

PENERAPAN VECTOR AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (VARIMA) PADA PREDIKSI INDEKS STANDAR PENCEMARAN UDARA DI KABUPATEN GRESIK

Istighfarin Djuana Putri*

Program Studi Matematika, FMIPA, Universitas Negeri Surabaya, Surabaya, Indonesia

e-mail : istighfarin.20030@mhs.unesa.ac.id

Affiati Oktaviarina

Program Studi Matematika, FMIPA, Universitas Negeri Surabaya, Surabaya, Indonesia

e-mail : affiatioktaviarina@unesa.ac.id

Abstrak

Vector Autoregressive Integrated Moving Average (VARIMA) adalah suatu pendekatan peramalan data *multivariate time series* yang digunakan untuk mengetahui keterkaitan antara beberapa variabel pada suatu waktu dengan variabel periode sebelumnya. Salah satu implementasi dari model VARIMA yaitu untuk meramalkan Indeks Standar Pencemaran Udara (ISPU). Meningkatnya jumlah pabrik dan aktivitas industri menjadi penyebab utama terjadinya pencemaran udara di Kabupaten Gresik. Oleh karena itu, penelitian ini bertujuan untuk menerapkan model dan memprediksi konsentrasi ISPU dengan menggunakan VARIMA. Data yang digunakan dalam penelitian ini adalah data harian konsentrasi $PM_{2.5}$, NO_2 , dan O_3 pada bulan November 2023 sampai dengan Februari 2024. Hasil dari penelitian ini diperoleh bahwa model terbaik berdasarkan nilai RMSE dan MAD terkecil model VARIMA (2,1,1), dimana konsentrasi $PM_{2.5}$ dan O_3 pada saat t dipengaruhi oleh ketiga variabel pada saat ke $(t-1)$, $(t-2)$, dan $(t-3)$ serta dipengaruhi oleh nilai residual ketiga variabel pada saat $(t-1)$. Sedangkan variabel NO_2 pada saat ke- t hanya dipengaruhi oleh variabel $PM_{2.5}$ dan O_3 pada saat ke $(t-1)$, $(t-2)$, dan $(t-3)$ serta dipengaruhi oleh nilai residual ketiga variabel pada saat $(t-1)$.

Kata Kunci: Pencemaran Udara, Peramalan, VARIMA.

Abstract

Vector Autoregressive Integrated Moving Average (VARIMA) is a data forecasting approach *multivariate time series* which is used to determine the relationship between several variables at one time and variables in the previous period. One implementation of the VARIMA model is to predict the Air Pollution Standard Index (ISPU). The increasing number of factories and industrial activities is the main cause of air pollution in Gresik Regency. Therefore, this study aims to apply the model and predict ISPU concentrations using VARIMA. The data used in this research is daily data on concentrations of $PM_{2.5}$, NO_2 , and O_3 from November 2023 to February 2024. The results of this research show that the best model is based on the smallest RMSE and MAD values, the VARIMA model (2,1,1), where the concentrations $PM_{2.5}$ and O_3 at time t are influenced by the three variables at time $(t-1)$, $(t-2)$, and $(t-3)$ and is influenced by the residual values of the three variables at time $(t-1)$. Meanwhile, the NO_2 variable at time t is only influenced by the $PM_{2.5}$ and O_3 variables at time $(t-1)$, $(t-2)$, and $(t-3)$ and is influenced by the residual values of the three variables at time $(t-1)$.

Keywords: Air Pollution, Forecasting, VARIMA.

PENDAHULUAN

Pencemaran udara merupakan salah satu permasalahan utama yang setiap tahunnya selalu terjadi. Hal ini terjadi karena adanya kemajuan teknologi. Kemajuan teknologi yang dimaksud antara lain semakin banyaknya industri pabrik, pembangkit listrik, dan kendaraan bermotor yang setiap harinya menghasilkan polusi. Akibatnya udara segar yang menjadi sumber pernapasan menjadi tercemar sehingga dapat menimbulkan gangguan kesehatan bagi manusia dan juga

berbahaya bagi ekosistem lingkungan (Abidin dan Hasibuan, 2019).

Berdasarkan data IQAir, pada tahun 2022, Indonesia pernah menduduki peringkat pertama sebagai negara dengan tingkat polusi terburuk di Asia Tenggara. Tinggal di Indonesia terutama di kota-kota besar tentunya tidak akan lepas dari permasalahan polusi udara. Tingginya tingkat polusi udara di kota-kota besar disebabkan oleh meningkatnya penggunaan kendaraan bermotor, bertambahnya industri pabrik, dan semakin berkurangnya ruang hijau sebagai penyaring udara, serta pengelolaan limbah yang kurang tepat. Salah

satunya yang terjadi dalam beberapa pekan terakhir, yaitu meningkatnya polusi udara di Kabupaten Gresik yang memprihatinkan. Dikutip dari Radar Gresik, pada tanggal 28 Agustus 2023 hasil pemantau kualitas udara menunjukkan bahwa indeks kualitas udara di Kabupaten Gresik berwarna merah atau berarti tidak sehat.

Gresik merupakan salah satu kabupaten di Provinsi Jawa Timur yang terkenal sebagai kota industri karena Kabupaten Gresik menjadi kawasan industri-industri besar, seperti industri pupuk dari PT Petrokimia Gresik, industri semen dari PT Semen Gresik, industri cat dari PT Nippon Paint, dan lain-lain. Dilihat dari hal tersebut, tentu saja menimbulkan banyak dampak, terutama terhadap kualitas udara. Meningkatnya jumlah pabrik dan aktivitas industri menjadi penyebab utama terjadinya pencemaran udara. Selain industri pabrik, menurut Kepala Dinas Lingkungan Hidup (DLH) Kabupaten Gresik, Siti Zubaidah mengatakan bahwa kendaraan bermotor juga menjadi salah satu penyebab terjadinya pencemaran udara, dikarenakan kurang lebih sekitar 70% proses pembakaran bahan bakar berasal dari kendaraan bermotor.

