to the ARIMAX model.

RESULT AND DISCUSSION

Descriptive Statistics

The data used starts from July 2018 to June 2023. The total data in each station is 60 months, where 54 months are used as training data and 6 months as testing data.

Table 1. Descriptive Statistics of Data

Variable	Maximum Value	Minimum Value	Average	Standard Deviation
Kemayoran Station				
Rainfall	1043.2	0	178.04	184.34
Humidity	84.07	63.37	75.95	4.34
Temperature	29.64	27.16	28.56	0.6

Tanjung Priok Station				
Rainfall	784.5	0	163.01	159.55
Humidity	84.42	69.53	77.58	4.19
Temperature	29.72	27.33	28.70	0.56

Based on the results of Table 1, it can be seen that Kemayoran Station has a higher average rainfall and a much larger standard deviation than Tanjung Priok Station. Both stations show relatively similar humidity and temperature levels.

Data Stationarity

By visualizing the data plot, we can identify whether the time series meets the stationary assumption or exhibits time-dependent behavior.

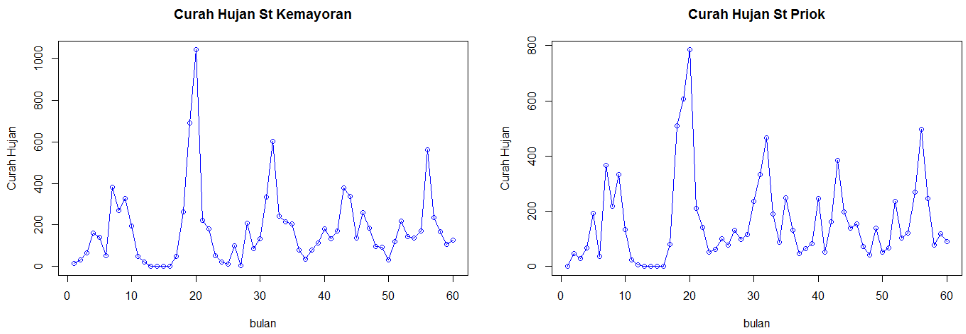


Figure 1. Plot of Rainfall Data

Based on Figure 1 above, it can be seen that the rainfall plots at both stations show characteristics that are not stationarity in variance and do not show seasonal patterns. To ensure stationarity more accurately, a stationary test on variance using the Box-Cox test will be conducted, as well as a stationary test on the mean using the Augmented Dickey-Fuller (ADF) test. Since there is a value of 0 in the rainfall data, the Box-Cox test begins by adding a constant 1 to all rainfall data. This is done to ensure that the Box-Cox test can be performed without causing mathematical problems caused by the presence of the value 0 in the data (Box et al., 2015).

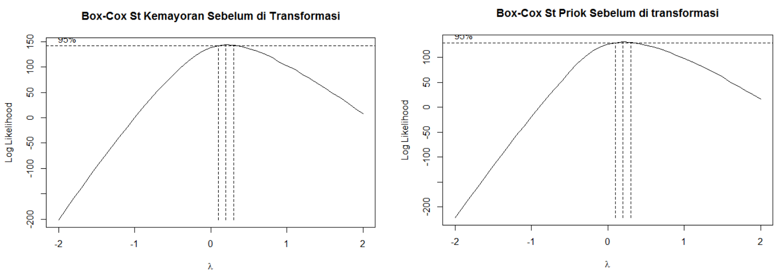


Figure 2. Box-Cox Testing Plot of Rainfall Data

The parameter λ=1 in the Box-Cox transformation can be considered variance stationary, meaning the variance value remains constant over time (Box et al., 2015). Based on Figure 2, it can be seen that the value of λ at both stations is not equal to 1, indicating that Box-Cox transformation is needed to make the value of λ become 1 at both stations. After the Box-Cox transformation, the Box-Cox test results in a value of λ = 1, indicating that the data is stationary in variance. Furthermore, data stationarity in the mean is tested using the ADF test.

Table 2. Result of Data Stationarity Test in Mean

Variable	P-value	Description
Kemayoran Station		
Rainfall	0.01	Significant
Humidity	0.01	Significant
Temperature	0.01	Significant
Tanjung Priok Station		
Rainfall	0.01	Significant
Humidity	0.01	Significant
Temperature	0.015	Significant

Based on the results in Table 2, the stationarity of the three variables at each station using the ADF test shows that the p-value of each variable is less than 0.05. This indicates that all variables have met stationary in the mean. Thus, rainfall, humidity, and temperature data at Kemayoran Station and Tanjung Priok Station show stability in the mean during the observed period.

