

Peramalan Kualitas Udara di Kota Surabaya Menggunakan VAR, VARX dan LSTM

Nadya Yuniar Desi Prameswari¹, Adatul Mukarromah²

Departemen Statistika, Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

e-mail: yuniarnadya4@gmail.com⁽¹⁾, adatul@statistika.its.ac.id⁽²⁾

Abstrak— Kota Surabaya merupakan pusat industri dan perdagangan di Jawa Timur. Sebagai ibukota dari Provinsi Jawa Timur dan kota metropolitan, Kota Surabaya mengalami peningkatan jumlah penduduk dan volume kendaraan dari tahun ke tahun. Semakin padatnya Kota Surabaya mengakibatkan munculnya berbagai permasalahan, salah satunya ialah yang berkaitan dengan pencemaran udara. Pencemaran udara yang terjadi dapat mengakibatkan munculnya berbagai permasalahan, salah satunya yang berkaitan dengan kesehatan. Berdasarkan masalah yang ditimbulkan oleh pencemaran udara, maka kualitas dari udara perlu diperhatikan. Kualitas dari udara dapat dilihat melalui Indeks Standar Pencemar Udara (ISPU) yang didapatkan melalui AQMS yang berada di stasiun SUF dengan beberapa indikator yang digunakan sebagai penentu kualitas udara yaitu CO, PM₁₀, dan NO₂. Tindakan preventif yang dapat dilakukan untuk monitoring kualitas udara ialah dengan melakukan peramalan terhadap kadar indikator penentu kualitas udara. Sehingga penelitian ini bertujuan untuk meramalkan CO, PM₁₀, dan NO₂ di stasiun SUF 7 Kota Surabaya menggunakan VAR, VARX dan LSTM. Didapati hasil bahwa ketiga indikator memiliki pola seasonal harian per setengah jam dan jumlah kadar ketiga indikator cenderung rendah saat sore hingga malam hari dan tinggi saat pagi hingga siang hari. Pada Model VAR yang terbentuk, asumsi residual white noise dan berdistribusi multivariate normal tidak terpenuhi, hal ini dikarenakan terdapat outlier pada data yang digunakan. Selain itu, hasil penelitian menunjukkan bahwa pada data train metode VARX menghasilkan model terbaik untuk indikator CO, PM₁₀, dan NO₂.

Kata Kunci— Kualitas Udara, LSTM, Peramalan, VAR, VARX

I. PENDAHULUAN

Kota Surabaya merupakan ibukota dari Provinsi Jawa Timur dan merupakan pusat industri dan perdagangan di Jawa Timur. Sebagai kota metropolitan, volume lalu lintas di Kota Surabaya menjadi tinggi. Hal ini dibuktikan dengan jumlah kendaraan di Kota Surabaya dari tahun ke tahun selalu mengalami peningkatan cukup tajam [1]. Peningkatan penggunaan kendaraan bermotor di Kota Surabaya disebabkan oleh pertumbuhan kota dan penduduknya. Semakin padatnya Kota Surabaya mengakibatkan munculnya berbagai permasalahan, salah satunya ialah yang berkaitan dengan pencemaran udara.

Salah satu penyebab utama terjadinya pencemaran udara di Kota Surabaya adalah sektor transportasi, di mana sektor transportasi menghasilkan 60 persen karbon monoksida (CO) [2], meningkatkan jumlah partikulat matter (PM₁₀) [3] dan NO₂ di udara yang dapat memberikan efek negatif pada kesehatan manusia [4]. Polutan penyebab polusi udara diperkirakan telah

menyebabkan 350 kematian di Kota Surabaya pada tahun 2021 [5]. Berdasarkan hal ini dapat diketahui bahwa tingkat pencemaran udara yang terjadi di Kota Surabaya sudah mencapai titik yang mengkhawatirkan dan menjadi permasalahan yang harus segera diselesaikan.

Kualitas dari udara juga secara langsung maupun tidak langsung berdampak bagi kehidupan makhluk hidup. Sehingga kontrol pada kualitas udara perlu dilakukan, salah satunya dengan mempersiapkan tindakan preventif yaitu dengan melakukan peramalan polutan-polutan penentu kualitas udara. Peramalan mengenai kualitas udara telah banyak dilakukan, beberapa diantaranya yaitu ARIMA [6] [7], DSARIMA [8] [9], FFNN [7], MGSTAR [6] [9], dan LSTM [7]. Beberapa penelitian ini telah meramalkan polutan udara secara *univariate*, padahal polutan-polutan pencemar udara yang ada dapat membentuk suatu campuran yang memungkinkan keterkaitan antar polutan-polutan pencemar udara, sehingga diperlukan metode pemodelan yang dapat menangkap efek perbedaan parameter pencemar udara. Pada penelitian ini, diduga terdapat korelasi antar polutan-polutan udara penentu kualitas udara, sehingga peramalan polutan udara di stasiun SUF 7 akan dilakukan secara *multivariate* dan metode yang akan digunakan ialah *Vector Autoregressive Moving Average* (VAR), *Vector Autoregressive with Exogenous Variables* (VARX) dan *Long Short-Term Memory* (LSTM). Dipilihnya lokasi ini karena stasiun pemantau tetap (SUF) 7 terletak di wilayah Kebonsari yang merupakan wilayah dengan beberapa sekolah, pemukiman serta perkantoran/perusahaan yang padat. Selain itu, juga terdapat beberapa jalan raya utama serta dilalui oleh jalan tol Surabaya-Porong, sehingga wilayah ini hampir selalu ramai oleh aktivitas lalu lintas. Metode VAR digunakan karena model ini dapat menjelaskan hubungan linier antara pengamatan dengan variabel yang berbeda, yang dapat dilihat melalui *Cross-Correlation Function* atau CCF [10], data *time series* kerap dipengaruhi oleh suatu kejadian yang menyebabkan asumsi distribusi multivariat normal tidak terpenuhi sehingga digunakan metode VARX serta kelebihan dari LSTM yakni dapat mengatasi *vanishing gradient* pada RNN [11].

Selanjutnya model terbaik dipilih berdasarkan *Root Mean Square Error* (RMSE) yang kemudian digunakan untuk meramalkan polutan udara penentu ISPU yang digunakan pada stasiun SUF 7, sehingga dapat memberikan informasi mengenai nilai polutan udara penentu ISPU serta memberi bahan pertimbangan dalam menanggulangi masalah pencemaran udara kepada Dinas Lingkungan Hidup Kota Surabaya atau Pemerintah Kota Surabaya.

II. TINJAUAN PUSTAKA

A. Expectation Maximization (EM)

Expectation Maximization merupakan metode imputasi yang didasarkan pada kuadrat terkecil dan estimasi likelihood maksimum [12]. Maksimum likelihood menggunakan seluruh data yang tersedia, baik yang lengkap maupun yang tidak untuk mendapatkan nilai estimasi parameter yang mempunyai peluang tertinggi dalam menghasilkan data sampel. EM terdiri dari *iterative* perhitungan, yaitu prediksi dan estimasi.

B. Analisis Korelasi

Analisis korelasi merupakan metode yang digunakan untuk mengetahui ada atau tidaknya hubungan antara satu variabel dengan variabel lainnya. Metode ini memiliki nilai yang bervariasi mulai dari -1 hingga 1, dimana -1 memiliki arti korelasi minus total, 0 menandakan tidak ada korelasi dan 1 sebagai korelasi positif total [13].

$$r_{Z_1 Z_2} = \frac{\sum_{i=1}^n (Z_1 - \bar{Z}_1)(Z_2 - \bar{Z}_2)}{\sqrt{\sum_{i=1}^n (Z_1 - \bar{Z}_1)^2} \sqrt{\sum_{i=1}^n (Z_2 - \bar{Z}_2)^2}} \quad (1)$$

dengan

Z_1 : variabel ke-1

Z_2 : variabel ke-2

n : banyaknya data

r : koefisien korelasi dengan ketentuan $-1 < r < 1$

Hipotesis pada pengujian ini adalah sebagai berikut.

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

Statistik uji yang digunakan adalah sebagai berikut.

$$t = r \sqrt{\frac{n-2}{1-r^2}} \quad (2)$$

H_0 akan ditolak jika $t_{hitung} > t_{\frac{\alpha}{2}, n-2}$ atau jika $t_{hitung} < -t_{\frac{\alpha}{2}, n-2}$.

C. Vector Autoregressive (VAR)

Model *Vector Autoregressive* (VAR) merupakan perluasan dari gabungan beberapa model AR, dimana model-model ini membentuk sebuah vektor yang antar variabelnya saling mempengaruhi. Bentuk umum model AR (p) dapat dituliskan seperti pada persamaan berikut [14].

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t \quad (3)$$

dengan,

$Z_t = Z_t - \mu$,

a_t : residual pada waktu ke- t ,

p : orde AR,

$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$

$\phi_1, \phi_2, \dots, \phi_p$: koefisien AR orde p .

Sebuah *multivariate time series* dapat dikatakan sebagai model VAR proses dari orde 1 jika mengikuti persamaan berikut.

$$\dot{Z}_t = \Phi_0 + \Phi \dot{Z}_{t-1} + a_t \quad (4)$$

dengan

Φ_0 : vektor konstanta berukuran $m \times 1$

\dot{Z}_t : $Z_t - \mu$

μ : $E(Z_t)$

\dot{Z}_t : vektor $m \times 1$ dari variabel pada waktu ke- t

\dot{Z}_{t-1} : vektor $m \times 1$ dari variabel pada waktu ke- $(t-1)$

a_t : vektor $m \times 1$ dari residual pada waktu ke- t

1. Stasioneritas Data

Data-data yang tidak stasioner sering dijumpai dalam analisis *time series*. Mengatasi ketidakstasioneritas dalam *varians* dapat menggunakan transformasi Box-Cox. Rumus transformasi Box-Cox dapat dilihat pada Persamaan 5.

$$T(Z_t) = \begin{cases} \frac{Z_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \lim_{\lambda \rightarrow 0} \frac{Z_t^\lambda - 1}{\lambda} = \ln Z_t, & \lambda = 0 \end{cases} \quad (5)$$

dengan λ merupakan parameter transformasi Box-Cox. Salah satu uji yang dapat digunakan untuk menguji stasioneritas dalam *mean* dapat dilakukan dengan uji *Augmented Dickey Fuller*. Uji ini didasarkan pada persamaan model berikut.

$$Z_t = \phi Z_{t-1} + a_t \quad (6)$$

dengan a_t adalah residual yang bersifat *white noise*. Pengujian yang dilakukan adalah sebagai berikut.

$H_0 : \phi = 1$

$H_1 : \phi < 1$

Statistik uji yang digunakan sebagai berikut.

$$DF = t \text{ ratio} = \frac{\hat{\phi} - 1}{std(\hat{\phi})} = \frac{\sum_{t=1}^n Z_{t-1} + a_t}{\hat{\sigma} \sqrt{\sum_{t=1}^n Z_{t-1}^2}} \quad (7)$$

pada taraf signifikansi 5%, $\hat{\phi}$ dan $\hat{\sigma}$ diperoleh dari hasil estimasi metode *least square*.

2. Matrix Cross Correlation Function (MCCF)

Sebagai contoh jika terdapat sebuah vektor *time series* dengan observasi sebanyak n , yaitu Z_1, Z_2, \dots, Z_n , maka persamaan matriks korelasi sampelnya adalah sebagai berikut.

$$\hat{\rho}(k) = [\hat{\rho}_{ij}(k)] \quad (8)$$

dengan $\hat{\rho}_{ij}(k)$ merupakan korelasi silang sampel untuk komponen *series* ke- i dan ke- j .

