



Peramalan Kualitas Udara di Semarang Menggunakan Metode *Autoregressive Integrated Moving Average (ARIMA)*

Forecasting Air Quality in Semarang Using the Autoregressive Integrated Moving Average (ARIMA) Method

Nida Faoziatun Khusna^{1,2}, Syifa Aulia^{1,2}, Shinta Amaria^{1,2}, Alfidha Rahmah^{1,2}, Safril Ahmadi Sanmas^{1,2}, Fatkhurokhman Fauzi^{1,3}

¹Department of Statistics, Faculty of Science Technology and Agriculture, Universitas Muhammadiyah Semarang, Semarang

²Health and Environmental Research Laboratory, Faculty of Science Technology and Agriculture, Universitas Muhammadiyah Semarang, Semarang

³Computing and Data Science Research Laboratory, Faculty of Science Technology and Agriculture, Universitas Muhammadiyah Semarang, Semarang

Corresponding author: nidafaok@gmail.com

Abstrak

Pencemaran udara menjadi suatu permasalahan yang dialami oleh seluruh wilayah di Indonesia. Kota Semarang menjadi salah satu kota dengan polusi udara terburuk ke-10 di Indonesia. Untuk meminimalisir dampak dari pencemaran udara, perlu dilakukan peramalan nilai $PM_{2.5}$ yaitu partikel udara super kecil yang apabila terhirup dapat membahayakan kesehatan. Metode ARIMA adalah salah satu metode yang digunakan untuk memodelkan data *time series* dan bertujuan untuk meramalkan data pada waktu yang akan datang. Data pengamatan yang digunakan adalah data harian $PM_{2.5}$ di Semarang mulai tanggal 3 Januari sampai 22 September 2023. Hasil analisis pada penelitian ini mendapatkan model yang sesuai untuk data harian $PM_{2.5}$ yaitu model ARIMA (1,1,1) dengan *Mean Absolute Percentage Error (MAPE)* sebesar 13.71377 atau 13.714%. Hasil peramalan menunjukkan terjadinya peningkatan nilai $PM_{2.5}$ dari waktu sebelumnya. Dengan demikian, nilai kualitas udara di Semarang untuk waktu yang akan datang berada dalam tahap tidak sehat dan terjadi peningkatan pencemaran udara.

Kata Kunci : ARIMA, Kualitas udara, Peramalan.

Abstract

Air pollution has become a problem experienced by all regions in Indonesia. Semarang City is one of the top 10 cities with the worst air pollution in Indonesia. To minimize the impact of air pollution, it is necessary to forecast the $PM_{2.5}$ value, which is a very small airborne particle that can be harmful when inhaled. The ARIMA method is one of the methods used to model time series data and aims to predict future data. The observational data used is the daily $PM_{2.5}$ data in Semarang from January 3 to September 22, 2023. The analysis in this study found a suitable model for daily $PM_{2.5}$ data, which is the ARIMA (1,1,1) model with a Mean Absolute Percentage Error (MAPE) of 13.71377 or 13.714%. The forecast results indicate an increase in the $PM_{2.5}$ value compared to previous times. Therefore, the air quality in Semarang for the future is expected to be unhealthy, with an increase in air pollution.

Keywords : Air quality, ARIMA, Forecasting.

PENDAHULUAN

Udara adalah campuran gas pada atmosfer bumi yang mempengaruhi kehidupan manusia serta makhluk hidup lainnya. Udara merupakan sumber tidak terbatas milik bersama yang mempengaruhi kehidupan manusia serta makhluk hidup lainnya. Semua makhluk hidup membutuhkan udara untuk bernapas. Oleh karena itu, pemeliharaan terhadap fungsi udara harus dilakukan demi menjaga kesehatan dan kesejahteraan manusia serta melindungi makhluk hidup lainnya. Udara yang baik ialah udara yang memiliki kualitas baik dan tidak mengandung unsur berbahaya. Turunnya kualitas udara sehingga udara mengalami penurunan mutu dalam penggunaannya dan akhirnya tidak

dapat digunakan kembali sebagaimana mestinya disebut sebagai pencemaran udara (Putri Imas et al., 2020). Beberapa gas pencemar udara adalah karbon monoksida (CO), karbon dioksida (CO₂), nitrogen oksida (NO_x), nitrogen dioksida (NO₂), Particulate Matter (PM_{2.5}) dan sebagainya (Ari Pani, 2015).

