

Prediksi Tingkat Polusi Udara Dengan *Data Mining*

Hikmah Dwiyanti Nasir¹⁾, Dahlia²⁾, Zawiyah Saharuna³⁾

¹ Teknik Elektro, Politeknik Negeri Ujung Pandang
hikmahdwi9@gmail.com

² Teknik Elektro, Politeknik Negeri Ujung Pandang
dahlia@poliupg.ac.id

³ Teknik Elektro, Politeknik Negeri Ujung Pandang
zawiyah@poliupg.ac.id

Abstrak

Polusi udara di kota besar menjadi masalah bagi lingkungan. *World Health Organization* (WHO) menyimpulkan bahwa 4,2 juta orang meninggal setiap tahun karena polusi udara. Polusi udara tidak hanya berdampak buruk pada manusia tetapi juga pada lingkungan yang dapat menyebabkan hujan asam, kabut asap, memburuknya lapisan ozon dan pemanasan global. Oleh karena itu, salah satu upaya pencegahan yang dapat dilakukan adalah memprediksi polusi udara. Pada penelitian ini, dilakukan prediksi polusi udara dengan memanfaatkan data historis parameter PM2.5 dan PM10 untuk kota Jakarta Pusat dari tahun 2015-2020 milik *Air Quality Index* (AQICN). Data kemudian diproses melalui tahap *preprocessing* dengan mengisi nilai yang kosong (*null*). Data diproses ke dalam Model *Autoregressive Integrated Moving Average* (ARIMA) menggunakan metode *Time Series Analysis*. Proses ARIMA melalui beberapa tahap yang perlu dilakukan untuk mendapatkan model yang paling sesuai, dengan melakukan *differencing*, melihat plot ACF dan PACF, dan menguji beberapa model. Sehingga diperoleh akurasi tertinggi pada prediksi PM2.5 tahun 2020 dan 2021 dengan model SARIMA(5,1,5)(1,1,1)₁₂ yaitu 90,05% dan 94,67%, sedangkan untuk PM10 prediksi pertama menggunakan model SARIMA(5,1,4)(1,1,1)₁₂ dengan akurasi 85,36 % dan prediksi kedua dengan model SARIMA(4,1,4)(1,1,1)₁₂ dengan akurasi 88,49%. Dari penelitian ini disimpulkan bahwa model ARIMA terbaik akan didapatkan dengan melakukan “multipel hipotesis” dengan menguji beberapa model yang memiliki skor AIC terkecil.

Keywords: *Data Mining*, ARIMA, SARIMA, Prediksi Polusi Udara, Polusi Udara, *Time Series Analysis*.

I. PENDAHULUAN

Masalah lingkungan dan kesehatan yang disebabkan oleh polusi udara di kota-kota besar menjadi sebuah tantangan [1]. *World Health Organization* (WHO) menyimpulkan bahwa 4,2 juta orang meninggal setiap tahun karena polusi udara. Polusi udara tidak hanya berdampak buruk pada kesehatan manusia tetapi juga pada lingkungan yang dapat menyebabkan hujan asam, kabut asap, memburuknya lapisan ozon dan pemanasan global [2].

Polusi udara meningkat dengan cepat karena berbagai aktivitas manusia, dan hal tersebut akan berdampak langsung ke atmosfer bumi yang biasanya berasal dari bahan kimia seperti aktivitas industri, partikel, atau bahan biologis yang menyebabkan gangguan, penyakit, atau kematian bagi manusia terutama bagi orang tua dan anak-anak [3].

Kota Jakarta merupakan salah satu kota besar dan terpadat di Indonesia yang rawan akan terjadinya polusi udara. Berdasarkan data dari Badan Pusat Statistik Kota Jakarta, jumlah kendaraan terus mengalami peningkatan tiap tahun terutama jumlah sepeda motor dengan rata-rata pertumbuhan 5,3% per tahun [4]. Salah satu penyebab terjadinya pencemaran udara adalah jumlah yang berlebih pada sektor transportasi [5].

