

IMPLEMENTASI METODE SARIMAX UNTUK PREDIKSI CURAH HUJAN JANGKA PENDEK DI PAGERAGEUNG, TASIKMALAYA

IMPLEMENTATION SHORT-TERM RAINFALL PREDICTION IN PAGERAGEUNG, TASIKMALAYA

Ari Azhar Maulana^{1)*}, Harnita Rosalina¹⁾

¹⁾ Balai Besar Wilayah Sungai Citanduy, Kementerian PUPR

Jl. Ir. H. Sutami No. 1 Kota Banjar, Indonesia

*Corresponding email: ti3.ari170197@gmail.com

Diterima: 18 September 2023; Direvisi: 5 Maret 2024; Disetujui: 16 Mei 2024

ABSTRACT

Rainfall plays a crucial role in shaping the weather in Indonesia, influenced by factors such as latitude, elevation, wind patterns, land and water distribution, as well as topography. Rising temperatures contribute to the increased intensity of extreme rainfall, amplifying the potential risk of disasters. Therefore, it is necessary to conduct analyses to predict weather based on historical time series data. This study aims to identify short-term rainfall patterns and trends using the SARIMAX method. The initial stage involves data processing and splitting the data into training data (2005-2014) and test data (2015-2019). Time series decomposition is then performed to identify patterns, followed by period identification and stationarity testing using the ADF test. The SARIMAX model is selected based on the lowest AIC value, followed by parameter estimation and diagnostic tests. Rainfall predictions are evaluated using model performance evaluation methods, while inferential statistics are used to describe population attributes through confidence intervals. The forecasting results show that using the SARIMAX (0, 0, 1)(0, 1, 1, 12)¹² model for short-term rainfall prediction achieved the best performance with the smallest MSE, MAE, and MAPE values on the test sample with a 95% confidence level. Repeating rainfall patterns with a mid-year decline and significant variation in magnitude were identified from the sample data for August 2020-2027, with the lowest rainfall at 39.05 mm and the highest at 397.03 mm in December. The implications of this study support efforts to mitigate natural disasters due to unexpected weather changes by integrating this model into early warning systems and water resource planning.

Keywords : hydrology, precipitation, time series, short-term, SARIMAX

ABSTRAK

Hujan memiliki peran penting dalam pembentukan cuaca di Indonesia, hujan selain dipengaruhi oleh lintang, elevasi, pola angin, distribusi tanah dan air, serta topografi. Peningkatan temperatur berkontribusi pada peningkatan intensitas hujan ekstrem dan memperbesar potensi risiko bencana. Sehingga perlu dilakukan analisis untuk mengetahui prakiraan cuaca berdasarkan kumpulan data dalam deret waktu sebelumnya. Penelitian ini bertujuan untuk mengidentifikasi pola dan tren curah hujan jangka pendek menggunakan metode SARIMAX. Tahap awal pemrosesan data serta pembagian data menjadi data latih (2005-2014) dan data uji (2015-2019). Kemudian dilakukan dekomposisi deret waktu untuk mengidentifikasi pola dalam deret waktu, diikuti dengan identifikasi periode dan uji stasioneritas menggunakan tes ADF. Pemilihan model SARIMAX didasarkan pada nilai AIC terendah, diikuti dengan estimasi parameter dan tes diagnostik. Prediksi curah hujan dievaluasi menggunakan metode evaluasi performa model, sementara statistik inferensia digunakan untuk menggambarkan atribut populasi menggunakan interval kepercayaan. Hasil peramalan menunjukkan bahwa penggunaan metode SARIMAX (0, 0, 1)(0, 1, 1, 12)¹² dalam meramalkan curah hujan jangka pendek mendapatkan kinerja terbaik dengan nilai MSE, MAE, dan MAPE terkecil pada uji sampel dengan tingkat kepercayaan 95%. Pola berulang curah hujan dengan penurunan pada pertengahan tahun dan variasi besaran yang signifikan teridentifikasi dari data sampel bulan Agustus 2020-2027, dengan curah hujan terendah 39,05 mm dan tertinggi 397,03 mm pada bulan Desember. Implikasinya, hasil penelitian ini mendukung upaya mitigasi bencana alam akibat perubahan cuaca yang tidak terduga melalui integrasi model ini dalam sistem peringatan dini dan perencanaan sumber daya air.

Kata kunci : hidrologi, curah hujan, deret waktu, peramalan jangka pendek, SARIMA

PENDAHULUAN

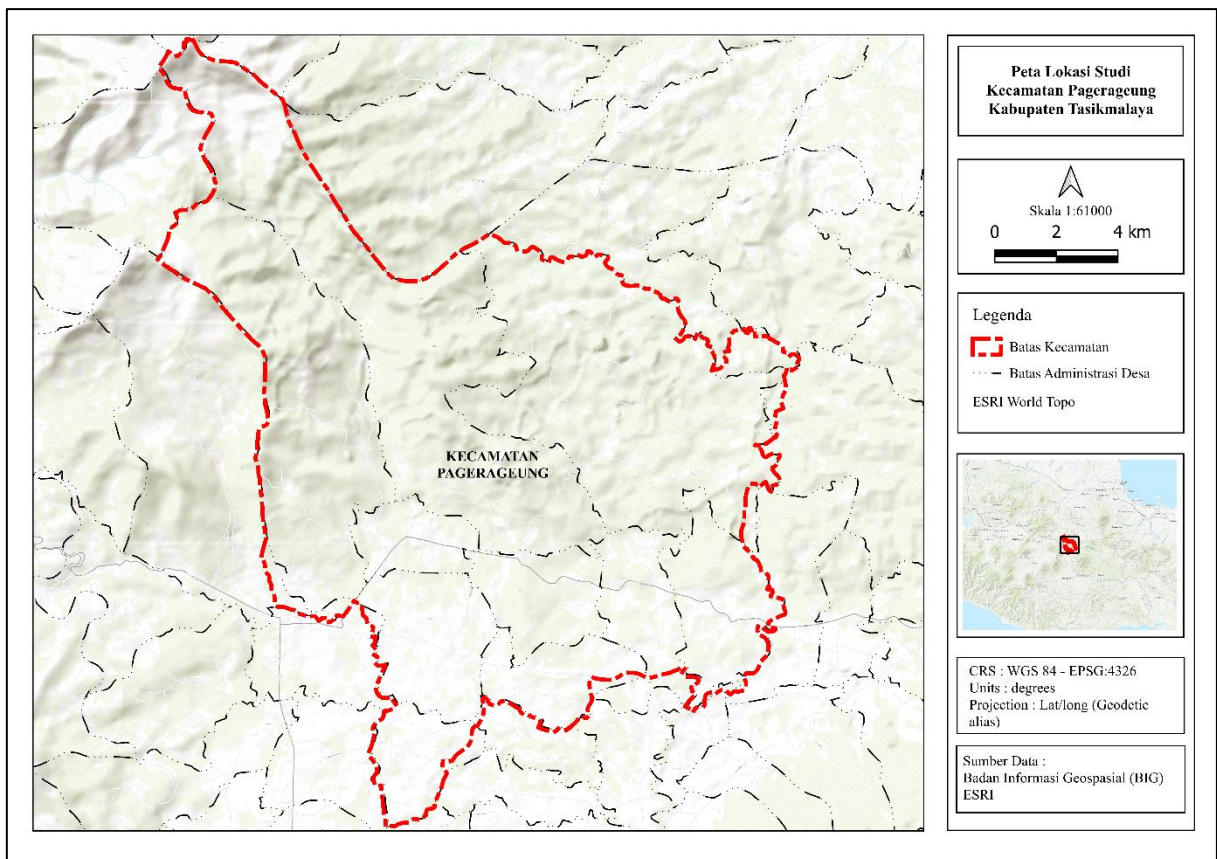
Hujan merupakan proses fisik yang disebabkan oleh fenomena meteorologi. Cuaca adalah suatu sistem yang kompleks, sehingga dapat dipahami bahwa “pemodel cuaca” atau “peramal cuaca” terkadang salah perhitungan (Tukidi, 2010). Sehingga, hujan merupakan salah satu komponen pembentuk cuaca tersebut dan saling berkaitan antara komponen lainnya. Banyaknya curah hujan yang mencapai permukaan bumi atau tanah selama selang waktu tertentu dapat diukur dengan jalan mengukur tinggi air hujan dengan curah hujan, yaitu tanpa mengingat macam atau bentuk pada saat mencapai permukaan bumi dan tidak memperhitungkan endapan yang meresap ke dalam tanah, hilang karena penguapan, atau pun mengalir (Subagyo Swarinoto & Sugiyono, 2011).

