

PENGARUH VARIABEL EKSOGEN TERHADAP PEMODELAN CURAH HUJAN DI KOTA BANDUNG

Mahmudi¹, Hilmah Amalia Mehri², Suma Inna³

^{1,2,3} Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia

mahmudi@uinjkt.ac.id

ABSTRACT *This study is motivated by the significant role of rainfall as a natural phenomenon that impacts various sectors. The main objective of this research is to evaluate the influence of exogenous variables on rainfall forecasting accuracy by comparing the SARIMA, SARIMAX, hybrid SARIMA-ANN, and hybrid SARIMAX-ANN models. Initial modeling without involving exogenous variables resulted in a MAPE value of 43.65% for the SARIMA model and 42.86% for the hybrid SARIMA-ANN model. The findings show that after incorporating the exogenous variable of sunshine duration, the accuracy of rainfall forecasting improved, with the SARIMAX model achieving a MAPE value of 28.69% and the hybrid SARIMAX-ANN model achieving a MAPE value of 27.03%. This study contributes to the development of rainfall forecasting methods by incorporating other influencing variables.*

Keywords: hybrid sarimax-ann, rainfall forecasting, exogenous variables, sarimax model

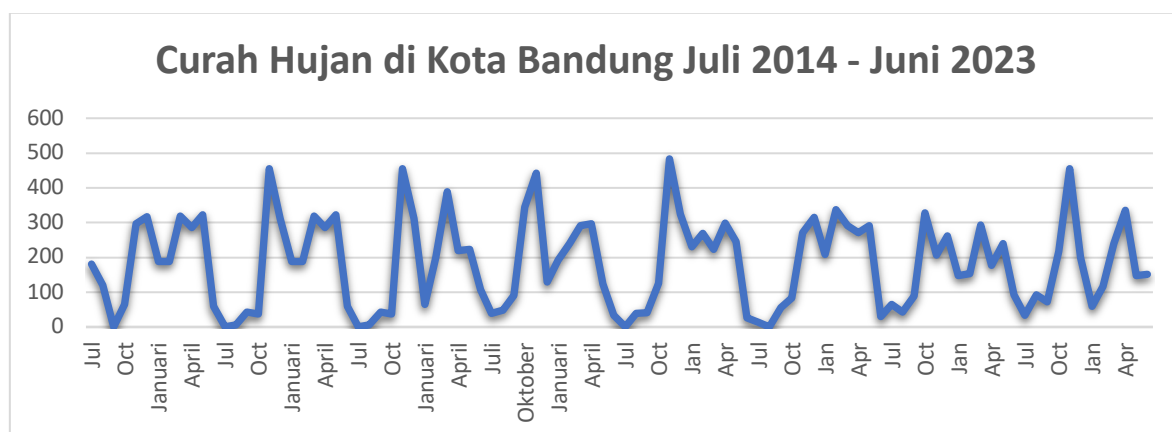
ABSTRAK Penelitian ini dilatarbelakangi oleh peran penting curah hujan sebagai fenomena alam yang memiliki dampak signifikan pada berbagai sektor. Tujuan utama penelitian ini adalah mengevaluasi pengaruh variabel eksogen dalam peramalan curah hujan dengan membandingkan akurasi model SARIMA, SARIMAX, hybrid SARIMA-ANN, dan hybrid SARIMAX-ANN. Pemodelan awal tanpa melibatkan variabel eksogen menghasilkan nilai MAPE sebesar 43,65% untuk model SARIMA dan 42,86% untuk model hybrid SARIMA-ANN. Hasil temuan setelah melibatkan variabel eksogen, yaitu lamanya penyinaran matahari, menunjukkan peningkatan akurasi peramalan curah hujan dengan model SARIMAX yang memiliki nilai MAPE sebesar 28,69% dan model hybrid SARIMAX-ANN dengan nilai MAPE sebesar 27,03%. Penelitian ini memberikan kontribusi dalam pengembangan metode peramalan curah hujan dengan melibatkan variabel lain yang mempengaruhi.

Kata-kata Kunci: hybrid sarimax-ann, peramalan curah hujan, variabel eksogen, model sarimax

PENDAHULUAN

Curah hujan merupakan faktor meteorologi yang sangat penting dalam berbagai aspek kehidupan sehari-hari. Implikasi sosial-ekonomi dari curah hujan sangat signifikan, termasuk kerusakan infrastruktur selama banjir dan gangguan pada jaringan transportasi (Le et al, 2019). Kota Bandung, sebagai salah satu ibu kota Provinsi di Indonesia, memiliki pola curah hujan yang dinamis sepanjang tahun. Memahami pola curah hujan di daerah ini sangat penting untuk perencanaan perkotaan yang efisien, manajemen sumber daya air, dan persiapan darurat.

Berbagai penelitian peramalan curah hujan telah dilakukan dengan berbagai metode. Misalnya, penelitian di Kota Ambon menggunakan model SARIMA menghasilkan nilai MAPE sebesar 76% (Kafara, Rumlawang, & Sinay, 2017). Studi perbandingan antara model SARIMA dan SARIMAX menunjukkan kinerja yang lebih baik pada model SARIMAX (Ibrahim & Musa, 2016). Model SARIMAX juga diterapkan dalam memprediksi jumlah penumpang kapal domestik di Pelabuhan Makassar (Fitriyati, Wijaya, Pagri, & Inayah, 2022). Penelitian yang dilakukan oleh Fiskin et al. (2022) menegaskan bahwa metode hybrid SARIMAX-ANN mengungguli SARIMAX dalam hal akurasi. Konsep pengembangan model hybrid analisis runtun waktu pertama kali diusulkan pada tahun 2003 dengan menggabungkan komponen linier dan nonlinier dalam proses peramalan (Zhang, 2003). Penelitian mengenai peramalan inflasi di Indonesia pada tahun 2022 juga menerapkan metode hybrid ini (Fitriyati, Mahmudi, Wijaya, & Maysun, 2022), menunjukkan bahwa penerapan model hybrid masih relevan dan efektif.



