

Perbandingan Model Peramalan Konsentrasi Partikulat ($PM_{2,5}$) di Stasiun Bundaran HI Dengan Metode ARIMA dan *Simple Moving Average* (SMA)

Nurliana, Chelsea Fatihah Rahma
11210940000026, 11210940000035

Program Studi Matematika, FST, Universitas Islam Negeri Syarif Hidayatullah Jakarta

Abstrak

Konsentrasi partikulat ($PM_{2,5}$) merupakan salah satu penyebab polusi udara di daerah Bundaran HI. Penelitian ini mengkaji model peramalan konsentrasi partikulat ($PM_{2,5}$) di Stasiun Bundaran HI menggunakan metode ARIMA dan *Simple Moving Average* (SMA). Model ARIMA(0,1,1) menghasilkan nilai MAPE sebesar 11,6%. Sebagai perbandingan, metode SMA dengan 8 periode digunakan dan menghasilkan MAPE sebesar 11,05%. Hasil menunjukkan bahwa metode SMA dengan 8 periode memberikan performa yang lebih baik dalam hal keakuratan peramalan dibandingkan dengan model ARIMA(0,1,1). Penelitian ini memberikan wawasan mengenai efektivitas kedua metode dalam peramalan kualitas udara.

Kata kunci: ARIMA, Metode, Peramalan, Perbandingan, SMA

I. PENDAHULUAN

Jakarta masih menjadi salah satu kota dengan kualitas udara terburuk di dunia. Salah satu komponen utama penyebab udara berkualitas buruk adalah partikulat halus atau $PM_{2,5}$ (*particulate matter* dengan diameter kurang dari 2,5 mikrometer). Tingginya angka $PM_{2,5}$ memiliki banyak dampak negatif baik pada kesehatan masyarakat maupun pada lingkungan. Adapun kriteria ambang batas dari $PM_{2,5}$ yakni, baik sebesar $0-15 \mu\text{gr}/\text{m}^3$, sedang sebesar $16-65 \mu\text{gr}/\text{m}^3$, tidak sehat sebesar $66-150 \mu\text{gr}/\text{m}^3$, sangat tidak sehat sebesar $151-250 \mu\text{gr}/\text{m}^3$, dan berbahaya $> 225 \mu\text{gr}/\text{m}^3$ [1].

Berdasarkan data yang diambil, daerah Bundaran HI berada di angka rata-rata tidak sehat dan angka tersebut tetap berada di kisaran tidak sehat setiap harinya. Oleh karena itu, peramalan $PM_{2,5}$ menjadi sangat penting dalam upaya pengendalian kualitas udara dan perlindungan kesehatan masyarakat. Metode peramalan yang digunakan dalam penelitian ini ialah metode ARIMA(*Autoregressive Integrated Moving Average*) dan SMA(*Simple Moving Average*). Dalam melakukan peramalan ada beberapa tantangan saat proses

dilakukan yaitu, ketidak lengkapan data harian yang ada sehingga menimbulkan masalah dalam melakukan peramalan dan keterbatasan metode yang digunakan sehingga hasil yang didapat menjadi kompleks.

Pembandingan metode ARIMA dan SMA dilakukan untuk mendapatkan nilai akurasi terbaik dari salah satu metode. Metode (ARIMA) yaitu suatu metode yang didasarkan pada nilai-nilai suatu peubah yang telah terjadi pada waktu lampau, lalu digunakan untuk menentukan pola historis data, dan kemudian digunakan untuk mengekstrapolasikan pola tersebut pada masa yang akan datang [2]. Metode SMA merupakan metode *forecasting* dengan menghitung nilai rata-rata dari jumlah periode data yang digunakan [3]. Metode SMA ini diyakini memiliki hasil yang lebih akurat karena proses yang lebih sederhana. Namun perlu dikaji nilai MAPE (*Mean Absolute Percentage Error*) untuk mengetahui seberapa akurat metode ini dalam melakukan peramalan.

Sebagai referensi, terdapat penelitian terlebih dahulu mengenai peramalan konsentrasi partikulat ($PM_{2,5}$). Elok melakukan penelitian yang memodelkan $PM_{2,5}$ dengan periode harian menggunakan metode ARIMA berdasarkan hasilnya nilai MAPE terbaiknya yaitu 30.35% yang mana menunjukkan keakratan yang cukup baik [4].

II. METODE PENELITIAN

2.1. Data

Penelitian ini menggunakan data time series konsentrasi partikulat ($PM_{2,5}$) di stasiun Bundaran HI yang bersumber dari artikel data.go dengan periode 1 Oktober 2023- 30 November 2023 atau sebanyak 61 data.