ARIMA Model

Before the ARIMA model, an important step is to check the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). These two functions are essential in the identification stage of the ARIMA model. We can identify the optimal Autoregressive (AR) and Moving Average (MA) orders for developing the ARIMA model.

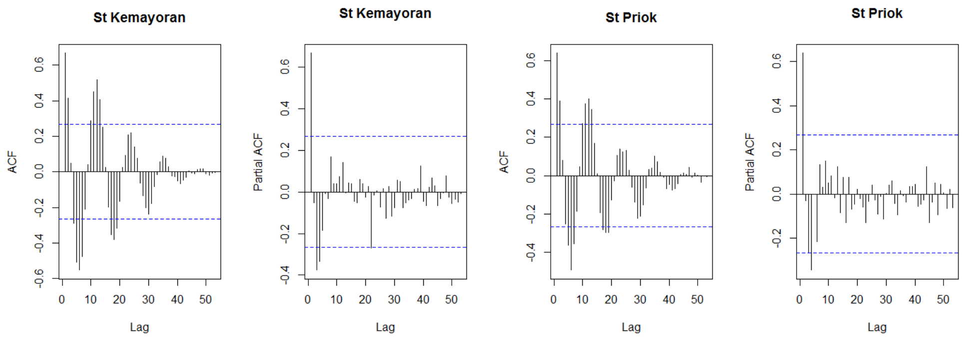


Figure 3. ACF and PACF Plot of Rainfall Data

Based on Figure 3 above, candidates for the ARIMA model of Kemayoran Station and Tanjung Priok Station are obtained. At Kemayoran Station, the ACF plot occurs cut-off at lag 1, 2, and 4, while the PACF plot cut-off at lag 1, 3, and 4. At Tanjung Priok Station, the ACF plot occurs cut-off at lag 1 and 2, while the PACF plot cut-off occurs at lag 1, 3, and 4. This study ignores the cut-off above lag 4 because it produces a poor model. Furthermore, model parameter estimation is carried out based on the results of the previous model identification. The following presents the estimation results of significant model candidates at both stations.

Table 3. ARIMA Modeling Results of Rainfall Data

Model	Parameter	Estimation Value	P-value	Decision	AIC	BIC
Kemayoran Station						
ARIMA (0,0,1)	MA(1)	0.496	0.000	Significant	93.711	99.678
ARIMA (0,0,2)	MA(1)	0.514	0.000	Significant	85.129	93.085
	MA(2)	0.438	0.002	Significant		
ARIMA (0,0,3)	MA(1)	0.594	0.000	Significant	80.462	90.406
	MA(2)	0.580	0.000	Significant		
	MA(3)	0.322	0.006	Significant		
ARIMA (0,0,4)	MA(1)	0.638	0.000	Significant	78.435	90.369
	MA(2)	0.710	0.000	Significant		
	MA(3)	0.602	0.000	Significant		
	MA(4)	0.312	0.019	Significant		
ARIMA (1,0,0)	AR(1)	0.673	0.000	Significant	81.905	87.872
ARIMA (2,0,2)	AR(1)	1.628	0.000	Significant	70.846	82.780
	AR(2)	-0.893	0.000	Significant		
	MA(1)	-1.120	0.000	Significant		
	MA(2)	0.503	0.000	Significant		
Tanjung Priok Station						

ARIMA (0,0,1)	MA(1)	0.660	0.000	Significant	88.121	94.088
ARIMA (0,0,2)	MA(1) MA(2)	0.544 0.457	0.000 0.000	Significant Significant	84.238	92.194
ARIMA (1,0,0)	AR(1)	0.698	0.000	Significant	79.760	85.727
ARIMA (2,0,2)	AR(1) AR(2) MA(1) MA(2)	1.600 -0.851 -1.029 0.429	0.000 0.000 0.000 0.004	Significant Significant Significant Significant	75.337	87.271
ARIMA (3,0,2)	AR(1) AR(2) AR(3) MA(1) MA(2)	2.158 -1.737 0.434 -1.576 0.834	0.000 0.000 0.016 0.000 0.000	Significant Significant Significant Significant Significant	76.228	90.151