Gambar 1. Plot Data Curah Hujan Kota Bandung Tahun Juli 2014-Juni 2023

Ilustrasi pada gambar 1 di atas menunjukkan bahwa curah hujan di Kota Bandung terjadi secara periodik yang mendorong penelitian ini menggunakan model SARIMA. Penelitian ini dilakukan untuk menyelidiki pengaruh variabel eksogen dalam meramalkan curah hujan di Kota Bandung menggunakan metode SARIMAX dan hybrid SARIMAX-ANN. Metode hybrid SARIMAX-ANN ini merupakan perkembangan dari SARIMAX dan Artificial Neural Network (ANN) yang menggunakan pendekatan

lebih kompleks guna meningkatkan akurasi peramalan. SARIMAX, sebagai pengembangan model ARIMA, merupakan pemodelan yang melibatkan komponen musiman dengan penambahan variabel eksogen. Selain itu, penelitian ini bertujuan untuk mengevaluasi apakah penambahan variabel eksogen dapat meningkatkan akurasi peramalan.

METODE PENELITIAN

Pada penelitian ini data yang digunakan adalah data sekunder, yaitu data curah hujan bulanan (Y), kelembapan (X_1), dan lamanya penyinaran matahari (X_2) Kota Bandung tahun 2014 hingga 2023 yang diperoleh dari Stasiun Geofisika Bandung Kelas I. Data tersebut dibagi menjadi data training dan testing. Data training terdiri dari data bulan Juli 2014 – Juni 2022 dan data testing terdiri dari bulan Juli 2022 – Juni 2023.

Pemodelan SARIMA

Pemodelan SARIMA diterapkan terhadap data curah hujan untuk menyeleksi model yang digunakan pada tahapan berikutnya. Data curah hujan diuji untuk memastikan bahwa data memenuhi asumsi kestasioneran. Uji asumsi ini melibatkan dua langkah, yakni uji kestasioneran terhadap varians dengan menggunakan uji Box-Cox, serta uji kestasioneran terhadap rata-rata melalui uji Augmented Dickey Fuller (ADF).

Hipotesis uji Box-Cox (Mistawati, Yasnani, & Lestari, 2021):

H_0 : Data tidak stasioner terhadap rata-rata.

H_1 : Data stasioner terhadap rata-rata.

Statistik uji : Apabila nilai $\lambda < 1$, maka data tidak stasioner terhadap varians. Data yang tidak stasioner terhadap varians akan dilakukan transformasi Box-Cox hingga stasioner dengan persamaan umum (Wei, 2006):

$$T(X_t) = \begin{cases} \log x, & \lambda = 0 \\ \frac{x_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \end{cases} \quad \dots (1)$$

Hipotesis uji Augmented Dickey Fuller (ADF) (Saz, 2011):

H_0 : Data tidak stasioner terhadap rata-rata.

H_1 : Data stasioner terhadap rata-rata.

Statistik uji : $\Delta y_t = \beta_0 + (\alpha - 1)y_{t-1} + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + \varepsilon_i$... (2)

Keputusan : Apabila nilai p-value kurang dari tingkat signifikansi $\alpha = 0,05$ maka tolak H_0 , yang artinya data stasioner terhadap rata-rata.

Data yang tidak stasioner terhadap rata-rata dilakukan differencing hingga data stasioner.

Setelah memastikan bahwa data memenuhi asumsi kestasioneran, tahapan berikutnya adalah mengidentifikasi model SARIMA (p, d, q)(P, D, Q)s yang mungkin berdasarkan plot Autocorrelation (ACF) dan Partial Autocorrelation (PACF). Persamaan model SARIMA (p, d, q)(P, D, Q)s yang digunakan untuk memprediksi curah hujan dinyatakan sebagai berikut (Geurts, 1977):

$$\varphi_p(B)\phi_p(B^S)(1-B)^d(1-B^S)^D Z_t = \theta_q(B)\theta_q(B^S)\alpha_t \quad \dots (3)$$

Kemudian, dilakukan estimasi dan uji signifikansi parameter model untuk menilai keberartian masing-masing parameter. Selain itu, dilakukan uji asumsi distribusi normal dengan uji Shapiro-Wilk dan uji Ljung-Box terhadap residu model. Uji residu model dilakukan untuk menilai apakah model yang diidentifikasi layak atau tidak. Model dikatakan layak apabila residu model berdistribusi normal dan tidak terdapat autokorelasi.

Hipotesis uji signifikansi parameter (Rahayu, Marwati, & Rachmatin, 2022):

H_0 : Parameter model tidak signifikan.

H_1 : Parameter model signifikan.

Statistik uji : $t_{hitung} = \frac{estimator}{SE(estimator)} \quad \dots (4)$

Keputusan : Apabila $t_{hitung} > t_{tabel}$ atau nilai p-value kurang dari tingkat signifikansi $\alpha = 0,05$ maka tolak H_0 , yang artinya parameter model signifikan.

Hipotesis Uji Shapiro-Wilk (Shapiro & Wilk, 1965):

H_0 : Data residu berdistribusi normal

H_1 : Data residu tidak berdistribusi normal

Statistik uji : $W = \frac{(\sum_{i=1}^n \alpha_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \dots (5)$

Keputusan : Apabila nilai p-value kurang dari tingkat signifikansi $\alpha = 0,05$ maka tolak H_0 , yang artinya data tidak berdistribusi normal.

Hipotesis Uji Ljung-Box (Msofe & Mbago, 2019):

H_0 : Data residu tidak terdapat autokorelasi

H_1 : Data residu terdapat autokorelasi

Statistik uji : $Q = n(n+2) \sum_{i=1}^m \frac{r_i^2}{n-1} \sim \chi_{m-k}^2 \quad \dots (6)$

Keputusan : Apabila nilai p-value kurang dari tingkat signifikansi $\alpha = 0,05$ maka tolak H_0 , yang artinya residu terdapat autokorelasi.

Apabila terdapat beberapa model yang signifikan dan memenuhi diagnostik uji, model dipilih berdasarkan nilai Akaike Info Criterion (AIC) terkecil. Rumus nilai AIC (Msofe & Mbago, 2019):

$$AIC(k) = T \log(\hat{\sigma}^2(k)) + 2k \quad \dots (7)$$

Pemodelan SARIMAX

Setelah model SARIMA dipilih, langkah berikutnya adalah menentukan variabel eksogen mana yang layak untuk selanjutnya dimasukkan ke dalam model SARIMAX. Model Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) adalah model prediksi analisis runtun waktu yang menggabungkan model Seasonal Autoregressive Integrated Moving Average (SARIMA) dengan variabel eksogen, juga dikenal sebagai regresor (Arunraj, Ahrens, & Fernandes, 2016). Variabel eksogen dapat dimodelkan dengan persamaan regresi linier berganda yang dinyatakan dalam persamaan berikut:

$$y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \dots + \beta_k X_{k,t} + \omega_t \quad \dots (8)$$

dimana β_0 adalah parameter konstan dan $\beta_1, \beta_2, \dots, \beta_k$ adalah parameter koefisien variabel eksogen ke- k , $X_{1,t}, X_{2,t}, \dots, X_{k,t}$ adalah pengamatan dari variabel eksogen ke- k yang sesuai dengan variabel dependen y_t ; ω_t adalah residu.