3. Matrix Partial Cross Correlation Function (MPCCF)

Persamaan untuk matriks autokorelasi *lag* parsial pada *lag s* yang didapatkan oleh Heyse dan Wei ialah sebagai berikut [14].

$$P(s) = [D_v(s)]^{-1} V_{vu}(s) [D_u(s)]^{-1} \quad (9)$$

dengan

$D_v(s)$: matriks diagonal dimana elemen ke- i merupakan akar dari elemen diagonal ke- i dari $V_v(s)$

$D_u(s)$: matriks diagonal dimana elemen ke- i merupakan akar dari elemen diagonal ke- i dari $V_u(s)$.

4. Identifikasi Model VAR

Identifikasi model VAR dapat dilakukan dengan melihat plot *time series* dan plot MCCF, MPCCF untuk menentukan orde p dan q pada model. Selain itu, menentukan orde VAR juga dapat dilakukan melalui *Akaike's Information Criterion* (AIC) yang paling minimum. Rumus AIC adalah sebagai berikut.

$$AIC_{(p+q)} = \ln |\hat{\Sigma}_{(p+q)}| + \frac{2m^2(p+q)}{n} \quad (10)$$

dengan

$\hat{\Sigma}_{(p+q)}$: estimasi dari matriks *varians* kovarians

p : orde AR

q : orde MA
 $2m^2(p + q)$: jumlah parameter dari AR dan MA
 n : jumlah data

5. Estimasi Parameter

Setelah model sementara telah terbentuk, selanjutnya dilakukan estimasi parameter model VARMA dengan menggunakan LS (*Least Square*). Sebagai contoh terdapat model VARMA sebagai berikut.

$$Y = X\beta + U \quad (11)$$

dengan

$$\begin{aligned}
 Y &= (Z_1, Z_2, \dots, Z_n) \\
 B &= [\Phi : \Theta] \\
 U &= (u_1, u_2, \dots, u_n) \\
 y &= \text{Vec}(Y) \\
 \beta &= \text{Vec}(B) \\
 u &= \text{Vec}(U)
 \end{aligned}$$

6. Uji Signifikansi Parameter

Uji signifikansi parameter dilakukan untuk mengetahui parameter yang signifikan terhadap model yang dilakukan dengan menggunakan uji t. Hipotesis pengujian yang digunakan adalah sebagai berikut.

$$H_0 : \phi_{ijk} = 0$$

$$H_1 : \phi_{ijk} \neq 0$$

Statistik uji yang digunakan dituliskan pada Persamaan 12.

$$t = \frac{\hat{\phi}_{ijk}}{SE(\hat{\phi}_{ijk})} \quad (12)$$

H_0 akan ditolak jika $|t_{hitung}| > t_{\frac{\alpha}{2}, n-p-1}$ atau jika nilai nilai p -value bernilai kurang dari α , dimana p menunjukkan jumlah parameter yang berarti bahwa parameter telah signifikan.

7. Pemeriksaan Diagnosa

Model VAR dengan parameter yang signifikan harus memenuhi asumsi *white noise* dan berdistribusi *multivariate normal*. Hipotesis yang akan diuji untuk pengujian asumsi *white noise* menggunakan uji Portmanteau adalah sebagai berikut.

$$H_0 : \text{vektor residual model memenuhi asumsi white noise}$$

$$H_1 : \text{vektor residual model tidak memenuhi asumsi white noise}$$

Statistik uji yang digunakan dituliskan pada Persamaan 13.

$$Q_h = n \sum_{i=1}^h \text{tr}(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1}) \quad (13)$$

\hat{C}_i diperoleh dari $\hat{C}_i = n^{-1} \sum_{t=i+1}^n \hat{a}_t \hat{a}_{t-i}'$.

H_0 akan ditolak jika $Q_h > X^2$ atau jika nilai p -value bernilai kurang dari α [15]. Selanjutnya pemeriksaan apakah residual telah memenuhi asumsi berdistribusi *multivariate normal* yang dilakukan dengan melakukan uji hipotesis Shapiro-Wilk. Hipotesis yang akan diuji adalah sebagai berikut.

$$H_0 : \text{vektor residual model memenuhi asumsi distribusi multivariate normal}$$

$$H_1 : \text{vektor residual model tidak memenuhi asumsi distribusi multivariate normal}$$

Statistik uji yang digunakan dituliskan pada Persamaan 14.

$$W = \frac{(\sum_{i=1}^n a_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})} \quad (14)$$

dengan

$$\begin{aligned}
 Y &= (Z_1, Z_2, \dots, Z_n) \\
 B &= [\Phi : \Theta]
 \end{aligned}$$

$$\begin{aligned}
 U &= (u_1, u_2, \dots, u_n) \\
 y &= \text{Vec}(Y)
 \end{aligned}$$

Gagal tolak H_0 jika p -value bernilai lebih dari α [16].

8. Kriteria Kebaikan Model

Salah satu kriteria nilai *error* yang terkecil pada data *training* dan data *testing* yang dapat digunakan untuk pemilihan model terbaik ialah *Root Mean Square Error* (RMSE). Rumus untuk menghitung RMSE adalah sebagai berikut.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Z_{i,t} - \hat{Z}_{i,t})^2} \quad (15)$$

dengan $Z_{i,t}$ menyatakan *real value* dan $\hat{Z}_{i,t}$ menyatakan *forecast*.

D. Diagram Kontrol Residual

Kejadian eksternal sering kali mempengaruhi data *time series*. Apabila penyebab dari hal tersebut tidak diketahui, maka disebut sebagai *outlier*. Akan tetapi, apabila penyebab dari kejadian tersebut diketahui maka disebut sebagai intervensi. Diagram kontrol residual merupakan salah satu cara yang dapat digunakan untuk mendeteksi keberadaan dari *outlier* yang dilakukan dengan cara melakukan kontrol pada proses target dan klasifikasi pengamatan individu menggunakan T^2 Hotelling [17].

$$J_{i,h} = \omega_{i,h}' \Sigma_{i,h}^{-1} \omega_{i,h} \quad (16)$$

dengan $\omega_{i,h}$ merupakan residual dari setiap pengamatan dalam vektor kolom.

E. Vector Autoregressive with Exogenous Variable (VARX)

Metode VARX merupakan pengembangan dari model VAR dengan menambah variabel eksogen di sebelah kanan persamaan. Variabel X sebagai variabel eksogen masuk ke dalam model VAR sehingga model dari VARX (p,s) ialah sebagai berikut [18].

$$\Phi_p(B)Z_t = \gamma_r(B)X_t \quad (17)$$

dengan

$$\Phi_p(B) = I_k - \Phi_1 B - \dots - \Phi_p B^p$$

$$\gamma_r(B) = \gamma_0 + \gamma_1 B - \dots - \gamma_r B^r$$

Φ_p merupakan matriks berukuran $(m \times m)$, dan γ_r merupakan matriks yang berukuran $(m \times r)$.

Dengan pemodelan VARX $Z_t = X\beta + D\delta + \varepsilon$, dan estimasi parameter yang diperoleh untuk δ dan β adalah sebagai berikut [19].

$$\hat{\delta} = D^{-1}(1 - M)(Z - X[X^T \Sigma^{-1} M X]^{-1} [X^T \Sigma^{-1} M Z])$$

$$\hat{\beta} = [X^T \Sigma^{-1} M X]^{-1} [X^T \Sigma^{-1} M Z]$$

dimana $\hat{\beta}$ merupakan estimator untuk model VARX, $\hat{\delta}$ adalah estimator untuk *exogenous variables*, serta $M = [I - D(D^T \Sigma^{-1} D)^{-1} D^T \Sigma^{-1}]$, dengan memenuhi syarat varian minimum, berdistribusi normal serta tidak bias.

F. Long Short Term Memory (LSTM)

Long short-term memory merupakan metode pengembangan dari *recurrent neural network* (RNN) dan merupakan salah satu metode yang terdapat dalam *neural network* [20]. Pengembangan pada metode RNN dilakukan karena terdapat kelemahan pada metode RNN, yaitu dalam sebuah arsitektur kelambatan jangka panjang tidak dapat diakses. Selain itu, *vanishing gradient* pada RNN juga dapat diatasi menggunakan

LSTM [11]. Terdapat beberapa *layer* penyusun struktur LSTM, yaitu *input layer*, *recurrent hidden layer* dan *output layer* [21]. Nilai *output* dari LSTM dengan persamaan sebagai berikut.

$$\hat{Z}_{i,t} = W_y h_{i,t} + b_y, \quad (18)$$

dengan

$\hat{Z}_{i,t}$: nilai dugaan dari variabel *output*

W_y : bobot dari *hidden state*

b_y : bobot dari *output*.

G. Pencemaran Udara

Pencemaran udara adalah pencemaran yang disebabkan oleh aktivitas manusia seperti pencemaran dari pabrik, kendaraan bermotor, pembakaran sampah, sisa pertanian, dan peristiwa alam seperti kebakaran hutan, letusan gunung berapi yang mengeluarkan debu, gas, dan awan panas [22]. Polutan udara yang dipakai untuk menghitung Indeks Standar Pencemar Udara (ISPU) adalah PM₁₀, CO, SO₂, NO₂ dan O₃.

III. METODE PENELITIAN

A. Sumber Data

Data yang digunakan dalam penelitian adalah data sekunder yang merupakan data hasil pemantauan kualitas udara di stasiun SUF 7 di Kota Surabaya per setengah jam yang dilakukan oleh Dinas Lingkungan Hidup Kota Surabaya. Data bulan September 2020 digunakan sebagai data *train* dan data minggu pertama bulan Oktober 2020 digunakan sebagai data *test*.

B. Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini adalah sebagai berikut.

Tabel 1. Variabel Penelitian

Simbol	Variabel	Keterangan
Z ₁	CO	Karbon Monoksida
Z ₂	PM ₁₀	Particulate Matter
Z ₃	NO ₂	Nitrogen Dioksida

C. Struktur Data

Adapun struktur data dari variabel yang digunakan adalah sebagai berikut.

Tabel 2. Struktur Data

t	Jam	Bulan	Z _{1,t}	Z _{2,t}	Z _{3,t}
1	0:30	September	Z _{1,1}	Z _{2,1}	Z _{3,1}
2	1:00	September	Z _{1,2}	Z _{2,2}	Z _{3,2}
⋮	⋮	⋮	⋮	⋮	⋮
1776	00:00	Oktober	Z _{1,1776}	Z _{2,1776}	Z _{3,1776}

Indeks yang digunakan dalam penelitian ini adalah Z_{i,t} dengan:

i : indeks dari jenis indikator penentu kualitas udara ISPU dengan $i = 1, 2, 3$. 1 merupakan kadar CO, 2 merupakan kadar PM₁₀ dan 3 merupakan kadar NO₂,

t : indeks dari urutan waktu kejadian dengan $t = 1, 2, \dots, 1776$.

D. Langkah Analisis

Adapun langkah-langkah analisis untuk mendapatkan hasil peramalan indikator penentu kualitas udara adalah sebagai berikut.