Based on the results listed in Table 3 above, it can be concluded that there are several significant ARIMA models for Kemayoran Station and Tanjung Priok Station. At Kemayoran Station, the significant ARIMA models are ARIMA (0,0,1), ARIMA (0,0,2), ARIMA (0,0,3), ARIMA (0,0,4), ARIMA (1,0,0), and ARIMA (2,0,2). Meanwhile, at Tanjung Priok Station, the significant ARIMA models are ARIMA (0,0,1), ARIMA (0,0,2), ARIMA (1,0,0), ARIMA (2,0,2), and ARIMA (3,0,2). Furthermore, significant ARIMA models are tested for normality with the Jarque Bera and Autocorrelation tests with Ljung Box on residual data.

Table 4. Normality Test and Autocorrelation Test on the ARIMA Model

Model	Normality Test		Autocorrelation Test	
	p-value	Decision	p-value	Decision
Kemayoran Station				
ARIMA (0,0,1)	0.524	Normally Distributed	0.094	There is no Autocorrelation
ARIMA (0,0,2)	0.707	Normally Distributed	0.261	There is no Autocorrelation
ARIMA (0,0,3)	0.362	Normally Distributed	0.711	There is no Autocorrelation
ARIMA (0,0,4)	0.863	Normally Distributed	0.910	There is no Autocorrelation
ARIMA (1,0,0)	0.950	Normally Distributed	0.831	There is no Autocorrelation
ARIMA (2,0,2)	0.522	Normally Distributed	0.765	There is no Autocorrelation
Tanjung Priok Station				

Kontinu: Jurnal Penelitian Didaktik Matematika
E-ISSN: 2656-5544
P-ISSN: 2715-7326
Vol. 8, No.1, Mei 2024
Hal. 1-19

ARIMA (0,0,1)	0.651	Normally Distributed	0.566	There is no Autocorrelation
ARIMA (0,0,2)	0.668	Normally Distributed	0.544	There is no Autocorrelation
ARIMA (1,0,0)	0.605	Normally Distributed	0.924	There is no Autocorrelation
ARIMA (2,0,2)	0.590	Normally Distributed	0.730	There is no Autocorrelation
ARIMA (3,0,2)	0.611	Normally Distributed	0.457	There is no Autocorrelation

Based on the results listed in Table 4, it can be seen that all significant ARIMA model residuals at both Kemayoran Station and Tanjung Priok Station are normally distributed, and there is no autocorrelation. The best ARIMA model is selected based on the smallest AIC and BIC values. At Kemayoran Station and Tanjung Priok Station, the best ARIMA model of rainfall data is the ARIMA (2,0,2) model. Here is the best ARIMA model obtained.

- ARIMA Model (2,0,2) Kemayoran Station

$$Y_t^{(0.2)} = 2.539 + 1.628Y_{t-1}^{(0.2)} - 0.893Y_{t-2}^{(0.2)} + e_t + 1.120e_{t-1} - 0.503e_{t-2}$$

- ARIMA Model (2,0,2) Tanjung Priok Station

$$Y_t^{(0.2)} = 2.491 + 1.560Y_{t-1}^{(0.2)} - 0.851Y_{t-2}^{(0.2)} + e_t + 1.029e_{t-1} - 0.429e_{t-2}$$

ARIMAX Model

After identifying and building a significant ARIMA model for rainfall data at Kemayoran Station and Tanjung Priok Station, this research continues with the ARIMAX model approach. At this stage, the ARIMA model that has been formed will add exogenous data in the form of humidity and temperature data. The following are the estimation results of significant model candidates from Kemayoran Station and Tanjung Priok Station.