$$\omega_t = \frac{\theta_q(B)\theta_Q(B^S)}{\varphi_p(B)\phi_P(B^S)(1-B)^d(1-B^S)^D} \varepsilon_t \quad \dots (9)$$

Secara umum persamaan model SARIMAX dapat diperoleh dengan mensubstitusi Persamaan (2) ke dalam Persamaan (3) (Cools, Moons, & Wets, 2009):

$$y_t = \beta_0 + \sum_{i=1}^k \beta_i X_{i,t} + \frac{\theta_q(B)\theta_Q(B^S)}{\varphi_p(B)\phi_P(B^S)(1-B)^d(1-B^S)^D} \varepsilon_t \quad \dots (10)$$

Pemodelan ANN

Artificial Neural Network adalah algoritma komputasi yang dapat memecahkan masalah kompleks yang terdiri dari neuron atau node buatan yang merupakan informasi unit pemrosesan dengan disusun berlapis-lapis dan saling berhubungan oleh *synaptic weights* (Nontapa, Kesamoon, Kaewhawong, & Intrapai boon, 2021). *Neural Network* memiliki tiga layer dalam pemrosesan informasi, yaitu *input layer*, *hidden layer*, dan *output layer*. *Input layer* adalah node-node yang membawa data masuk untuk diproses pada layer selanjutnya, *hidden layer* adalah node-node dengan penambahan bobot untuk pemrosesan informasi, serta *output layer* adalah node-node untuk informasi yang telah diproses. Pada penelitian ini model ANN digunakan untuk memodelkan residu dari model SARIMA dan SARIMAX yang telah didapat dengan input residu waktu ke- $t - 1$ dan output residu waktu ke- t .

Pemodelan Hybrid

Model SARIMA dan SARIMAX keduanya merupakan model linier, sehingga tidak dapat menangani komponen data non linier yang terdapat di dalamnya. Disisi lain, model ANN mampu untuk memodelkan komponen data non linier. Ide dari pemodelan hybrid ini adalah untuk memodelkan komponen non linier yaitu residu dari model SARIMA(p,d,q)(P,D,Q)s dan SARIMAX(p,d,q)(P,D,Q)s dengan ANN dan menggabungkannya pada komponen linear (Zhang, 2003).

Hasil dari model linier SARIMA atau SARIMAX dan model residu ANN yang didapat digabungkan dengan metode hybrid SARIMA-ANN atau hybrid SARIMAX-ANN dengan persamaan berikut:

$$y_t = L_t + N_t \quad \dots (11)$$

dimana L_t adalah komponen linier dan N_t menunjukkan komponen non linier. Residu pada waktu ke- t dari model linear dapat dituliskan sebagai berikut:

$$\varepsilon_t = y_t - \hat{L}_t \quad \dots (12)$$

Residu ε_t dimodelkan dengan ANN hingga diperoleh persamaan:

$$\hat{N}_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-n}) + e_t \quad \dots (13)$$

Sehingga kombinasi peramalannya menjadi:

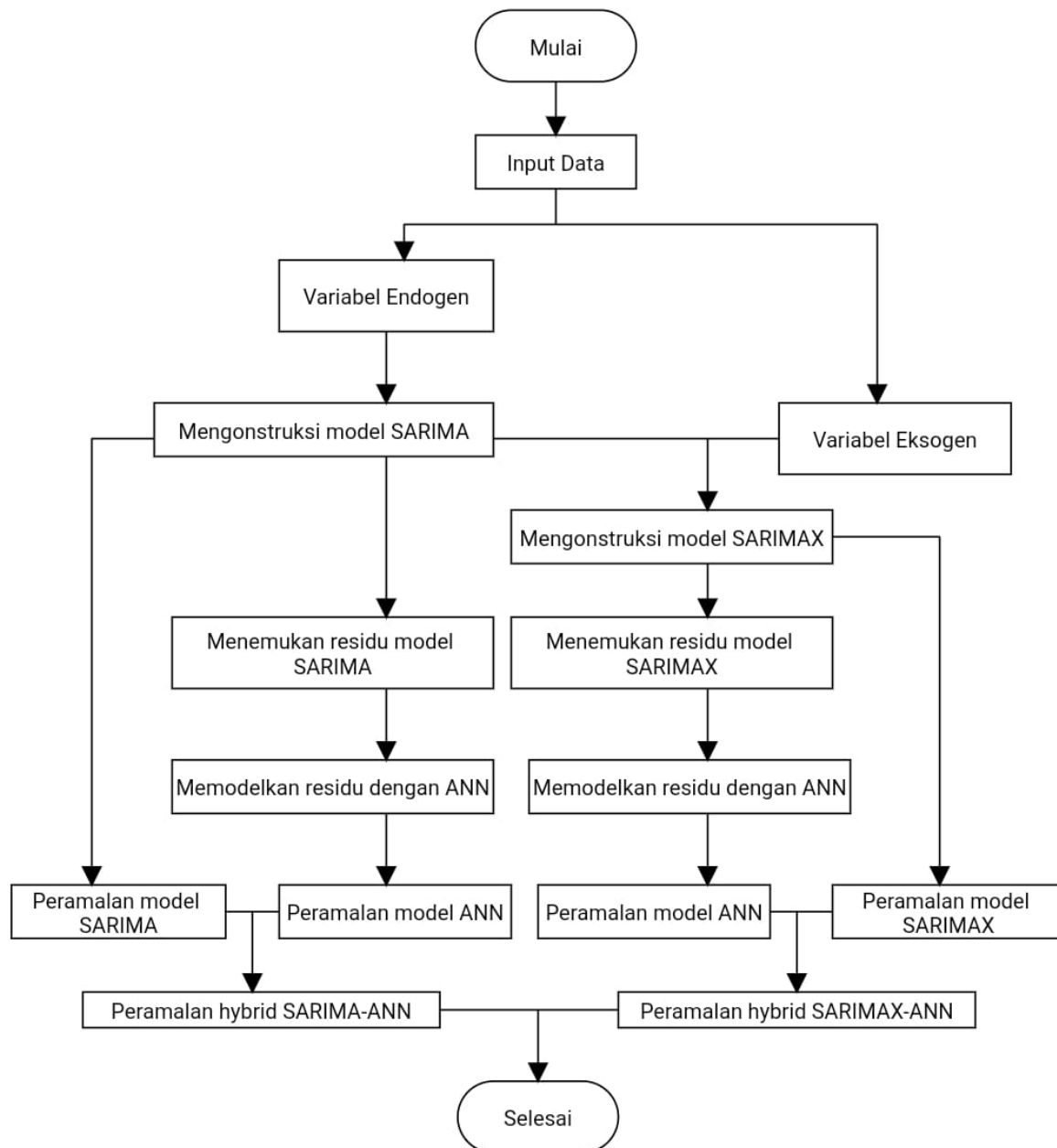
$$\hat{y}_t = \hat{L}_t + \hat{N}_t \quad \dots (14)$$

Secara umum pemodelan hybrid dapat dilihat pada diagram alir pada gambar 2.