Peraturan Menteri Lingkungan Hidup dan Kehutanan nomor 14 tahun 2020 tentang Indeks Standar Pencemaran Udara (ISPU) menyebutkan bahwa kualitas udara dapat diketahui menggunakan Indeks Standar Pencemaran Udara (ISPU). ISPU terdiri dari 7 parameter yaitu parameter particular (PM_{10}), particular ($PM_{2.5}$), nitrogen dioksida (NO_2), sulfur dioksida (SO_2), karbon monoksida (CO), ozon (O_3), dan hidrokarbon (HC). Peramalan Indeks Standar Pencemaran Udara (ISPU) sangat penting dilakukan karena sebagai acuan untuk membantu pemerintah dalam melakukan upaya pengendalian pencemaran udara baik bagi pemerintah pusat maupun pemerintah daerah dan dalam jangka pendek maupun jangka panjang.

Peramalan adalah proses perkiraan jumlah sesuatu pada waktu yang akan datang berdasarkan data pada masa lampau (Sudjana, 2005). Salah satu metode yang cocok untuk menganalisis data *time series* dengan lebih dari satu variabel adalah *Vector Autoregressive Integrated Moving Average* (VARIMA). *Vector Autoregressive Integrated Moving Average* (VARIMA) adalah suatu pendekatan peramalan data *multivariate time series* yang digunakan untuk mengetahui keterkaitan antara beberapa variabel

pada suatu waktu dengan variabel periode sebelumnya. Model VARIMA merupakan gabungan dari model *Autoregressive Integrated Moving Average* (ARIMA) yang membentuk sebuah vektor dimana antar variabel saling mempengaruhi (Wei, 2006). Keunggulan VARIMA adalah peneliti tidak perlu menentukan mana variabel endogen dan mana variabel eksogen.

Berdasarkan pemaparan diatas, belum ada penelitian tentang pencemaran udara menggunakan metode VARIMA. Oleh karena itu, penulis akan meneliti Indeks Standar Pencemaran Udara (ISPU) di Kabupaten Gresik dengan menggunakan metode *Vector Autoregressive Integrated Moving Average* (VARIMA) dan menggunakan variabel konsentrasi $PM_{2.5}$, NO_2 , dan O_3 . Pemodelan VARIMA dilakukan karena adanya indikasi bahwa setiap konsentrasi ISPU terdapat hubungan timbal balik. Dalam melakukan pemodelan VARIMA, penulis menggunakan *Ordinary Least Square* (OLS) sebagai estimasi parameter serta menggunakan nilai RMSE dan MAD sebagai evaluasi ketepatan model.

KAJIAN TEORI

ANALISIS MULTIVARIATE TIME SERIES

Analisis *time series* adalah salah satu prosedur statistika yang digunakan untuk meramalkan kondisi yang akan terjadi di masa yang akan datang. Jika ditinjau dari variabel penelitian, analisis *time series* dibedakan menjadi 2 yaitu *univariate* dan *multivariate*. Data *time series univariate* merupakan analisis *time series* dengan satu variabel yang memiliki autokorelasi. Sedangkan *time series multivariate* merupakan analisis *time series* dengan lebih dari satu variabel yang memiliki korelasi dan juga memiliki hubungan dengan variabel lainnya. Salah satu model yang dapat digunakan untuk melakukan analisis *time series multivariate* adalah model *Vector Autoregressive Integrated Moving Average* (VARIMA) (Wei, 2006).

STASIONERITAS

Stasioneritas berarti bahwa tidak terdapat perubahan yang drastis pada data. Fluktuasi data berada disekitar suatu nilai rata-rata yang konstan, tidak tergantung pada waktu dan variansi dari fluktuasi tersebut (Makridakis dkk., 1999). Stasioneritas data deret waktu terdiri dari dua, yaitu:

1. Stasioner terhadap rata-rata

Data stasioneritas terhadap rata-rata dapat diuji secara formal menggunakan uji *Augmented Dickey-Fuller* (ADF). Uji ini digunakan untuk melihat apakah data deret waktu stasioner terhadap rata-rata atau tidak. Untuk menstasionerkan data terhadap rata-rata, maka langkah yang diambil adalah melakukan *differencing*. *Differencing* dapat dilanjutkan ke tingkat yang lebih tinggi apabila setelah dilakukan *differencing* pertama data tetap tidak stasioner.

2. Stasioner terhadap varians

Stasioneritas terhadap ragam dapat diperiksa melalui plot *Box-Cox*. Menurut Wei (2006) jika nilai parameter transformasi (λ) sama dengan atau mendekati satu maka data tersebut telah stasioner terhadap ragam. Data yang tidak stasioner terhadap ragam dapat diatasi dengan transformasi *Box-Cox* sebagai berikut:

$$g(y_t) = \begin{cases} \frac{y_t^\lambda - 1}{\lambda} & \text{untuk } \lambda \neq 0 \\ \ln(y_t) & \text{untuk } \lambda = 0 \end{cases} \quad (1)$$

ARIMA

Model ARIMA merupakan model yang terdiri dari penggabungan dua proses yaitu proses AR dan MA dengan *differencing* orde d. Bentuk umum model *autoregressive integrated moving average* (ARIMA) dengan orde (p,d,q) dapat dituliskan sebagai berikut (Wei, 2006).

$$\Phi_p(1 - B)^d Z_t = \Theta_q(B) \alpha_t \quad (2)$$

Dengan:

Z_t : data pada periode ke-t

B : operator *back shift*

d : orde *differencing*

$(1 - B)^d$: deret waktu yang stasioner pada pembedaan ke-d

Φ_i : koefisien *autoregressive*, $i=1,2,\dots,p$

Θ_i : koefisien *moving average*, $i=1,2,\dots,q$

α_t = Nilai *error* pada waktu ke-t

VARIMA

Model *Vector Autoregressive Integrated Moving Average* (VARIMA) adalah model deret waktu multivariat yang telah melalui proses *differencing* dan digunakan untuk mengetahui hubungan diantara beberapa variabel pada waktu dengan periode sebelumnya. Sehingga suatu variabel dipengaruhi oleh variabel-variabel lain pada lag tertentu. Menurut