Table 5. Results of ARIMAX Modeling with Exogenous Variables of Humidity and Temperature

Model	Parameter	Estimation Value	P-value	Decision
		Kemayoran Station		

ARIMAX (0,0,1)	MA(1)	0.318	0.004	Significant
	Humidity	0.116	0.000	Significant
	Temperature	-0.058	0.636	Not Significant
ARIMAX (1,0,0)	AR(1)	0.381	0.003	Significant
	Humidity	0.115	0.000	Significant
	Temperature	-0.031	0.799	Not Significant
Tanjung Priok Station				
ARIMAX (0,0,1)	MA(1)	0.475	0.001	Significant
	Humidity	0.113	0.000	Significant
	Temperature	-0.014	0.927	Not Significant
ARIMAX (1,0,0)	AR(1)	0.483	0.000	Significant
	Humidity	0.110	0.000	Significant
	Temperature	-0.012	0.932	Not Significant
ARIMAX (2,0,2)	AR(1)	1.096	0.000	Significant
	AR(2)	-0.775	0.000	Significant
	MA(1)	-0.838	0.000	Significant
	MA(2)	1.000	0.000	Significant
	Humidity	0.099	0.000	Significant
	Temperature	-0.101	0.170	No Significant

Based on the results in Table 5 above, the ARIMAX model with the exogenous humidity variable significantly affects rainfall. In contrast, the exogenous variable of temperature does not have a significant impact on rainfall. Since the exogenous temperature variable is insignificant in rainfall, the research continues to form the ARIMAX model with the exogenous humidity variable.

Table 6. Results of ARIMAX Modeling with Exogenous Variable of Humidity

Model	Parameter	Estimation Value	P-value	Decision
Kemayoran Station				
ARIMAX (0,0,1)	MA(1)	0.310	0.004	Significant
	Humidity	0.121	0.000	Significant
ARIMAX (1,0,0)	AR(1)	0.381	0.003	Significant
	Humidity	0.117	0.000	Significant
Tanjung Priok Station				
ARIMAX (0,0,1)	MA(1)	0.474	0.001	Significant
	Humidity	0.114	0.000	Significant
ARIMAX (1,0,0)	AR(1)	0.482	0.000	Significant
	Humidity	0.111	0.000	Significant
ARIMAX (2,0,2)	AR(1)	1.088	0.000	Significant
	AR(2)	-0.642	0.002	Significant
	MA(1)	-0.679	0.002	Significant
	MA(2)	0.602	0.032	Significant
	Humidity	0.107	0.000	Significant

Based on the results listed in Table 6 above, it can be seen that the ARIMAX model with exogenous humidity variable significantly influences rainfall. The models that show significance at Kemayoran Station are ARIMAX (0,0,1) and ARIMAX (1,0,0). Meanwhile, at Tanjung Priok Station are ARIMAX (0,0,1), ARIMAX (1,0,0), and ARIMAX (2,0,2). Furthermore, a significant ARIMAX model is tested for normality with the Jarque Bera and Autocorrelation tests with Ljung Box on residual data.

Table 7. Normality Test and Autocorrelation Test on ARIMAX Model

Model	Normality Test		Autocorrelation Test	
	p-value	Decision	p-value	Decision
Kemayoran Station				
ARIMAX (0,0,1)	0.067	Normally Distributed	0.631	There is no Autocorrelation
ARIMAX (1,0,0)	0.071	Normally Distributed	0.994	There is no Autocorrelation
Tanjung Priok Station				
ARIMAX (0,0,1)	0.363	Normally Distributed	0.957	There is no Autocorrelation
ARIMAX (1,0,0)	0.424	Normally Distributed	0.847	There is no Autocorrelation
ARIMAX (2,0,2)	0.502	Normally Distributed	0.998	There is no Autocorrelation

Based on the results listed in Table 7, it can be seen that all significant ARIMAX model residuals are normally distributed, and there is no autocorrelation. To determine the best model for Kemayoran Station and Tanjung Priok Station, the ARIMAX models will be compared based on each station's AIC and BIC values.

Table 8. AIC and BIC Values of the ARIMAX Model With Exogenous Variable Humidity

Model	AIC	BIC
Kemayoran Station		
ARIMAX (0,0,1)	48.169	56.125
ARIMAX (1,0,0)	46.713	54.669
Tanjung Priok Station		
ARIMAX (0,0,1)	61.204	69.160
ARIMAX (1,0,0)	59.201	67.157
ARIMAX (2,0,2)	61.311	75.234

Kontinu: Jurnal Penelitian Didaktik Matematika
E-ISSN: 2656-5544
P-ISSN: 2715-7326
Vol. 8, No.1, Mei 2024
Hal. 1-19

Based on the results listed in Table 8 above, it can be seen that the smallest AIC and BIC values at both Kemayoran Station and Tanjung Priok Station are the ARIMAX (1,0,0) model. Here is the best ARIMAX model obtained.