Penelitian yang menggunakan *data mining* dengan menerapkan teknik *time series analysis* yaitu untuk menganalisis tren polusi udara yang ada di Delhi, India dengan membuat prediksi tentang tren polusi udara dimasa depan dengan polutan udara seperti sulfur dioksida (SO₂), nitrogen dioksida (NO₂), partikel (PM), karbon

monoksida (CO), ozon (O₃) [6]. *Autoregressive Integrated Moving Average* (ARIMA) adalah salah satu model *time series* yang sangat populer dan telah digunakan secara luas [7]. Selain itu, model ARIMA dapat menunjukkan kinerja yang lebih unggul dalam presisi akurasi dalam memprediksi jeda (*lag*) waktu berikutnya [8].

Berdasarkan permasalahan polusi udara, maka pada penelitian ini dibuat prediksi polusi udara untuk kota Jakarta Pusat dengan menggunakan teknik *data mining* yaitu analisis *time series* dengan model ARIMA dan data hasil prediksi divisualisasikan dalam bentuk grafik.

II. KAJIAN LITERATUR

A. *Data Mining*

Data mining menggunakan kombinasi dari basis pengetahuan eksplisit, keterampilan analitis yang canggih untuk mendapatkan tren dan pola tersembunyi dari suatu data. Pada umumnya *Data mining* terbagi menjadi 2 model yakni *supervised* dan *unsupervised*.

- Supervised Data Mining* adalah mempelajari perilaku dari sebuah data, dengan tujuan menghubungkan pola perilaku dengan histori dan informasi lainnya yang direkam [9]. *Supervised Learning* juga diartikan sebagai model prediktif.
- Unsupervised Data Mining* sering digunakan pertama kali untuk mempelajari pola dan mencari pola yang sebelumnya tersembunyi untuk memahami, mengklasifikasikan, melambungkan, dan mengkode objek penelitian sebelum menerapkan sebuah teori [9].

B. Time Series Analisis

Analisis deret waktu atau *time series analysis* adalah bidang penelitian untuk mempelajari pengamatan deret waktu dan membangun model untuk menggambarkan struktur data dan memprediksi nilai deret waktu di masa mendatang [10]. Dalam Time series terdiri dari beberapa komponen yaitu didefinisikan sebagai berikut:

- a. *Level (L)*: Nilai rata-rata dalam *time series*.
- b. *Trend (T)*: Nilai kenaikan atau penurunan dalam *time series*.
- c. *Seasonality (S)*: Siklus jangka pendek berulang dalam *time series*.
- d. *Noise (N)*: Variasi acak dalam *time series*.

Keempat komponen tersebut dapat mempengaruhi sifat *time-series* terutama kestasioneran data. Apabila data mengandung *seasonality*, *trend*, dan *level* yang terus berubah seiring waktu maka perlu untuk dilakukan *differencing*, karena seringkali data *time series* yang ditemukan tidak stationer [11][12].

C. ARMA (Autoregressive Moving Average)

Model *Autoregressive Moving Average* (ARMA) merupakan model peramalan *univariate time series* yang merupakan gabungan antara model *autoregressive* (AR) dan model *Moving Average* (MA) [13]. Model ARMA memiliki orde p dan q sehingga ditulis ARMA (p,q) atau ARIMA (p,0,q) yang memiliki formulasi sebagai berikut [11].

$$Y_t = \mu + \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \varepsilon - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \dots \dots \dots (3)$$

Dimana Y_t adalah variabel tidak bebas; μ adalah konstanta; $\varphi_1, \varphi_2, \dots \varphi_p$ adalah parameter *autoregressive*; $\theta_1, \theta_2 \dots \theta_p$ adalah koefisien parameter *moving average*; $Y_{t-1}, Y_{t-2} \dots Y_{t-p}$ adalah variabel bebas; ε_{t-q} adalah sisaan pada saat ke t-q.