Wei (2006) bentuk umum model VARIMA (p,d,q) adalah

$$\Phi_p(B)D(B)Z_t = \Theta_q(B)\alpha_t \quad (3)$$

Dimana,

$$\Phi_p(B) = \Phi_0 - \Phi_1 B - \dots - \Phi_p B^p \quad (4)$$

$$\Theta_q(B) = \Theta_0 - \Theta_1 B - \dots - \Theta_p B^p \quad (5)$$

Dengan:

Z_t : vector pengamatan dengan $t=1,2,\dots,n$

B : operator *back shift*

Φ_p : matriks yang berisi koefisien *autoregressive* orde ke-p

Θ_q : matriks yang berisi koefisien *moving average* orde ke-q

α_t = Nilai *error* pada waktu ke-t

MACF

Matrix Autocorelation Function merupakan suatu korelasi dari beberapa variabel yang disusun dalam matriks. Jika terdapat sebuah vektor *time series* dengan variabel pengamatan sebanyak n, yaitu $Z_1, Z_2, Z_3, \dots, Z_n$, maka persamaan matriks korelasi sampelnya adalah sebagai berikut (Wei, 2006)

$$\hat{\rho}(k) = [\hat{\rho}_{i,j}(k)] \quad (6)$$

$\hat{\rho}(k)$ merupakan korelasi silang sampel untuk komponen series ke-i dan ke-j yang dinyatakan dalam persamaan berikut:

$$\hat{\rho}_{i,j}(k) = \frac{\sum_{t=1}^{n-k} (Z_{i,t} - \bar{Z}_i)(Z_{j,t+k} - \bar{Z}_j)}{[\sum_{t=1}^{n-k} (Z_{i,t} - \bar{Z}_i)^2 \sum_{t=1}^{n-k} (Z_{j,t+k} - \bar{Z}_j)^2]^{\frac{1}{2}}} \quad (7)$$

Dengan:

$\hat{\rho}_{i,j}(k)$: korelasi silang sampel variabel ke-I dan ke-j pada lag ke-k

$\bar{Z}_i$: rata-rata sampel dari variabel komponen ke-i

$\bar{Z}_j$: rata-rata sampel dari variabel komponen ke-j

Menurut Wei (2006) MACF digunakan untuk mengidentifikasi orde model MA. MACF memiliki sifat *cut off* setelah lag q pada model MA(q). Tiao dan Box (1981) dalam Wei (2006) memperkenalkan sebuah metode yang sesuai untuk meringkas hasil korelasi sampel dengan menggunakan simbol (+), (-), dan (·) sebagai berikut:

1. Simbol (+) menotasikan nilai $\hat{\rho}(k)$ lebih besar dari 2 kali *standard error* dan menunjukkan adanya korelasi positif.

2. Simbol (−) menotasikan nilai $\hat{\rho}(k)$ lebih kecil dari -2 kali *standard error* dan menunjukkan adanya korelasi negatif.

3. Simbol (·) menotasikan nilai $\hat{\rho}(k)$ berada di antara -2 sampai 2 kali *standard error* dan menunjukkan tidak adanya korelasi.

MPACF

Matrix Partial Autocorrelation Function merupakan generalisasi dari PACF ke dalam bentuk vektor deret waktu yang dilakukan oleh Tiao dan Box (1981) dalam Wei (2006), matriks ini digunakan untuk mengidentifikasi orde model dugaan sementara. Menurut Wei (2006) MPACF digunakan untuk mengidentifikasi orde model AR. MPACF memiliki sifat cut off setelah lag p pada model AR(p). MPACF pada lag ke-k dinotasikan dengan $\rho(k)$

$$\rho(k) = \begin{cases} \Gamma'(1)[\Gamma(0)]^{-1} \\ \{\Gamma'(k) - c'(k)[A(k)^{-1}b(k)]\}[\Gamma(0) - b'(k)[A(k)^{-1}b(k)] \end{cases} \quad (8)$$

Untuk $k \geq 1$ maka nilai $A(k)$, $b(k)$, dan $c(k)$ adalah sebagai berikut:

$$A(k) = \begin{bmatrix} \Gamma(0) & \Gamma'(0) & \dots & \Gamma'(k-2) \\ \Gamma(1) & \Gamma(0) & \dots & \Gamma'(k-3) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma(k-2) & \Gamma(k-3) & \dots & \Gamma(0) \end{bmatrix}, b(k) = \begin{bmatrix} \Gamma'(k-1) \\ \Gamma'(k-2) \\ \vdots \\ \Gamma'(1) \end{bmatrix}, c(k) = \begin{bmatrix} \Gamma(1) \\ \Gamma(2) \\ \vdots \\ \Gamma(k-1) \end{bmatrix}$$

Dengan $\Gamma(k)$ adalah matriks kovarian lag-k

ESTIMASI DAN UJI SIGNIFIKANSI PARAMETER

1. Estimasi parameter merupakan suatu metode untuk menduga nilai karakteristik populasi berdasarkan nilai karakteristik sampel. Alasan penulis menggunakan *Ordinary Least Square* (OLS) sebagai estimasi parameter adalah karena OLS bekerja dengan cara meminimumkan jumlah kuadrat error tanpa mengasumsikan data tersebut berdistribusi normal atau tidak.
2. Menurut Wei (2006) uji signifikan parameter dilakukan setelah diperoleh nilai estimasi dari beberapa parameter yang terdapat dalam uji model. Uji signifikansi parameter dilakukan dengan menggunakan signifikansi level 5%. Hipotesis yang diuji adalah sebagai berikut:
 $H_0 = 0$ (parameter tidak signifikan)
 $H_1 \neq 0$ (parameter signifikan)
 Dengan statistik uji:

$$t_{hitung} = \frac{\hat{\Phi}_{p(ij)}}{SE\hat{\Phi}_{p(ij)}} \quad (10)$$

$$t_{hitung} = \frac{\hat{\Theta}_{q(ij)}}{SE\hat{\Theta}_{q(ij)}} \quad (11)$$

Dengan:

$\hat{\Phi}_{p(ij)}$: koefisien *vector autoregressive* (VAR)

$\hat{\Theta}_{q(ij)}$: koefisien *vector moving average* (VMA)

Kriteria keputusan yang digunakan yaitu tolak H_0 jika $|t_{hitung}| > t_{\frac{\alpha}{2}, n-p}$ atau p-value $< \alpha$.