PM_{2.5} merupakan partikel udara yang berukuran lebih kecil dari 2.5 μm (mikrometer). Partikel ini dapat diukur menggunakan metode penyinaran sinar Beta (*Beta Attenuation Monitoring*) dengan satuan μgr/m³ (mikrogram per meter kubik). Badan Meteorologi, Klimatologi, dan Geofisika mengklasifikasikan nilai PM_{2.5} menjadi lima kategori yaitu baik (0-15,5 μgr/m³), sedang (15,6-55,4 μgr/m³), tidak sehat (55,5-150,4 μgr/m³), sangat tidak sehat (150,5-250,4 μgr/m³), dan berbahaya (> 250,5 μgr/m³).

Paparan pencemaran udara bagi manusia dapat menyebabkan timbulnya berbagai penyakit. Dalam jangka pendek, polusi udara dapat menyebabkan gangguan eksim, jerawat, penuaan dini, serangan jantung, flu, rinitis, dan gangguan ADHD yang terjadi pada bayi sejak dalam kandungan hingga masa kanak-kanak. Dalam jangka panjang, polusi udara dapat menyebabkan penyumbatan pembuluh darah, kelahiran prematur, pneumonia, kanker paru-paru, asma, alzheimer, parkinson, stroke dan penurunan kognitif.

Pada tahun 2022, Indonesia menempati peringkat pertama di Asia Tenggara dengan kualitas udara terburuk. Laporan tersebut dibuat oleh IQAir menggunakan metode analisis pengukuran kualitas udara PM_{2.5}. Pencemaran udara menjadi suatu permasalahan yang dialami oleh seluruh wilayah di Indonesia, salah satunya Kota Semarang. Polusi udara Kota Semarang berada dalam kategori sedang dan berada di urutan ke-10 sebagai kota dengan polusi udara terburuk sepanjang Agustus 2023.

Dampak dari pencemaran udara dapat diminimalisir dengan pengontrolan kualitas udara melalui peramalan untuk masa mendatang. Metode peramalan yang sering digunakan adalah ARIMA. *Autoregressive Integrated Moving Average* (ARIMA) merupakan model matematika yang digunakan untuk peramalan deret waktu univariat berdasarkan *autoregressive* (AR), *differencing* (I), dan *moving average* (MA). Model yang dihasilkan dapat digunakan untuk data tidak stasioner dengan menambahkan *differencing* (Samal et al., 2019).

Penelitian mengenai ARIMA telah banyak dilakukan sebelumnya. Penelitian oleh Atika Tresna Arianto dkk (2021) dengan judul “Peramalan Konsentrasi Particulate Matter 2.5 (PM_{2.5}) menggunakan Model *Vector Autoregressive* dengan Metode *Maximum Likelihood Estimation*” menghasilkan bahwa nilai MAPE dari penerapan model VAR(1) dengan metode MLE pada data Konsentrasi PM_{2.5} di Kab. Cirebon, Kab. Kuningan, Kab. Majalengka adalah 9.37, 8.75, dan 5.72. Liangxiao Zhao, Zhiyang Li, dan Leilei Qu (2022) melakukan penelitian dengan judul “*Forecasting of Beijing PM_{2.5} with a hybrid ARIMA model based on integrated AIC and improved GS fixed-order methods and seasonal decomposition.*” Hasil penelitian menunjukkan bahwa model hybrid ARIMA (ARIMA-add dan ARIMA-mul) memiliki kinerja prediksi yang baik dengan nilai RMSE 99.23%, MAE sebesar 99.20%, R² sebesar 118.61%, TIC 99.28%, NMAE sebesar 98.71%, NMSE sebesar 99.97%, OPC 43.13%, dan indeks CEC sebesar 99.25%. Kemudian, penelitian oleh Jiahui Duan, Yaping Gong, dan Zhiyao Zhao (2023) dengan judul “*Air-quality prediction based on the ARIMA-CNN-LSTM combination model optimized by dung beetle optimizer*” menghasilkan nilai RMSE pada keempat kota penelitian adalah 7.59, 14.94, 7.84, dan 5.50, nilai MAE sebesar 5.29, 10.84, 5.12, dan 3.77, serta nilai R² sebesar 0.99, 0.96, 0.95, dan 0.95.