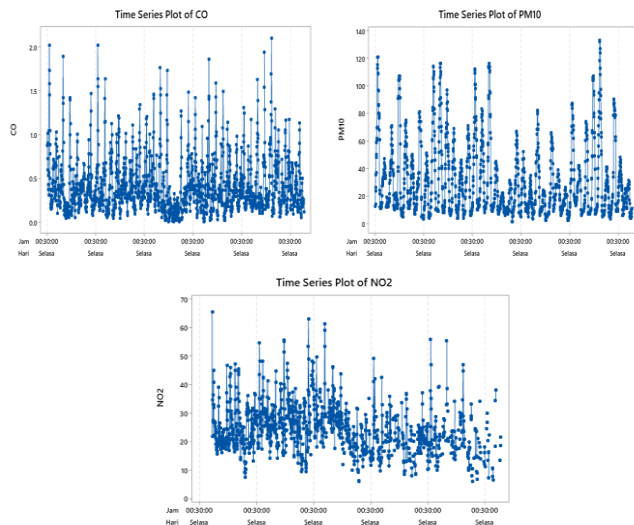
1. Melakukan *preprocessing* data, jika terdapat *missing value* maka akan dilakukan imputasi pada *missing value* dengan menggunakan *Expectation Maximization*.
2. Mendeskripsikan karakteristik data dari kualitas udara pada indikator CO, PM₁₀ dan NO₂ di stasiun SUF 7 di Kota Surabaya.
3. Melakukan uji korelasi.
4. Membagi data menjadi data *train* dan data *test*.
5. Melakukan pemodelan dan peramalan pada indikator CO, PM₁₀ dan NO₂ di stasiun SUF 7 di Kota Surabaya dengan menggunakan metode *Vector Autoregressive* (VAR) dengan langkah-langkah sebagai berikut.
 - (i) Melakukan identifikasi stasioneritas dalam *mean* dan *varians* pada ketiga indikator kualitas udara yang digunakan.
 - (ii) Melakukan identifikasi orde berdasarkan plot MCCF dan MPCCF.
 - (iii) Melakukan estimasi parameter model dan uji signifikansi parameter model.
 - (iv) Melakukan pemeriksaan diagnosis asumsi *white noise* dan berdistribusi normal *multivariate* pada residual data yang digunakan.
 - (v) Melakukan peramalan kualitas udara parameter PM₁₀, CO dan NO₂ di stasiun SUF 7 di Kota Surabaya dengan menggunakan metode VAR.
 - (vi) Pemilihan model VAR terbaik.
6. Melakukan pemodelan dan peramalan pada indikator CO, PM₁₀ dan NO₂ di stasiun SUF 7 di Kota Surabaya dengan menggunakan metode *Vector Autoregressive with Exogenous Variables* (VARX) dengan langkah-langkah sebagai berikut.
 - (i) Mendeteksi *outlier* menggunakan diagram kontrol residual.
 - (ii) Membuat variabel *dummy* sebanyak *outlier* yang ditemukan pada deteksi *outlier*.
 - (iii) Melakukan identifikasi orde berdasarkan plot MCCF dan MPCCF.
 - (iv) Melakukan estimasi parameter model dan uji signifikansi parameter model.
 - (v) Melakukan pemeriksaan diagnosis asumsi *white noise* dan berdistribusi normal *multivariate* pada residual data yang digunakan.
 - (vi) Melakukan peramalan kualitas udara parameter PM₁₀, CO dan NO₂ di stasiun SUF 7 di Kota Surabaya dengan menggunakan metode VARX.
 - (vii) Pemilihan model VARX terbaik
7. Melakukan pemodelan dan peramalan pada kualitas udara parameter PM₁₀, CO dan NO₂ di stasiun SUF 7 di Kota Surabaya dengan menggunakan metode *Long Short-Term Memory* (LSTM) dengan langkah-langkah sebagai berikut.
 - (i) Menentukan jumlah *time steps* yang akan digunakan sebagai variabel *input*.
 - (ii) Melakukan *preprocessing* dengan normalisasi data dengan persamaan $Z_{i,t}^* = \frac{Z_{i,t} - \min(Z_{i,t})}{\max(Z_{i,t}) - \min(Z_{i,t})}$
 - (iii) Mengubah variabel *input* menjadi bentuk tiga dimensi dengan bentuk jumlah sampel, *time steps*, dan *feature*.
 - (iv) Menentukan jumlah *neuron* dan *hidden layer* pada LSTM. Jumlah *neuron* yang digunakan adalah 1 hingga 5 dan jumlah *hidden layer* yang digunakan ialah 1.

- (v) Melakukan peramalan kualitas udara parameter PM₁₀, CO dan NO₂ di stasiun SUF 7 di Kota Surabaya dengan menggunakan metode LSTM.
- (vi) Melakukan *postprocessing* data hasil ramalan.
- (vii) Pemilihan model LSTM terbaik.
8. Melakukan peramalan pada kualitas udara parameter PM₁₀, CO dan NO₂ di stasiun SUF 7 di Kota Surabaya berdasarkan model terbaik.
9. Menarik kesimpulan dan saran.

IV. ANALISIS DAN PEMBAHASAN

A. Karakteristik Indikator Penentu Kualitas Udara di Stasiun SUF 7 Kota Surabaya

Karakteristik indikator CO, PM₁₀ dan NO₂ yang diukur menggunakan AQMS pada Stasiun SUF 7 yang terletak di Kebonsari dapat dilihat sebagai berikut.



Gambar 1. Time Series Plot CO, PM₁₀ dan NO₂ di Stasiun SUF 7

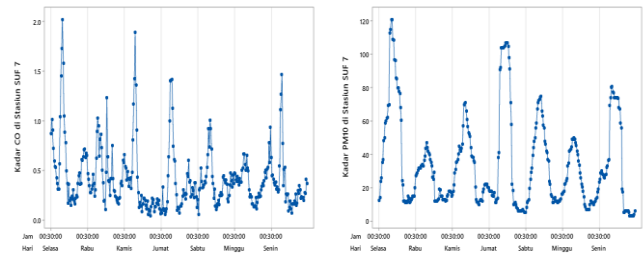
Gambar 1 menunjukkan plot kadar CO, PM₁₀ dan NO₂ di Stasiun SUF 7 Kota Surabaya yang dicatat tiap setengah jam mulai dari periode 1 September 2020 hingga 7 Oktober 2020. Gambar 1 juga menunjukkan bahwa terdapat data yang hilang pada data yang digunakan, sehingga perlu dilakukan imputasi *missing value* sebelum dilakukan analisis lebih lanjut.

Pada penelitian ini, metode imputasi *missing value* yang digunakan ialah *Expectation Maximization*. Karakteristik dari data yang digunakan setelah dilakukan proses imputasi *missing value* dapat dilihat pada Tabel 3.

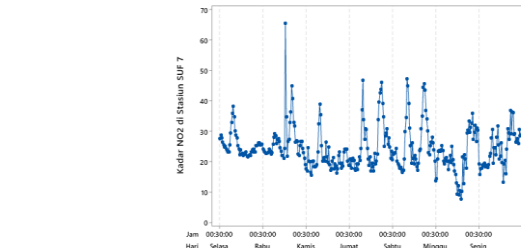
Tabel 3. Karakteristik Indikator Penentu Kualitas Udara

Variabel	Mean	Minimum	Maksimum
CO	0,413343	0,00135	2,1071
PM ₁₀	30,0911	0,06666	133
NO ₂	24,0575	5,8684	65,551

Tabel 3 menunjukkan bahwa kadar indikator PM₁₀ memiliki rata-rata kadar tertinggi yaitu 30,0911. Kadar dari indikator CO memiliki nilai minimum paling kecil dibandingkan indikator lain yaitu 0,00135, sedangkan untuk nilai maksimum kadar indikator NO₂ mencapai 65,551. Kemudian untuk lebih jelasnya dalam melihat pola data indikator penentu kualitas udara dapat ditinjau melalui Gambar 2.



Gambar 2. Time Series Plot Kadar CO, PM₁₀ dan NO₂ dalam 7 Hari



Berdasarkan Gambar 2, kadar ketiga indikator penentu kualitas udara pada hari Minggu cenderung memiliki jumlah kadar yang lebih rendah dibandingkan jumlah kadar di hari lain. Selain itu, kadar ketiga indikator penentu kualitas udara cenderung tinggi saat pukul 08.00. Besarnya jumlah kadar pada pukul ini seiring dengan dimulainya aktivitas warga Kota Surabaya. Pada pukul 16.00 kadar ketiga indikator penentu kualitas udara memiliki fluktuasi yang rendah yang berbanding lurus dengan berakhirnya aktivitas warga Kota Surabaya. Berdasarkan hal ini, diketahui bahwa kadar indikator penentu kualitas udara CO, PM₁₀ dan NO₂ di Stasiun SUF 7 Kota Surabaya memiliki pola *seasonal* yaitu musiman harian per setengah jam.

B. Analisis Korelasi

Pada penelitian ini, dilakukan pengujian korelasi pada ketiga variabel penentu kualitas udara di Stasiun SUF 7 Kota Surabaya guna mengetahui apakah terdapat hubungan antar variabel yang digunakan dalam penelitian. Hasil pengujian korelasi pada penelitian ini dapat dilihat pada Tabel 4.

Tabel 4. Hasil Analisis Korelasi

Variabel	P-value	Keputusan	Tingkat Korelasi
CO - PM ₁₀	$1,69 \times 10^{-22}$	Tolak H ₀	0,18
CO - NO ₂	$1,23 \times 10^{-229}$	Tolak H ₀	0,55
PM ₁₀ - NO ₂	$3,41 \times 10^{-14}$	Tolak H ₀	0,14

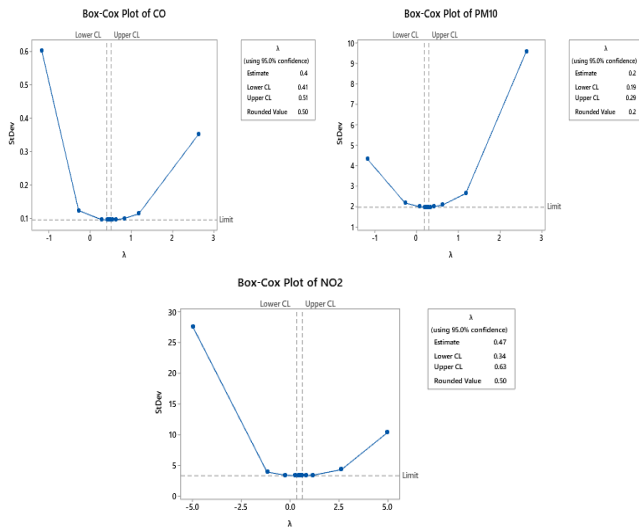
Berdasarkan Tabel 4, terdapat korelasi pada variabel-variabel yang digunakan. Tingkat keeratan antara indikator CO dan indikator PM₁₀ serta indikator PM₁₀ dan NO₂ memiliki tingkat keeratan yang sangat lemah, hal ini dibuktikan dengan nilai korelasinya yang bernilai 0,18 dan 0,14. Sedangkan pada indikator CO dan NO₂ didapati nilai korelasi sebesar 0,55 atau memiliki hubungan yang kuat. Berdasarkan adanya hubungan antar indikator ini, maka metode analisis yang akan digunakan untuk pemodelan ialah VAR, VARX, dan LSTM.

C. Pemodelan Indikator Penentu Kualitas Udara di Stasiun SUF 7 Kota Surabaya dengan VAR

Tahapan-tahapan yang dilakukan pada pemodelan menggunakan VAR dapat dilihat pada subsubbab pembahasan berikut.