UJI RESIDUAL WHITE NOISE

Residual pada model *Vector Autoregressive Integrated Moving Average* (VARIMA) yang terbentuk harus bersifat *white noise* dan berdistribusi normal multivariat.

1. White Noise

Model dikatakan bersifat *white noise* jika residual dari model telah memenuhi asumsi identik atau variansi residual homogen serta independen (tidak terdapat korelasi antar residual). Lutkepohl (2005) menyatakan bahwa pengujian yang bisa digunakan adalah uji *portmanteau* dengan menggunakan signifikansi level 5%. Hipotesis yang digunakan adalah:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0$$

(residual memenuhi asumsi *white noise*)

$$H_1 : \text{minimal terdapat satu } \rho_k \neq 0$$

(residual tidak memenuhi asumsi *white noise*)

Dengan statistik uji yang digunakan sebagai berikut:

$$Q_h = n \sum_{j=1}^n \text{tr}(\hat{C}_j' \hat{C}_0^{-1} \hat{C}_j \hat{C}_0^{-1}) \quad (12)$$

Dengan :

$\hat{C}_j$: matriks autokovarians dari residual $\hat{a}_t$

$\hat{C}_0$: matriks $\hat{C}_j$ Ketika $j=0$

n : banyaknya data pengamatan

Kriteria Keputusan yang digunakan yaitu tolak H_0 jika $Q_h > \chi^2$ atau p-value $< \alpha$.

2. Berdistribusi Normal

Pemeriksaan residual berdistribusi normal dapat dilakukan dengan menggunakan *Jarque-Bera Normality Test* (Tsay, 2005). Uji *Jarque-Bera* dilakukan dengan menggunakan signifikansi level 5%. Hipotesis yang digunakan adalah:

H_0 : residual berdistribusi normal

H_1 : residual tidak berdistribusi normal

Rumus *Jarque-Bera* (Brocklebank dan Dickey, 2003) dinyatakan sebagai berikut:

$$JB = n \left(\frac{S^2}{6} + \frac{(K-3)^2}{24} \right) \quad (13)$$

Dimana,

$$S = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \right)^{3/2}} \quad (14)$$

$$K = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \right)^2} \quad (15)$$

Dengan :

S : kemiringan (*skewness*)

K : kurtosis

n : banyaknya data pengamatan

Kriteria Keputusan yang digunakan yaitu tolak H_0 jika $JB > \chi^2_{\alpha,2}$ atau $p\text{-value} \leq \alpha$.

PEMILIHAN MODEL TERBAIK

Model *Vector Autoregressive Integrated Moving Average* (VARIMA) yang terbaik dapat diketahui berdasarkan nilai RMSE (*Root Mean Square Error*). RMSE merupakan metode yang dapat digunakan untuk mengevaluasi ketepatan model *time series* dengan mempertimbangkan sisa perhitungan ramalan pada data *out-sample* (Wei, 2006). RMSE dapat dicari menggunakan rumus sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Z_t - \bar{Z}_t)^2} \quad (16)$$

Dengan,

n : banyaknya data testing

Z_t : data aktual periode ke- t

$\bar{Z}_t$: data peramalan periode ke- t

Selain RMSE, untuk mengukur kesalahan nilai dugaan model dapat menggunakan *Mean Absolute Deviation* (MAD). MAD adalah metode yang digunakan untuk mengukur tingkat akurasi dari suatu peramalan data selama periode tertentu tanpa memperhatikan apakah hasil peramalan lebih besar atau lebih kecil dibandingkan kenyataannya. Rumus MAD sebagai berikut:

$$MAD = \sum_{t=1}^n \left| \frac{Z_t - \bar{Z}_t}{n} \right| \quad (17)$$

Dengan,

n : banyaknya data testing

Z_t : data aktual periode ke- t

$\bar{Z}_t$: data peramalan periode ke- t

METODE PENELITIAN

JENIS DAN RANCANGAN PENELITIAN

Penelitian ini menggunakan jenis pendekatan deskriptif kuantitatif dengan studi literatur. Deskriptif kuantitatif dilakukan dengan mengumpulkan data dan menganalisis data sesuai dengan metode statistik untuk menguji hipotesis yang telah ditentukan. Sedangkan studi literatur dilakukan dengan mengumpulkan kajian pustaka yang relevan sebagai panduan dalam menyelesaikan penelitian.

VARIABEL PENELITIAN

Variabel dalam penelitian ini ada tiga, yaitu:

1. $Z_{(1,t)}$: PM_{2.5} ($\mu\text{g}/\text{m}^3$)
2. $Z_{(2,t)}$: NO₂ ($\mu\text{g}/\text{m}^3$)
3. $Z_{(3,t)}$: O₃ ($\mu\text{g}/\text{m}^3$)

TEKNIK PENGUMPULAN DATA

Data yang digunakan dalam penelitian ini adalah data sekunder berupa kualitas udara di Kabupaten Gresik yang diidentifikasi menggunakan tiga parameter Indeks Standar Pencemaran Udara (ISPU) yaitu PM_{2.5}, NO₂, dan O₃ yang diperoleh dari <https://plumelabs.com/>. Data yang digunakan adalah data harian periode November 2023 sampai dengan Februari 2024.

TEKNIK ANALISIS DATA

Adapun langkah analisis dalam penelitian ini berdasarkan tujuan penelitian adalah sebagai berikut:

1. Langkah analisis untuk mendapatkan model VARIMA terbaik.
 - Mendeskripsikan variabel-variabel penelitian dengan menggunakan statistika deskriptif. Tahap ini akan melihat nilai mean, standar deviasi, nilai maksimum, dan nilai minimum dari setiap variabel.