- ARIMAX Model (1,0,0) Kemayoran Station

$$Y_t^{(0.2)} = -6.358 + 0.381Y_{t-1}^{(0.2)} + 0.117X_{1,t} + e_t$$

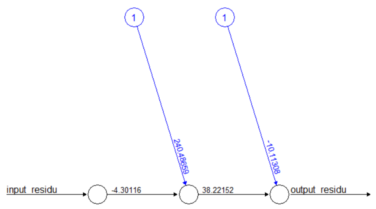
- ARIMAX Model (1,0,0) Tanjung Priok Station

$$Y_t^{(0.2)} = -6.112 + 0.482Y_{t-1}^{(0.2)} + 0.111X_{1,t} + e_t$$

Hybrid ARIMAX-ANN Modeling

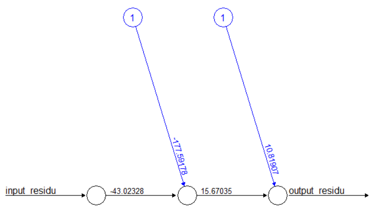
After modeling ARIMAX, the next step is to model Hybrid ARIMAX-AN. The forecasting results from the ARIMAX model will be used as the linear component. Meanwhile, the residuals from the ARIMAX model will be modeled using an Artificial Neural Network (ANN) and used as a non-linear component. The input to ANN modeling is the residual data at time t-1, while the output is the residual data at time t. This research uses hidden layers 1, 2, and 3. The Hybrid ARIMAX-ANN model is a method that combines the ARIMAX method with ANN, which is expected to provide more accurate prediction results. The results of the ANN architecture formed at Kemayoran Station and Tanjung Priok Station are as follows:

ANN Architecture of Kemayoran Station with 1 Hidden Layer



ANN Architecture of Kemayoran Station with 2 Hidden Layer

ANN Architecture of Tanjung Priok Station with 1 Hidden Layer



ANN Architecture of Tanjung Priok Station with 2 Hidden Layer

ANN Architecture of Kemayoran Station with 3 Hidden Layer

ANN Architecture of Tanjung Priok Station with 3 Hidden Layer

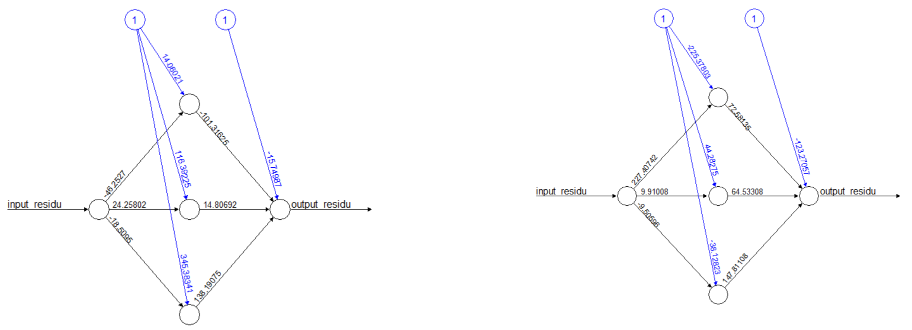


Figure 4. ANN Architecture on ARIMAX Residual Data

Based on Figure 4 above, the Hybrid ARIMAX-ANN equation is as follows.