Tabel 1 Data AQI

Hari		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Bulan	Okt	122	122	86	95	89	86	107	92	85	78	74	93	89	102	98	96
	Nov	86	84	89	97	63	73	85	74	89	107	89	79	81	79	66	81

Hari		17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Bulan	Okt	93	108	93	78	88	91	86	96	81	99	102	97	94	83	95
	Nov	75	120	80	75	87	69	93	101	83	75	76	96	82	67	

2.2. Metode

Metode yang digunakan adalah *Autoregressive Integrated Moving Average* (ARIMA) dan *Simple Moving Average* (SMA). Proses pengolahan data akan menggunakan software R studio

2.3. Alur Penyelesaian

- 1) Mencari data konsentrasi partikulat (PM_{2,5}) di stasiun Bundaran HI dari beberapa artikel.
- 2) Membagi data menjadi data training yang akan digunakan untuk membangun model ARIMA pada data periode 1 Oktober 2023-19 November 2023 dan data testing untuk menguji akurasi model pada data periode 20 November 2023-30 November 2023.
- 3) Membuat plot dari data trainnig yang bertujuan untuk melihat kestasioneran pada pola data. Data dapat dikatakan stasioner jika stasioner dalam mean dan varians, stasioner dalam mean jika pergerakan data berada pada bidang horizontal dan stasioner dalam varians jika pergerakan berada dibatas yang sama atau berada di bidang segiempat.
- 4) Mengatasi data yang belum stasioner dalam variansi dengan tranformasi Box-Cox.
- 5) Memeriksa data apakah sudah stasioner dalam mean menggunakan uji KPSS (Kwiatkowski, Phillips, Schmidt and Shin) jika $p\text{-value} < 0.05$, artinya data tidak stasioner . Untuk mengatasi data yang belum stasioner dalam mean dapat dilakukan dengan *differencing* (trend stokastik) atau *de-trending* (trend deterministik).
- 6) Identifikasi model untuk mencari nilai p dan q dengan membuat plot-plot *Autocorrelation Function* (ACF), *Partial Autocorrelation Function* (PACF), *Extended Autocorrelation Function* (EACF), dan *Bayesian Information Criterion* (BIC).
- 7) Estimasi model dapat dilakukan dengan beberapa metode diantaranya ialah metode moment, metode least square, dan metode maksimum likelihood.
- 8) Setelah melakukan identifikasi dan estimasi model tahap berikutnya melakukan pengecekan atau Dignosis model dengan metode analisis residual yang akan dibuat plot residual jika model fit maka plot akan berada dalam bidang segi empat; uji normalitas residual dengan membuat histogram, QQ-plot, uji Shapiro-Wilk, dan uji Jarque-Bera jika pada kedua uji ini memiliki nilai $p\text{-value} > 0.05$ maka residual memenuhi asumsi normal; autokorelasi residual menggunakan uji Box-Ljung untuk mengetahui apakah residual memenuhi asumsi independent. Metode lain untuk diagnosis model ialah analisis overparameterisasi/*overfitting*. Diagnosis model dilakukan untuk

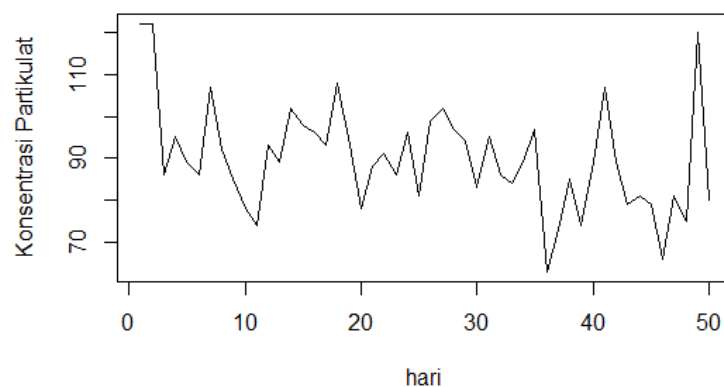
memperoleh model yang fit atau cocok sehingga tahap selanjutnya yaitu forecasting dapat dilakukan.

- 9) Melakukan forecasting atau peramalan untuk 11 hari kedepan menggunakan data testing untuk melakukan perhitungan nilai *Mean Square Error* (MSE) dan *Mean Absolute Percentage Error* (MAPE). Nilai MAPE memiliki beberapa kriteria yakni <10% sangat baik, 10-20% baik, 20-50% cukup baik, dan >50% buruk.
- 10) Melakukan perbandingan menggunakan metode SMA (*Simpel Moving Average*) dengan membandingkan hasil nilai MAPE yang terkecil.