Pengaruh faktor geografis wilayah Indonesia dan lingkungan sekitarnya terhadap faktor iklim/ cuaca mengakibatkan 3 (tiga) jenis curah hujan, yaitu : khatulistiwa, muson dan lokal. Beberapa faktor fisik penting yang berperan dalam pembentukan hujan di Indonesia, antara lain : lintang, elevasi, pola angin (angin pasat dan muson), distribusi tanah dan air, serta pegunungan yang tinggi (Tukidi, 2010). Indonesia, yang terletak di antara dua benua dan dua samudera, memiliki iklim yang khas dan rentan terhadap perubahan iklim baik secara lokal maupun

global (Lubis & Delinom, 2010). Di Indonesia, terjadi perubahan musim setiap enam bulan, dimulai dari bulan Oktober hingga Maret adalah musim hujan, sementara musim kemarau berlangsung dari bulan April hingga September (Sukmawati, 2023).

Kecamatan Pagerageung, Kabupaten Tasikmalaya, Provinsi Jawa Barat memiliki koordinat geografis pada -7.0966994 LS, 108.0922783 BT. Secara geografis wilayah ini berbatasan dengan Gunung Cakrabuana di bagian selatan, bersebelahan dengan Kecamatan Panjalu (Ciamis) di bagian timur, serta berbatasan dengan Kecamatan Panumbangan (Ciamis), dan Kecamatan Malangbong (Garut) di bagian utara.

Adapun menurut laporan prakiraan potensi banjir Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), Kabupaten Tasikmalaya memiliki beberapa wilayah kecamatan yang memiliki risiko banjir rendah hingga sedang, termasuk diantaranya adalah Kecamatan Pagerageung (Nuraini, 2023). Menurut keterangan dari Badan Penanggulangan Bencana Daerah (BPBD) Kabupaten Tasikmalaya, situasi saat ini akan memiliki dampak pada daerah sekitarnya. Karena dua sungai meluap, banjir mengakibatkan ratusan rumah di Desa Tanjungsari, Tasikmalaya, Jawa Barat terendam (Purnama, 2023). Salah satu



Gambar 1 Peta Lokasi Studi

penyebabnya merupakan curah hujan yang tinggi dan kapasitas sungai yang tidak mampu menampung air sehingga terjadi limpasan di luar badan sungai. Adapun salah satu pos curah hujan milik Balai Besar Wilayah Sungai Citanduy di Pagerageung, Kabupaten Tasikmalaya mampu menjadi sumber data guna memperoleh informasi curah hujan dengan radius tertentu untuk daerah sekitarnya. Guna mengurangi dampak jika terjadi bencana alam karena mitigasi kebencanaan kurang siap, diperlukan pengambilan keputusan yang tepat.

Kabupaten Tasikmalaya, Jawa Barat, memiliki potensi besar dalam berbagai bidang, seperti pertanian, pariwisata, dan industri. Namun, dalam pemanfaatan data dan prediksi, khususnya dengan metode forecasting dengan banyak manfaat, Kabupaten Tasikmalaya masih tertinggal. Meskipun metode forecasting memiliki banyak manfaat, belum ada penelitian atau aplikasi yang menggunakannya di Kabupaten Tasikmalaya. Hal ini menunjukkan peluang besar untuk penelitian dan pengembangan pada banyak hal, khususnya bidang pengelolaan sumber daya air. Adapun skala implementasi metode forecasting periode harian, bulanan hingga musiman, hasilnya membantu dalam pengambilan keputusan dan perencanaan pembangunan jangka panjang.

Data yang digunakan untuk metode prediksi yang sudah ada selama ini umumnya membutuhkan komputasi dengan performa yang lebih tinggi. Hal ini dapat dilihat dari kebutuhan RAM dan CPU yang besar, serta waktu training yang lama, oleh karena itu perlu dilakukan analisis prakiraan untuk mengetahui prakiraan komponen utama penyusun cuaca berdasarkan kumpulan data dalam deret waktu tertentu sebelumnya. Analisis deret waktu adalah seperangkat data yang diperlukan untuk memprediksi masa depan berdasarkan pengamatan masa lalu. Data yang digunakan bisa perjam, harian, mingguan, bulanan dan tahunan (Gujarati, 2003). Dalam penelitian ini, data bulanan digunakan sebagai variabel. Data yang menunjukkan fluktuasi dalam periode waktu tertentu disebut pola musiman (Hayati et al., 2021). Dalam konteks penelitian ini, metode Seasonal Autoregressive Integrated Moving Average (SARIMA) dapat digunakan sebagai alat analisis yang cukup relevan (Hyndman & Athanasopoulos, 2018). Model ARIMAX dan SARIMAX digunakan untuk analisis pembentukan model curah hujan harian kemudian hasil dari keduanya dibandingkan dengan melihat nilai AIC terkecil. Dari hasil analisis diketahui bahwa model SARIMAX lebih cocok digunakan untuk model curah hujan dibanding dengan model ARIMAX dengan perolehan nilai AIC terkecil sebesar 1250.86 (Rahmayani, 2013).

Dalam pendekatan analisis data berdasarkan urutan waktu, terdapat faktor-faktor lain yang memiliki pengaruh pada perilaku data. Oleh karena itu, untuk data dengan pola musiman dan adanya variabel tambahan yang mempengaruhi, metode yang digunakan adalah *Seasonal Autoregressive Integrated Moving Average with Exogenous* (SARIMAX) (Virati et al., 2020). Dengan demikian, SARIMAX dapat diterapkan pada data dengan pola musiman, dengan penambahan faktor X yang mencerminkan fluktuasi yang terjadi saat musim hujan tiba. Oleh karena itu, metode yang sesuai untuk memprediksi cuaca dalam jangka pendek adalah pendekatan deret waktu menggunakan *Seasonal Autoregressive Integrated Moving Average With Exogenous Variables* (SARIMAX). Studi ini bertujuan untuk menerapkan serta mengidentifikasi pola dan tren curah hujan jangka pendek menggunakan metode SARIMAX berdasarkan data historis.