- ARIMAX-ANN (1,0,0) Model with 1 Hidden Layer at Kemayoran Station

$$Y_t = L_t + N_t$$

with

$$L_t = -6.358 + 0.381Y_{t-1}^{(0.2)} + 0.117X_{1,t} + e_t$$

$$N_t = -10.113 + 38.221f_1^h$$

- ARIMAX-ANN (1,0,0) Model with 2 Hidden Layer at Kemayoran Station

$$Y_t = L_t + N_t$$

with

$$L_t = -6.358 + 0.381Y_{t-1}^{(0.2)} + 0.117X_{1,t} + e_t$$

$$N_t = -10.113 + 60.913f_1^h - 29.219f_2^h$$

- ARIMAX-ANN (1,0,0) Model with 3 Hidden Layer at Kemayoran Station

$$Y_t = L_t + N_t$$

with

$$L_t = -6.358 + 0.381Y_{t-1}^{(0.2)} + 0.117X_{1,t} + e_t$$

$$N_t = -15.750 - 101.316f_1^h + 14.807f_2^h + 138.191f_3^h$$

- ARIMAX-ANN (1,0,0) Model with 1 Hidden Layer at Tanjung Priok Station

$$Y_t = L_t + N_t$$

with

$$L_t = -6.112 + 0.482Y_{t-1}^{(0.2)} + 0.111X_{1,t} + e_t$$

$$N_t = 10.819 + 15.670f_1^h$$

- ARIMAX-ANN (1,0,0) Model with 2 Hidden Layer at Tanjung Priok Station

$$Y_t = L_t + N_t$$

with

$$L_t = -6.112 + 0.482Y_{t-1}^{(0.2)} + 0.111X_{1,t} + e_t$$

$$N_t = 107.183 - 91.235f_1^h - 159.125f_2^h$$

- ARIMAX-ANN (1,0,0) Model with 3 Hidden Layer at Tanjung Priok Station

$$Y_t = L_t + N_t$$

with

$$L_t = -6.112 + 0.482Y_{t-1}^{(0.2)} + 0.111X_{1,t} + e_t$$

$$N_t = -123.271 + 72.581f_1^h + 64.533f_2^h + 147.811f_3^h$$

Model Evaluation

The best models obtained from the ARIMAX and Hybrid ARIMAX-ANN models will be evaluated based on the model accuracy of the testing data. This process aims to determine the extent of each model's ability to forecast rainfall. Model accuracy is based on the Mean Absolute Percentage Error (MAPE) value. The best model is the model with the smallest MAPE value. The following are the results of calculating the MAPE value of the testing data at Kemayoran Station and Tanjung Priok Station.

Table 9. MAPE Value Results on Testing Data

Model	MAPE
Kemayoran Station	
ARIMAX (1,0,0)	25.096%

Hybrid ARIMAX-ANN (1,0,0) with 1 hidden layer	21.145%
Hybrid ARIMAX-ANN (1,0,0) with 2 hidden layer	21.575%
Hybrid ARIMAX-ANN (1,0,0) with 3 hidden layer	23.208%
Tanjung Priok Station	
ARIMAX (1,0,0)	40.245%
Hybrid ARIMAX-ANN (1,0,0) with 1 hidden layer	44.044%
Hybrid ARIMAX-ANN (1,0,0) with 2 hidden layer	37.416%
Hybrid ARIMAX-ANN (1,0,0) with 3 hidden layer	44.126%

Based on the results listed in Table 9 above, at Kemayoran Station, the smallest MAPE value is generated from the Hybrid ARIMAX-ANN model with 1 hidden layer. Meanwhile, Tanjung Priok Station's smallest MAPE value is generated from the Hybrid ARIMAX-ANN model with 2 hidden layers. This shows that the Hybrid ARIMAX-ANN model produces better accuracy than the ARIMAX model in rainfall modeling in Jakarta.

The above results are in line with research conducted by (Putera, 2020) on forecasting non-cash transaction payments, research conducted by (Supriya, 2020) on forecasting rice plant damage caused by Yellow Stem pests, and research conducted by (Neog et al., 2022) on forecasting rice production in the fall. The three studies resulted in the same conclusion that the Hybrid ARIMAX-ANN model is better than the ARIMAX model.

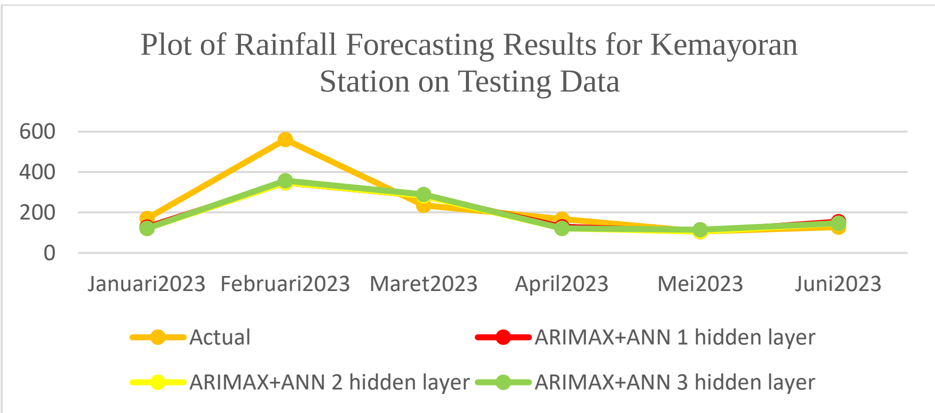


Figure 5. The plot of Rainfall Forecasting Results for Kemayoran Station on Testing Data