Evaluasi Model

Untuk menilai apakah variabel eksogen yang diperoleh memiliki pengaruh terhadap peramalan dilakukan evaluasi model dengan MAPE pada model SARIMA, SARIMAX, hybrid SARIMA-ANN, dan hybrid SARIMAX-ANN. Mean Absolute Percentage Error (MAPE) adalah ukuran ketepatan prediksi dari metode prediksi dalam statistik (Nontapa et al., 2021). Tingkat akurasi meningkat seiring dengan penurunan hasil perhitungan MAPE (Hanuwati, Mukid, & Prahutama, 2016). Rumus untuk menghitung MAPE, dengan t adalah data waktu ke- t :

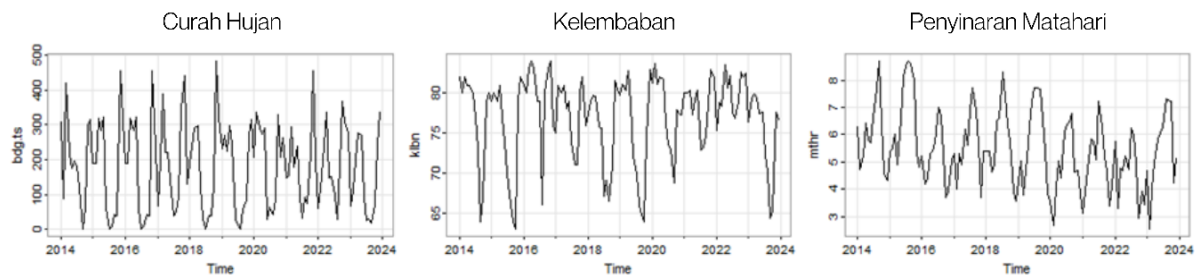
$$MAPE = \frac{1}{n} \sum \frac{|Z_t - F_t|}{Z_t} \times 100\% \quad \dots (15)$$



Gambar 2. Diagram Alir Pemodelan Hybrid

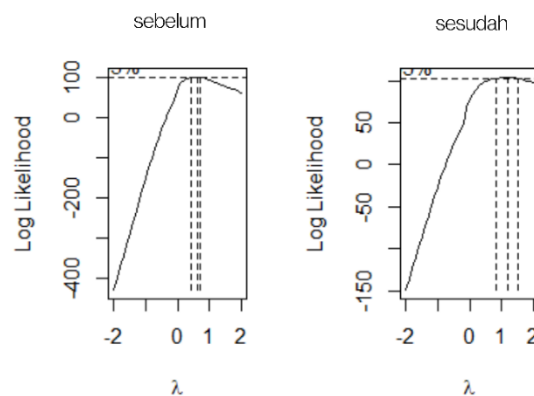
HASIL PENELITIAN DAN PEMBAHASAN

Data yang digunakan pada penelitian ini disajikan pada gambar di bawah. Berdasarkan Gambar 3 dapat terlihat bahwa data curah hujan, kelembapan, dan lama penyinaran memiliki pola musiman.



Gambar 3. Plot Data Curah Hujan, Kelembaban, Lamanya Penyinaran Matahari di Kota Bandung

Masing-masing data dibagi menjadi data training dan data testing. Pada pemodelan pertama yaitu SARIMA, data training curah hujan dilakukan uji stasioneritas terhadap varians hujan dan didapati $\lambda < 1$, sehingga data curah hujan perlu ditransformasi. Berikut plot Box-Cox sebelum dan sesudah data ditransformasi:



Gambar 4. Plot BoxCox Curah Hujan

Setelah data training curah hujan ditransformasi Box-Cox, data menjadi stasioner terhadap varians seperti yang terdapat pada Gambar 4. Uji stasioneritas selanjutnya yaitu uji stasioner data terhadap rata-rata pada data hasil transformasi dengan uji ADF.

Tabel 1. Hasil Perhitungan Uji ADF Data Transformasi Curah Hujan

Dickey-Fuller	P-value
-1,4672	0,7222

Berdasarkan Tabel 1, nilai p-value $> 0,05$ maka data tidak stasioner terhadap rata-rata sehingga dilakukan *differencing* data terhadap non-musiman 1 kali dan dilanjutkan differencing terhadap musiman 1 kali. Differencing yang dilakukan dapat didefinisikan pada persamaan berikut:

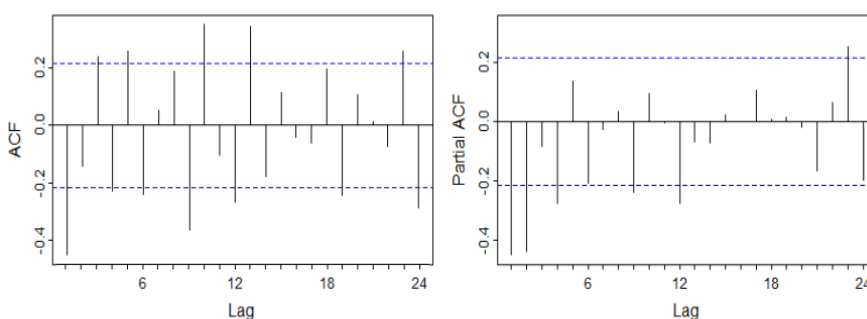
$$\begin{cases} (1 - Z^{12})X_t = X_t - X_{t-12}, & \text{untuk differencing musiman} \\ (1 - Z)X_t = X_t - X_{t-1}, & \text{untuk differencing non musiman} \end{cases} \dots (15)$$

dengan Z adalah operator backshift sedemikian sehingga $Z^j X_t = X_{t-j}, j = 0, 1, 2, \dots$

Tabel 2. Hasil Perhitungan Uji ADF Data Differencing

Dickey-Fuller	P-value
-5,1993	0,01

Setelah differencing, hasil uji ADF pada data menunjukkan p-value < 0,05 yang artinya data stasioner terhadap rata-rata. Data yang telah stasioner terhadap varians dan rata-rata dilakukan identifikasi model SARIMA(p,d,q)(P,D,Q)s melalui plot Autocorrelation (ACF) dan Partial Autocorrelation (PACF).