SARIMAX menawarkan solusi yang lebih efisien dalam hal komputasi. Meskipun membutuhkan waktu training yang lebih lama, SARIMAX dapat dijalankan dengan sumber daya komputasi yang lebih terbatas dan hemat energi. Oleh karena itu, SARIMAX menjadi pilihan yang tepat untuk penelitian di lokasi yang memiliki keterbatasan komputasi.

Metode SARIMAX telah terbukti efektif dalam memprediksi curah hujan di berbagai wilayah di Indonesia. Penelitian di Jawa Barat menunjukkan kinerja model SARIMAX yang memuaskan dengan RMSE 13.5 mm dan Indeks Nash-Sutcliffe 0.78 (Marita, 2019).

METODOLOGI

Data sekunder dalam penelitian ini diperoleh dari Balai Besar Wilayah Sungai Citanduy dengan mengambil total curah hujan bulanan dari Pos Curah Hujan Pagerageung, Kabupaten Tasikmalaya, menggunakan sejumlah 180 data. Metode yang digunakan dalam penelitian ini adalah metode *Seasonal Autoregressive Integrated Moving Average With Exogenous Variables* (SARIMAX). Metode yang dipilih menentukan model, prediksi dan jenis perangkat lunak yang akurat dalam langkah penemuan data menggunakan *Python*.

Dimulai dengan *data preprocessing* yang terdiri dari pengumpulan sumber data, pembersihan data, pemilihan atribut, transformasi data dan penggabungan data yaitu mereplikasi data dari lebih dari *dataset* untuk menjadi *dataset* yang baru (Henderi & Wanda, 2017).

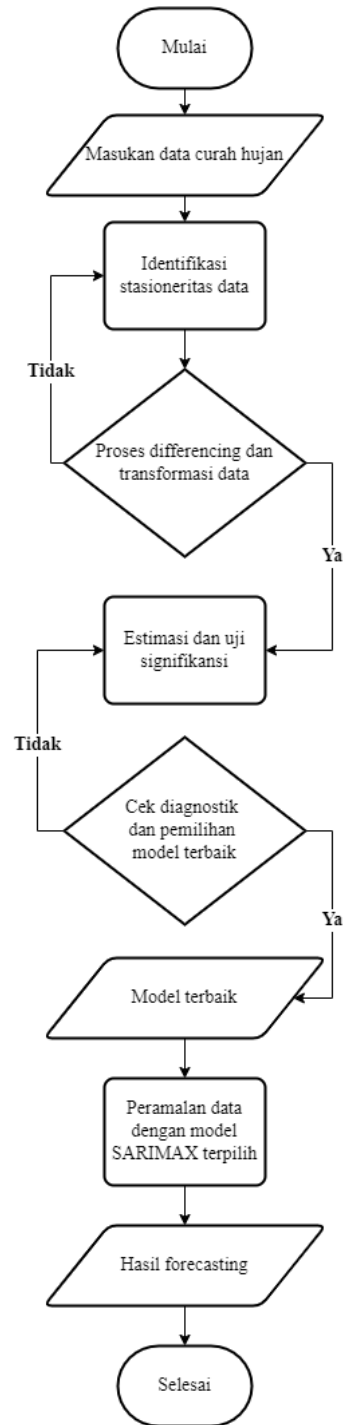
Berikut adalah pemodelan *Seasonal Autoregressive Integrated Moving Average With Exogenous Variables* (SARIMAX), yang dimulai dengan membagi data menjadi data latih (data curah hujan tahun 2005-2014) dan data uji (data curah hujan tahun 2015-2019), perlu dilakukan dekomposisi deret waktu untuk menggambarkan dan memeriksa pola pada dalam suatu deret waktu menjadi beberapa komponen utama, yaitu trend, musiman, dan residual (Hyndman & Athanasopoulos, 2018) Kemudian identifikasi *rata* periode 4,6,8, dan 12 bulan untuk membandingkan data eksisting dan data yang bergulir rata-rata secara visual, sedangkan tes *Augmented Dickey-Fuller* (ADF) diperlukan untuk menguji stasioneritas dalam uji deret waktu dan setelah itu, evaluasi dilakukan dengan mengukur hubungan kolerasi antara variabel dan data dalam konteks waktu menggunakan fungsi *Autocorrelation Function* (ACF) sedangkan fungsi *Partial Autocorrelation Function* (PACF) digunakan untuk menilai kolerasi antara variabel tersebut dan data pada titik tertentu, mempertimbangkan kontrol variabel lainnya. Pilihan model SARIMA yang paling sesuai ditentukan berdasarkan nilai AIC yang terendah.

Selanjutnya kita menjalankan tes diagnostic pada model SARIMA untuk menilai apakah ada jejak residu yang menunjukkan *white noise* atau tidak. Kemudian, kita kembali ke langkah estimasi parameter dan uji diagnostik dengan menggunakan model SARIMAX. Di PCH Pagerageung, Kabupaten Tasikmalaya, prakiraaan curah hujan dilakukan dengan menggunakan model SARIMAX. Selanjutnya, peforma algoritma prediksi dievaluasi menggunakan metode *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE), dan *Mean Percentage Absolute Error* (MAPE) untuk menilai apakah hasil ramalan dalam kategori yang memuaskan atau tidak (Hayati et al., 2021). Selanjutnya, dalam konteks statistik inferensia (induktif), menggunakan rentang keyakinan (*confidence interval*) untuk menggambarkan atribut atau sifat-sifat dari populasi, seperti rerata dan standar deviasi (Damanik & Simamora, 2019). Berikut ini *flowchart* untuk menggambarkan tahapan prosedur model SARIMAX pada **Gambar 2**.

Identifikasi Stasioneritas Data

Sebelum menentukan model SARIMAX yang optimal, Langkah awalnya adalah melakukan pengujian atas stasioneritas data curah hujan, baik dalam hal *varians dan mean*. Oleh karena itu, pada data yang telah stasioner, uji *varians* menggunakan

metode *Box-cox* dilakukan, sementara *mean* diuji dengan *Dickey Fuller Test* (Mubarok et al., 2015).