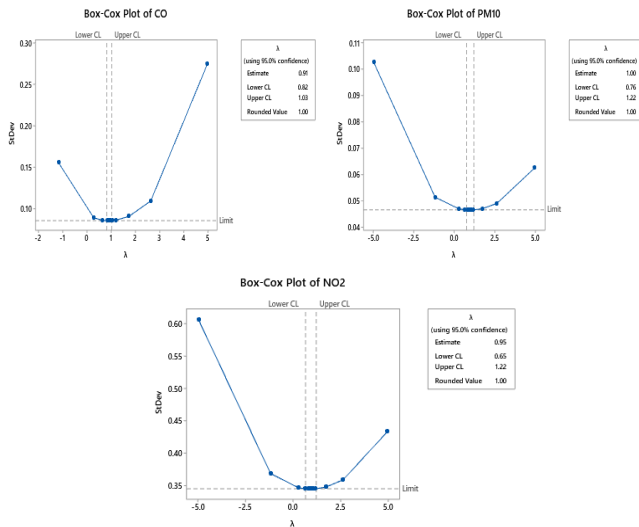
1. Stasioneritas Data

Pada identifikasi stasioneritas data, data yang digunakan harus stasioner terhadap *varians* dan stasioner terhadap *mean*. Hasil identifikasi stasioneritas data dapat dilihat sebagai berikut.



Gambar 3. Box-Cox Plot Sebelum Transformasi

Berdasarkan Gambar 3, dari ketiga indikator penentu kualitas udara tidak ada satupun indikator yang stasioner terhadap *varians*. Hal ini dapat dilihat melalui *rounded value* pada masing-masing variabel bernilai kurang dari 1,00 sehingga selanjutnya dilakukan transformasi.



Gambar 4. Box-Cox Plot Setelah Transformasi

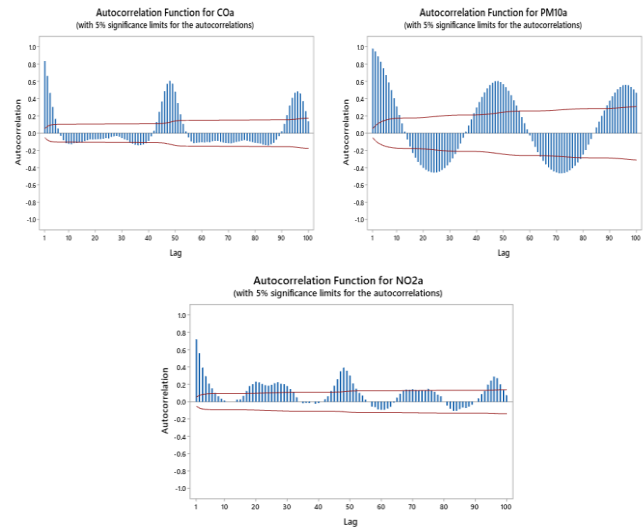
Setelah dilakukan proses transformasi, ketiga indikator penentu kualitas udara telah stasioner terhadap *varians*. Sehingga dapat dilakukan identifikasi ke tahap selanjutnya, yaitu identifikasi stasioneritas data terhadap *mean* yang dapat dilakukan secara visual menggunakan plot MCCF.

Tabel 5. Hasil Pengujian Augmented Dickey Fuller Non-Musiman

Variabel	P-value	Kesimpulan
CO	0,01	Stasioner
PM ₁₀	0,01	Stasioner
NO ₂	0,01	Stasioner

Berdasarkan Tabel 5 data yang digunakan pada penelitian kali ini telah stasioner terhadap *mean*. Hal ini dapat dilihat melalui nilai *p-value* dari tiap indikator yang digunakan telah

bernilai kurang dari alpha (0,05). Pada Gambar 2 data yang digunakan mengindikasikan adanya dugaan pola musiman. Pola musiman akan diidentifikasi secara *univariate* menggunakan plot ACF.



Gambar 5. ACF Plot Indikator Penentu Kualitas Udara di Stasiun SUF 7 Kota Surabaya

Sebelum dilakukan *differencing* dapat dilihat bahwa setelah lag 1, lag berikutnya pada ketiga indikator cenderung terlihat *dies down* dan memiliki pola musiman 48. Dikarenakan data memiliki pola musiman 48, maka selanjutnya dilakukan *differencing* 1 dan *differencing* 48. Hasil pengujian Augmented Dickey Fuller musiman yang telah dilakukan *differencing* dapat dilihat pada Tabel 6.

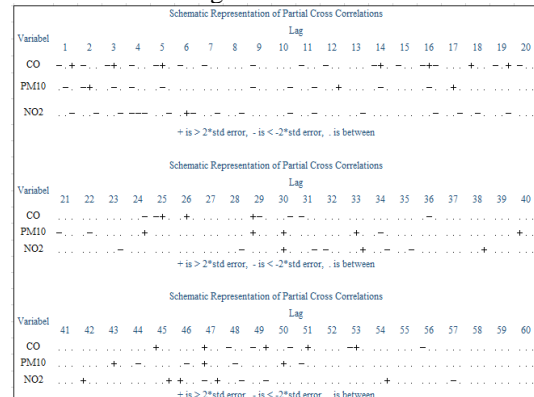
Tabel 6. Hasil Pengujian Augmented Dickey Fuller Non-Musiman

Variabel	P-value	Kesimpulan
CO	0,01	Stasioner
PM ₁₀	0,01	Stasioner
NO ₂	0,01	Stasioner

Setelah dilakukan *differencing*, nilai *p-value* dari hasil ADF musiman bernilai kurang dari alpha, hal ini mengindikasikan bahwa data indikator penentu kualitas udara di stasiun SUF 7 Kota Surabaya telah stasioner terhadap *mean* dan dapat digunakan untuk analisis ke tahap selanjutnya.

2. Identifikasi Model VAR

Identifikasi model VAR dapat dilihat melalui *Matrix Partial Cross Correlation Function* (MPCCF) plot. MPCCF plot pada penelitian ini adalah sebagai berikut.



Gambar 6. MPCCF Plot Indikator Penentu Kualitas Udara

Plot MPCCF dari indikator penentu kualitas udara di Stasiun SUF 7 Kota Surabaya signifikan pada lag ke-1 hingga lag ke-5. Sedangkan lag yang menunjukkan pola musiman terdapat pada lag ke-48. Sehingga model VAR yang terbentuk ialah $(5,1,0)(1,1,0)^{48}$.

3. Estimasi Parameter dan Uji Signifikansi

Estimasi parameter model VAR dilakukan terhadap model VAR $(5,1,0)(1,1,0)^{48}$ dan didapati sebanyak 171 parameter. Selanjutnya dilakukan uji signifikansi menggunakan *restrict* untuk mengetahui parameter yang signifikan terhadap model.

Tabel 7. Hasil Estimasi Parameter Model VAR $(5,1,0)(1,1,0)^{48}$

Equation	Parameter	Estimasi	Std Error	t-Value	P-Value	Variabel
CO	$\emptyset_{(1,1,1)}$	-0,351	0,025	-13,7	0,0001	CO _(t-1)
	$\emptyset_{(2,1,1)}$	-0,146	0,023	-6,2	0,0001	CO _(t-2)
	$\emptyset_{(3,1,1)}$	-0,159	0,023	-6,81	0,0001	CO _(t-3)
	$\emptyset_{(4,1,1)}$	-0,125	0,023	-5,4	0,0001	CO _(t-4)
	$\emptyset_{(4,1,2)}$	-0,106	0,035	-2,97	0,003	PM _{10(t-4)}
	$\emptyset_{(48,1,1)}$	-0,563	0,025	-22,54	0,0001	CO _(t-48)
	$\emptyset_{(49,1,1)}$	-0,123	0,025	-4,93	0,0001	CO _(t-49)
	$\emptyset_{(54,1,1)}$	0,043	0,020	2,12	0,0339	CO _(t-54)
PM ₁₀	$\emptyset_{(96,1,1)}$	-0,306	0,022	-13,56	0,0001	CO _(t-96)
	$\emptyset_{(1,2,2)}$	0,092	0,027	3,37	0,0008	PM _{10(t-1)}
	$\emptyset_{(2,2,1)}$	0,036	0,011	3,09	0,002	CO _(t-2)
	$\emptyset_{(2,2,2)}$	0,252	0,026	9,38	0,0001	PM _{10(t-2)}
	$\emptyset_{(3,2,1)}$	0,045	0,013	3,36	0,0008	CO _(t-3)
	$\emptyset_{(3,2,2)}$	-0,047	0,020	-2,27	0,0236	PM _{10(t-3)}
	$\emptyset_{(4,2,1)}$	0,037	0,013	2,83	0,0047	CO _(t-4)
	$\emptyset_{(4,2,3)}$	-0,005	0,002	-1,97	0,0496	NO _{2(t-4)}
	$\emptyset_{(5,2,1)}$	0,026	0,012	2,22	0,0268	CO _(t-5)
	$\emptyset_{(48,2,2)}$	-0,683	0,023	-29,33	0,0001	PM _{10(t-48)}
	$\emptyset_{(49,2,2)}$	0,082	0,028	2,87	0,0041	PM _{10(t-49)}
	$\emptyset_{(50,2,2)}$	0,186	0,029	6,28	0,0001	PM _{10(t-50)}
	$\emptyset_{(51,2,1)}$	0,029	0,012	2,44	0,0147	CO _(t-51)
	$\emptyset_{(54,2,1)}$	-0,035	0,012	-2,84	0,0046	CO _(t-54)
	$\emptyset_{(96,2,2)}$	-0,511	0,023	-22,15	0,0001	PM _{10(t-96)}
	$\emptyset_{(97,2,2)}$	0,141	0,026	5,37	0,0001	PM _{10(t-97)}
	$\emptyset_{(98,2,2)}$	0,068	0,027	2,52	0,0117	PM _{10(t-98)}
	$\emptyset_{(102,2,1)}$	-0,027	0,012	-2,25	0,0249	CO _(t-102)
NO ₂	$\emptyset_{(1,3,1)}$	0,348	0,103	3,37	0,0008	CO _(t-1)
	$\emptyset_{(1,3,3)}$	-0,459	0,026	-17,33	0,0001	NO _{2(t-1)}
	$\emptyset_{(2,3,3)}$	-0,178	0,023	-7,46	0,0001	NO _{2(t-2)}
	$\emptyset_{(3,3,3)}$	-0,157	0,024	-6,54	0,0001	NO _{2(t-3)}
	$\emptyset_{(4,3,1)}$	-0,231	0,091	-2,51	0,0121	CO _(t-4)
	$\emptyset_{(4,3,3)}$	-0,078	0,024	-3,16	0,0016	NO _{2(t-4)}
	$\emptyset_{(5,3,3)}$	-0,062	0,021	-2,85	0,0045	NO _{2(t-5)}
	$\emptyset_{(48,3,1)}$	0,282	0,097	2,88	0,004	CO _(t-48)
	$\emptyset_{(48,3,3)}$	-0,630	0,025	-24,55	0,0001	NO _{2(t-48)}
	$\emptyset_{(49,3,1)}$	0,261	0,114	2,29	0,0225	CO _(t-49)
	$\emptyset_{(49,3,3)}$	-0,249	0,030	-8,3	0,0001	NO _{2(t-49)}
	$\emptyset_{(50,3,1)}$	-0,223	0,088	-2,53	0,0116	CO _(t-50)
	$\emptyset_{(54,3,3)}$	0,076	0,020	3,76	0,0002	NO _{2(t-54)}
	$\emptyset_{(96,3,3)}$	-0,310	0,024	-12,76	0,0001	NO _{2(t-96)}
	$\emptyset_{(97,3,3)}$	-0,082	0,025	-3,25	0,0012	NO _{2(t-97)}