III. HASIL DAN PEMBAHASAN

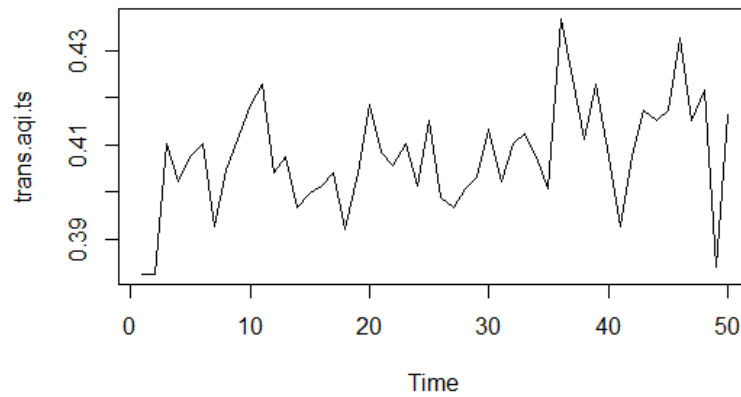
3.1 Analisis Data Menggunakan Metode ARIMA

Berdasarkan Tabel 1 di atas, bagi data menjadi data training dan data testing. Data training dari hari ke-1 sampai hari ke 50 sedangkan data testing dari hari 51 sampai hari ke-61. Data training digunakan untuk membangun model sedangkan data testing digunakan untuk validasi model. Tahap pertama buat plot grafik dari data training untuk dianalisa apakah data sudah stasioner dalam variansi dan mean.



Gambar 1 Plot runtun waktu data AQI

Berdasarkan Gambar 1 terlihat jelas bahwa data tidak stasioner dalam variansi dan mean. Pertama transformasi data agar data stasioner dalam variansi. Dengan menggunakan transformasi Box-Cox didapatkan $\lambda = -0,2$ sehingga data ditransformasi menjadi $y_t^{-0,2}$.



Gambar 2 Plot runtun waktu data AQI setelah transformasi

Selanjutnya kedua periksa kestasioneran dalam mean dengan melakukan uji KPSS, dengan hipotesis:

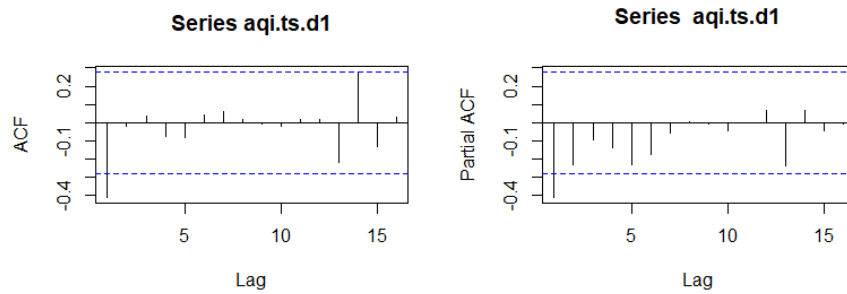
$$H_0 : \text{Stasioner} \quad H_1: \text{Tidak Stasioner}$$

Didapatkan $p\text{-value} < \alpha = 0.03972 < 0,05$ maka tolak H_0 sehingga perlu dilakukan differencing. Setelah di *difrencing* lalu diuji kembali dengan uji KPSS didapatkan $p\text{-value} > \alpha = 0,1 > 0,05$ maka terima H_0 . Karena sudah stasioner dalam variansi dan mean tahap selanjutnya adalah memodelkan data runtun waktu. Langkah-langkah memodelkan data runtun waktu sebelum masuk tahap peramalan (forecasting) adalah identifikasi model, estimasi model, dan diagnosis model.

Langkah pertama identifikasi model, untuk menentukan orde p dan q dari model ARIMA (menentukan model terbaik sementara) pada suatu data runtun waktu dilakukan dengan mengidentifikasi plot ACF, PACF, EACF, dan BIC dari data yang sudah stasioner dalam variansi dan mean.