Gambar 5. Plot ACF dan PACF Curah Hujan

Untuk menentukan orde p, dapat merujuk pada plot PACF. Plot PACF terjadi cut-off pada lag ke-4, menunjukkan kemungkinan orde p adalah AR(0), AR(1), AR(2), AR(3), dan AR(4). Sementara itu, orde P pada komponen musiman dapat ditentukan dengan melihat plot PACF pada kelipatan 12 lag, terdapat 1 lag yang signifikan, maka orde P yang mungkin adalah SAR(0) atau SAR(1).

Penentuan orde q dapat dilakukan dengan memeriksa plot ACF. Plot ACF terjadi cut-off pada lag ke-4, sehingga orde q yang mungkin adalah MA(0), MA(1), MA(2), MA(3), dan MA(4). Selanjutnya, untuk orde Q pada komponen musiman, dilakukan dengan memeriksa plot ACF pada lag musiman yang signifikan maka orde Q mungkin adalah SMA(0), SMA(1), atau SMA(2). Pada tahap sebelumnya, data telah di differencing terhadap musiman dan non-musiman sehingga untuk orde d dan D adalah 1. Berdasarkan Prinsip Parsimoni yang menyatakan model yang lebih sederhana lebih disenangi daripada model dengan parameter yang banyak (Sitrorus, Wayuningsih, & Hayati, 2017), sehingga pada penelitian ini dibatasi hingga orde ke-2, model SARIMA yang diteliti adalah model dengan orde p = 0,1, atau 2; orde q = 0,1, atau 2; orde P = 0 atau 1; orde Q = 0,1, atau 2. Sebagai contoh model SARIMA(2,1,2)(1,1,2)12, SARIMA(2,1,2)(1,1,1)12, SARIMA(2,1,2)(1,1,0)12, SARIMA(2,1,0)(1,1,1)12, dan yang lainnya.

Model-model tersebut akan diestimasi dan diuji signifikansi parameternya. Selain itu, dilakukan uji diagnostik residu berdistribusi normal, dan uji autokorelasi terhadap residu model.

Tabel 3. Hasil Uji Model SARIMA(p,d,q)(P,D,Q)s

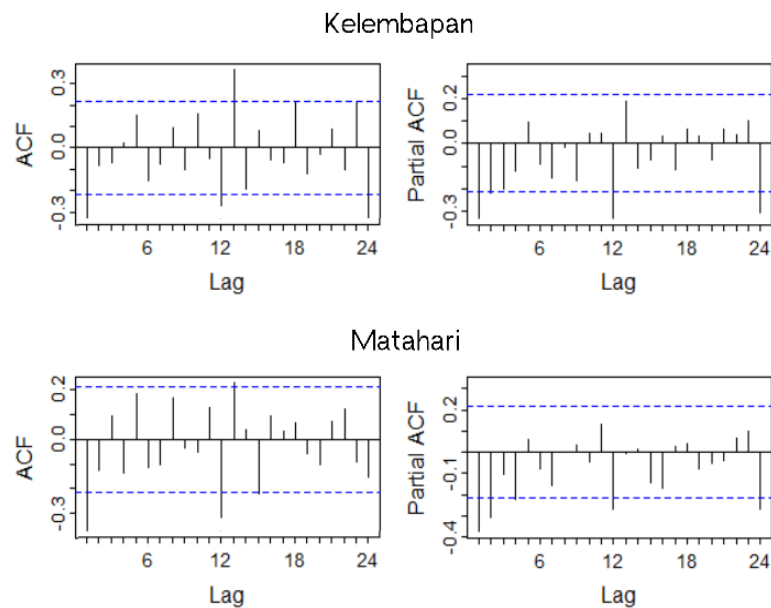
Model	Parameter	Uji Signifikansi	Uji Shapiro-Wilk	Uji Ljung-Box
		P-value		
SARIMA(2,1,0)(1,1,0)12	ar_1	0,0000	0,0979	0,7709
	ar_2	0,0000		
	sar_1	0,0256		
SARIMA(2,1,0)(0,1,1)12	ar_1	0,0000	0,0005	0,7353
	ar_2	0,0002		
	sma_1	0,0000		
SARIMA(2,1,0)(0,1,0)12	ar_1	0,0000	0,0024	0,07431
	ar_2	0,0000		
SARIMA(1,1,0)(1,1,0)12	ar_1	0,0000	0,0213	0,0651
	sar_1	0,0054		

Tabel 3 menyajikan model dengan keseluruhan parameter signifikan, sedangkan model lain yang terdapat parameter tidak signifikan tidak ditampilkan. Berdasarkan Tabel 3, model yang memenuhi semua uji adalah model SARIMA (2,1,0)(1,1,0)12. Model tersebut memiliki parameter yang signifikan karena nilai p-value dari uji signifikansi < 0,05. Selain itu, residu model tersebut terdistribusi normal serta tidak terdapat autokorelasi karena nilai p-value dari masing-masing uji > 0,05. Model ini nantinya akan ditambahkan variabel eksogen pada model SARIMAX. Pada tahapan metode hybrid SARIMA-ANN, residu model SARIMA(2,1,0)(1,1,0)12 juga digunakan sebagai input dari pemodelan ANN.

Pemodelan SARIMAX

Pemodelan SARIMAX memerlukan variabel eksogen yang terdapat pola musiman dan layak untuk dimasukkan ke dalam pemodelannya. Hal tersebut perlu dilakukan melalui pengecekan terhadap model variabel eksogen yang digunakan. Dengan tahapan yang sama seperti data curah hujan, variabel-variabel eksogen ini diuji kestasionerannya. Setelah differencing non-musiman dan musiman masing-masing 1 kali, data kelembapan dan penyinaran matahari telah stasioner terhadap rataan maupun varians.

Pengidentifikasi variabel eksogen layak atau tidak untuk pemodelan SARIMAX terlebih dahulu dilakukan dengan cara memodelkan masing-masing variabel eksogen menggunakan model SARIMA. Plot ACF dan PACF diperlukan untuk mengidentifikasi model tersebut.