- Membagi data menjadi dua, yaitu data training dan data testing dengan perbandingan data 80:20.
 - Identifikasi data menggunakan plot *time series*.
 - Melakukan uji stasioneritas data. Stasioneritas terhadap varians menggunakan transformasi *box-cox* dan terhadap mean menggunakan uji ADF.
 - Identifikasi model VARIMA dengan menggunakan plot MACF dan MPACF.
 - Melakukan estimasi parameter dan uji signifikansi parameter pada model VARIMA menggunakan metode *Ordinary Least Square* (OLS).
 - Memeriksa diagnostik residual. Tahap ini akan mengecek residual dengan menggunakan uji *white noise* dan uji berdistribusi normal.
 - Memilih model terbaik. Pemilihan model terbaik dilakukan dengan membandingkan nilai RMSE dan nilai MAD yang terkecil.
2. Langkah analisis untuk mendapatkan hasil peramalan konsentrasi PM_{2.5}, NO₂, dan O₃.
- Peramalan Indeks Standar Pencemaran Udara (ISPU). Melakukan peramalan dengan menggunakan model VARIMA terbaik yang telah didapatkan dari tahap sebelumnya.

HASIL DAN PEMBAHASAN

Hasil analisis dan pembahasan akan dijelaskan pada bagian ini mengenai peramalan Indeks Standar Pencemaran Udara (ISPU) di Kabupaten Gresik dengan menggunakan metode VARIMA yang terdiri atas karakteristik data, analisis model terbaik, dan peramalan data.

KARAKTERISTIK DATA

Variabel yang digunakan dalam penelitian ini ada tiga, yaitu konsentrasi PM_{2.5}, NO₂, dan O₃. Data yang digunakan adalah data harian periode November 2023 sampai dengan Februari 2024. Penelitian ini menggunakan data training sebanyak 89 data, yaitu dari tanggal 01 November 2023 sampai dengan 29 Januari 2024 dan data testing sebanyak 22 data, yaitu dari tanggal 30 Januari 2024 sampai dengan 20

Februari 2024. Karakteristik konsentrasi PM_{2.5}, NO₂, dan O₃ di Kabupaten Gresik pada bulan November 2023 hingga Februari 2024 dengan menggunakan statistika deskriptif sebagai berikut.

Tabel 1. Statistika Deskriptif ISPU di Kabupaten Gresik

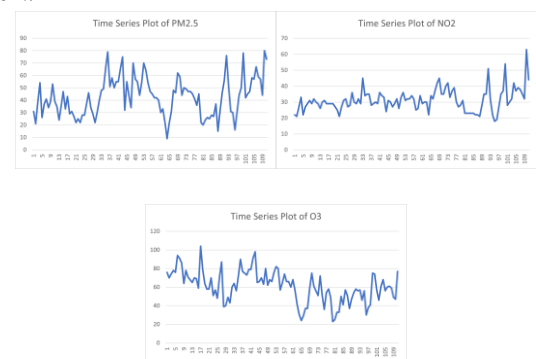
Variabel	Mean	S.Deviasi	Min	Max
PM _{2.5}	42,70	15,46	9	80
NO ₂	31,189	16,73	18	63
O ₃	60,87	16,73	23	104

Berdasarkan Tabel 1 diketahui bahwa konsentrasi O₃ memiliki rata-rata diatas konsentrasi PM_{2.5} dan NO₂ yaitu sebesar 60,87 µg/m³. Konsentrasi PM_{2.5} memiliki rata-rata sebesar 42,70 µg/m³. Sedangkan konsentrasi NO₂ memiliki rata-rata sebesar 31,189 µg/m³. Selain itu, untuk nilai terendah konsentrasi PM_{2.5} adalah 9 µg/m³ yang terjadi pada tanggal 04 Januari 2024. Sedangkan untuk nilai tertinggi konsentrasi PM_{2.5} adalah 80 µg/m³ yang terjadi pada tanggal 19 Februari 2024. Untuk nilai terendah konsentrasi NO₂ adalah 18 µg/m³ yang terjadi pada tanggal 04 Februari 2024. Sedangkan untuk nilai tertinggi konsentrasi NO₂ adalah 63 µg/m³ yang terjadi pada tanggal 19 Februari 2024. Untuk nilai terendah konsentrasi O₃ adalah 23 µg/m³ yang terjadi pada tanggal 20 Januari 2024. Sedangkan untuk nilai tertinggi konsentrasi O₃ adalah 104 µg/m³ yang terjadi pada tanggal 17 November 2023.

PEMODELAN VARIMA

1. Identifikasi Plot Time Series

Dalam melakukan analisis VARIMA, syarat pertama adalah data yang akan dianalisis harus stasioner. Oleh karena itu, perlu dilakukan identifikasi plot time series dengan tujuan untuk mengetahui data tersebut telah stasioner atau tidak.



Gambar 1. Plot Time Series

Berdasarkan Gambar diatas menunjukkan bahwa variabel PM_{2.5}, NO₂, dan O₃ belum

stasioner terhadap mean dan varians karena pada plot *time series* cenderung meningkat dan menurun secara acak. Maka, perlu dilakukan uji stasioner untuk memastikan lebih lanjut.

2. Uji Stasioneritas

1. Stasioneritas Terhadap Varians

Stasioner terhadap varians dapat dilihat berdasarkan nilai λ yang dihasilkan dari plot *box-cox*. Menurut Wei (2006) Jika nilai λ sama dengan 1 dapat disimpulkan bahwa data telah stasioner terhadap varians dan sebaliknya. Tabel 2 berikut menunjukkan hasil nilai λ dari ketiga variabel.

Tabel 2. Nilai λ Ketiga Variabel

Variabel	Nilai λ
PM _{2.5}	0,5
NO ₂	0
O ₃	1

Berdasarkan Tabel 2 didapatkan nilai λ untuk konsentrasi PM_{2.5} adalah 0,5 dan NO₂ adalah 0 sehingga kedua variabel tersebut belum stasioner terhadap varians, maka perlu dilakukan proses transformasi data pada data tersebut.

2. Stasioneritas Terhadap Mean

Selanjutnya stasioner terhadap mean dapat dilihat berdasarkan nilai p-value yang dihasilkan dari uji *augmented dickey fuller*. Jika nilai p-value $< \alpha$ dapat disimpulkan bahwa data telah stasioner terhadap mean dan sebaliknya. Tabel 3 berikut menunjukkan hasil uji ADF dari ketiga variabel.