Gambar 2 Flowchart forecasting model SARIMAX

Jika terdapat ketidakstasioneran dalam data terkait *varians*, maka Langkah selanjutnya adalah menerapkan transformasi *Box-cox* menggunakan rumus umum sebagai berikut (Wei, 2006) :

$$Z^* = \frac{Z^{\lambda-1}}{\lambda}, \text{ berlaku } \lambda \neq 0 \dots\dots\dots (1)$$

Dimana :

Z^* : Transformasi Box Cox

λ : Nilai koefisien dari transformasi Box Cox

Dickey-Fuller Test merupakan metode dalam analisis deret waktu yang menguji apakah suatu deret waktu stasioner atau tidak. Bentuk matematisnya adalah sebagai berikut:

$$H_0: \phi = 0 \text{ (Terdapat akar unit)}$$

$$H_1: < 0 \text{ (Tidak terdapat akar unit)}$$

Pada persamaan Dickey-Fuller, variabel dependennya merupakan *time series* yang diuji, apakah stasioner atau tidak (yt), sedangkan variabel independennya adalah waktu (t) dan lag dari *time series* tersebut ($yt-1$). Berikut persamaan regresi Dickey-Fuller adalah sebagai berikut :

$$yt = \alpha + \beta t + \gamma yt - 1 + \varepsilon \dots\dots\dots (2)$$

Dimana :

yt : deret waktu yang akan diuji stasioner atau tidak

α : *intercept*

β : koefisien regresi terhadap waktu

γ : koefisien regresi terhadap lag dari deret waktu

ε : *error term* (residual)

Model prediksi *Box-Jenkins* diidentifikasi dengan memeriksa pola *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF). Secara prinsip, baik ACF maupun PACF memiliki tujuan yang sama, yaitu mengungkap korelasi linier antara data dalam deret waktu yang dipisahkan oleh lag- k , yang mengacu pada keterkaitan data dalam deret waktu yang terpisah oleh lag- k (Hadiansyah, 2020).

Autocorrelation Function (ACF)

Autocorrelation Function (ACF) merupakan cara untuk melihat keterkaitan yang antara elemen-elemen dalam dataset deret waktu melalui pengamatan (Spyros et al., 1993) Apabila data tidak menunjukkan tanda-tanda autokorelasi, maka dapat disimpulkan bahwa data tersebut acak atau tidak memiliki pola yang jelas. Nilai ACF pada lag- k dapat diformulasikan sebagai berikut :

$$\rho k = \frac{\sum_{t=1}^{N-k} (Z_t - \bar{z})(Z_{t+k} - \bar{z})}{\sum_{t=1}^{N-k} (Z_t - \bar{z})^2} \dots\dots\dots (3)$$

Dimana :

ρk : fungsi autokorelasi

Z_t : data waktu- t

Z : rata-rata data ke- t

Z_{t+k} : data waktu $t+k$

Partial Autocorrelation Function (PACF)

Fungsi *Partial Autocorrelation Function* (PACF) mengindikasikan sejauh mana keterkaitan antara Z_t dan Z_{t+k} dengan memperhitungkan penghapusan pengaruh lag 1,2, dan seterusnya hingga $k-1$. Nilai PACF pada lag- k dapat ditentukan sebagai berikut.

$$\phi k K = \frac{\rho k - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_{k-j}} \dots\dots\dots (4)$$

Dimana :

$\phi k K$: fungsi autokorelasi parsial

ρk : fungsi autokorelasi

Peramalan Data dengan Model *Seasonal Autoregressive Integrated Moving Average With Variabel Exogen* (SARIMAX)

SARIMA dengan variabel bebas adalah SARIMAX. $(p, d, q)(P, D, Q)^s(X)$ di mana, X adalah variabel bebas. Variabel independen dapat disimulasikan dengan menggunakan sejumlah persamaan linier berganda. Bentuk umum model SARIMAX $(p, d, q)(P, D, Q)^s(X)$ dinyatakan (Arunraj et al., 2016) :

$$Y_t = \beta_0 + \beta_1 V_{1,t} + \beta_2 V_{2,t} + \dots + \beta_i V_{i,t} + \omega_T \dots\dots\dots (5)$$

Dimana :

Y_t : data pada waktu ke- t

$V_{i,t}$: variabel *dummy* ke- j pada saat t dengan $i = 1,2,3,\dots,i$

β_0, β_k : koefisien regresi dari variabel *dummy*

ω_T : residual pada waktu ke- t dari proses deret waktu

Residual ω_T dinyatakan dalam model ARIMA (Arunraj et al., 2016) :

$$\omega_T = \frac{\theta_q(B)\theta_q(B^S)e_t}{\Phi_p(B^S)\phi_p(B)(1-B)^d(1-B^S)^D} \dots\dots\dots (6)$$

Selanjutnya, persamaan (5) substitusikan ke dalam persamaan (4). Model umum SARIMAX dapat ditulis (Arunraj et al., 2016) :

$$Y_t = \beta_0 + \beta_1 V_{1,t} + \beta_2 V_{2,t} + \dots + \beta_i V_{i,t} + \frac{\theta_q(B)\theta_q(B^S)e_t}{\Phi_p(B^S)\phi_p(B)(1-B)^d(1-B^S)^D} \dots\dots\dots (7)$$

$$\Phi_p(B^S)\phi_p(B)(1-B)^d(1-B^S)^D Y_t = \theta_q(B)\Theta_q(B^S)e_t + \beta_0 + \beta_1 V_{1,t} + \beta_2 V_{2,t} + \dots + \beta_i V_{i,t} \dots \quad (8)$$

Dimana :

p, d, q : orde AR, differencing, MA non musiman

P, D, Q : orde AR, differencing, MA musiman

S : musiman

$\Phi_p(B^S) = (1 - \Phi_1 B^{1S} - \dots - \Phi_p B^S)$ koefisien komponen AR musiman S dengan orde P

$\Theta_q(B^S) = (1 - \theta_1 B^{1S} - \dots - \theta_q B^S)$ koefisien komponen MA musiman S dengan orde Q

$(1 - B^S)^D$: differencing musiman S dengan orde D

e_t : nilai error pada waktu ke-t

B : operator backward shift (operator langkah mundur)

Pengukuran Kinerja Algoritma

Pemilihan model terbaik memperhitungkan nilai residu. Nilai residu adalah selisih antara hasil taksiran dengan nilai sebenarnya. Oleh karena itu itu, penting untuk melakukan evaluasi terhadap akurasi dari penerapan metode yang digunakan. Pada tahap ini, beberapa metode yang harus diterapkan untuk mengukur kinerja suatu algoritma, yaitu Mean Square Error (MSE), Mean Absolute Error (MAE), dan Mean Percentage Absolute Error (MAPE) (Khalimi, 2021).