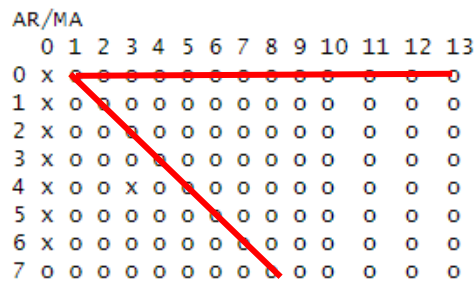
Tabel 2 Penentuan p dan q dengan plot ACF & PACF

Proses	ACF	PACF
AR(p)	Turun cepat secara eksponensial / sinusoidal	Terputus setelah lag p
MA(q)	Terputus setelah lag q	Turun cepat secara eksponensial / sinusoidal
ARMA(p,q)	Turun cepat secara eksponensial / sinusoidal	Turun cepat secara eksponensial / sinusoidal
AR(p) atau MA(q)	Terputus setelah lag q	Terputus setelah lag p
<i>White noise</i> (acak)	Tidak ada yang signifikan (tidak ada yang keluar batas)	Tidak ada yang signifikan (tidak ada yang keluar batas)

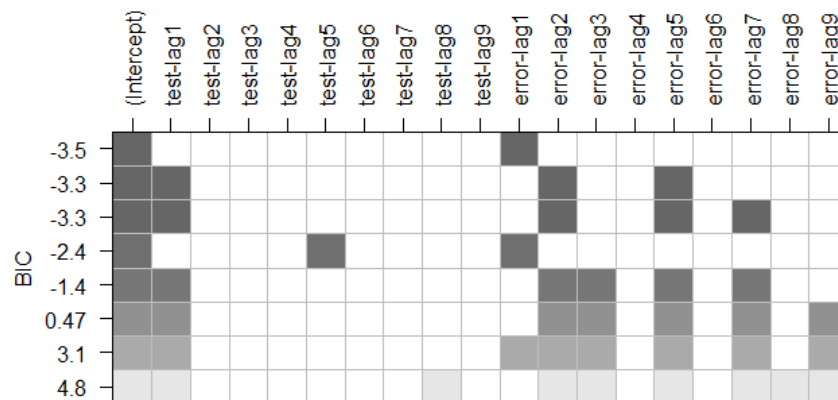


Gambar 3 Plot ACF dan PACF

Dari Gambar 3 menunjukkan PACF menurun secara eksponensial menuju nol dan ACF terputus setelah lag ke-1, sehingga kita kandidat modelnya adalah MA(1). Identifikasi model dengan EACF, Pada Gambar 4 sudut segitiga garis merah berada di MA(1) maka kandidat modelnya MA(1). Dan identifikasi model dengan BIC dengan kandidat model dilihat dari nilai terkecil BIC. Test lag(p) merupakan AR(p) sedangkan error lag(q) merupakan MA(q). Gambar 5 menunjukkan hanya terdapat intercept dan error lag1 maka kandidat modelnya juga MA(1).



Gambar 4 Tabel EACF



Gambar 5 Kriteria Informasi BIC

Dari indentifikasi model diatas dapat kita simpulan $p=0$, $q=1$ (karena kandidat modelnya MA(1), dan $d=1$ (karena di differencing satu kali) maka model ARIMA(0,1,1)=IMA(1,1) adalah model terbaik kita sementara. Langkah kedua

estimasi model, didapatkan nilai $ma1$ sebesar $-0,8663$ dan standard error $0,0960$. Uji rasio $ma1$ dengan hipotesis:

$$H_0: \text{Signifikan} \quad H_1: \text{Tidak signifikan}$$

$$z = \frac{\text{nilai } ma1}{s.e \text{ } ma1} = \frac{-0,8663}{0,0960} = -9,023, \text{ maka } |z| > z_{\frac{\alpha}{2}} = 9,023 > 2 \text{ terima } H_0 \text{ sehingga}$$

$ma1$ signifikan. Sehingga estimasi modelnya adalah sebagai berikut:

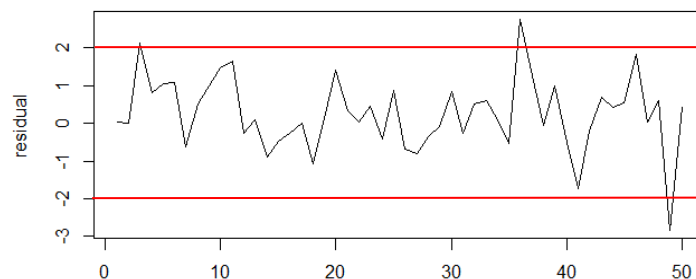
$$y_t^{-0,2} \sim IMA(1,1)$$

$$\Delta y_t^{-0,2} \sim MA(1)$$

$$\Delta y_t^{-0,2} = e_t - 0,8663e_{t-1} \quad (1)$$

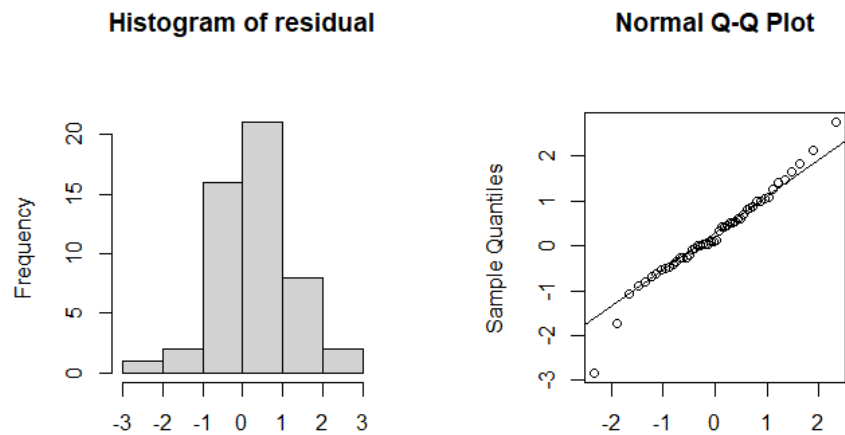
$$y_t^{-0,2} - y_{t-1}^{-0,2} = e_t - 0,8663e_{t-1} \quad (2)$$

Langkah ketiga diagnosis model, tujuannya adalah untuk menunjukkan apakah model $ARIMA(0,1,1)$ model yang terbaik atau tidak. Jika iya maka kita bisa masuk ke tahap selanjutnya yaitu peramalan (*forecasting*). Diagnosis model terbagi menjadi dua metode yaitu analisis residual dan overparameterisasi. Metode pertama analisis residual, terdapat plot residual, uji normalitas residual, dan uji autokorelasi residual.



Gambar 6 Plot Residual

Pada Gambar 6 menunjukkan pergerakan residual berada di bidang persegi panjang dari -2 sampai 2 meskipun terdapat beberapa outlier. Kita dapat menyimpulkan bahwa variansi residual stasioner. Menggunakan uji normalitas residual yaitu dengan histogram, QQ plot, Uji Shapiro Wilk, dan Uji Jarque Bera untuk mengetahui apakah residual normal atau tidak. Pada Gambar 7 histogram menunjukkan bahwa residual cenderung normal dan QQ plot menunjukkan titik-titik residual menyebar berdekatan dengan garis lurus diagonal. Dari histogram dan QQ plot dapat kita simpulkan bahwa residual berdistribusi normal.



Gambar 7 Histogram dan *QQ* plot

Untuk hasil yang lebih meyakinkan maka gunakan uji Shapiro Wilk dan Jarque Bera, dengan hipotesis:

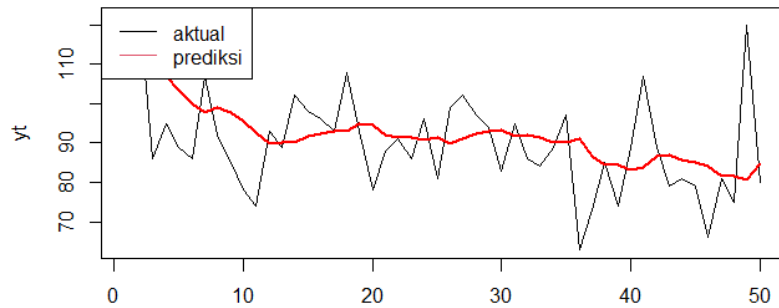
$$H_0 : \text{Normal} \quad H_1: \text{Tidak Normal}$$

Didapatkan p-value kedua tes $> 0,05$ (alpha) maka terima H_0 dapat disimpulkan bahwa residual normal. Menggunakan uji autokorelasi residual dengan Ljung Box untuk mengetahui apakah residual independet atau tidak. Pada plot ACF lag menunjukkan sampai lag ke-16 maka hipotesisnya adalah:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_{16} = 0 \quad H_1: \exists \rho_i \neq 0, i = 1, 2, \dots, 16$$

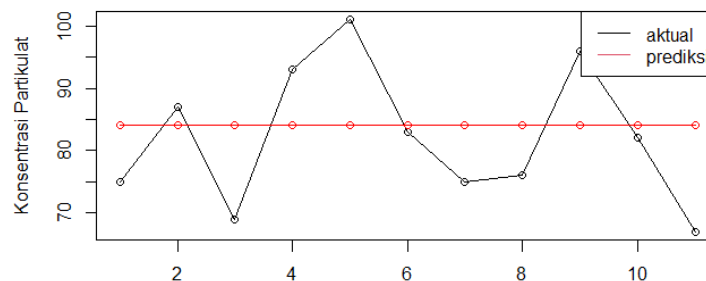
Didapatkan p-value $> \alpha = 0.8717 > 0,05$ maka terima H_0 dapat disimpulkan bahwa residual independent. Metode kedua overparameterisasi, membuat model dengan menaikkan order satu tingkat. Terdapat dua kandidat yaitu dengan model ARIMA(0,1,2) dan ARIMA(1,1,1). Pada model ARIMA(0,1,2) ma2 tidak signifikan dan model ARIMA(1,1,1) ar1 tidak signifikan. Kita juga dapat melihat dari nilai AIC, nilai AIC pada model ARIMA(0,1,1) yaitu -295,29 lebih kecil dibandingkan dengan model ARIMA(0,1,2) yaitu -294,65 dan ARIMA(1,1,1) yaitu -294,73. Dengan demikian model terbaik yang kita pilih atau model FIT adalah ARIMA(0,1,1).