D. ARIMA (Autoregressive Integrated Moving Average)

Model *Autoregressive Integrated Moving Average* (ARIMA) merupakan model peramalan *univariate time series* yang menggabungkan antara model *Autoregressive* (AR) dan model *Moving Average* (MA) untuk data yang stasioner (I)[14]. Secara matematis, model ARIMA (p, d, q) dapat nyatakan dalam bentuk.

$$Y_t = \mu + \sum_{i=1}^p \varphi_m y_{t-1} + \sum_{j=1}^q \theta_n \varepsilon_{t-j}$$

Dimana, μ adalah istilah konstan dan mewakili rata-rata dari *time series stochastic*; φ_m adalah koefisien dari parameter *autoregressive*; θ_n adalah koefisien dari parameter rata-rata *moving average*; Y_t adalah nilai prediksi sinyal pada waktu t; y_{t-1} adalah nilai *time series* pada waktu (t - 1); ε_{t-j} adalah eror dalam nilai yang diprediksi sebagai perbandingan dengan sinyal aktual pada waktu (t - j).

E. SARIMA (Seasonal Autoregressive Integrated Moving Average)

SARIMA model berguna untuk memodelkan deret waktu musiman di mana rata-rata dan statistik lainnya untuk musim tertentu tidak stasioner sepanjang tahun. Model SARIMA merupakan lanjutan langsung dari model ARMA *non seasonal* dan ARIMA[15]. Bentuk umum

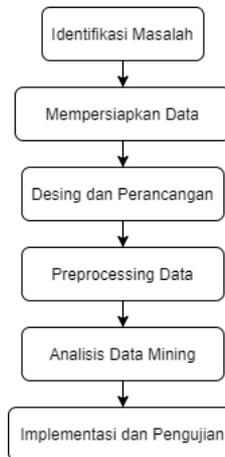
model ARIMA *seasonal* dilambangkan sebagai ARIMA (p, d, q) × (P, D, Q) s, di mana p adalah urutan AR non-musiman, d adalah *differencing* non-musiman, q adalah MA non-musiman, P adalah AR musiman, D adalah *differencing* musiman, Q adalah MA musiman, dan s adalah rentang waktu dari pola musiman yang berulang[10].

F. Pengukuran Error

Mengukur tingkat eror atau keakuratan suatu model dapat dilakukan dengan beberapa metrik pengukuran seperti MSE (*Mean Square Error*), RMSE (*Root Mean Square Error*), dan MAPE (*Mean Absolute Percentage Error*). MSE dan RMSE memberikan rata-rata satu set kesalahan, sedangkan MAPE adalah rata-rata kesalahan persentase absolut dari peramalan. Kesalahan atau eror didefinisikan sebagai nilai aktual atau nilai pengamatan dikurangi nilai peramalan [16].

III. METODE PENELITIAN

Metode penelitian ini digambarkan dengan diagram alir karena mudah untuk dipahami serta memberikan gambaran yang jelas mengenai tahapan-tahapan yang dilakukan. Berikut adalah tahapan penelitian yang dilakukan.



Gambar 1. Tahapan Metode Penelitian

A. Indentifikasi Masalah

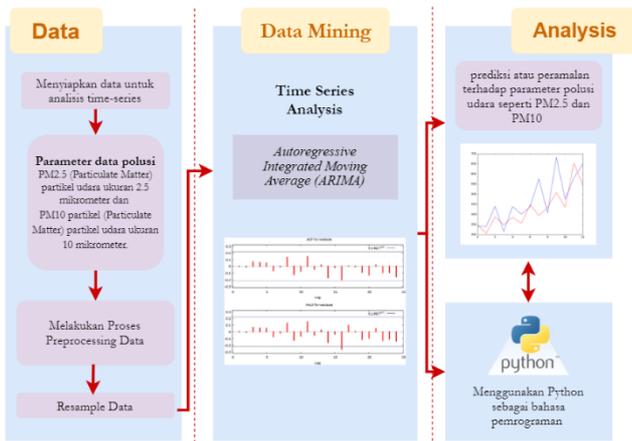
Identifikasi masalah dilakukan dengan tujuan untuk memahami konsep dari *Data Mining* terutama analisis *time series* yang menjadi tugas utama dalam penelitian ini serta penggunaan model ARIMA dalam melakukan peramalan atau prediksi terhadap tren polusi udara di kota Jakarta.