Penilaian model terbaik pada kriteria MSE (Mean Square Error), dapat dihitung menggunakan rumus berikut, dimana e_i merupakan residu pengamatan ke-i (Susanto & Ulama, 2016) :

$$MSE = \sqrt{\frac{(Y' - Y)^2}{n}} \dots \dots \dots (9)$$

Dimana :

Y' : Nilai prediksi

Y : Nilai eksisting

n : Jumlah data

Dalam perhitungan Mean Absolute Error (MAE), perlu menghasilkan nilai kesalahan rata-rata yang merupakan perbedaan antara nilai yang sebenarnya dan nilai yang prediksi. MAE sering digunakan untuk mengukur tingkat kesalahan dalam analisis deret waktu.

$$MAE = \sum \frac{|Y' - Y|}{n} \dots \dots \dots (10)$$

Dimana :

Y' : Nilai prediksi

Y : Nilai eksisting

n : Jumlah data

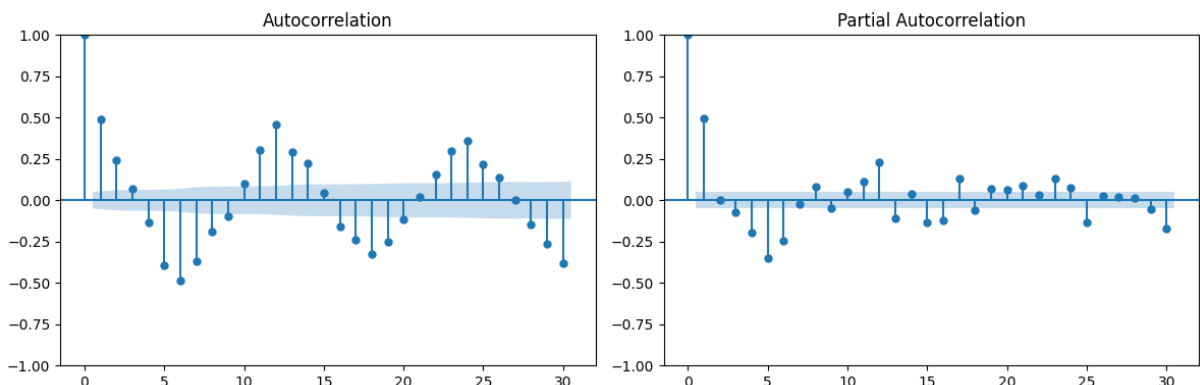
Dalam konteks ini, Mean Percentage Absolute Error (MAPE) adalah hasil perhitungan persentase berdasarkan nilai MAE, yang kemudian digunakan untuk menentukan nilai selanjutnya.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|\hat{Y}_t - Y_t|}{Y_t} \times 100 \dots \dots \dots (11)$$

Rentang Keyakinan (Confidence Interval)

Nilai parameter dalam rentang kepercayaan diestimasi menggunakan estimasi interval (interval estimation), yaitu interval yang menunjukkan rentang di mana suatu parameter populasi mungkin berbeda.

Ini bertujuan untuk membentuk landasan bagi proses membuat kesimpulan mengenai parameter-parameter populasi berdasarkan data yang diperoleh secara empiris. Oleh karena itu, setiap kesimpulan yang diambil terkait dengan parameter-parameter dari sampel yang diamati harus mempertimbangkan pemahaman tentang distribusi sampel tersebut (Boediono & Koster, 2002).



Gambar 3 a) ACF dan b) PACF Data Curah Hujan Pos Pagerageung Tahun 2005-2022

Umumnya, ketika tidak memiliki pengetahuan tentang parameter-parameter dalam suatu populasi, kita melakukan estimasi berdasarkan sampel statistik yang diambil dari populasi tersebut. Tujuan estimasi adalah untuk mendapatkan pemahaman yang lebih baik tentang karakteristik-karakteristik populasi yang belum diketahui, dengan menggunakan informasi yang diperoleh dari sampel atau alat estimasi (Sudjana, 2002). Bentuk umum untuk semua rentang keyakinan :

$$C.I = \text{Point Estimate} \pm (\text{Crit. Value})(\text{Standard Error}) \dots\dots\dots (11)$$

Tentukan probabilitas yang diinginkan (sampaikan μ dalam interval tertentu) dengan tingkat kepercayaan sebesar 95%, lalu lakukan luas interval yang memiliki peluang 95% untuk mengandung nilai sebenarnya.

$$P(\bar{X} - c \leq \mu \leq \bar{X} + c) = 0.95 \dots\dots\dots (12)$$

Berikut adalah bentuk umum dari interval kepercayaan (*confidence interval*) untuk nilai rata-rata populasi Ketika deviasi standar σ sudah diketahui.

$$\bar{X} \pm z \frac{\sigma}{\sqrt{n}} \dots\dots\dots (13)$$

Dimana :

- \bar{X} : nilai rata-rata sampel
- z : nilai untuk *confidence level* tertentu
- σ : standar deviasi populasi n
- n : jumlah sampel yang diamati

HASIL DAN PEMBAHASAN

Karakteristik Curah Hujan

Data pemantauan curah hujan di Pos Curah Hujan Pagerageung dari Januari 2005 hingga Desember 2019 menunjukkan bahwa pola curah hujan bersifat musiman dan mengalami fluktuasi. Curah hujan tinggi tercatat pada bulan Desember tahun 2005, 2006, 2012-2014 dan 2016, yang jauh lebih tinggi dibandingkan dengan bulan-bulan lainnya. Rentang nilai curah hujan tertinggi berkisar antara 375,5 hingga 833,5 mm, sementara yang mencapai 0 mm.

Walapun faktor x yang mempengaruhi curah hujan adalah musim kemarau dan musim hujan, namun besarnya curah hujan akan berubah secara drastis pada setiap periode tersebut (Virati et al., 2020). Pada penelitian (Lee & Suhartono, 2006) memperhitungkan efek variasi kalender dengan memasukkan faktor x .

Stasioneritas Data

Dalam proses pemodelan SARIMAX, penting untuk tetap mematuhi asumsi telah ditetapkan. Salah satu asumsi penting adalah stasioneritas data, yang dapat diuji dengan memeriksa grafik rata-rata bergerak (*varians/ moving average*) dan mengamati bagaimana data berubah seiring waktu. Untuk mengonfirmasi asumsi ini, dilakukan uji stasioneritas menggunakan uji *Augmented Dicky Fuller* (ADF) dan menerapkan diferensiasi data selama 12 periode. Hasil uji menunjukkan bahwa nilai *p-value* yang diperoleh adalah 0.000002, yang jelas lebih kecil dari tingkat signifikansi $\alpha = 5\%$ (0,05). **Gambar 3** menampilkan hasil dari proses differencing. Hasil pengujian juga berdampak pada nilai *Akaike's Information Criterion* (AIC) yang mencapai skor terendah sebesar 2105,0038 menunjukkan bahwa ini adalah model prediksi terbaik. Oleh karena itu, dapat disimpulkan bahwa data curah hujan stasiun Pagerageung dari tahun 2005 - 2022 adalah stasioner.

Pemodelan SARIMAX

Setelah memastikan bahwa data telah mencapai stasioneritas, langkah berikutnya adalah mengidentifikasi model SARIMAX. Ini dapat dilakukan dengan melihat histogram ACF dan PACF seperti yang ditampilkan dalam **Gambar 3**. Hasil analisis menunjukkan bahwa dalam plot ACF terdapat indikasi pola musiman yang jelas, tercermin dalam kelipatan perpindahan 6 yang lebih menonjol daripada yang lainnya.

Beberapa pilihan model alternatif telah diidentifikasi dengan mempertimbangkan parameter-parameter penting, dan telah memilih 5 model yang memenuhi kriteria tersebut. Untuk mengukur tingkat kesalahan prediksi menggunakan *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE). MSE digunakan untuk mengukur kesalahan prediksi dalam bentuk kuadrat, sehingga rentang nilai yang optimal adalah dari 0 hingga tak terhingga. Berikut ini nilai *Mean Squared Error* (MSE) pada **Tabel 1**.