Berdasarkan Tabel 7, hanya terdapat 41 dari 171 parameter yang signifikan terhadap model VAR $(5,1,0)(1,1,0)^{48}$. Selanjutnya, model VAR $(5,1,0)(1,1,0)^{48}$ dapat ditulis ke dalam bentuk persamaan berikut.

$$Z_{1,t} = Z_{1,t-1} - 0,351Z_{1,t-1} - 0,146Z_{1,t-2} - 0,159Z_{1,t-3} - 0,125Z_{1,t-4} - 0,106Z_{2,t-4} - 0,563Z_{1,t-48} - 0,123Z_{1,t-49} + 0,043Z_{1,t-54} - 0,306Z_{1,t-96} + a_{1,t}.$$

$$Z_{2,t} = Z_{2,t-1} + 0,092Z_{1,t-2} + 0,036Z_{1,t-2} + 0,252Z_{2,t-2} + 0,045Z_{1,t-3} - 0,047Z_{2,t-3} + 0,037Z_{1,t-4} - 0,005Z_{3,t-4} + 0,026Z_{1,t-5} - 0,683Z_{2,t-48} + 0,082Z_{2,t-49} + 0,186Z_{2,t-50} + 0,029Z_{1,t-51} - 0,035Z_{1,t-54} - 0,511Z_{2,t-96} + 0,141Z_{2,t-97} + 0,068Z_{2,t-98} - 0,027Z_{1,t-102} + a_{2,t}.$$

$$Z_{3,t} = Z_{3,t-1} + 0,348Z_{1,t-1} - 0,459Z_{3,t-1} - 0,178Z_{3,t-2} - 0,157Z_{3,t-3} - 0,231Z_{1,t-4} - 0,078Z_{3,t-4} - 0,062Z_{3,t-5} + 0,282Z_{1,t-48} - 0,630Z_{3,t-48} + 0,261Z_{1,t-49} - 0,249Z_{3,t-49} - 0,223Z_{1,t-50} + 0,076Z_{3,t-54} - 0,310Z_{3,t-96} - 0,082Z_{3,t-97} + a_{3,t}.$$

Berdasarkan persamaan di atas, dapat diketahui bahwa hasil peramalan jumlah kadar indikator pada suatu variabel tidak hanya dipengaruhi oleh variabel itu sendiri, melainkan juga dipengaruhi oleh variabel-variabel lain.

4. Pemeriksaan Diagnosa Model VAR

Pada model VAR yang terbentuk dengan parameter-parameter yang signifikan selanjutnya perlu dilakukan pengujian asumsi residual *white noise* dan asumsi residual berdistribusi multivariat normal. Hasil pemeriksaan diagnosa pada penelitian ini ialah sebagai berikut.

Tabel 8. Hasil Pengujian *Portmanteau*

Lag	Chi-Square	P-value	Lag	Chi-Square	P-value
1	3,1989	0,96203	53	658,4496	0,00099
2	9,9358	0,93706	54	664,3346	0,00099
3	12,9177	0,98901	96	1076,171	0,00099
4	18,1319	0,99400	97	1083,4327	0,00099
5	24,7305	0,99500	98	1090,0658	0,00099
48	640,0940	0,00099	99	1100,9041	0,00099
49	642,8955	0,00099	100	1108,1198	0,00099
50	645,6256	0,00099	101	1113,1186	0,00099
51	649,2034	0,00099	102	1121,5494	0,00099
52	654,2498	0,00099			

Berdasarkan Tabel 8 residual dari data belum memenuhi asumsi residual *white noise*. Hal ini dapat dilihat melalui nilai *p-value* pada beberapa *lag* yang bernilai kurang dari 0,05. Selanjutnya ialah pemeriksaan residual berdistribusi *multivariate normal*.

Tabel 9. Hasil Pengujian Shapiro-Wilk

W	P-Value
0,88339	$< 2,2 \times 10^{-16}$

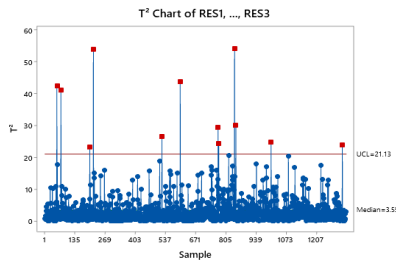
Hasil pengujian menggunakan Shapiro-Wilk menghasilkan nilai *p-value* yang sangat kecil atau kurang dari 0,05, sehingga dapat disimpulkan bahwa residual dari data tidak memenuhi asumsi residual berdistribusi multivariat normal.

Model VAR $(5,1,0)(1,1,0)^{48}$ tidak memenuhi asumsi residual *white noise* dan berdistribusi multivariat normal, sehingga tidak dapat digunakan untuk meramalkan dan mengetahui keterkaitan antar indikator penentu kualitas udara di stasiun SUF 7 Kota Surabaya. Berdasarkan hal ini, maka selanjutnya akan dilakukan deteksi *outlier*.

D. Deteksi Outlier

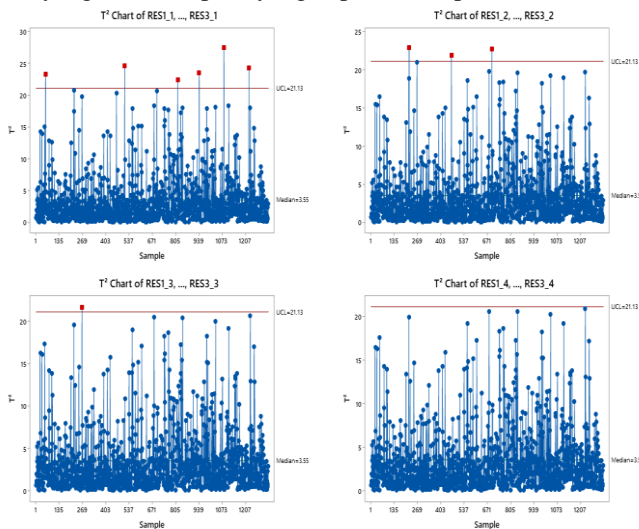
Pendeteksian *outlier* dapat dilakukan menggunakan diagram kontrol residual yang dihitung dengan statistik T^2 -*Hotelling* dan dilakukan secara *iterative* hingga didapati kondisi

in control. Hasil deteksi *outlier* pada data indikator penentu kualitas udara di stasiun SUF 7 Kota Surabaya dapat dilihat sebagai berikut.



Gambar 7. Diagram Kontrol Residual Sebelum Iterasi

Pada diagram kontrol residual sebelum dilakukan iterasi terdeteksi bahwa terdapat 12 sinyal *out of control* yaitu pada residual data ke-55, 73, 201, 217, 521, 602, 770, 773, 843, 849, 1005 dan 1321 dan selanjutnya dilakukan iterasi dengan membuang residual data yang *outlier* seperti yang dapat dilihat pada Gambar 8.



Gambar 8. (a) Diagram Kontrol Residual Setelah Iterasi Pertama dan (b) Iterasi Kedua (c) Iterasi Ketiga dan (d) Iterasi Keempat

Setelah dilakukan iterasi pertama hingga iterasi keempat, diagram kontrol residual mendeteksi adanya total 22 sinyal *out of control*, yaitu pada residual data ke-57, 511, 817, 940, 1082, 1226, 221, 464, 697, dan 266. Hasil diagram kontrol residual setelah dilakukan iterasi keempat dapat dilihat pada Gambar 8 bagian (d). Karena residual sudah *in control*, maka selanjutnya dilakukan pembentukan variabel *dummy* pada data yang digunakan.

Tabel 10. Variabel *Dummy* Model VAR (5,1,0)(1,1,0)⁴⁸

Residual	D ₁	D ₂	D ₃	...	D ₂₂
55	1	0	0	...	0
73	0	1	0	...	0
201	0	0	1	...	0
⋮	⋮	⋮	⋮	⋮	⋮
266	0	0	0	...	1

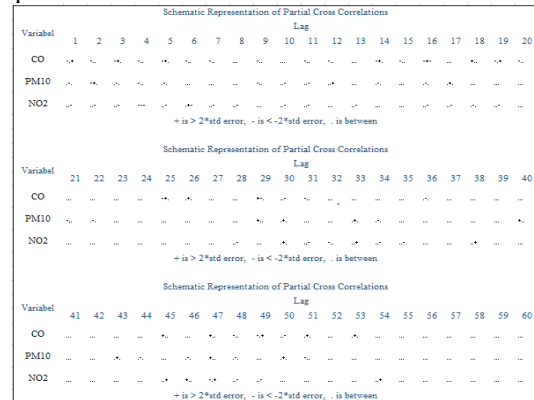
Setelah variabel *dummy* selesai dibuat, selanjutnya dibuat model baru untuk mendapatkan model yang sesuai menggunakan VARX.

E. Pemodelan Indikator Penentu Kualitas Udara di Stasiun SUF 7 Kota Surabaya dengan Variabel Eksogen (VARX)

Tahapan-tahapan yang dilakukan pada pemodelan menggunakan VARX dapat dilihat pada pembahasan berikut.

1. Identifikasi Model VARX

Identifikasi model VARX dapat dilihat melalui *Matrix Partial Cross Correlation Function* (MPCCF) plot. Plot MPCCF pada penelitian ini dengan variabel *dummy* dapat dilihat pada Gambar 9.



Gambar 9. MPCCF Plot Tiga Variabel dan Variabel *Dummy*

Berdasarkan Gambar 9 indikator penentu kualitas udara di Stasiun SUF 7 Kota Surabaya signifikan pada lag ke-1 hingga lag ke-5. Sehingga model VARX yang terbentuk ialah (5,5).

2. Estimasi Parameter dan Uji Signifikansi

Estimasi parameter model VARX dilakukan terhadap model VAR (5,5) dan didapati sebanyak 297 parameter. Selanjutnya dilakukan uji signifikansi menggunakan *restrict* untuk mengetahui parameter yang signifikan terhadap model.