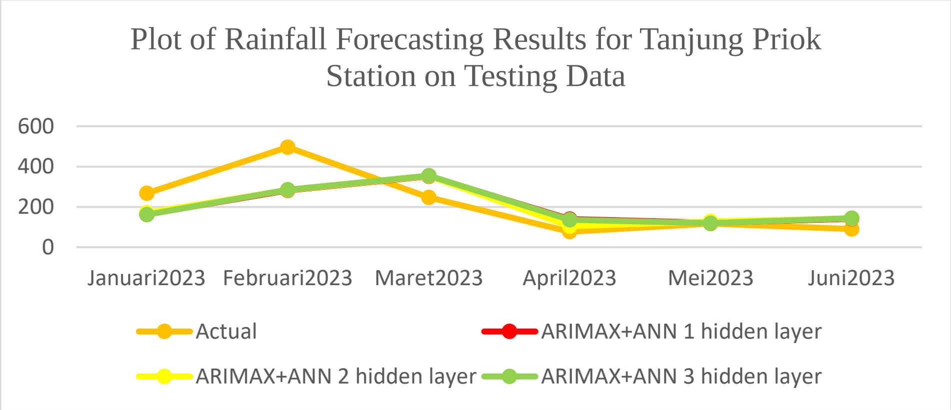


Figure 6. Plot of Rainfall Forecasting Results for Tanjung Priok Station on Testing Data

Figure 5 shows that the Hybrid ARIMAX-ANN model with 1 hidden layer at Kemayoran Station produces forecasts closest to the actual value compared to other models. Meanwhile, Figure 6 shows that the Hybrid ARIMAX-ANN model with 2 hidden layers at Tanjung Priok Station produces forecasts that are closest to actual values compared to other models.

CONCLUSION

Based on the results of this study, a significant exogenous variable in rainfall modeling using the ARIMAX model is the humidity variable. The temperature variable is not included in the ARIMAX modeling because it does not significantly affect rainfall modeling in Jakarta. The Hybrid ARIMAX-ANN model with a certain number of hidden layers produces better accuracy than the ARIMAX model in modeling rainfall in Jakarta. The number of hidden layers that produce better accuracy at Kemayoran Station is 1, while at Tanjung Priok Station, there are 2 hidden layers. Hybrid ARIMAX-ANN modeling at Kemayoran Station produces better accuracy with a MAPE value of 21.145% than at Tanjung Priok Station with a MAPE value of 37.416%. The following research can involve other exogenous variables that may significantly affect the modeling.

REFERENCES

Amelia, R., Dalimunthe, D. Y., Kustiawan, E., & Sulistiana, I. (2021). ARIMAX Model For Rainfall Forecasting In Pangkalpinang, Indonesia. *IOP Conf. Series: Earth and Environmental Science, 926(1)*. https://doi.org/10.1088/1755-1315/926/1/012034

Badri, A. Al, Nandarie, A. C. A., & Haryanto, Y. D. (2023). Optimasi Model ARIMA Dalam Prakiraan Curah Hujan Di Jambi. *Geographia : Jurnal Pendidikan Dan Penelitian Geografi*, *4*(1), 39–43. https://doi.org/10.53682/gjppg.v4i1.7041

Box, G. E. ., Jenkins, G. ., Reinsel, G. ., & Ljung, G. . (2015). *Time Series Analysis Forecasting and Control* (5th ed.).

Fitriyanti. (2022). Aplikasi Jaringan Syaraf Tiruan Backpropagation Dalam Prediksi Curah Hujan Bulanan Di Kabupaten Wajo Sulawesi Selatan. *Jurnal Pendidikan Fisika*, *11*(1), 44–55. https://doi.org/10.24252/jpf.v11i1.33142

Juliana, A., Junaid, M. T., Yuniati, N., & Malik, A. D. (2023). Studi efisiensi pasar bentuk lemah pada pasar modal indonesia : sebelum dan sesudah covid 19. *AKUNTABEL: Jurnal Akuntansi Dan Keuangan*, *20*(1), 18–30.