Berdasarkan penjelasan sebelumnya, maka perlu dilakukan peramalan mengenai data runtun waktu PM_{2.5} di Kota Semarang untuk waktu yang akan datang. Peramalan akan dilakukan dengan menggunakan metode ARIMA. Penelitian ini bertujuan untuk menentukan satu bentuk model statistik yang sesuai dan meramalkan untuk data pencemaran udara PM_{2.5} di Kota Semarang dengan menggunakan metode ARIMA. Oleh sebab itu, dengan adanya hasil peramalan ini, maka Pemerintah Kota Semarang dapat menjadikan penelitian ini sebagai pedoman untuk mengambil kebijakan di masa yang akan datang dalam mengatasi pencemaran udara.

METODE

1. Sumber Data

Penelitian ini merupakan penelitian kuantitatif dengan menggunakan data sekunder yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Data yang digunakan berbentuk *time series* dalam periode waktu 3 Januari sampai 22 September 2023. Data yang diolah adalah data per hari dengan jumlah sebanyak 263 hari.

2. Metode ARIMA

Metode peramalan yang digunakan dalam penelitian ini adalah metode Autoregressive Integrated Moving Average (ARIMA). Metode ini juga dikenal sebagai metode *Box Jenkins*. Menurut *Box-Jenkins* metode ARIMA terdiri dari empat tahap yaitu identifikasi metode *time series*, pendugaan parameter-parameter bagi metode alternatif, pengujian metode, dan perkiraan nilai *time series*. Metode ARIMA dikembangkan oleh Wold dengan mengembangkan metode ARMA pada tiga arah identifikasi efisien dan prosedur penaksiran (untuk proses AR(Autoregressive), MA (*Moving Average*), dan ARMA (campuran)), perluasan hasil tersebut untuk mencari deret berskala musiman dan pengembangan sederhana yang mencakup proses-proses non-stasioner (ARIMA) (Hartati, 2017).

Asumsi yang digunakan pada metode ini adalah data *time series* yang bersifat stasioner. Dalam hal ini artinya *mean* dan varian dari data bersifat konstan. Apabila data dari *mean* tidak stasioner maka perlu dilakukan *differencing* dan apabila data dari varian maka dilakukan *transformasi*. Bentuk umum model ARIMA adalah sebagai berikut.

$$Z_t = b_0 + b_1 Z_{t-1} + \dots + b_p Z_{t-p} + e_t - c_1 e_{t-1} - \dots - c_q e_{t-q} \quad (1)$$

Berdasarkan persamaan diatas Z_t merupakan data deret waktu sebagai variabel dependen waktu ke- t . Z_{t-p} merupakan data deret waktu saat kurun waktu ke- $(t-p)$. b_1, b_p, c_1, c_q merupakan parameter model dan e_{t-q} adalah nilai *error* pada saat waktu ke- $(t-q)$.