Tabel 11. Hasil Estimasi Parameter Model VARX (5,5)

Equation	Parameter	Estimasi	Std Error	t-Value	P-Value	Variabel
CO	$\delta_{(0,1,7)}$	-0,584	0,109	-5,34	0,0001	D _{7(t)}
	$\delta_{(1,1,3)}$	-0,220	0,109	-2,02	0,0436	D _{3(t-1)}
	$\delta_{(2,1,3)}$	-0,290	0,117	-2,47	0,0137	D _{3(t-2)}
	$\delta_{(3,1,7)}$	-0,558	0,109	-5,09	0,0001	D _{7(t-3)}
	$\delta_{(4,1,6)}$	0,255	0,109	2,35	0,0191	D _{6(t-4)}
	$\delta_{(4,1,7)}$	0,324	0,110	2,95	0,0033	D _{7(t-4)}
	$\delta_{(5,1,3)}$	-0,222	0,108	-2,04	0,0413	D _{3(t-5)}
	$\delta_{(5,1,7)}$	-0,372	0,110	-3,38	0,0007	D _{7(t-5)}
	$\phi_{(1,1,1)}$	0,641	0,026	24,39	0,0001	CO _(t-1)
	$\phi_{(2,1,1)}$	0,191	0,029	6,47	0,0001	CO _(t-2)
	$\phi_{(3,1,1)}$	-0,102	0,025	-4,02	0,0001	CO _(t-3)
	$\phi_{(48,1,1)}$	0,179	0,019	9,26	0,0001	CO _(t-48)
	$\phi_{(52,1,2)}$	0,108	0,039	2,75	0,0061	PM _{10(t-52)}
	$\phi_{(53,1,1)}$	-0,087	0,020	-4,27	0,0001	CO _(t-53)
	$\phi_{(53,1,2)}$	-0,097	0,037	-2,59	0,0098	PM _{10(t-53)}
	$\phi_{(54,1,3)}$	0,012	0,003	4,02	0,0001	NO _{2(t-54)}
	$\phi_{(96,1,1)}$	0,074	0,024	3,05	0,0024	CO _(t-96)
	$\phi_{(96,1,2)}$	0,109	0,037	2,89	0,0039	PM _{10(t-96)}
	$\phi_{(97,1,1)}$	0,064	0,028	2,24	0,0256	CO _(t-97)
	$\phi_{(97,1,2)}$	-0,093	0,038	-2,44	0,0148	PM _{10(t-97)}
	$\phi_{(98,1,1)}$	-0,116	0,025	-4,52	0,0001	CO _(t-98)
PM ₁₀	$\phi_{(102,1,1)}$	-0,045	0,018	-2,47	0,0138	CO _(t-102)
	$\delta_{(0,2,1)}$	0,437	0,058	7,49	0,0001	D _{1(t)}
	$\delta_{(0,2,2)}$	-0,440	0,058	-7,55	0,0001	D _{2(t)}
	$\delta_{(0,2,3)}$	0,309	0,058	5,33	0,0001	D _{3(t)}
	$\delta_{(0,2,4)}$	-0,487	0,058	-8,38	0,0001	D _{4(t)}
	$\delta_{(0,2,5)}$	0,263	0,058	4,53	0,0001	D _{5(t)}
	$\delta_{(0,2,6)}$	-0,422	0,058	-7,26	0,0001	D _{6(t)}

Tabel 11. Hasil Estimasi Parameter Model VARX (5,5) (Lanjutan)

Equation	Parameter	Estimasi	Std Error	t-Value	P-Value	Variabel
NO2	$\delta_{(2,2,1)}$	0,297	0,059	5,03	0,0001	D _{1(t-2)}
	$\delta_{(2,2,2)}$	-0,132	0,059	-2,22	0,0269	D _{2(t-2)}
	$\delta_{(2,2,4)}$	-0,259	0,059	-4,36	0,0001	D _{4(t-2)}
	$\delta_{(4,2,2)}$	0,152	0,059	2,58	0,0101	D _{2(t-4)}
	$\delta_{(4,2,4)}$	0,275	0,059	4,61	0,0001	D _{4(t-4)}
	$\delta_{(4,2,6)}$	0,127	0,059	2,15	0,0316	D _{6(t-4)}
	$\phi_{(1,2,1)}$	0,064	0,008	7,29	0,0001	CO _(t-1)
	$\phi_{(1,2,2)}$	1,093	0,024	45,42	0,0001	PM _{10(t-1)}
	$\phi_{(2,2,2)}$	0,166	0,036	4,55	0,0001	PM _{10(t-2)}
	$\phi_{(3,2,2)}$	-0,388	0,034	-11,12	0,0001	PM _{10(t-3)}
	$\phi_{(4,2,2)}$	0,161	0,036	4,47	0,0001	PM _{10(t-4)}
	$\phi_{(5,2,2)}$	-0,071	0,023	-2,98	0,0029	PM _{10(t-5)}
	$\phi_{(48,2,2)}$	0,067	0,014	4,65	0,0001	PM _{10(t-48)}
	$\phi_{(50,2,2)}$	-0,065	0,014	-4,57	0,0001	PM _{10(t-50)}
	$\phi_{(52,2,1)}$	-0,023	0,009	-2,37	0,0182	CO _(t-52)
	$\phi_{(54,2,3)}$	0,004	0,001	3,06	0,0023	NO _{2(t-54)}
	$\phi_{(96,2,1)}$	0,034	0,008	3,8	0,0002	CO _(t-96)
	$\phi_{(97,2,2)}$	0,173	0,021	8,03	0,0001	PM _{10(t-97)}
	$\phi_{(98,2,2)}$	-0,171	0,021	-7,96	0,0001	PM _{10(t-98)}
	$\phi_{(99,2,1)}$	-0,039	0,010	-3,67	0,0003	CO _(t-99)
	$\phi_{(102,2,1)}$	0,048	0,009	5,01	0,0001	CO _(t-102)
	$\delta_{(0,3,5)}$	-1,527	0,451	-3,38	0,0007	D _{5(t)}
	$\delta_{(2,3,2)}$	1,152	0,452	2,55	0,011	D _{2(t-2)}
	$\delta_{(2,3,3)}$	-1,044	0,487	-2,14	0,0323	D _{3(t-2)}
	$\delta_{(3,3,1)}$	0,897	0,452	1,98	0,0476	D _{1(t-3)}
	$\delta_{(3,3,2)}$	0,936	0,452	2,07	0,0387	D _{2(t-3)}
	$\phi_{(1,3,1)}$	0,458	0,086	5,27	0,0001	CO _(t-1)
	$\phi_{(1,3,3)}$	0,477	0,026	18,22	0,0001	NO _{2(t-1)}
	$\phi_{(2,3,3)}$	0,199	0,026	7,4	0,0001	NO _{2(t-2)}
	$\phi_{(4,3,1)}$	-0,344	0,089	-3,87	0,0001	CO _(t-4)
	$\phi_{(4,3,3)}$	0,068	0,022	2,99	0,0028	NO _{2(t-4)}
	$\phi_{(48,3,3)}$	0,136	0,020	6,65	0,0001	NO _{2(t-48)}
	$\phi_{(49,3,2)}$	0,361	0,170	2,12	0,0341	PM _{10(t-49)}
	$\phi_{(50,3,2)}$	-0,490	0,198	-2,48	0,0133	PM _{10(t-50)}
	$\phi_{(51,3,1)}$	0,288	0,100	2,87	0,0042	CO _(t-51)
	$\phi_{(51,3,3)}$	-0,056	0,023	-2,44	0,0146	NO _{2(t-51)}
	$\phi_{(53,3,1)}$	-0,213	0,096	-2,22	0,0264	CO _(t-53)
	$\phi_{(54,3,2)}$	0,217	0,069	3,14	0,0017	PM _{10(t-54)}
	$\phi_{(54,3,3)}$	0,065	0,019	3,3	0,001	NO _{2(t-54)}
	$\phi_{(96,3,3)}$	0,098	0,022	4,47	0,0001	NO _{2(t-96)}
	$\phi_{(98,3,1)}$	0,306	0,094	3,23	0,0013	CO _(t-98)
	$\phi_{(98,3,3)}$	-0,062	0,023	-2,7	0,007	NO _{2(t-98)}
	$\phi_{(99,3,2)}$	-0,385	0,163	-2,36	0,0184	PM _{10(t-99)}
	$\phi_{(100,3,1)}$	-0,197	0,089	-2,2	0,028	CO _(t-100)
	$\phi_{(100,3,3)}$	0,377	0,161	2,34	0,0195	PM _{10(t-100)}

Berdasarkan Tabel 11 dapat diketahui bahwa hanya terdapat 73 dari 297 parameter yang signifikan terhadap model VARX (5,5) setelah dilakukan *restrict* sebanyak 224 parameter. Selanjutnya, model VARX (5,5) dapat ditulis ke dalam bentuk persamaan berikut.

$$Z_{1,t} = -0,584D_{7,t} - 0,220D_{3,t-1} - 0,290D_{3,t-2} - 0,558D_{7,t-3} + 0,255D_{6,t-4} + 0,324D_{7,t-4} - 0,222D_{3,t-5} - 0,372D_{7,t-5} + Z_{1,t-1} + 0,641Z_{1,t-1} + 0,191Z_{1,t-2} - 0,102Z_{1,t-3} + 0,179Z_{1,t-4} + 0,108Z_{2,t-52} - 0,087Z_{1,t-53} - 0,097Z_{2,t-53} + 0,012Z_{3,t-54} +$$

$$0,074Z_{1,t-96} + 0,109Z_{2,t-96} + 0,064Z_{1,t-97} - 0,093Z_{2,t-97} - 0,116Z_{1,t-98} - 0,045Z_{1,t-102} + a_{1,t}.$$

$$Z_{2,t} = 0,437D_{1,t} - 0,440D_{2,t} + 0,309D_{3,t} - 0,487D_{4,t} + 0,263D_{5,t} - 0,422D_{6,t} + 0,297D_{1,t-2} - 0,132D_{2,t-2} - 0,259D_{4,t-2} + 0,152D_{2,t-4} + 0,275D_{4,t-4} + 0,127D_{6,t-4} + Z_{2,t-1} + 0,064Z_{1,t-1} + 1,093Z_{2,t-1} + 0,166Z_{2,t-2} - 0,388Z_{2,t-3} + 0,161Z_{2,t-4} - 0,071Z_{2,t-5} + 0,067Z_{2,t-48} - 0,065Z_{2,t-50} - 0,023Z_{1,t-52} + 0,004Z_{3,t-54} + 0,034Z_{1,t-96} + 0,173Z_{2,t-97} - 0,171Z_{2,t-98} - 0,039Z_{1,t-99} + 0,048Z_{1,t-102} + a_{2,t}.$$

$$Z_{3,t} = -1,527D_{5,t} + 1,152D_{2,t-2} - 1,044D_{3,t-2} + 0,897D_{1,t-3} + 0,936D_{2,t-3} + Z_{3,t-1} + 0,458Z_{1,t-1} + 0,477Z_{3,t-1} + 0,199Z_{3,t-2} - 0,344Z_{1,t-4} + 0,068Z_{3,t-4} + 0,136Z_{3,t-48} + 0,361Z_{2,t-49} - 0,490Z_{2,t-50} + 0,288Z_{1,t-51} - 0,056Z_{3,t-51} - 0,213Z_{1,t-53} + 0,217Z_{2,t-54} + 0,065Z_{3,t-54} + 0,098Z_{2,t-96} + 0,306Z_{1,t-98} - 0,062Z_{3,t-98} - 0,385Z_{2,t-99} - 0,197Z_{1,t-100} + 0,377Z_{2,t-100} + a_{3,t}.$$

Berdasarkan persamaan di atas, dapat diketahui bahwa hasil peramalan jumlah kadar indikator pada suatu variabel tidak hanya dipengaruhi oleh variabel itu sendiri, melainkan juga dipengaruhi oleh variabel lain serta variabel *dummy*.

3. Pemeriksaan Diagnosa Model VARX

Pada model VARX yang terbentuk selanjutnya dilakukan pengujian asumsi residual *white noise* dan berdistribusi multivariat normal. Hasil pemeriksaan diagnosa pada penelitian ini adalah sebagai berikut.