Gambar 6. Plot ACF dan PACF Kelembapan & Lamanya Penyinaran Matahari

Berdasarkan Gambar 6, identifikasi model SARIMA pada variabel kelembapan dan lamanya sinar matahari dilakukan seperti pada data curah hujan diperoleh sebagai berikut:

Tabel 4. Identifikasi Variabel Eksogen

Variabel Eksogen	Parameter	Orde
Kelembapan	p	0,1, atau 2
	d	1
	q	0 atau 1
	P	0, 1, atau 2
	D	1
	Q	0 atau 1
Penyinaran Matahari	p	0,1, atau 2
	d	1
	q	0 atau 1
	P	0, 1, atau 2
	D	1
	Q	0 atau 1

Sebagai contoh pada variabel kelembapan, model yang mungkin adalah model SARIMA (2,1,1) (2,1,1)₁₂, SARIMA (2,1,1)(1,1,1)₁₂, SARIMA (2,1,1)(1,1,0)₁₂, SARIMA(2,1,0)(1,1,1)₁₂, dan yang lainnya. Begitu juga pada variabel penyinaran matahari. Kemudian dilakukan uji signifikansi parameter, uji diagnostik residu berdistribusi normal, dan uji autokorelasi terhadap residu model. Pada pemodelan SARIMA dengan data variabel kelembapan tidak diperoleh model yang memenuhi uji

signifikansi parameter dan diagnostik model secara bersamaan, sehingga variabel kelembapan tidak bisa dimasukkan ke dalam pemodelan SARIMAX. Sedangkan pada variabel eksogen penyinaran matahari dengan hasil seluruh parameternya signifikan serta residu model berdistribusi normal dan tidak terdapat autokorelasi disajikan pada Tabel 5.

Tabel 5. Hasil Uji Model Variabel Eksogen Penyinaran Matahari

Model	Parameter	Uji Signifikansi	Uji Shapiro-Wilk	Uji Ljung-Box	AIC
		P-value			
SARIMA(2,1,0) (2,1,0) ¹²	ar_1	0,0000	0,4081	0,7002	-19,0906
	ar_2	0,0196			
	sar_1	0,0002			
	sma_1	0,0003			
SARIMA(2,1,0) (1,1,0) ¹²	ar_1	0,0001	0,3984	0,507	-10,711
	ar_2	0,0046			
	sar_1	0,0056			
SARIMA(1,1,1) (0,1,1) ¹²	ar_1	0,0014	0,3382	0,9866	-42,6064
	ma_1	0,0000			
	sma_1	0,0000			

Dari ketiga model yang memenuhi asumsi dipilih model terbaik untuk variabel penyinaran matahari dengan nilai AIC terkecil yaitu SARIMA(1,1,1)(0,1,1)¹².

Pada tahapan pemodelan SARIMAX variabel eksogen yang digunakan hanya variabel yang memenuhi yaitu variabel penyinaran matahari. Setelah proses penambahan variabel eksogen penyinaran matahari pada model curah hujan SARIMA(2,1,0)(1,1,0)¹², selanjutnya dilakukan uji asumsi terhadap model.

Tabel 6. Hasil Uji Model SARIMAX (2,1,0)(1,1,0)¹²

Model	Parameter	Uji Signifikansi	Uji Shapiro-Wilk	Uji Ljung-Box
		P-value		
SARIMAX(2,1,0)(1,1,0) ¹² dengan variabel eksogen penyinaran matahari.	ar_1	0,0000	0,2649	0,5434
	ar_2	0,0000		
	sar_1	0,0000		
	xreg	0,0000		

Berdasarkan Tabel 6, dapat disimpulkan bahwa model SARIMAX(2,1,0)(1,1,0)¹² dengan variabel eksogen penyinaran matahari dapat digunakan untuk melakukan peramalan curah hujan di Kota Bandung karena parameternya signifikan, residu berdistribusi normal, dan tidak terdapat autokorelasi pada residu model.

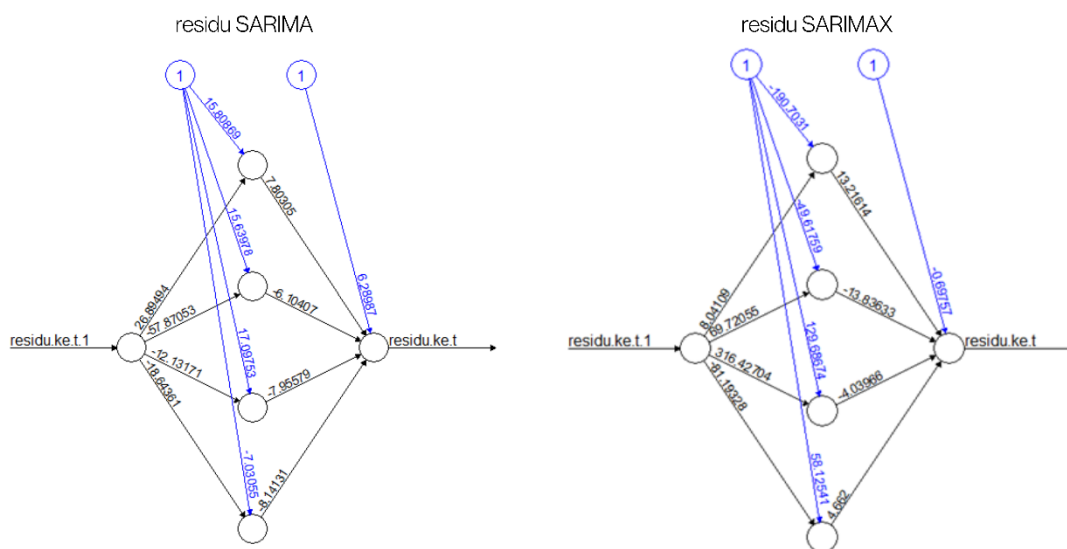
Model SARIMAX(2,1,0)(1,1,0)₁₂ dengan variabel eksogen matahari digunakan untuk melakukan peramalan terhadap curah hujan di Kota Bandung selama 12 bulan mendatang. Pada tahapan metode hybrid SARIMAX-ANN, residu model SARIMAX(2,1,0)(1,1,0)₁₂ juga digunakan sebagai input pemodelan ANN.