Tabel 1 Nilai MSE Model SARIMAX

Model	In sample	Out sample
(0, 0, 1)(0, 1, 1, 12) ¹²	15566.32	110.89
(1, 1, 1)(0, 1, 1, 12) ¹²	15523.52	116.83
(1, 1, 1)(1, 2, 1, 12) ¹²	64316.46	191.24
(0, 1, 2)(1, 0, 2, 12) ¹²	36805878.30	8671.20
(1, 1, 1)(2, 2, 1, 12) ¹²	105086.43	171.54

Nilai MSE yang lebih rendah menunjukkan model yang lebih baik dalam meminimalkan kesalahan

prediksi secara keseluruhan. Berdasarkan **Tabel 1** maka model SARIMAX terbaik adalah $(0, 0, 1)(0, 1, 1, 12)^{12}$ dengan nilai *in sample* MSE sebesar 15566.32 dan *out sample* 110.89 lebih kecil dibandingkan dengan kombinasi model lainnya.

MAE digunakan untuk mengukur kesalahan prediksi secara absolut, sehingga rentang nilai terbaik adalah 0 hingga tak terbatas. Berikut ini nilai *Mean Absolute Error* (MAE) pada **Tabel 2**.

Tabel 2 Nilai MAE Model SARIMAX

Model	In sample	Out sample
$(0, 0, 1)(0, 1, 1, 12)^{12}$	92.70	86.22
$(1, 1, 1)(0, 1, 1, 12)^{12}$	91.32	91.59
$(1, 1, 1)(1, 2, 1, 12)^{12}$	116.52	153.54
$(0, 1, 2)(1, 0, 2, 12)^{12}$	4915.52	8657.68
$(1, 1, 1)(2, 2, 1, 12)^{12}$	136.38	137.49

Semakin kecil nilai MAE, semakin tinggi akurasi prediksi model tersebut. Walaupun demikian, nilai optimal MAE akan bervariasi tergantung pada skala variabel target yang digunakan. Berdasarkan informasi yang tercantum pada **Tabel 2** model SARIMAX terunggul adalah $(0, 0, 1)(0, 1, 1, 12)^{12}$ yang memiliki nilai MAE *in sample* sebesar 92.70 dan nilai MAE *out sample* sebesar 86.22. Kombinasi model ini menunjukkan hasil yang lebih baik dibandingkan dengan model-model lainnya.

MAPE berfungsi sebagai alat untuk mengukur kesalahan prediksi dalam bentuk persentase, sehingga rentang nilai ideal berkisar dari 0% hingga tak terbatas. Berikut ini nilai *Mean Absolute Percentage Error* (MAPE) pada **Tabel 3**.

Semakin kecil nilai MAPE, semakin baik model dalam memberikan prediksi yang lebih akurat relatif terhadap nilai aktual. Tetapi, perlu diperhatikan bahwa MAPE memiliki kelemahan

yaitu tidak dapat dihitung jika terdapat nilai aktual yang sama dengan nol. Berdasarkan informasi yang tertera pada **Tabel 3** dapat disimpulkan bahwa model SARIMAX terbaik adalah $(0, 0, 1)(0, 1, 1, 12)^{12}$ yang menunjukkan nilai MAPE *in sample* sebesar 84.83% dan MAPE *out sample* 44.05%, yang lebih rendah dibandingkan dengan kombinasi model lainnya.

Tabel 3 Nilai MAPE Model SARIMAX

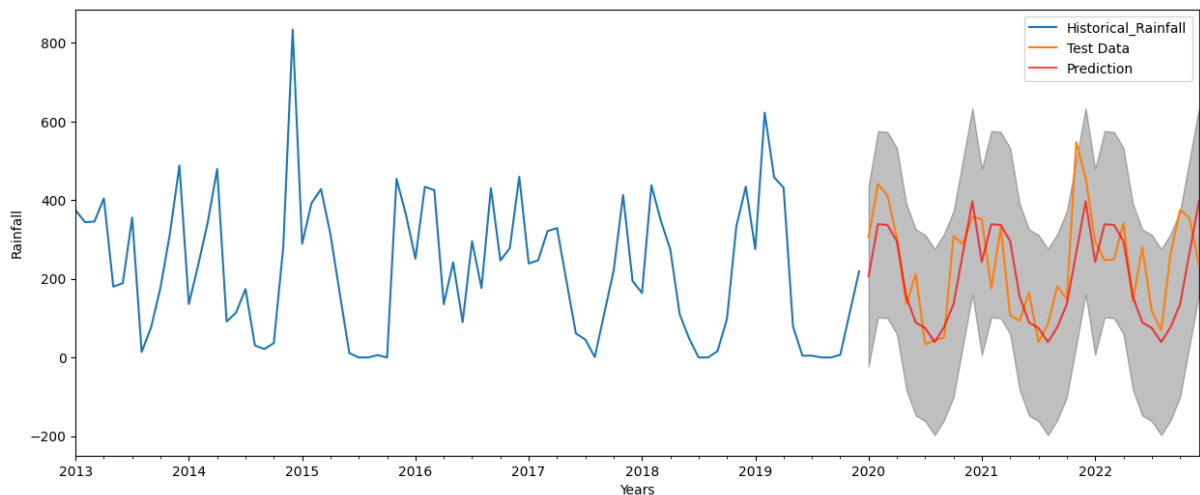
Model	In sample (%)	Out sample (%)
$(0, 0, 1)(0, 1, 1, 12)^{12}$	84.83	44.05
$(1, 1, 1)(0, 1, 1, 12)^{12}$	85.17	44.22
$(1, 1, 1)(1, 2, 1, 12)^{12}$	79.40	79.93
$(0, 1, 2)(1, 0, 2, 12)^{12}$	-104.13	6072.97
$(1, 1, 1)(2, 2, 1, 12)^{12}$	61.54	72.23

Rangkuman lengkap dari pengukuran kesalahan prediksi terdapat dalam

Tabel 4 dan model yang memiliki kinerja baik dapat ditemukan dalam tabel tersebut. Dengan demikian, dapat disimpulkan bahwa model terbaik yang digunakan adalah SARIMAX $(0, 0, 1)(0, 1, 1, 12)^{12}$.

Tabel 4 Rekapitulasi Uji Pengukuran Kesalahan Prediksi

Model SARIMAX	Kriteria	Status Performa
$(0, 0, 1)(0, 1, 1, 12)[12]$	MSE	110.89
	MAE	86.22
	MAPE	44.05
$(1, 1, 1)(0, 1, 1, 12)[12]$	MSE	116.83
	MAE	91.59
	MAPE	44.22
$(1, 1, 1)(1, 2, 1, 12)[12]$	MSE	191.24
	MAE	153.54
	MAPE	79.93
$(0, 1, 2)(1, 0, 2, 12)[12]$	MSE	8671.20



Gambar 1 Visualisasi nilai eksisting dan prediksi data curah hujan tahun 2013-2022

Model SARIMAX	Kriteria	Status Performa
(1, 1, 1)(2, 2, 1, 12)[12]	MAE	8657.68
	MAPE	6072.97
	MSE	171.54
	MAE	137.49
	MAPE	72.23

Peramalan Data Curah Hujan

Hasil penelitian ini menunjukkan bahwa hasil peramalan dan pemodelan prediktif untuk data deret waktu adalah mempertimbangkan kategori model data. Keempat model data tersebut meliputi data horizontal, trending, musiman dan siklis (Hanke, 1981). Pola dalam data, termasuk tren dan komponen musiman, dapat diidentifikasi dengan menggunakan *autocorrelation analysis* (Hanke, 1981).