Tabel 3. Hasil Uji ADF Ketiga Variabel

Variabel	Data Asli		First Differencing	
	T-Statistic	P-Value	T-Statistic	P-Value
PM _{2.5}	-2,3141	0,4469	-5,2358	0,01
NO ₂	-2,3282	0,4411	-4,9415	0,01
O ₃	-2,897	0,207	-4,801	0,01

Berdasarkan Tabel 3 nilai p-value dari ketiga variabel pada data asli tidak menunjukkan stasioner terhadap *mean*. Oleh karena itu, diperlukan proses differencing agar data menjadi stasioner terhadap *mean*. Pada *first differencing* nilai p-value dari ketiga variabel sebesar $0,01 < \alpha$ yang berarti bahwa data pada *first differencing* telah stasioner terhadap *mean*.

3. Identifikasi Model VARIMA

Tahap selanjutnya adalah identifikasi model VARIMA dengan menggunakan plot MACF dan MPACF. Berikut merupakan plot MACF dan MPACF dari data yang telah stasioner.

Schematic Representation of Cross Correlations											
Variable/Lag	0	1	2	3	4	5	6	7	8	9	10
PM25	+++	...	...	...	...	...	...	...	...	...	...
NO2	++	...	...	...	...	...	...	...	...	...	...
O3	++	...	...	...	...	...	...	...	...	...	...
+ is > 2*std error, - is < -2*std error, . is between											

Gambar 2. Plot MACF data *first differencing*

Schematic Representation of Partial Cross Correlations										
Variable/Lag	1	2	3	4	5	6	7	8	9	10
PM25	...	...	...	...	...	...	...	...	...	...
NO2	...	...	...	...	...	...	...	...	...	...
O3	...	...	...	...	...	...	...	...	...	...

Gambar 3. Plot MPACF data *first differencing*

Plot MACF dan MPACF diatas digunakan untuk menentukan orde MA(q) dan AR(p). Berdasarkan Gambar 2 terlihat bahwa korelasi yang paling signifikan dari ketiga variabel terjadi pada lag 0 dan 1 karena pada lag tersebut menunjukkan adanya simbol (+) atau (-). Sedangkan pada lag lainnya dianggap tidak signifikan karena hanya satu variabel yang menunjukkan signifikan. Sedangkan berdasarkan Gambar 3 terlihat bahwa korelasi yang paling signifikan dari ketiga variabel terjadi pada lag 1 dan 2. Oleh karena itu, dugaan model sementara yang terbentuk adalah VARIMA (1,1,0), VARIMA (1,1,1), VARIMA (2,1,0), dan VARIMA (2,1,1).

4. Estimasi Parameter

Selanjutnya akan dilakukan estimasi parameter pada dugaan model sementara menggunakan metode *Ordinary Least Square* (OLS) dan α yang digunakan adalah 5%. Hasil estimasi dan uji signifikansi model ditampilkan pada Tabel berikut.

Tabel 4. Uji Signifikansi Parameter

Model	Variabel	Parameter	Estimasi	P-Value
VARIMA (1,1,0)	PM _{2.5}	Φ_{11}^1	-0,35014	0,0248
		Φ_{12}^1	0,08859	0,9100
		Φ_{13}^1	0,00241	0,7989
	NO ₂	Φ_{21}^2	0,00570	0,8232
		Φ_{22}^2	-0,36569	0,0060
		Φ_{23}^2	0,00002	0,9912
	O ₃	Φ_{31}^3	-0,58480	0,7982
		Φ_{32}^3	-0,78017	0,9467
		Φ_{33}^3	-0,20356	0,1506
VARIMA (1,1,1)	PM _{2.5}	Φ_{11}^1	0,20104	0,4426
		Φ_{12}^1	1,01681	0,4649
		Φ_{13}^1	0,02077	0,1001

VARIMA (2,1,0)		θ_{11}^1	0,74981	0,0001
		θ_{12}^1	0,76051	0,0743
		θ_{13}^1	0,02424	0,0071
	NO ₂	Φ_{11}^2	0,04430	0,2613
		Φ_{12}^2	0,15705	0,4412
		θ_{11}^2	0,03299	0,2240
		θ_{12}^2	0,98209	0,0001
		Φ_{11}^3	-2,66631	0,4959
		Φ_{12}^3	14,12957	0,4793
	O ₃	Φ_{13}^3	0,52183	0,0072
		θ_{11}^3	-4,54918	0,0091
		θ_{12}^3	10,46243	0,1164
		θ_{13}^3	1,26063	0,0001
	PM _{2.5}	Φ_{11}^1	-0,43414	0,0092
		Φ_{12}^1	0,45533	0,5924
		Φ_{13}^1	0,00480	0,6212
		Φ_{21}^1	-0,29935	0,0771
		Φ_{22}^1	0,80281	0,3474
		Φ_{23}^1	0,00559	0,5640
VARIMA (2,1,1)	PM _{2.5}	Φ_{11}^2	0,01061	0,6840
		Φ_{12}^2	-0,48669	0,0006
		Φ_{23}^2	-0,00031	0,8418
		Φ_{21}^2	-0,00886	0,7407
		Φ_{22}^2	-0,30757	0,0260
		Φ_{23}^2	0,00001	0,9962
	NO ₂	Φ_{11}^3	-1,12190	0,6325
		Φ_{12}^3	9,43989	0,4403
		Φ_{13}^3	-0,22692	0,1063
		Φ_{21}^3	0,14777	0,9511
		Φ_{22}^3	-20,6979	0,0938
		Φ_{23}^3	-0,26511	0,0592
	O ₃	Φ_{11}^1	0,07843	0,8581
		Φ_{12}^1	-4,80376	0,1663
		Φ_{13}^1	0,01349	0,5340
		Φ_{21}^1	-0,00756	0,9778
		Φ_{22}^1	-1,12451	0,5140
		Φ_{23}^1	0,00631	0,6478

Berdasarkan Tabel 4 dapat disimpulkan bahwa tidak semua parameter dari model signifikan terhadap α . Meskipun demikian,

Kostenko dan Hyndman (2008) menyatakan bahwa meskipun variabel tidak signifikan, tetap memungkinkan untuk digunakan dalam melakukan peramalan. Oleh karena itu, dalam penelitian ini tetap menggunakan parameter model VARIMA yang tidak signifikan untuk dilakukan pengujian lebih lanjut.