3. Tahapan Penelitian

Langkah-langkah yang dilakukan dalam melakukan peramalan kualitas udara di Kota Semarang menggunakan metode ARIMA adalah sebagai berikut:

- a. Analisis Deskriptif
Pada tahap awal ini data yang diperoleh dicari analisis deskriptifnya berupa rata-rata, median, standar deviasi, nilai maksimum dan nilai minimum. Kemudian, membuat plot *time series* pada data dan memeriksa kestasioneran data.
- b. Uji Kestasioneran Data

Model ARIMA hanya digunakan pada data yang stasioner maka dilihat dari plot time series dapat memeriksa kestasioneran data. Apabila data bersifat non-stasioner dalam rata-rata maka diperlukan proses *differencing* hingga data stasioner.

- c. Identifikasi Model
Selanjutnya membuat plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) yang digunakan untuk mengidentifikasi model ARIMA sementara dan dilakukan uji signifikansi
- d. Pemilihan Model Terbaik
Dalam memilih model terbaik dapat dilihat dari nilai AIC (*Akaike's Information Criterion*) yang terkecil.
- e. Uji Asumsi Residual
Setelah diperoleh model terbaik, selanjutnya uji asumsi residual dari model terbaik tersebut yang terdiri dari Uji *White Noise* dan Uji Normalitas.
- f. Peramalan
Setelah model terbaik tersebut dapat memenuhi uji asumsi residual, maka dapat dilakukan peramalan.

HASIL DAN PEMBAHASAN

1. Analisis Deskriptif

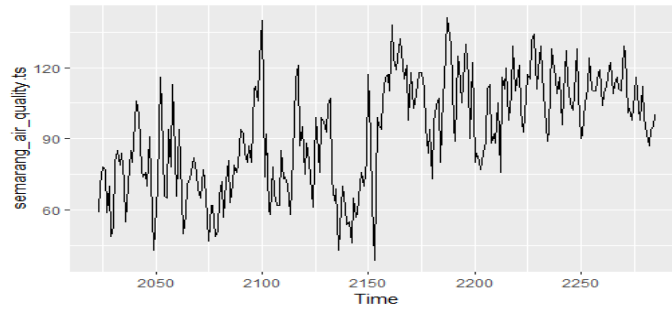
Sebelum dilakukan analisis data untuk peramalan, diperlukan analisis deskriptif terlebih dahulu. Analisis tersebut digunakan untuk menggali informasi dari data. Berikut merupakan analisis deskriptif yang digunakan berupa rata-rata, median, standar deviasi, nilai maksimum, dan nilai minimum.

Tabel 1.
Analisis Deskriptif

Analisis	Nilai
Rata-rata	92,64
Median	94,00
Standar Deviasi	22,94
Maksimum	141,00
Minimum	39,00

Tabel 1 menunjukkan nilai rata-rata data kualitas udara sebesar 92,64 yang dapat dikategorikan sebagai kualitas udara yang tidak sehat. Selain itu, memiliki nilai median sebesar 94; standar deviasi sebesar 22,94; nilai maksimum sebesar 141 dan nilai minimum sebesar 39. Berikut diberikan grafik dari data aktual kualitas udara di Kota Semarang.

Gambar 1:
Plot Data PM_{2.5}



Berdasarkan Gambar 1 terlihat bahwa pola data kualitas udara bersifat tidak stasioner karena data memiliki pola trend dan musiman. Pola *trend* dapat dideteksi dari pola data yang terus mengalami kenaikan, sedangkan pola musiman terlihat dari pola data yang menunjukkan naik dan turun dalam jangka waktu yang tetap.

2. Uji Kestasioneran Data

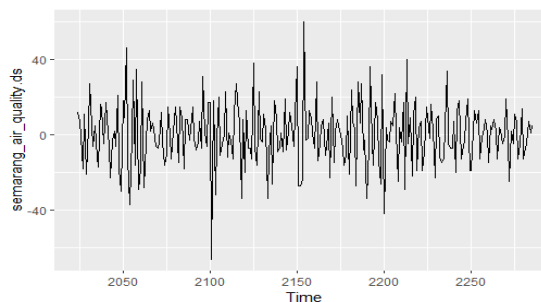
Terlihat pada Gambar 1 bahwa data tidak berada disekitar rata-rata sehingga tidak stasioner sehingga perlu dilakukan proses *differencing*. Kemudian dapat dilakukan pengujian kembali kestasioneran data menggunakan uji *Augmented Dickey-Fuller* (ADF). Berikut merupakan hasil uji ADF setelah dilakukan proses *differencing* sebanyak satu kali.