Hasil prediksi data curah hujan untuk periode 2020-2022 dapat dijadikan sebagai referensi untuk memproyeksikan interval kepercayaan. Dalam konteks penelitian ini, untuk membangun keyakinan yang lebih tepat dengan tingkat kepercayaan 95%, diperlukan interval kepercayaan yang lebih ketat. Ini berarti bahwa zona penolakan dalam uji hipotesis normal akan menjadi lebih besar. Oleh karena itu, dapat disimpulkan bahwa interval kepercayaan yang lebih pendek menunjukkan tingkat akurasi yang lebih tinggi (Damanik & Simamora, 2019). Berdasarkan prediksi interval kepercayaan dalam penelitian ini, terlihat jelas bahwa sampel data bersifat musiman karena fluktuasi data terjadi secara berkala dalam satuan waktu tahunan.

Prediksi adalah upaya untuk memproyeksikan peristiwa yang akan terjadi di masa depan. Ketika melihat data historis dan hasil prediksi, terlihat adanya keterkaitan yang menunjukkan bahwa prediksi yang telah dilakukan dapat digunakan sebagai panduan atau referensi (Wahyuni et al., 2014). Dengan curah hujan pagerageung rentang tahun 2013 – 2019 digunakan sebagai acuan untuk menguji pola variatif antara data eksisting dan prakiraan untuk mencari model prediksi data curah hujan di masa mendatang sampai dengan 96 (sembilan puluh enam) periode atau 8 (delapan) tahun terhitung mulai dari tahun 2020 hingga 2022.

Dalam penelitian ini, guna membangun keyakinan yang tepat dengan tingkat kepercayaan 95%, digunakan metode inferensi statistik untuk menghasilkan rentang nilai atas dan bawah. Berdasarkan .

Tabel 5 menampilkan data *confidence interval* peramalan batas atas dan batas bawah data curah

hujan periode tahun 2020 – 2022 pada tingkat keyakinan 95%, sedangkan kondisi tersebut menjadikan kepercayaan lebih proporsional pada penelitian ini telah cukup terpenuhi. Karena data sampel yang merupakan hasil perhitungan dan peluang keberhasilannya tetap sama dari satu percobaan ke percobaan lainnya sehingga kondisi syarat binominal terpenuhi. Bisa dilihat pada **Error! Reference source not found.** yang merupakan hasil dari model peramalan menunjukkan pola curah hujan berulang yang terjadi penurunan pertengahan tahun dengan besaran bervariasi, diambil sampel data pada bulan agustus berulang setiap tahun mulai dari 2020 – 2022 dengan nilai terendah 39.05 mm sedangkan untuk nilai paling tinggi berada di bulan desember sebesar 397.03 mm.

Tabel 5 Perbandingan *Confidence Interval* Pada Tingkat Kepercayaan 95% hasil Peramalan Curah Hujan 8 tahun ke depan

Date	Forecast	95% Confidence Interval	
		Lower	Upper
01/01/2020	206.69	-23.37	436.75
01/02/2020	338.69	101.81	575.56
01/03/2020	336.85	99.98	573.72
01/04/2020	295.84	58.97	532.72
01/05/2020	154.54	-82.34	391.41
...
01/08/2022	39.05	-197.89	275.99
01/09/2022	77.28	-159.66	314.22
01/10/2022	135.73	-101.21	372.67
01/11/2022	266.31	29.37	503.25
01/12/2022	397.03	160.11	633.94

Hasil prediksi curah hujan mulai dari januari 2020 hingga desember 2022 menunjukkan pola perubahan yang signifikan dalam nilai prediksi dari bulan ke bulan, mencerminkan pola musiman dalam kondisi cuaca atau faktor-faktor lingkungan lainnya yang memengaruhi curah hujan.

Oleh karena itu, hasil prediksi dari model SARIMAX untuk data curah hujan di Pagerageung menunjukkan bahwa curah hujan tertinggi akan terjadi pada bulan Desember, sementara yang terendah akan terjadi pada bulan Agustus. Pola ini konsisten berulang setiap tahunnya dalam rentang tahun 2020 – 2022. Hal ini memberikan indikasi tentang pola peramalan untuk masa depan. Di Indonesia, musim hujan aktif biasanya dimulai pada bulan Oktober dan berakhir pada bulan Maret, sedangkan musim kemarau akan berlanjut dari bulan April hingga bulan September.

Model peramalan menggunakan metode SARIMAX memiliki dampak yang signifikan dalam meningkatkan kesadaran terhadap mitigasi kebencanaan banjir. Dampak tersebut mencakup berbagai aspek lingkungan yang saling berhubungan untuk membantu mengatasi ancaman bencana hidrometeorologi dan mengurangi dampak negatif, khususnya di sekitar wilayah Pagerageung, Kabupaten Tasikmalaya hingga secara umum di seluruh dunia.

KESIMPULAN

Kesimpulan

Pada penelitian ini menunjukkan bahwa metode SARIMAX dalam meramalkan curah hujan jangka pendek di Pagerageung, Kabupaten Tasikmalaya dapat memperoleh hasil kinerja MSE, MAE, dan MAPE dengan *out sample* terkecil dengan tingkat kepercayaan 95%.

Terjadi pola berulang curah hujan dengan penurunan yang terjadi pada pertengahan tahun dan variasi besaran yang signifikan. Data sampel setiap bulan Agustus dari tahun 2020 hingga 2022 menunjukkan bahwa curah hujan rendah tercatat 39,05 mm, sementara curah hujan tertinggi pada bulan Desember sebesar 397,03 mm.

Implikasinya, hasil penelitian ini dapat mendukung upaya mitigasi bencana alam yang disebabkan oleh perubahan cuaca tidak terduga mencakup integrasi model ini dalam sistem peringatan dini dan perencanaan sumber daya air untuk meningkatkan kewaspadaan dan mitigasi risiko terkait curah hujan ekstrem.

Rekomendasi

Untuk penelitian selanjutnya disarankan melakukan, pertimbangan integrasi variabel lain yang relevan untuk memperkuat model SARIMAX dan menghasilkan peramalan yang lebih presisi. Selain itu, dalam rangka meningkatkan akurasi peramalan, perlu diupayakan pengumpulan data curah hujan yang lebih lengkap dan berkualitas tinggi. Data yang lebih rinci dan konsisten akan membantu meningkatkan keandalan model dan meminimalkan potensi bias dalam peramalan.

Diperlukan pula mempertimbangkan analisis perbandingan dengan metode peramalan lainnya, perbandingan ini akan memberikan wawasan lebih mendalam tentang kelebihan dan kelemahan masing-masing metode peramalan, serta membantu dalam penentuan metode terbaik untuk situasi atau lokasi geografis tertentu.