B. Mempersiapkan Data

Dalam penelitian ini data yang diperlukan terdiri dari parameter pencemar udara seperti PM2.5, PM10 dan tanggal untuk daerah Jakarta Pusat. Data parameter polusi udara tersebut diperoleh dari sumber yang menyediakan data kualitas udara yaitu *Air Quality Index* (AQICN). AQICN adalah sebuah website yang menyediakan data polusi udara untuk berbagai kota diseluruh dunia.

C. Design dan Perancangan

Desain dan perancangan sistem digunakan sebagai acuan dalam penelitian untuk menggambarkan proses maupun kebutuhan dalam penelitian, desain dan perancangan penelitian ini digambarkan pada Gambar 2.



Gambar 2. Design dan Perancangan

Data yang telah terkumpul akan melalui proses *preprocessing*. Pada penelitian ini prediksi akan dilakukan dengan menghasilkan data prediksi per minggu, oleh karena itu data per hari yang didapatkan akan dilakukan proses *downsampling* dengan mengambil data per minggu. Apabila terdapat data *NaN* dalam proses *resample*, maka metode yang dapat dilakukan adalah interpolasi linear.

Data hasil *resample* yang telah di dapatkan akan diolah dengan melakukan analisis *time series*. *Tools* yang digunakan adalah *python* dan *library statsmodels* dan *matplotlib*. Model yang digunakan untuk analisis *time series* adalah ARIMA (*Autoregressive Integrated Moving Average*). Tujuan akhir dari penelitian ini yaitu prediksi polusi udara tahun 2021 untuk masing-masing parameter PM2.5 dan PM10, prediksi dilakukan secara terpisah untuk kedua parameter tersebut karena sifat data yang *independent* atau tidak berhubungan.

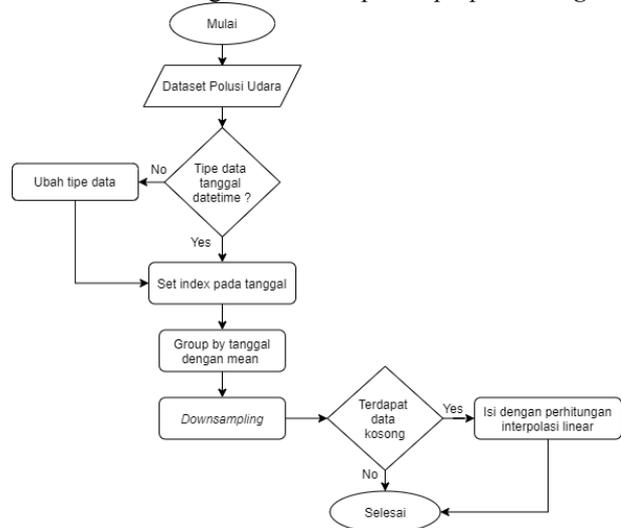
Data yang tersedia dalam penelitian ini terdiri data polusi udara bulan Juli 2015 hingga Juni 2020, sehingga perlu untuk mendapatkan data polusi udara bulan Juli hingga Desember 2020. Prediksi dilakukan dengan menganalisis setiap parameter PM2.5 dan PM10 dengan konsisten melakukan prediksi per lima tahun atau 60 bulan data *time series*. Prediksi pertama untuk menghasilkan data polusi Juli - Desember 2020 digunakan data bulan Juli 2015 – Juni 2020. Prediksi kedua digunakan data *time series* Januari 2016 – Desember 2020 untuk mendapatkan hasil akhir peramalan polusi udara pada tahun 2021.

D. Preprocessing Data

Preprocessing adalah tahapan yang berfungsi untuk melakukan pembersihan data, menghapus noise serta mengisi nilai yang hilang. Berikut adalah beberapa hal yang dilakukan dalam melakukan *preprocessing* data polusi udara untuk *time series*.