Pemodelan Hybrid SARIMA-ANN dan SARIMAX-ANN

Setelah didapatkan model SARIMA atau SARIMAX untuk data curah hujan di Kota Bandung, maka didapatkan residu dari model tersebut. Hasil ramalan dari SARIMA atau SARIMAX dijadikan komponen linier. Adapun residu dari kedua model tersebut dimodelkan dengan model ANN yang hasil peramalannya dijadikan komponen non linier.

Pada pemodelan ANN, residu dari model SARIMA atau SARIMAX waktu ke- t dijadikan sebagai nilai output, sedangkan residu model SARIMA atau SARIMAX waktu ke- $t - 1$ dijadikan sebagai input. Dalam metode ANN, jumlah hidden layer dan neuron dapat ditentukan. Dalam hal ini, peneliti memilih untuk menggunakan 1 hidden layer dan 4 neuron (Rahayu et al., 2022). Belum tentu dengan jumlah neuron yang lebih banyak menghasilkan akurasi yang lebih baik (Supriyanto, Sunardi, & Riadi, 2022).

Berikut arsitektur ANN (1,4,1) yang terbentuk:



Gambar 7. Struktur Jaringan ANN (1,4,1) pada Model SARIMA atau SARIMAX

Melalui software R yang sudah tersedia library untuk melakukan backpropagation, didapatkan struktur jaringan ANN seperti pada Gambar 7. Persamaan model ANN (1,4,1) residu model SARIMA sebagai berikut:

$$N_t = 6,289 - 7,803 f^h(X_{1,1}) - 6,104 f^h(X_{1,2}) - 7,9557 f^h(X_{1,3}) - 8,1413 f^h(X_{1,4}) \dots (16)$$

dengan:

$$f^h = \frac{1}{1+e^{-X_{j,t}}} \dots (17)$$

dimana:

$$j = 1$$

$$X_{1,1} = 26,894 + 15,808$$

$$X_{1,2} = -57,87 + 15,639$$

$$X_{1,3} = -12,131 + 17,097$$

$$X_{1,4} = -18,643 - 7,03$$

Untuk ANN dari residu model SARIMAX:

$$N_t = -0,697 + 13,216 f^h(X_{1,1}) - 13,836 f^h(X_{1,2}) - 4,039 f^h(X_{1,3}) + 4,662 f^h(X_{1,4}) \dots (18)$$

dengan:

$$f^h = \frac{1}{1+e^{-X_{j,t}}} \dots (19)$$

dimana:

$$j = 1$$

$$X_{1,1} = 8,041 - 190,703$$

$$X_{1,2} = -69,72 - 49,617$$

$$X_{1,3} = 316,42 + 129,68$$

$$X_{1,4} = -81,193 + 58,125$$

Setelah mendapatkan model ANN (1,4,1), peramalan dengan model hybrid diperoleh dengan menjumlahkan hasil peramalan model linier SARIMA (2,1,0)(1,1,0)₁₂ atau SARIMAX(2,1,0)(1,1,0)₁₂ dengan model nonlinier ANN-nya.

Keempat model yang diperoleh (SARIMA, SARIMAX, hybrid SARIMA-ANN, hybrid SARIMAX-ANN) digunakan untuk peramalan terhadap curah hujan di Kota Bandung pada data testing mulai dari Juli 2022 hingga Juni 2023 yang hasilnya tersaji pada Tabel 7. Hasil peramalan dengan keempat model dibandingkan dengan data asli menghasilkan nilai yang beragam.

Tabel 7. Hasil Peramalan Curah Hujan Kota Bandung

Tahun	Bulan	Data Asli	Hasil Peramalan			
			SARIMA	Hybrid SARIMA-ANN	SARIMAX	Hybrid SARIMAX-ANN
2022	Juli	82,2	99,95	114,05	105,76	100,41
	Agustus	29,9	71,76	55,85	58,36	62,33
	September	18,2	119,22	103,30	139,57	143,54

Tahun	Bulan	Data Asli	Hasil Peramalan			
			SARIMA	Hybrid SARIMA-ANN	SARIMAX	Hybrid SARIMAX-ANN
2023	Oktober	366,7	289,27	303,36	358,68	353,32
	November	307,2	411,47	425,56	345,92	327,35
	Desember	277,7	250,86	234,94	262,65	266,62
	Januari	68,3	115,61	129,70	109,67	91,10
	Februari	111,1	154,98	139,06	180,22	184,19
	Maret	199,8	286,72	270,81	279,80	283,76
	April	275,5	329,10	313,19	277,10	281,07
	Mei	268,5	201,69	185,78	180,43	184,40
	Juni	89,8	167,50	181,59	91,54	86,19

Selanjutnya akan dihitung nilai MAPE dari masing-masing model untuk mengukur akurasi hasil ramalan dengan data asli dan melihat apakah penambahan variabel eksogen lamanya penyinaran matahari memiliki pengaruh terhadap akurasi model.

Tabel 8. Nilai MAPE

Model	MAPE
SARIMA(2,1,0)(1,1,0) ₁₂	43,65%
Hybrid SARIMA-ANN	42,86%
SARIMAX(2,1,0)(1,1,0) ₁₂	28,69%
Hybrid SARIMAX-ANN	27,03%

Hasil MAPE dari Tabel 8 menunjukkan bahwa model dengan penambahan variabel eksogen, baik pada model SARIMA maupun hybrid SARIMA-ANN memiliki pengaruh terhadap akurasi model. Hal ini terbukti dari nilai MAPE SARIMAX(2,1,0)(1,1,0)₁₂ dan hybrid SARIMAX-ANN lebih rendah dibandingkan SARIMA(2,1,0)(1,1,0)₁₂ dan hybrid SARIMA-ANN. Model yang memiliki akurasi tertinggi adalah model hybrid SARIMAX-ANN dengan nilai MAPE 27,03%. Jika ditinjau dari penambahan pemodelan residu dengan ANN terhadap peningkatan akurasi model, maka diperoleh kenaikan akurasi walaupun kurang dari 2%. Hal ini terlihat dari nilai MAPE hybrid SARIMA-ANN lebih kecil 0,79% bila dibandingkan dengan nilai MAPE SARIMA(2,1,0)(1,1,0)₁₂, begitu juga nilai MAPE hybrid SARIMAX-ANN lebih kecil 1,66% bila dibandingkan dengan nilai MAPE SARIMAX(2,1,0)(1,1,0)₁₂.