5. Pemeriksaan Diagnostik Residual

1. Uji *White Noise*

Pemeriksaan residual *white noise* pada model dugaan VARIMA dilakukan dengan menggunakan uji *Portmanteau*. Hasil perhitungan statistik uji pada uji *Portmanteau* sebagai berikut:

Tabel 5. Uji *Portmanteau*

VARIMA (1,1,0)			VARIMA (1,1,1)		
Lag	Chi-Square	P-Value	Lag	Chi-Square	P-Value
6	83,62	0,0004	6	52,97	0,0338
7	102,49	0,0001	7	64,47	0,0399
8	111,80	0,0001	8	71,83	0,0527
9	127,79	0,0001	9	84,62	0,0360
10	133,93	0,0002	10	88,94	0,0855
VARIMA (2,1,0)			VARIMA (2,1,1)		
6	54,06	0,0270	6	39,75	0,0540
7	67,01	0,0183	7	51,44	0,0459
8	74,51	0,0336	8	63,11	0,0385
9	89,49	0,0158	9	71,27	0,0577
10	95,75	0,0322	10	75,92	0,1273

Dari Tabel 5 terlihat bahwa p-value yang $> \alpha$ hanya residual pada lag tertentu dari VARIMA (1,1,1) dan VARIMA (2,1,1) maka gagal menolak H_0 yang berarti bahwa hanya residual model VARIMA (1,1,1) dan VARIMA (2,1,1) yang memenuhi asumsi *white noise*.

2. Uji Berdistribusi Normal

Pemeriksaan residual berdistribusi normal pada model dugaan VARIMA dilakukan dengan menggunakan uji *Jarque Bera Normality*. Hasil perhitungan statistik uji pada *Jarque-Bera Normality* sebagai berikut:

Tabel 6. Uji *Jarque Berra Normality*

Model	Variabel	Uji Jarque Berra	
		Chi-Square	P-Value
VARIMA (1,1,0)	PM _{2.5}	1,80	0,4068
	NO ₂	0,06	0,9692
	O ₃	3,38	0,1843
VARIMA (1,1,1)	PM _{2.5}	1,63	0,4429
	NO ₂	1,20	0,5494
	O ₃	1,55	0,4610
VARIMA	PM _{2.5}	0,84	0,6556

(2,1,0)	NO ₂	1,37	0,5033
	O ₃	1,83	0,4000
VARIMA (2,1,1)	PM _{2.5}	6,09	0,0475
	NO ₂	0,07	0,9644
	O ₃	0,52	0,7692

Berdasarkan Tabel 6 semua model VARIMA memiliki nilai p-value yang $> \alpha$ kecuali konsentrasi PM_{2.5} pada VARIMA (2,1,1), sehingga gagal menolak H_0 yang artinya bahwa semua model VARIMA mempunyai residual yang berdistribusi normal.

6. Pemilihan Model VARIMA Terbaik

Selanjutnya adalah pemilihan model VARIMA terbaik. Berdasarkan hasil diagnostik residual diperoleh hanya model VARIMA (1,1,1) dan VARIMA (2,1,1) yang memenuhi kriteria. Pemilihan model VARIMA terbaik dilakukan dengan membandingkan nilai RMSE dan MAD yang terkecil dari data testing. Nilai RMSE dan MAD juga dapat mengukur keakuratan dari suatu peramalan. Hasil nilai RMSE dan MAD disajikan pada Tabel 7 sebagai berikut:

Tabel 7. Hasil Nilai RMSE dan MAD

Model	Variabel	RMSE	RMSE Total	MAD	MAD Total
VARIMA (1,1,1)	PM _{2.5}	35,28	26,56	31,39	23,27
	NO ₂	16,63		13,71	
	O ₃	27,77		24,70	
VARIMA (1,1,2)	PM _{2.5}	30,32	22,97	26,33	19,68
	NO ₂	15,62		12,78	
	O ₃	22,97		19,93	

Berdasarkan Tabel 7 model VARIMA yang memiliki nilai RMSE dan MAD terkecil adalah VARIMA (2,1,1) artinya model yang terpilih untuk meramalkan konsentrasi PM_{2.5}, NO₂, dan O₃ di Kabupaten Gresik adalah VARIMA (2,1,1). Secara sistematis model VARIMA (2,1,1) untuk Indeks Standar Pencemaran Udara (ISPU) di Kabupaten Gresik adalah sebagai berikut:

$$\begin{pmatrix} \Delta Z_{1,t} \\ \Delta Z_{2,t} \\ \Delta Z_{3,t} \end{pmatrix} = \begin{pmatrix} 0,078 & -4,80 & 0,013 \\ 0,027 & -0,59 & 0 \\ -2,59 & 28,67 & 0,474 \end{pmatrix} \begin{pmatrix} \Delta Z_{1,t-1} \\ \Delta Z_{2,t-1} \\ \Delta Z_{3,t-1} \end{pmatrix} + \begin{pmatrix} -0,007 & -1,124 & 0,006 \\ 0,0282 & -0,303 & 0 \\ -0,302 & 29,487 & -0,091 \end{pmatrix} \begin{pmatrix} \Delta Z_{1,t-2} \\ \Delta Z_{2,t-2} \\ \Delta Z_{3,t-2} \end{pmatrix} + \begin{pmatrix} a_{1,t} \\ a_{2,t} \\ a_{3,t} \end{pmatrix} - \begin{pmatrix} 0,5644 & -5,520 & 0,028 \\ 0,0143 & -0,097 & 0 \\ -4,964 & 31,729 & 1,240 \end{pmatrix} \begin{pmatrix} a_{1,t-1} \\ a_{2,t-1} \\ a_{3,t-1} \end{pmatrix}$$

PERAMALAN ISPU

Hasil Peramalan Indeks Standar Pencemaran Udara (ISPU) di Kabupaten Gresik menggunakan

model VARIMA (2,1,1) pada 30 hari selanjutnya, yaitu pada tanggal 21 Februari 2024 sampai dengan 21 Maret 2024 disajikan pada Tabel 8 sebagai berikut.