Tahap terakhir adalah peramalan (forecasting) dengan model FIT yang kita dapatkan. Pertama kita prediksi nilai y_t dengan perintah fitted() lalu bandingkan hasil prediksi dengan data aktual. Pada Gambar 8 menunjukkan bahwa hasil prediksi bergerak tidak jauh dari data aktual.



Gambar 8 Perbandingan hasil prediksi dengan data aktual

Gunakan perintah `forecast()` untuk memvalidasi model dengan cara memprediksi konsentrasi partikulat ($PM_{2,5}$) 11 hari kedepan. Didapatkan nilai prediksi dari hari ke-51 sampai hari ke-61 sama yaitu 84,13273. Gambar 9 merupakan perbandingan data aktual dan nilai prediksi.

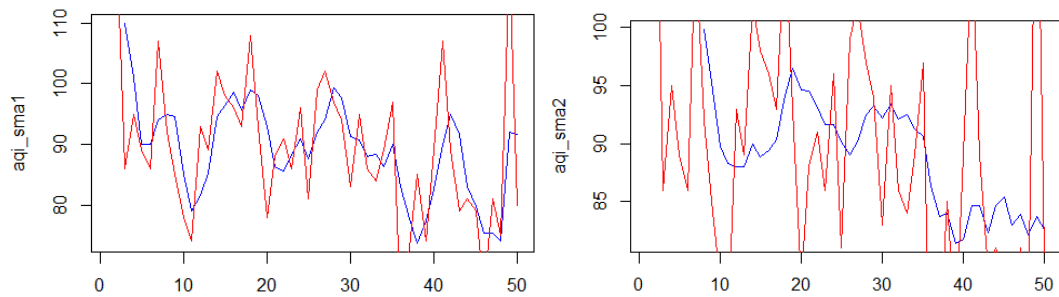


Gambar 9 Perbandingan data testing (data aktual) dan nilai prediksi

Bandingkan data testing dengan hasil prediksi dengan dua cara yaitu menghitung MSE (*Mean Square Error*) atau menghitung MAPE (*Mean Absolute Percentage Error*). Didapatkan nilai MSE sebesar 115.773 dan nilai MAPE sebesar 0,1167 atau 11,67% dimana lebih kecil dari 20% ini menunjukkan bahwa model peramalan dengan ARIMA(0,1,1) baik.

3.2 Analisis Data Menggunakan Metode SMA

Menggunakan data training dengan n periode SMA yaitu merupakan perhitungan rata-rata dari sejumlah n periode ke belakang. Kita pilih $n = 3$ yaitu dengan perhitungan rata-rata dari sejumlah 3 periode ke belakang sedangkan $n = 8$ yaitu dengan perhitungan rata-rata dari sejumlah 8 periode ke belakang. Didapatkan plot perbandingan data aktual dan nilai prediksi dengan pada Gambar 10.



Gambar 10 Plot $n = 3$ (kiri) dan $n = 8$ (kanan)

Selanjutnya gunakan perintah `forecast()` dengan memprediksi 11 hari kedepan dengan $n = 3$ diperoleh nilai MSE sebesar 201.9299 dan nilai MAPE sebesar 16,13% sedangkan dengan $n = 8$ diperoleh nilai MSE sebesar 112.1635 dan nilai MAPE sebesar 11,05%.

IV. KESIMPULAN DAN SARAN

Nilai MAPE dengan metode SMA dengan 8 periode lebih kecil dibandingkan nilai MAPE dengan metode ARIMA(0,1,1) sehingga metode SMA dengan 8 periode memberikan performa yang lebih baik dalam hal keakuratan peramalan dibandingkan dengan model ARIMA(0,1,1).