Lavers, D. A., & Villarini, G. (2013). The nexus between atmospheric rivers and extreme precipitation across Europe. *Geophysical Research Letters*, *40*(12), 3259–3264. https://doi.org/10.1002/grl.50636

Mailisa, P. D., Hasibuan, L. H., Nur, R. A., & Asfa'ani, E. (2023). Optimasi Prediksi Curah Hujan Kota Padang dengan Model ARIMA. *Mathematics and Applications Journal*, *Vol 5*(2), 177–185.

Neog, B., Gogoi, B., & Patowary, A. N. (2022). Development Of Hybrid Time Series Models For Forecasting Autumn Rice Using ARIMAX-ANN And ARIMAX-SVM. *Annals Of Forest Research*, *65*(1), 9119–9133.

Nur, R. A., & Putri, D. M. (2024). Penerapan ARIMA pada Data Curah Hujan di Stasiun Meteorologi Kelas II Minangkabau Padang Pariaman. *Jostech : Journal of Science and Technology*, *4*(1), 77–86. https://doi.org/ttps://doi.org/10.15548/jostech.v4i1.8361

Ospina, R., Gondim, J. A. M., Leiva, V., & Castro, C. (2023). An Overview of Forecast Analysis with ARIMA Models during the COVID-19 Pandemic : Methodology and Case Study in Brazil. *MDPI*, *2023*, 1–18. https://doi.org/https://doi.org/10.3390/math11143069

Putera, M. L. S. (2020). Peramalan Transaksi Pembayaran Non-Tunai Menggunakan Arimax-Ann Dengan Konfigurasi Kalender. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *14*(1), 135–146. https://doi.org/10.30598/barekengvol14iss1pp135-146

Putra, R. M., & Rani, N. A. (2020). Prediksi Curah Hujan Harian di Stasiun Meteorologi Kemayoran Menggunakan Artificial Neural Network (ANN). *Buletin GAW Bariri*, *1*(2), 101–108. https://doi.org/10.31172/bgb.v1i2.35

Shumway, R. H., & Stoffer, D. S. (2016). *Time Series Analysis and Its Applications With R Examples* (4th ed.). Springer.

Smith, A. B., & Katz, R. W. (2013). US billion-dollar weather and climate disasters: Data sources, trends, accuracy, and biases. *Natural Hazards*, *67*(2), 387–410. https://doi.org/10.1007/s11069-013-0566-5

Suhartanto, E., & Cahya, E. N. (2019). Analisa Limpasan Berdasarkan Curah Hujan

Menggunakan Model Artificial Neural Network (ANN) Di Sub DAS Brantas Hulu. *Jurnal Teknik Pengairan*, *10*(2), 134–144. https://doi.org/10.21776/ub.pengairan.2019.010.02.07

Sumarjaya, I. W., Ayu, I. G., & Srinadi, M. (2021). Penggunaan Model Arimax Untuk Meramalkan Data Curah Hujan Bulanan Di Bali. *E-Jurnal Matematika*, *10*(November), 186–191. https://doi.org/10.24843/MTK.2021.v10.i04.p341

Supriya, K. (2020). Comparative Study of ARIMAX-ANN Hybrid Model with ANN and ARIMAX Models to Forecast the Damage Caused by Yellow Stem Borer (Scirpophaga incertulas) in Telangana State. *International Journal of Current Microbiology and Applied Sciences*, *9*(7), 1490–1497. https://doi.org/10.20546/ijcmas.2020.907.171

Suryani, A. R., Hendikawati, P., Februari, B., Maret, B., Suryani, A. R., Peramalan, H. P., & Hujan, C. (2018). Peramalan Curah Hujan Dengan Metode Autoregressive Integrated Moving Average With Exogenous Input (Arimax). *Unnes Journal of Mathematics*, *7*(1), 120–129.

Wijayanti, K., Martha, S., & Debataraja, N. N. (2021). Perbandingan Model ARIMAX dan Fungsi Transfer pada Peramalan Curah Hujan. *Bimaster : Buletin Ilmiah Matematika, Statistika Dan Terapannya*, *10*(2), 233–242. http://dx.doi.org/10.26418/bbimst.v10i2.45877