Metode SARIMAX ini hanya menginterpretasikan pola stasioner dari bagian yang dipilih saja, perlu prediksi dengan metode lain agar mendapatkan hasil fluktuatif. Karena SARIMAX dengan hasil stasioner umumnya memiliki *lead time* prediksi pendek, sehingga untuk *lead time* yang lebih panjang cenderung akurasi prediksi menurun karena model tidak dapat menangkap perubahan terhadap tren dan pola data di masa depan. Sehingga memerlukan metode lain untuk melakukan prediksi jangka panjang.

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>
- Boediono, & Koster, W. (2002). *Teori dan Aplikasi Statistika dan Probabilitas: Sederhana, Lugas dan Mudah Dimengerti* (2nd ed.). PT. Remaja Rosdakarya.
- Damanik, E. O., & Simamora, E. (2019). Estimasi Interval Kepercayaan Parameter Selisih Rata-Rata IPK Kelas Pendidikan Reguler dan Ekstensi FMIPA Angkatan 2010 Unimed Dengan Bootstrap Persentil. *KARISMATIKA (Kumpulan Artikel Ilmiah, Informatika, Statistik, Matematika Dan Aplikasi)*, 5(3), 1–9.
- Gujarati, D. N. (2003). Basic Econometrics. In *The Economic Journal* (4th ed., Vol. 82, Issue 326). McGraw Hill Book Company. <https://doi.org/10.2307/2230043>
- Hadiansyah, A. (2020). *Peramalan Penjualan Cat Pada PT. HIJ Menggunakan Autoregressive Integrated Moving Average*. Fakultas Pascasarjana, Universitas Komputer Indonesia.
- Hanke, J. E. (1981). *Business Forecasting* (8th ed.). Pearson Addison Wesley.
- Hayati, Debetaraja, N. N., & Martha, S. (2021). Prediksi Data Jumlah Penumpang Kereta Dengan Efek Variasi Kalender Pada Model SARIMAX. *Buletin Ilmiah Matematika, Statistika Dan Terapannya (Bimaster)*, 10(4), 379–388. <https://doi.org/10.26418/bbimst.v10i04.49536>

- Henderi, & Wanda, R. L. (2017). Preprocessing Data Untuk Sistem Peramalan Tingkat Kedisiplinan Mahasiswa. *Innovative Creative and Information Technology*, 3(2), 296–308. <https://doi.org/10.33050/icit.v3i2.70>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting : Principles and Practice Chapter 1 Getting started* (2nd ed.). OTexts.
- Khalimi, A. M. (2021). *Cara Hitung RMSE, MSE, MAPE, dan MAE dengan Excel*. Pengalaman-Edukasi.Com. <https://www.pengalaman-edukasi.com/2021/01/cara-menghitung-rmse-root-mean-square.html>
- Lee, M. H., & Suhartono. (2006). Calendar Variation Model For Forecasting Time Series Data With Islamic Calendar Effect. *Jurnal Matematika Sains Dan Teknologi*, 7(2 SE-Articles), 85–94.
- Lubis, R. F., & Delinom, R. (2010). Perubahan Iklim dan Pemanasan Global di Indonesia; Dampaknya Terhadap Kondisi Bawah Permukaan Studi Kasus : DKI Jakarta. *Widyariset, Pusat Geoteknologi, Lembaga Ilmu Pengetahuan Indonesia*, 13(3), 65–70. <https://doi.org/10.14203/widyariset.13.3.2010.65-70>
- Marita, T. (2019). Pemodelan Curah Hujan Bulanan di Jawa Barat dengan Pendekatan SARIMA dan SARIMAX. *IOP Conference Series: Earth and Environmental Science*, 434.
- Mubarok, M. H., Atmono, D., Widodo, A., & Suprih, S. (2015). Peramalan Cuaca di Stasiun Meteorologi Klas I Juanda Surabaya Menggunakan Metode ARIMA dan Vector Autoregressive (VAR). *Fakultas Matematika Dan Ilmu Pengetahuan Alam, Institut Sepuluh November (ITS)*, 1–6.
- Nuraini, A. S. (2023). *Curah Hujan Tinggi, BMKG Sebut 12 Wilayah di Kabupaten Tasikmalaya Rawan Banjir*. Tribun Priangan. <https://priangan.tribunnews.com/2023/03/30/curah-hujan-tinggi-bmkg-sebut-12-wilayah-di-kabupaten-tasikmalaya-rawan-banjir>
- Purnama, F. (2023). *BPBD: Ratusan rumah warga di Tasikmalaya terdampak banjir*. ANTARANEWS.Com. <https://www.antaraneews.com/berita/3433980/bpbd-ratusan-rumah-warga-di-tasikmalaya-terdampak-banjir>
- Rahmayani, L. (2013). *Model ARIMAX dan SARIMAX Untuk Meramalkan Data Curah Hujan Hujan*. Jurusan Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Gunung Djati.
- Spyros, M., Untung Sus, A., Steven C., W., Basith, A., & Victor E., M. (1993). *Metode dan Aplikasi Peramalan* (2nd ed.). Erlangga.
- Subagyo Swarinoto, Y., & Sugiyono. (2011). Pemanfaatan Suhu Udara dan Kelembapan Udara Dalam Persamaan Regresi untuk Simulasi Prediksi Total Hujan Bulanan di Bandar Lampung. *Jurnal Meteorologi Dan Geofisika*, 12(3).
- Sudjana. (2002). *Metoda Statistika*. Tarsito.
- Sukmawati, A. (2023). *2 Jenis Musim di Indonesia*. Kompas.Com. <https://www.kompas.com/skola/read/2023/03/24/050000169/2-jenis-musim-di-indonesia?page=all#:~:text=Musim di Indonesia mempunyai waktu,dari bulan April hingga September.>
- Susanto, Y., & Ulama, B. S. S. (2016). Pemodelan Curah Hujan dengan Pendekatan Model ARIMA , Feed Forward Neural Network dan Hybrid (ARIMA-NN) di Banyuwangi. In *Jurnal Sains dan Seni ITS* (Vol. 5, Issue 2). Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh November (ITS). <https://doi.org/10.12962/j23373520.v5i2.16409>
- Tukidi. (2010). Karakter Curah Hujan Di Indonesia. *Jurnal Geografi*, 7(2), 136–145. <https://doi.org/10.15294/jg.v7i2.84>
- Virati, M. Q., Pamanik, D. P., & Pramana, S. (2020). Forecasting Number of Passengers of TransJakarta using SARIMAX Method. *Journal of Data Science and Its Applications, Telkom University*, 3(1), 31–037. <https://doi.org/10.34818/JDSA.2020.3.45>
- Wahyuni, S., Farikhin, & Joko Suprayitno, I. (2014). Peramalan Fungsi Transfer Single Input Indeks Harga Saham Gabungan Terhadap Saham Negara Terdekat. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 2(2), 49–56. <https://doi.org/10.26714/jsunimus.2.2.2014.%25p>

Wei, W. W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods*. (2nd ed.). Pearson Addison Wesley.
<https://doi.org/10.2307/2289741>