V. REFERENSI

- [1] Peraturan Badan Meteorologi, Klimatologi, Dan Geofisika Republik Indonesia Nomor 2 Tahun 2020 Tentang Penyediaan Dan Penyebaran Informasi Kualitas Udara.
- [2] R. Ayu Wulandari dan Rahmat Gernowo, “METODE AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) DAN METODE ADAPTIVE NEURO FUZZY INFERENCE SYSTEM (ANFIS) DALAM ANALISIS CURAH HUJAN,” 2019.
- [3] I. Solikin, S. Hardini, F. E. Citra Sari, and C. M. Chaiago, “Membangun Aplikasi Metode WMA dan Metode SMA Sebagai Support System Pengambilan Keputusan,” *Jurnal Bumigora Information Technology (BITe)*, vol. 4, no. 1, pp. 107–114, Jun. 2022, doi: 10.30812/bite.v4i1.1938.
- [4] Zulfa. Elok Indana, “PEMODELAN KANDUNGAN PARTICULATE MATTER 10 DAN 2.5 DALAM PENCEMARAN UDARA PROVINSI DKI JAKARTA MENGGUNAKAN METODE ARIMA” 2023

VI. APPENDIX

```
## Load package 'TSA' dan 'forecast'  
#install.packages("TSA")  
#install.packages("forecast")  
#install.packages("readxl")
```

```
library(readxl)  
library(TSA)  
library(forecast)  
library(tseries)
```

```
## Load data dan simpan sebagai objek aqi  
setwd("C:/Users/ADMINT/Documents/ARW")  
aqi <- read.csv('data_AQI.csv', header = TRUE)
```

```
## Ringkasan data Aqi  
head(aqi)  
dim(aqi)  
n <- length(aqi$PM_duakomalima);n
```

```
## Bagi data menjadi dua bagian: data training dan testing (70-90%, 30%-10%)  
## Data training untuk membangun model ARIMA (50 hari)  
## Data testing untuk validasi model (11 hari)  
aqi.training <- aqi$PM_duakomalima[1:(n-11)]  
aqi.testing <- aqi$PM_duakomalima[(n-10):n]
```

```
plot(aqi.training)  
## Gunakan perintah ts() untuk merubah konsentrasi partikulat (PM2,5) harian menjadi data  
runtun waktu  
aqi.ts <- ts(aqi.training)
```

```

## Buat plot runtun waktu dari data aqi
plot(aqi.ts, ylab = "Konsentrasi Partikulat", xlab = "hari")

## Cek perlu tidaknya melakukan transformasi data untuk mengatasi non-stasioner dalam
variansi
bc <- BoxCox.ar(aqi.ts)
bc
trans.aqi.ts <- (aqi.ts)^-0.2
plot(trans.aqi.ts)

## Cek apakah data sudah stasioner dalam mean (uji adf atau uji kpss)
## Uji trend stokastik dengan uji KPSS (H0: stasioner vs H1: non-stasioner)
kpss.test(trans.aqi.ts, null="Level")
#p-value hasilnya  $0.03 < 0.05$  Ho ditolak, artinya data tidak stasioner dan perlu differencing

## Plot diferensi ke-1 dari trans.aqi.ts
aqi.ts.d1 <- diff(trans.aqi.ts, d = 1)
plot(aqi.ts.d1)

##Uji stasioneritas dari diferensiasi ke-1 aqi.ts.d1: (H0: stasioner vs H1: non-stasioner)
kpss.test(aqi.ts.d1)
#p-value  $(0.1) > 0.05$ , Ho diterima, tidak ada indikasi non-stasioner dalam mean

##### TAHAP 1: IDENTIFIKASI MODEL #####
## Identifikasi Model dengan ACF dan PACF dari aqi.ts.d1
par(mfrow=c(1,2))
acf(aqi.ts.d1)
pacf(aqi.ts.d1)
# kandidat model:MA(1)

## Identifikasi Model dengan EACF
eacf(aqi.ts.d1)
# kandidat model:MA(1)