Tabel 2.
Hasil Pengujian ADF

Uji	<i>P-value</i>	Keterangan
ADF Test	0,01	Stasioner

Tabel 2 menunjukkan bahwa hasil uji ADF memiliki nilai *p-value* sebesar 0,01 atau lebih kecil dari signifikansinya 0,05. Hal tersebut dapat diartikan bahwa dengan proses *differencing* sebanyak satu kali, data telah menjadi stasioner. Kestasioneran data juga dapat dilihat pada hasil plot data sebagai berikut.

Gambar 2:
Plot Data Differencing

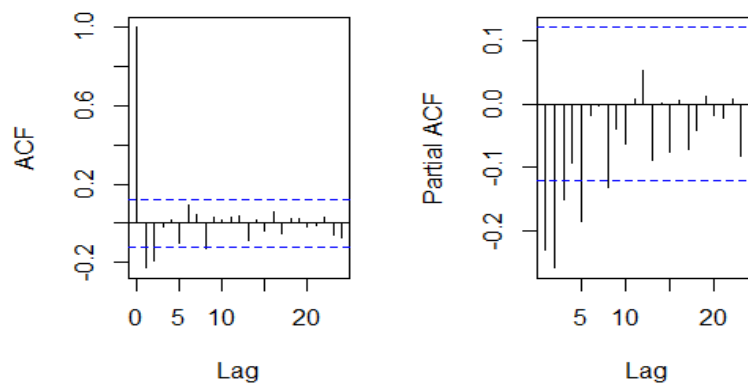


Dilihat pada Gambar 2 bahwa data berfluktuasi disekitar titik nol (konstan). Berdasarkan gambar tersebut dapat dikatakan bahwa data telah stasioner dalam rata-rata maupun variansi.

3. Identifikasi Model

Identifikasi model dari data dapat dilakukan dengan memplotkan data kualitas udara yang telah di differencing ke dalam plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF). Berikut merupakan plot ACF dan PACF.

Gambar 3:
ACF dan PACF Differencing



Dilihat dari gambar diatas bahwa pada plot PACF dari 4 lag pertama ada lag 1, lag 2 dan lag 3 yang berada diluar garis signifikan, sedangkan pada plot ACF dari 4 lag pertama ada lag 1, lag 2 dan lag 3 yang berada diluar garis signifikan yang menunjukkan maka menunjukkan adanya proses Autoregressive (AR) berorde 3 dan Moving Average (MA) berorde 3. Sehingga didapatkan model-model ARIMA sementara yang mungkin adalah sebagai berikut.

- Model 1 : ARIMA (0,1,2)
- Model 2 : ARIMA (1,1,0)
- Model 3 : ARIMA (1,1,1)
- Model 4 : ARIMA (3,1,0)
- Model 5 : ARIMA (3,1,3)

Setelah model sementara telah diperoleh, kemudian dilakukan uji signifikansi untuk menguji apakah parameter model layak masuk dalam model atau tidak. Berikut merupakan hasil uji signifikansi model ARIMA sementara.