Tabel 8. Ramalan ISPU di Kabupaten Gresik

Tanggal	PM _{2.5}	NO ₂	O ₃
21/02/2024	24,39807	23,63101	30,92112
22/02/2024	24,22274	23,57578	30,43562
23/02/2024	24,04794	23,52068	29,95007
24/02/2024	23,87378	23,46594	29,46453
25/02/2024	23,70025	23,41109	28,97902
26/02/2024	23,52735	23,35638	28,49351
27/02/2024	23,35509	23,30179	28,00798
28/02/2024	23,35509	23,24732	27,52246
29/02/2024	23,18355	23,19322	27,03694
01/03/2024	23,01255	23,13901	26,55142
02/03/2024	22,84219	23,08493	26,06589
03/03/2024	22,67245	23,03097	25,58037
04/03/2024	22,50335	22,97737	25,09485
05/03/2024	22,33489	22,92367	24,60933
06/03/2024	22,16705	22,87009	24,12381
07/03/2024	21,99985	22,81664	23,63829
08/03/2024	21,83328	22,76331	23,15276
09/03/2024	21,66735	22,71033	22,66724
10/03/2024	21,50214	22,65725	22,18172
11/03/2024	21,33747	22,6043	21,6962
12/03/2024	21,17343	22,55146	21,21068
13/03/2024	20,84726	22,49898	20,72516
14/03/2024	20,68512	22,44639	20,23963
15/03/2024	20,52362	22,39393	19,75411
16/03/2024	20,36275	22,34159	19,26859
17/03/2024	20,20251	22,28937	18,78307
18/03/2024	20,0429	22,2375	18,29755
19/03/2024	19,88402	22,18552	17,81203
20/03/2024	19,72568	22,13367	17,3265
21/03/2024	19,56797	22,08194	16,84098

Tabel 8 diatas menunjukkan hasil peramalan Indeks Standar Pencemaran Udara (ISPU) dengan konsentrasi PM_{2.5}, NO₂, dan O₃ di Kabupaten Gresik. Untuk nilai tertinggi konsentrasi PM_{2.5} sebesar 24,39807 µg/m³, NO₂ sebesar 23,63101 µg/m³ dan O₃ sebesar 30,92112 µg/m³ yang terjadi pada tanggal 21 Februari 2024.

PENUTUP

SIMPULAN

Berdasarkan hasil analisis yang telah dilakukan, dapat disimpulkan sebagai berikut:

1. Model VARIMA terbaik pada data Indeks

Standar Pencemaran Udara (ISPU) di Kabupaten Gresik dengan tiga variabel yaitu konsentrasi $PM_{2.5}$, NO_2 , dan O_3 adalah VARIMA (2,1,1) dengan variabel $PM_{2.5}$ dan O_3 pada saat ke-1 dipengaruhi oleh ketiga variabel pada saat ke-1, ke-2, dan ke-3 serta dipengaruhi oleh nilai residual ketiga variabel pada saat ke-1. Sedangkan variabel NO_2 pada saat ke-1 hanya dipengaruhi oleh variabel $PM_{2.5}$ dan O_3 pada saat ke-1, ke-2, dan ke-3 serta dipengaruhi oleh nilai residual ketiga variabel pada saat ke-1.

2. Hasil peramalan kualitas udara di Kabupaten Gresik pada tanggal 21 Februari 2024 sampai dengan 21 Maret 2024 menunjukkan penurunan dan berdasarkan kategori ISPU berada pada kategori baik, dimana tingkat kualitas udara masih dapat diterima pada kesehatan makhluk hidup. Hasil prediksi selengkapnya disajikan pada Tabel 8.

SARAN

Saran dari hasil penelitian ini adalah diharapkan peneliti selanjutnya dapat melakukan analisis model VARIMA dengan perluasan seperti VARIMAX dengan menggunakan variabel prediktor, VARIMA ARCH/GARCH dengan tujuan memodelkan error yang dihasilkan oleh model.

DAFTAR PUSTAKA

- Abidin, J., dan Hasibuan, F. A., 2019. Pengaruh dampak pencemaran udara terhadap kesehatan untuk menambah pemahaman masyarakat awam tentang bahaya dari polusi udara. *Prosiding Snfur*, 4(2), 3.
- Anggraeni, W., Vinarti, R. A., dan Kurniawati, Y. D., 2015. Performance comparisons between arima and arimax method in moslem kids clothes demand forecasting: Case study. *Procedia Computer Science*, 72, 630–637.
- Brocklebank, J. C., dan Dickey, D. A., 2003. *SAS for forecasting time series*. John Wiley & Sons.
- Febrianti, D. R., Tiro, M. A., dan Sudarmin, S., 2021. Metode vector autoregressive (var) dalam menganalisis pengaruh kurs mata uang terhadap ekspor dan impor di indonesia. *VARIANSI: Journal of Statistics and Its application on Teaching and Research*, 3(1), 23.
- Khikmah, N., 2021. Peramalan curah hujan kabupaten kudus dengan metode var (vektor autoregressive).
- Kostenko, A. V., dan Hyndman, R. J., 2008. *Forecasting without significance tests?* manuscript, Monash University, Australia.
- Lutkepohl, H., 2005. *New introduction to multiple time series analysis*. Springer Science & Business Media.
- Makridakis, S., Wheelwright, S. C., dan McGee, V. E., 1999. *Metode dan aplikasi peramalan*. Jakarta: Erlangga.
- Nugroho, A. A.-Z., , 2022. Pemodelan multivariate time series dengan vector autoregressive integrated moving average (varima). *Jurnal Riset Statistik*
- Sudjana, N., 2005. *Metode statistika edisi keenam*. Bandung: PT. Tarsito.
- Tsay, R. S., 2005. *Analysis of financial time series*. John wiley & sons.
- Wei, W., 2006. *Time Series Analysis: Univariate and Multivariate Methods*, 2nd edition, 2006.
- Wulandary, S., 2020. Metode vector autoregressive integrated (vari) untuk peramalan jumlah wisatawan mancanegara di batam dan jakarta. *JMSK*, 17(1), 94–108.