```

```
## Identifikasi Model dengan kriteria informasi BIC
```

```
par(mfrow=c(1,1))
```

```
res <- armasubsets(y=aqi.ts.d1,nar=9,nma=9,y.name='test',ar.method='ols')
```

```
plot(res)
```

```
# kandidat model: MA(1)
```

```
### KESIMPULAN --> p =0, q =1, d =1 (ARIMA(0,1,1)=IMA(1, 1))
```

```
##### TAHAP 2: ESTIMASI MODEL #####
```

```
## Estimasi parameter model dari trans.aqi.ts sebagai proses ARIMA dengan p= 0, q=1 dan d=1
```

```
aqi.arima.011 <- stats::arima(trans.aqi.ts,order=c(0,1,1))
```

```
aqi.arima.011
```

```
#Gunakan perintah auto.arima() dari package 'forecast' - Secara otomatis untuk menentukan order p, q, & d sekaligus memberikan estimasi parameter model
```

```
aqi.auto <- auto.arima(trans.aqi.ts)
```

```
aqi.auto
```

```
##### TAHAP 3: DIAGNOSIS MODEL #####
```

```
residual <- rstandard(aqi.arima.011)
```

```
## Plot residual
```

```
plot(residual)
```

```
## Cek normalitas dengan histogram dan QQ plot
```

```
par(mfrow=c(1,2))
```

```
hist(residual)
```

```
qqnorm(residual)
```

```
qqline(residual)
```

```

## Uji normalitas
shapiro.test(residual)
jarque.bera.test(residual)
# p-value shapiro tes > 0.05, Ho diterima, residual normal
# p-value jarque bera tes > 0.05, Ho diterima, residual normal

## Cek autokorelasi: individu
par(mfrow=c(1,1))
acf(residual)

## Uji autokorelasi dengan ljung.box: serentak
Box.test(residual,lag=15,type = c("Ljung-Box"))
# p-value > 0.05, tidak signifikan, Ho diterima, residual independent

## Diagnosis dengan overfitting
aqi.arima.011
aqi.arima.012 <- stats::arima(trans.aqi.ts,order=c(0,1,2))
aqi.arima.012
aqi.arima.111 <- stats::arima(trans.aqi.ts,order=c(1,1,1))
aqi.arima.111

# parameter tambahan ma2 tidak signifikan & parameter ma1 tidak berubah drastis
# parameter tambahan ar1 tidak signifikan & parameter ma1 tidak berubah drastis

##### TAHAP 4: FORECASTING #####

## Plot data sebenarnya (yt) dan hasil prediksi model ARIMA(0,1,1)
actual <- aqi.ts

## perintah fitted() digunakan untuk mengeluarkan hasil prediksi yt
fitted(aqi.arima.011)
pred <- fitted(aqi.arima.011)^(-10/2) # transformasi balik ke data aktual

```

```

par(mfrow=c(1,1))
plot(actual, ylab = "yt")
lines(pred, col = "red", lwd = 2)
legend("topleft", c("aktual", "prediksi"), col=c(1,2), lty = 1)

## Gunakan perintah forecast() untuk memvalidasi model
## dengan cara memprediksi konsentrasi partikulat (PM2,5) 11 hari kedepan
aqi.forecast <- forecast(aqi.arima.011, 11)
aqi.forecast
plot(aqi.forecast)
prediksi <- aqi.forecast$mean # simpan hasil prediksi 11 hari kedepan ke objek prediksi
prediksi2 <- prediksi^(-10/2) # transformasi hasil prediksi ke dalam skala data asli

## data testing
aqi.testing <- aqi$PM_duakomalima[(n-10):n]

## plot data sebenarnya (yt) dengan hasil peramalan
plot(aqi.testing,type='o', ylab = "Konsentrasi Partikulat", xlab = "waktu (hari)")
lines(ts(prediksi2),col='red',type='o')
legend("topright", c("aktual", "prediksi"), col=c(1,2), lty = 1)

## Bandingkan data testing dengan hasil prediksi
## Cara 1: hitung MSE (Mean Square Error)
mean((aqi.testing-prediksi2)^2) #MSE

## Cara 2: hitung MAPE (Mean Absolute Percentage Error)
mean(abs((aqi.testing - prediksi2)/aqi.testing)) #MAPE
#MAPE = 68% (model prediksi yang baik < 20%)

##### Analisis runtun waktu dengan metode SMA #####
#install.packages("TTR")
#install.packages("Metrics")

```

```

library(TTR)
library(Metrics)

#untuk n=3
aqi_sma1<-SMA(aqi.ts, n = 3)
plot(aqi_sma1)
lines(aqi_sma1, col="blue")
lines(aqi.ts, col="red")

prediksi1<-forecast(aqi_sma1, h=11)
autoplot(prediksi1)
lines(aqi.ts, col='red')
prediksi_mean<-prediksi1$mean
mean((aqi.testing-prediksi_mean)^2) #MSE
mean(abs((aqi.testing - prediksi_mean)/aqi.testing)) #MAPE

#untuk n=8
aqi_sma2<-SMA(aqi.ts, n =8)
plot(aqi_sma2)
lines(aqi_sma2, col="blue")
lines(aqi.ts, col="red")

prediksi2<-forecast(aqi_sma2, h=11)
autoplot(prediksi2)
prediksi_mean<-prediksi2$mean
mean((aqi.testing-prediksi_mean)^2) #MSE
mean(abs((aqi.testing - prediksi_mean)/aqi.testing)) #MAPE

```