KESIMPULAN DAN SARAN

Hasil analisis curah hujan di Kota Bandung menunjukkan bahwa lamanya penyinaran matahari memiliki pengaruh signifikan terhadap peramalan curah hujan. Dalam perbandingan nilai MAPE, dapat disimpulkan bahwa baik model SARIMAX maupun hybrid SARIMAX-ANN memberikan nilai MAPE yang lebih baik dibandingkan dengan model SARIMA dan hybrid SARIMA-ANN. Oleh karena itu, dapat disimpulkan bahwa faktor lamanya penyinaran matahari secara positif berkontribusi pada peningkatan akurasi peramalan curah hujan, terutama pada model SARIMAX dan hybrid SARIMAX-ANN. Penelitian ini tidak dimodelkan dengan melibatkan kedua variabel eksogen secara bersamaan. Hal ini memungkinkan untuk penelitian yang mengembangkan pemodelan dengan melibatkan dua atau lebih variabel eksogen secara bersamaan.

DAFTAR PUSTAKA

- Arunraj, N. S., Ahrens, D., & Fernandes, M. (2016). Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry. *International Journal of Operations Research and Information Systems*, 7(2), 1–21. <https://doi.org/10.4018/ijoris.2016040101>
- Cools, M., Moons, E., & Wets, G. (2009). Investigating the variability in daily traffic counts through use of ARIMAX and SARIMAX models. *Transportation Research Record*, (2136), 57–66. <https://doi.org/10.3141/2136-07>
- Fiskin, C. S., Turgut, O., Westgaard, S., & Cerit, A. G. (2022). Time series forecasting of domestic shipping market: Comparison of SARIMAX, ANN-based models and SARIMAX-ANN hybrid model. *International Journal of Shipping and Transport Logistics*, 14(3), 193–221. <https://doi.org/10.1504/IJSTL.2022.122409>
- Fitriyati, N., Wijaya, M., Pagri, M. I., & Inayah, N. (2022). Forecasting domestic ship passengers in the Makassar Port using feed-forward neural network and SARIMAX. *7th International Conference Onn Mathematics: Pure, Applied and Computuon:Mathematics of Quantumm Computing*.
- Fitriyati, Nina, Mahmudi, M., Wijaya, M. Y., & Maysun, M. (2022). Forecasting Indonesian inflation using a hybrid ARIMA-ANFIS. *Desimal: Jurnal Matematika*, 5(3), 289–304. <https://doi.org/10.24042/djm.v5i3.14093>
- Geurts, M. (1977). American Marketing Association. *Journal of Marketing Research*, 14(2), 269. <https://doi.org/https://doi:10.2307/3150485>
- Hanurowati, N., Mukid, M. A., & Prahutama, A. (2016). Pemodelan dan Peramalan Indeks Harga Saham Gabungan (IHSG), Jakarta Islamic Index (JII), dan Harga Minyak Dunia Brent Crude Oil Menggunakan Metode Vector Autoregressive Exogenous (VARX). *Jurnal Gaussian*, 5(4), 683–693. <https://doi.org/10.14710/j.gauss.5.4.683-693>

- Ibrahim, A., & Musa, A. O. (2016). On The Performance of Sarima and Sarimax Model in Forecasting Monthly Average Rainfall in Kogi State, Nigeria. *FUDMA: Journal of Sciences (FJS)*, 7(6), 24–31. <https://doi.org/10.33003/fjs-2023-0706-2095>
- Kafara, Z., Rumlwang, F. Y., & Sinay, L. J. (2017). Peramalan Curah Hujan Dengan Pendekatan Seasonal Autoregressive Integrated Moving Average (SARIMA) Studi Kasus: Curah Hujan Bulanan di Kota Ambon, Provinsi Maluku. *Jurnal Ilmu Matematika dan Terapan*, 11(1), 63–74. <https://doi.org/10.30598/barekengvol11iss1pp63-74>
- Le, T.-T., Pham, B. T., Ly, H.-B., Shirzadi, A., & Le, L. M. (2019). Development of 48-hour Precipitation Forecasting Model using Nonlinear Autoregressive Neural Network. *Innovation of Sustainable Infrastructure*, 54. https://doi.org/10.1007/978-981-15-0802-8_191
- Mistawati, M., Yasnani, Y., & Lestari, H. (2021). Forecasting prevalence of dengue hemorrhagic fever using ARIMA model in Sulawesi Tenggara Province, Indonesia. *Public Health of Indonesia*, 7(2), 75–86. <https://doi.org/10.36685/phi.v7i2.411>
- Msofe, Z. A., & Mbago, M. C. (2019). Forecasting international tourist arrivals in zanzibar using box – jenkins SARIMA model. *General Letters in Mathematics*, 7(2), 100–107. <https://doi.org/10.31559/glm2019.7.2.6>
- Nontapa, C., Kesamoon, C., Kaewhawong, N., & Intrapai boon, P. (2021). A New Hybrid Forecasting Using Decomposition Method with SARIMAX Model and Artificial Neural Network. *International Journal of Mathematics and Computer Science*, 16(4), 1341–1354.
- Rahayu, I., Marwati, R., & Rachmatin, D. (2022). Peramalan Jumlah Penderita DBD di Provinsi Jawa Barat dengan Metode Hybrid Sarimax-Ann. *JMT: Jurnal Matematika Dan Terapan*, 4(2), 9–19. <https://doi.org/10.21009/jmt.4.2.2>
- Saz, G. (2011). The efficacy of SARIMA models for forecasting inflation rates in developing countries: The case for Turkey. *International Research Journal of Finance and Economics*, 62(62), 111–142.
- Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4), 591. <https://doi.org/10.2307/2333709>
- Sitrorus, V. B., Wayuningsih, S., & Hayati, M. N. (2017). Peramalan dengan Metode Seasonal Autoregressive Integrated Moving Average (SARIMA) di Bidang Ekonomi (Studi Kasus: Inflasi Indonesia). *Jurnal EKPONENSIAL*, 8(1), 126–132.
- Supriyanto, S., Sunardi, S., & Riadi, I. (2022). Pengaruh Nilai Hidden layer dan Learning rate Terhadap Kecepatan Pelatihan Jaringan Saraf Tiruan Backpropagation. *JIKO (Jurnal Informatika Dan Komputer)*, 6(1), 27. <https://doi.org/10.26798/jiko.v6i1.508>

Wei, W. W. S. (2006). Time Series Analysis Univariate and Multivariate Methods (2nd Edition) Addison Wesley (2005).pdf. New Introduction to Multiple Time Series Analysis, pp. 1–764.

Zhang, G. P. (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing*, 50, 159–175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)