- Perlu untuk memperhatikan tipe data tanggal pada saat analisis *time series* karena waktu atau tanggal merupakan hal yang paling berpengaruh pada saat melakukan peramalan dengan *time series*. Selain itu memungkinkan untuk melihat pola seperti tren, musim, siklus, dan ketidakteraturan dalam data.
- Apabila dalam data terdapat tanggal yang memiliki beberapa objek (data *cross section*) maka lakukan rata-rata pada objek.

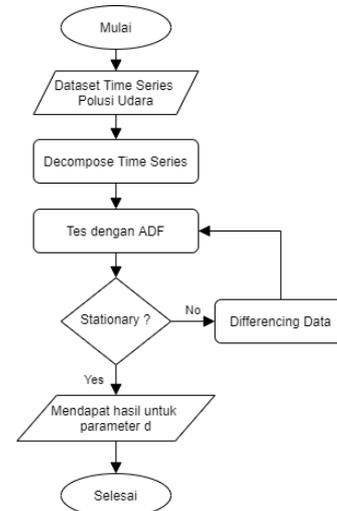
- Apabila terdapat data yang kosong setelah melakukan *resample* maka gunakan metode interpolasi linear untuk mendapatkan nilai yang kosong tersebut. Berikut adalah diagram alir dari proses *preprocessing*.



Gambar 3. Preprocessing Data Time Series

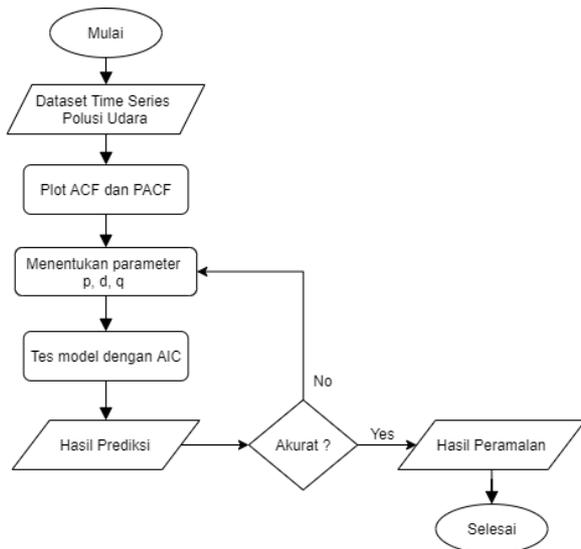
E. Analisis Data Mining

Pada tahap ini, metode yang digunakan dalam analisis *time series* adalah menggunakan model ARIMA (*Autoregressive Integrated Moving Average*). Diagram alir pada Gambar 4 menunjukkan proses awal dalam analisis *time series* dalam menunjukkan apakah *time series* stasioner atau tidak.



Gambar 4. Cek Stationarity Data

Data yang telah stationer akan dilakukan pemilihan model terbaik yang didapatkan dengan *multiple hipotesis* yaitu membandingkan beberapa model. Model terbaik dari ARIMA model adalah model yang memiliki nilai eror atau kesalahan yang paling kecil. Semakin kecil nilai eror tersebut maka semakin akurat hasil prediksi atau peramalan yang didapatkan. Menentukan kriteria dari nilai eror yang baik untuk hasil prediksi menggunakan ARIMA tidaklah absolut, tetapi semua tergantung dari variabel atau parameter yang diukur [17]. berikut adalah diagram alir tentang proses pemilihan model yang terbaik untuk ARIMA