Tabel 3.
Parameter Model

No	Model	Parameter	P-value	Keterangan
1	(0,1,2)	MA (1)	0,000	Signifikan
		MA (2)	0,000	Signifikan

2	(1,1,0)	AR (1)	0,000	Signifikan
		AR (1)	0,000	Signifikan
3	(1,1,1)	MA (1)	0,000	Signifikan
		AR (1)	0,000	Signifikan
4	(3,1,0)	AR (2)	0,000	Signifikan
		AR (3)	0,000	Signifikan
		AR (1)	0,023	Signifikan
		AR (2)	0,000	Signifikan
		AR (3)	0,000	Signifikan
5	(3,1,3)	MA (1)	0,000	Signifikan
		MA (2)	0,000	Signifikan
		MA (3)	0,000	Signifikan

4. Pemilihan Model Terbaik

Salah satu cara yang umum digunakan dalam penelitian model terbaik yaitu melihat dari nilai Akaike's Information Criterion (AIC) yang terkecil. Berikut merupakan hasil nilai AIC pada masing-masing model sementara.

Tabel 4.
Hasil AIC Model

Model	AIC
(0,1,2)	2166.64
(1,1,0)	2199.19
(1,1,1)	2164.69
(3,1,0)	2178.87
(3,1,3)	2165.18

Dari hasil di atas, diperoleh model dengan nilai AIC terkecil yaitu model ARIMA (1,1,1) dan menjadi model terbaik yang dapat digunakan untuk peramalan. Adapun persamaan yang diperoleh adalah sebagai berikut.

$$x_t = 0,6017x_{t-1} - 1x_{t-1} + 0,1813$$

5. Uji Asumsi Residual

Uji asumsi residual terdiri dari uji asumsi *white noise* dan uji normalitas. Pengujian dilakukan pada model terbaik yaitu ARIMA (1,1,1) untuk melihat kelayakan model dalam meramalkan kualitas udara. Berikut merupakan hasil pengujian asumsi *white noise*.

Tabel 5.
Uji Asumsi *White noise*

Uji	P-value	Keterangan
Ljung-Box	0.8809	White Noise

Setelah melakukan Uji Ljung-Box didapatkan bahwa model ARIMA (1,1,1) memenuhi asumsi *white noise*. Selanjutnya dilakukan pengujian residual distribusi normal pada model terbaik dan diperoleh hasil sebagai berikut.

Tabel 6.
Uji Asumsi Normalitas

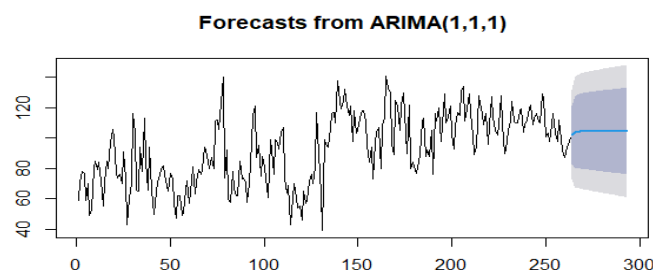
Uji	P-value	Keterangan
Shapiro Wilk	0.4447	Berdistribusi Normal

Setelah melakukan Uji Saphiro Wilk didapatkan hasil bahwa model ARIMA (1,1,1) memenuhi asumsi normalitas. Terpenuhinya semua asumsi sehingga model ARIMA (1,1,1) layak digunakan untuk meramalkan kualitas udara.

6. Peramalan

Diperoleh model terbaik yang akan digunakan untuk peramalan adalah ARIMA (1,1,1). Berikut merupakan grafik prediksi kualitas udara di Kota Semarang untuk 30 hari kedepan menggunakan model ARIMA (1,1,1).

Gambar 4:
Hasil Peramalan



Nilai *Mean Absolute Percentage Error* (MAPE) untuk model ARIMA (1,1,1) adalah 13.71377 atau 13.714% sehingga dapat dikatakan bahwa model tersebut memiliki kemampuan prediksi yang baik.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan dapat disimpulkan bahwa model peramalan kualitas udara dengan metode ARIMA didapatkan model terbaik yaitu ARIMA (1,1,1) dengan nilai kesalahan MAPE sebesar 13.71377 atau 13.714%. Nilai kualitas udara di Kota Semarang untuk waktu yang akan datang terjadi peningkatan pencemaran udara dan berada dalam tahap tidak sehat.