Tabel 12. Hasil Uji Portmanteau Model VARX (5,5)

Lag	Chi-Square	P-value	Lag	Chi-Square	P-value
1	6,2048	0,73926	53	706,1395	0,00099
2	17,4680	0,52247	54	713,4031	0,00099
3	22,4179	0,69430	96	1143,7571	0,00099
4	26,8656	0,85514	97	1148,6326	0,00099
5	31,8352	0,92307	98	1158,8571	0,00099
48	679,9916	0,00099	99	1170,2791	0,00099
49	685,3938	0,00099	100	1182,6872	0,00099
50	688,6889	0,00099	101	1187,9186	0,00099
51	693,8673	0,00099	102	1197,1581	0,00099
52	698,4971	0,00099			

Berdasarkan Tabel 12 residual dari data yang digunakan memenuhi asumsi residual *white noise* pada lima lag awal yang digunakan dan pada lag-lag selanjutnya tidak memenuhi asumsi residual *white noise*. Hal ini dapat dilihat melalui nilai *p-value* yang bernilai kurang dan/atau lebih dari 0,05. Sehingga residual data yang digunakan pada penelitian ini tidak memenuhi asumsi residual *white noise*, akan tetapi diasumsikan telah memenuhi asumsi residual *white noise*. Selanjutnya ialah pemeriksaan residual berdistribusi *multivariate normal*.

Tabel 13. Hasil Pengujian Shapiro-Wilk

W	P-Value
0,90967	$< 2,2 \times 10^{-16}$

Hasil pengujian menggunakan Shapiro-Wilk menghasilkan nilai *p-value* yang sangat kecil atau kurang dari 0,05, sehingga dapat disimpulkan bahwa residual dari data indikator penentu kualitas udara di stasiun SUF 7 Kota Surabaya tidak memenuhi asumsi residual berdistribusi multivariat normal. Akan tetapi pada penelitian ini, residual diasumsikan telah memenuhi asumsi residual berdistribusi multivariat normal dan dapat digunakan untuk tahapan analisis selanjutnya.

F. Pemodelan Indikator Penentu Kualitas Udara di Stasiun SUF 7 Kota Surabaya dengan LSTM

Pada penelitian ini, variabel yang digunakan sebagai variabel *input* pada jaringan LSTM didapatkan dari 19 lag signifikan pada model VAR, yaitu lag 1, 2, 3, 4, 5, 48, 49, 50, 51, 52, 53, 54, 96, 97, 98, 99, 100, 101, dan 102 pada ketiga variabel serta variabel *dummy* yang digunakan pada VARX. Selain itu juga dicobakan 1 *hidden layer*, 1 hingga 5 neuron, adam *optimization* untuk mem-perbaharui bobot dari jaringan LSTM. Model terbaik yang didapatkan pada tiap indikator dan *timestep* pada pemodelan LSTM tanpa variabel *dummy* dapat dilihat sebagai berikut.

Tabel 14. Nilai RMSE Terkecil di Tiap *Timesteps* Tanpa Variabel *Dummy*

Time-step	CO		PM ₁₀		NO ₂	
	Unit LSTM	RMSE	Unit LSTM	RMSE	Unit LSTM	RMSE
1	5	0,355	4	35,252	1	16,438
2	4	0,418	5	36,413	5	16,730
3	3	0,390	3	35,009	4	16,343
4	4	0,445	3	35,578	4	14,997
5	5	0,412	4	35,943	4	16,008
48	1	0,356	5	37,204	3	16,189
49	2	0,364	4	35,909	3	16,208
50	5	0,340	2	35,038	4	14,846
51	4	0,408	1	34,802	3	15,682
52	5	0,345	5	35,066	5	15,159
53	4	0,416	3	36,384	1	16,427
54	1	0,418	4	35,747	1	15,522
96	5	0,412	5	37,524	5	15,828
97	2	0,399	3	33,574	3	16,655
98	5	0,387	5	35,316	2	16,323
99	4	0,375	2	35,570	5	16,623
100	4	0,353	5	35,690	4	15,957
101	1	0,416	5	36,865	5	15,223
102	4	0,413	4	37,572	1	15,049

Berdasarkan Tabel 14 model terbaik yang akan dipilih ialah model yang menghasilkan nilai RMSE terkecil pada data *train*. Sehingga untuk indikator CO, diketahui bahwa pada data *train* model terbaik yang didapatkan ialah model yang menggunakan *timestep* 50 dan 5 unit LSTM. Model terbaik dari indikator PM₁₀ berasal dari model dengan 97 *timestep* dan 3 unit LSTM. Sedangkan untuk indikator NO₂ model terbaik diperoleh dari hasil perhitungan menggunakan 50 *timestep* 4 unit LSTM. Selanjutnya, model LSTM tanpa variabel *dummy* sebagai *input* dapat ditulis ke dalam bentuk persamaan berikut.

$$Z_{1,t}^* = 0,0321 + 0,855h_{1,t} - 0,855h_{2,t} - 0,571h_{3,t} + 0,023h_{4,t} - 0,093h_{5,t}.$$

$$Z_{2,t}^* = 0,0778 + 0,466h_{1,t} - 0,838h_{2,t} + 0,117h_{3,t}$$

$$Z_{3,t}^* = 0,0856 - 0,421h_{1,t} - 0,718h_{2,t} + 0,330h_{3,t} - 0,270h_{4,t}.$$

$$\text{dengan } Z_{i,t}^* = \frac{Z_{i,t} - \min(Z_{i,t})}{\max(Z_{i,t}) - \min(Z_{i,t})}.$$

Selanjutnya akan dibahas mengenai hasil pemodelan menggunakan metode LSTM yang juga mengikutsertakan variabel *dummy* sebagai variabel *input*. Berdasarkan jumlah *timestep* serta jumlah *unit* LSTM yang digunakan didapati 95 model untuk tiap-tiap indikator yang digunakan. Hasil dari pemodelan dapat dilihat sebagai berikut.

Tabel 15. Nilai RMSE Terkecil di Tiap *Timesteps* dengan Variabel *Dummy*

Time-step	CO		PM ₁₀		NO ₂	
	Unit LSTM	RMSE	Unit LSTM	RMSE	Unit LSTM	RMSE
1	5	0,368	3	35,922	5	16,388
2	5	0,485	5	35,548	5	15,081
3	4	0,470	5	36,465	5	14,844
4	3	0,487	2	34,902	4	14,856
5	3	0,463	1	34,892	2	15,183
48	1	0,363	5	37,112	1	15,297
49	2	0,370	5	36,259	2	15,411
50	5	0,483	2	35,678	1	16,583
51	4	0,464	5	33,765	1	15,744
52	4	0,476	5	34,841	3	15,054
53	4	0,490	4	34,957	5	14,876
54	2	0,393	1	37,425	5	15,879
96	5	0,481	5	36,538	1	15,601
97	5	0,480	5	36,468	4	16,356
98	4	0,342	1	35,057	1	15,232
99	5	0,481	2	36,312	4	15,997
100	5	0,387	5	37,276	4	15,455
101	3	0,477	3	34,976	4	15,487
102	2	0,480	5	36,549	1	14,929

Berdasarkan Tabel 15, pemodelan menggunakan LSTM dengan menyertakan variabel *dummy* sebagai variabel *input* didapati hasil bahwa untuk indikator CO ialah model yang menggunakan *timestep* 98 dan 4 *unit* LSTM. Model terbaik dari indikator PM₁₀ berasal dari model dengan 51 *timestep* dan 5 *unit* LSTM. Sedangkan untuk indikator NO₂ model terbaik diperoleh dari hasil perhitungan menggunakan 3 *timestep* 5 *unit* LSTM. Model terbaik yang terbentuk dari tiap indikator penentu kualitas udara di stasiun SUF 7 Kota Surabaya dapat dituliskan ke dalam bentuk persamaan.

$$Z_{1,t}^* = 0,1124 + 0,025h_{1,t} - 0,138h_{2,t} + 0,427h_{3,t} - 0,942h_{4,t}.$$

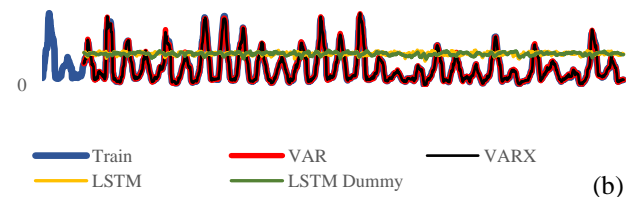
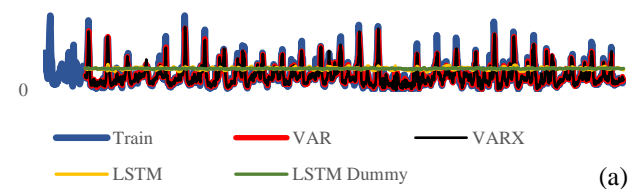
$$Z_{2,t}^* = 0,0394 + 0,927h_{1,t} + 0,653h_{2,t} + 0,122h_{3,t} + 0,914h_{4,t} + 0,615h_{5,t}.$$

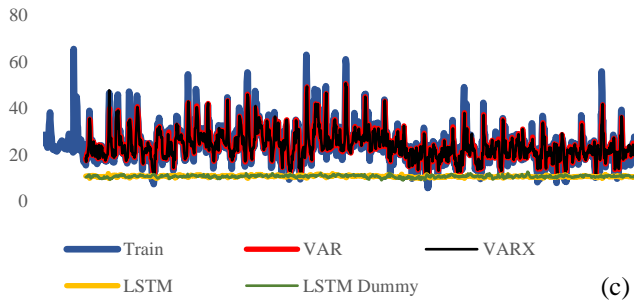
$$Z_{3,t}^* = 0,0674 - 0,394h_{1,t} - 0,295h_{2,t} + 0,295h_{3,t} + 0,682h_{4,t} + 0,223h_{5,t}.$$

$$\text{dengan } Z_{i,t}^* = \frac{Z_{i,t} - \min(Z_{i,t})}{\max(Z_{i,t}) - \min(Z_{i,t})}.$$

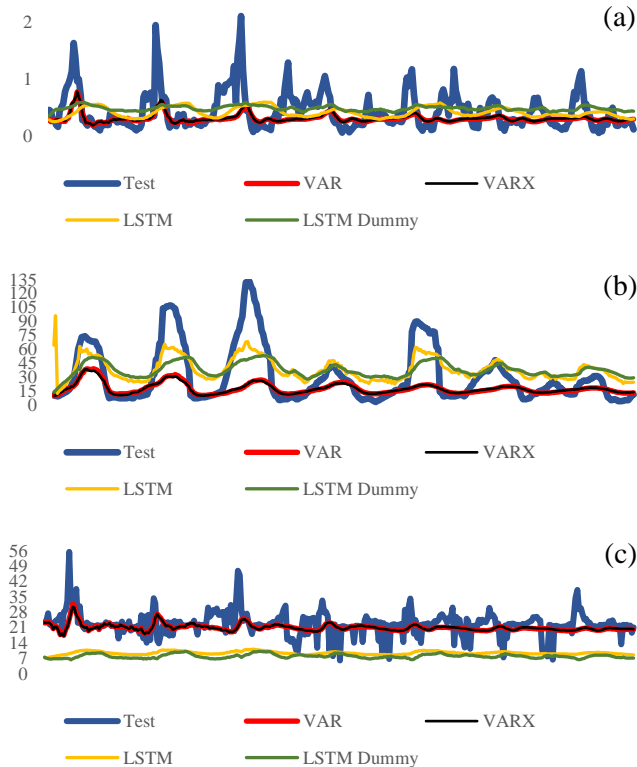
G. Hasil Peramalan dan Pemilihan Model Terbaik

Setelah dilakukan pemodelan, selanjutnya yang perlu dilakukan ialah melakukan peramalan. Plot hasil peramalan untuk indikator CO, PM₁₀ dan NO₂ pada metode VAR, VARX dan LSTM dapat dilihat sebagai berikut.