Gambar 5. Memilih Model ARIMA

F. Implementasi dan Pengujian

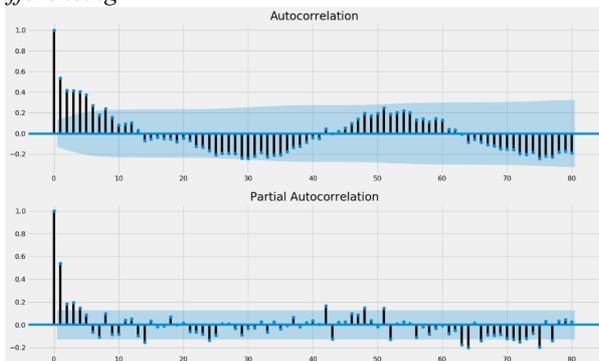
Metode pengujian yang dilakukan adalah dengan melakukan testing pada beberapa model ARIMA (p, d, q). Dari 10 model yang dipilih, dilihat berdasarkan nilai AIC terkecil. Terdapat 3 metode yang digunakan untuk melihat nilai eror dari suatu model tersebut, metode yang digunakan adalah MSE, RMSE, dan MAPE.

IV. HASIL DAN PEMBAHASAN

Data polusi udara yang telah melalui proses *preprocessing*, seperti *resample* dan menghapus nilai kosong (*NaN/null*) mendapatkan hasil data pengamatan untuk parameter PM2.5 dan PM10 dengan masing-masing jumlah observasi 261 dan 258. Melihat sifat *stationarity* dari parameter PM2.5 dan PM10 dilakukan dengan melihat grafik hasil *decompose*, dengan *decompose* komponen dari *time series* dapat terlihat khususnya dalam tren, *seasonality*, *noise* pada deret waktu PM2.5 dan PM10. Selain melakukan *decompose* hal yang perlu dilakukan adalah melakukan plot ACF dan PACF Hal ini dilakukan demi memahami dengan baik masalah dari data *time series* untuk analisis dan prediksi.

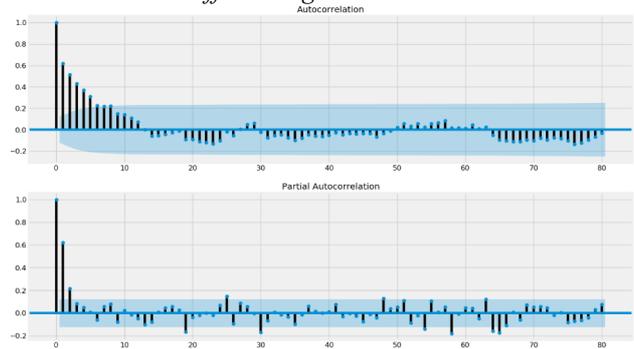
4.1. Implementasi Autokorelasi PM2.5 dan PM10

Pada percobaan menampilkan plot ACF dan PACF untuk parameter PM2.5 ditampilkan dalam Gambar 6 yang menunjukkan bahwa *time series* tersebut memiliki *seasonality* yang sangat kuat dibuktikan dengan pola *lag* yang bergelombang. Sehingga perlu untuk dilakukan *differencing*



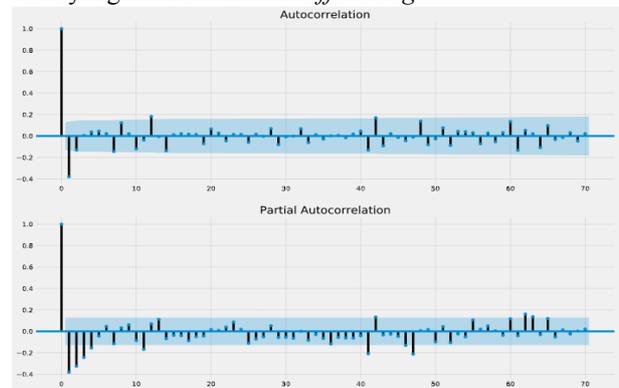
Gambar 6. Plot ACF dan PACF Time series PM2.5

Sama halnya dengan PM10 seperti Gambar 7 walaupun tidak sepenuhnya bergelombang seperti PM2.5 hubungan antar *lag* terlihat memiliki korelasi yang kuat, dan terdapat *seasonality* dalam *time series* tersebut sehingga perlu untuk dilakukan *differencing*.