SARAN

Penelitian ini menggunakan data kandungan jenis gas pencemar PM_{2.5} sebagai pengukur kualitas udara. Penelitian selanjutnya dapat menggunakan jenis gas pencemar yang lainnya. Metode peramalan memiliki banyak jenis dan kegunaannya, untuk penelitian ini digunakan metode ARIMA. Penelitian selanjutnya dapat menggunakan metode peramalan yang lainnya. Hal tersebut disebabkan oleh metode ARIMA tidak bisa meramalkan dalam jangka panjang.

DAFTAR PUSTAKA

- Ainiyah, L. & Bansori, M. 2021. "Prediksi Jumlah Kasus Covid-19 Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA) (Studi Kasus Kabupaten Sidoarjo)" dalam *Jurnal Sains Dasar* Vol. 10 No. 2 (Hal. 62-68). Yogyakarta: Universitas Negeri Yogyakarta.
- Duan, J. Gong, Y. Luo, J. Zhao, Z. 2023. "Air-quality prediction based on the ARIMA-CNN-LSTM combination model optimized by dung beetle optimizer". *Scientific Reports*. 13:12127, <https://doi.org/10.1038/s41598-023-36620-4>.
- Hartati. 2017. "Penggunaan Metode Arima Dalam Meramal Pergerakan Inflasi" dalam *Jurnal Matematika Sains dan Teknologi* Vol.18 No.1 (Hal.1-10). Banten: Universitas Terbuka.
- Imas A. P. Gusdini, N. Dewi D. & M, M. 2020. "Analisis Kualitas Udara Dengan Indeks Standar Pencemar Udara (Ispu) Dan Sebaran Kadar Polutannya Di Provinsi Dki Jakarta" dalam *Jurnal SEOI* Vol.2 Edisi 2. Jakarta: Universitas Sahid Jakarta.
- Koesoemaningroem, N., Endroyono, & Nugroho, S. M. S. 2021. "Peramalan Pencemaran Udara di Kota Surabaya Menggunakan Metode Dsarima Dengan Pendekatan Percentile Error Bootstrap (PEB)" dalam *Jurnal Teknologi Informasi dan Ilmu Komputer* Vol. 8 No. 5 (Hal. 987-994). Surabaya: Universitas Brawijaya.
- Pani D., A. 2015. "Peramalan Pencemaran Udara Oleh Particulate Matter (Pm10) Di Pekanbaru Dengan Metode Box-Jenkins" dalam *Prosiding Semirata 2015 bidang MIPA BKS-PTN Barat* (Hal. 63-73). Pontianak: Universitas Tanjungpura.
- Rezaldi, D. A., & Sugiman. 2021. "Peramalan Metode ARIMA Data Saham PT. Telekomunikasi Indonesia" dalam *Prosiding Seminar Nasional Matematika* Vol. 4 (Hal. 611-620). Semarang: Universitas Negeri Semarang.
- Samal, K. K. R., Babu, K. S., Das, S. K., & Acharaya, A. 2019. "Time series based air pollution forecasting using SARIMA and prophet model". *ACM International Conference Proceeding Series*, Hal. 80-85. <https://doi.org/10.1145/3355402.3355417>.



- Tresno A, A. Parmikanti, K. Suhandi, B. & Nurani R, B. 2021. “Peramalan Konsentrasi Particulate Matter 2.5 (PM2.5) menggunakan Model Vector Autoregressive dengan Metode Maximum Likelihood Estimation” dalam *Kubik: Jurnal Publikasi Ilmiah Matematika* Vol. 6 No. 1 (Hal. 1-12). Bandung: UIN Sunan Gunung Djati.
- Zhao, L. Li, Z. Qu, L. 2022. “Forecasting of Beijing PM2.5 with a hybrid ARIMA model based on integrated AIC and improved GS fixed-order methods and seasonal decomposition”. *Heliyon*, V.8, Issue 12, e12239.