Gambar 10. Hasil Peramalan Data *Train* Indikator (a) CO (b) PM₁₀ (c) NO₂**Gambar 10.** Hasil Peramalan Data *Train* Indikator (a) CO (b) PM₁₀ (c) NO₂ (Lanjutan)

Hasil peramalan data *train* menggunakan metode VAR dan VARX pada ketiga indikator penentu kualitas udara di stasiun SUF 7 Kota Surabaya yang digunakan memiliki hasil peramalan yang mengikuti data aktual. Sedangkan hasil peramalan menggunakan metode LSTM dan LSTM yang menyertakan variabel *dummy* sebagai variabel *input* belum mengikuti pola dari data aktual, tetapi telah memiliki pola data yang sama.

**Gambar 11.** Hasil Peramalan Data *Test* Indikator (a) CO (b) PM₁₀ (c) NO₂

Hasil peramalan data *test* menggunakan metode VAR, VARX, LSTM serta LSTM dengan variabel *dummy* pada indikator CO dan PM₁₀ cenderung mengikuti data aktual. Metode VAR dan VARX juga menghasilkan hasil peramalan indikator NO₂ yang mengikuti pola dari data aktual, sedangkan hasil peramalan menggunakan LSTM dan LSTM dengan variabel *dummy* kurang menghasilkan hasil peramalan yang mengikuti pola dari data aktual.

Pemilihan model terbaik dilakukan dengan melihat nilai RMSE terkecil yang dihasilkan oleh setiap model. Nilai RMSE

pada setiap indikator dan metode yang digunakan pada penelitian ini dapat dilihat pada Tabel 16.

Tabel 16. Nilai RMSE Terkecil pada Tiap Metode

Data	Indikator	Metode	RMSE
Train	CO	VAR	0.147470914
		VARX	0.14387047
		LSTM	0.340894105
		LSTM Dummy	0.342375885
	PM ₁₀	VAR	3.150165032
		VARX	2.6465331
		LSTM	33.57497798
		LSTM Dummy	33.76596748
	NO ₂	VAR	4.597958256
		VARX	4.54182874
		LSTM	14.84692605
		LSTM Dummy	14.84400652
Test	CO	VAR	0.322961028
		VARX	0.31755301
		LSTM	0.343840701
		LSTM Dummy	0.319201442
	PM ₁₀	VAR	29.35093274
		VARX	29.40259195
		LSTM	22.1012192
		LSTM Dummy	26.41394641
	NO ₂	VAR	5.72878423
		VARX	5.75267423
		LSTM	14.28049417
		LSTM Dummy	15.71237367

Hasil peramalan terbaik untuk data *train* pada indikator CO, PM₁₀ serta NO₂ ialah hasil peramalan yang berasal dari metode VARX, hal ini dapat diketahui berdasarkan nilai kebaikan model atau nilai RMSE yang dihasilkan yang memiliki nilai paling kecil dibandingkan dengan metode lain yang digunakan seperti yang dapat dilihat pada Tabel 16. Hasil peramalan pada data *test* menunjukkan bahwa metode VARX menghasilkan peramalan terbaik untuk indikator CO, metode LSTM menghasilkan peramalan terbaik untuk indikator PM₁₀, dan metode VAR menghasilkan peramalan terbaik untuk indikator NO₂.

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dijabarkan, diperoleh kesimpulan pada penelitian ini ialah sebagai berikut.

1. Ketiga indikator penentu kualitas udara yakni CO, PM₁₀ dan NO₂ di stasiun SUF 7 Kota Surabaya memiliki pola seasonal yaitu musiman harian per setengah jam. Jumlah kadar ketiga indikator cenderung rendah saat sore hingga malam hari dan tinggi saat pagi hingga siang hari.
2. Pada metode VAR diperoleh model VAR (5,1,0)(1,1,0)⁴⁸. Residual dari model VAR (5,1,0)(1,1,0)⁴⁸ tidak memenuhi pengujian asumsi residual *white noise* dan berdistribusi *multivariate normal* sehingga perlu dilakukan deteksi *outlier* pada data yang digunakan.
3. Deteksi *outlier* dilakukan dengan diagram kontrol residual menggunakan prosedur *iterative* dengan statistika T^2 *Hotelling* dan didapati 22 *outlier* setelah dilakukan iterasi

sebanyak empat kali. Selanjutnya dilakukan pembuatan variabel *dummy* sebanyak *outlier* pada data.

4. Metode VARX menghasilkan model VAR (5,5). Residual dari data pada VARX tidak memenuhi pengujian asumsi residual *white noise* dan berdistribusi *multivariate normal* akan tetapi diasumsikan telah memenuhi asumsi residual berdistribusi multivariat normal dan *white noise*.
5. Metode LSTM yang tidak menyertakan variabel *dummy* sebagai variabel *input* didapati hasil model terbaik untuk indikator CO ialah model yang menggunakan *timestep* 98 dan 4 *unit* LSTM. Model terbaik dari indikator PM₁₀ berasal dari model dengan 51 *timestep* dan 5 *unit* LSTM. Model terbaik untuk indikator NO₂ model terbaik diperoleh dari hasil perhitungan menggunakan 3 *timestep* 5 *unit* LSTM. Sedangkan pada metode LSTM yang menyertakan variabel *dummy* sebagai variabel *input* didapati hasil model terbaik untuk indikator CO ialah model yang menggunakan *timestep* 98 dan 4 *unit* LSTM. Model terbaik dari indikator PM₁₀ berasal dari model dengan 51 *timestep* dan 5 *unit* LSTM. Sedangkan untuk indikator NO₂ model terbaik diperoleh dari hasil perhitungan menggunakan 3 *timestep* 5 *unit* LSTM.
6. Model terbaik yang didapatkan pada penelitian ini berasal dari hasil pemodelan menggunakan VARX. Hal ini dapat dilihat melalui hasil peramalan dari metode VARX pada data *train* dipenelitian ini mampu menghasilkan nilai kebaikan model yang paling kecil untuk ketiga indikator penentu kualitas udara dibandingkan kedua metode lain yang digunakan.

B. Saran

Berdasarkan hasil penelitian yang telah diperoleh, saran dari penulis yang dapat diberikan sebagai bahan pertimbangan bagi penelitian selanjutnya adalah sebagai berikut.

1. Terdapat banyak data yang hilang, sehingga perlu diperbaiki kembali manajemen pengumpulan datanya agar dapat menghasilkan model dan ramalan yang lebih baik.
2. Melakukan deteksi *outlier* dengan menggunakan *Multivariate Cumulative Sum* (MCUSUM) dan *Multivariate Exponentially Weighted Moving Averag* (MEWMA).
3. Perlu adanya kebijakan-kebijakan yang dibuat oleh Pemerintah Kota Surabaya untuk mencegah pencemaran udara di Kota Surabaya.

DAFTAR PUSTAKA

- [1] Pemerintah Kota Surabaya, "Transportasi dan Pariwisata," in *Statistik Sektor Kota Surabaya Tahun 2019*, Surabaya, 2019, p. 432.
- [2] S. Fardiaz, *Polusi Air dan Udara*, Yogyakarta: Kanisius, 2003.
- [3] D. Muziansyah, S. Rahayu and S. Sebayang, "Model Emisi Gas Buangan Kendaraan Bermotor Akibat Aktivitas Transportasi (Studi Kasus: Terminal Pasar Bawah Ramayana Kota Bandar Lampung)," *Jurnal Rekayasa Sipil dan Desain*, pp. 57-70, 2015.
- [4] D. N. Wijayanti, *Gambaran dan analisis Risiko Nitrogen Dioksida (NO₂) per Kota/Kabupaten dan Provinsi di Indonesia (Hasil Pemantauan Kualitas Udara Ambien dengan Metode Pasif di Pusarpedal Tahun 2011)*, Jakarta: Skripsi Universitas Indonesia, 2012.
- [5] IQ Air, "IQ Air," 2021. [Online]. Available: <https://www.iqair.com/id/indonesia/east-java/surabaya>. [Accessed 23 February 2021].
- [6] R. A. Robles, J. C. Ortega, J. S. Fu and G. Reed, "A Hybrid ARIMA and Artificial Neural Network Model to Forecast Particulate Matter in Urban Areas: The Case of Temuco, Chile," *Journal Atmosphere Environment*, no. 42, pp. 8331-8440, 2008.
- [7] H. Prabowo and Suhartono, *Peramalan Kualitas Udara di Kota Surabaya untuk menentukan Kategori Indeks Standar Pencemar Udara*, Surabaya: Tugas Akhir Departemen Statistika ITS, 2019, pp. 1-5.
- [8] B. Chrisdayanti, and A. Suharsono, "Peramalan Kandungan Partikular Matter (PM₁₀) dalam Udara Ambien Kota Surabaya Menggunakan Double Seasonal ARIMA," *Jurnal Sains dan Seni ITS*, no. 4(2), pp. 242-247, 2015.
- [9] N. Nahdliyah, Suhartono and M. S. Akbar, *Model Multivariat Generalized Space Time Autoregressive (MGSTAR) untuk Monitoring Kualitas Udara di Surabaya*, Surabaya: Tugas Akhir Departemen Statistika ITS, 2019, pp. 79-104.
- [10] G. E. Box, G. M. Jenkins and G. C. Reinsel, *Time Series Analysis: Forecasting and Control Third Edition*, vol. 62, Englewood Cliffs: Prentice Hall, 1994, pp. 540-552.
- [11] A. Azzouni and G. Pujjole, "A Long-Short Term Memory Recurrent Neural Network Framework for Network Traffic Matrix Prediction," *Networking and Internet Architecture*, pp. 1-5, 2017.
- [12] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum Likelihood from Incomplete Data via The Expectation Maximization (EM) Algorithm," *Journal of The Royal Statistical Society Series B (Methodological)*, pp. 1-38, 1977.
- [13] D. Nettleton, "Selection of Variables and Factor Derivation," in *Commercial Data Mining*, 2014, pp. 79-104.
- [14] W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, New York: Pearson Education, 2006.
- [15] H. Lutkepohl, *New Introduction to Multiple Time Series Analysis*, New York: Springer, 2005.
- [16] S. S. Shapiro and M. B. Wilk, "An Analysis of Variance Test for Normality (Complete Sample)," *Biometrika*, pp. 591-611, 1965.
- [17] R. Odok, *Control Chart for Serially Dependent Multivariate Data*, Technical University Denmark, 2012.
- [18] A. Suharsono, S. Guritno and Subanar, "Autoregressive Vector Modelling Simulation with Innovative Outlier," *Journal of Basic and Applied Scientific Research*, vol. 1, no. 12, pp. 2535-2537, 2011.

- [19] A. Suharsono, *Pemodelan Vector Autoregressive dengan Adanya Outlier atau Pergeseran terhadap Rata-rata*, Yogyakarta: FMIPA Universitas Gajah Mada, 2012.
- [20] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Compute*, pp. 1735-1780, 1997.
- [21] T. Ma, Z. Tao, Y. Wang and H. Yu, "Long Short-Term Memory Neural Network for Traffic Speed Prediction using Remote Microwave Sensor Data," *Transportation Research Part C: Emerging Technologies*, pp. 187-197, 2015.
- [22] Menteri Negara Lingkungan Hidup, Keputusan Menteri Negara Lingkungan Hidup No. 45 Tahun 1977 tentang Indeks Standar Pencemar Udara, Jakarta, 1997.