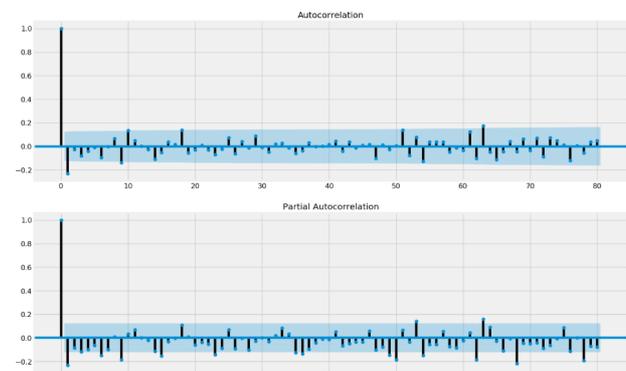


Gambar 7. Plot ACF dan PACF Time Series PM10

Gambar 8 dan 9 merupakan Plot ACF dan PACF *time series* yang telah dilakukan *differencing*.



Gambar 8. Differencing Plot ACF dan PACF PM2.5



Gambar 9. Differencing Plot PM2.5 dan PM10

4.2. Implementasi Model Terbaik

Menentukan parameter p, d, q serta P, D, dan Q dapat menggunakan AIC (*Akaike Information Criterion*) dengan memilih beberapa model dengan skor AIC terkecil kemudian diuji dengan melihat akurasi prediksi yang diukur melalui nilai MSE (*Mean Square Error*), RMSE (*Root Mean Square Error*), dan MAPE (*Mean Absolute Percentage Error*). Nilai eror yang terkecil dapat diambil keputusan menjadi model yang layak untuk digunakan. Tabel 1 menunjukkan beberapa model yang diuji untuk *time series* PM2.5.

Tabel 1. Tabel Hasil Uji Coba Beberapa Model SARIMA Time series PM2.5

Model SARIMA	AIC	MSE	RMSE	MAPE
(1,1,1) (1,1,1,12)	1867,04	269,31	16,41	10,6 %
(5,1,3) (1,1,1,12)	1844,92	274,65	16,57	11,1 %
(5,1,4) (1,1,1,12)	1843,21	259,83	16,12	10,29 %
(5,1,5) (1,1,1,12)	1838,57	236,74	15,39	9,95 %
(5,1,6) (1,1,1,12)	1833,00	260,05	16,13	10,36 %
(6,1,1) (1,1,1,12)	1836,33	268,87	16,4	10,66 %
(6,1,2) (1,1,1,12)	1834,45	275,57	16,6	10,84 %
(6,1,3) (1,1,1,12)	1836,89	287,2	16,95	11,25 %
(6,1,5) (1,1,1,12)	1838,02	269,07	16,4	10,86 %
(6,1,6) (1,1,1,12)	1829,63	254,08	15,94	10,17%

Tabel 2. Tabel Hasil Uji Coba Beberapa Model SARIMA Time Series PM10

Model SARIMA	AIC	MSE	RMSE	MAPE
(1,1,1) (1,1,1,12)	2031,53	75,37	8,68	14,00 %
(4,1,1) (1,1,1,12)	2017,44	77,12	8,78	14,66 %
(4,1,3) (1,1,1,12)	2018,63	74,19	8,61	14,35 %
(4,1,4) (1,1,1,12)	2016,86	80,9	8,99	14,15 %
(4,1,5) (1,1,1,12)	2009,18	77,52	8,8	14,89 %
(5,1,1) (1,1,1,12)	2011,52	76,83	8,77	14,62 %
(5,1,2) (1,1,1,12)	2015,36	77,09	8,78	14,75 %
(5,1,3) (1,1,1,12)	2009,88	73,55	8,58	14,02 %
(5,1,4) (1,1,1,12)	2021,72	66,87	8,18	13,70 %
(5,1,5) (1,1,1,12)	2003,76	80,68	8,98	14,63 %

Berdasarkan Tabel 1 model terbaik yang akan digunakan untuk prediksi pertama PM2.5 adalah SARIMA (5,1,5)×(1,1,1)12 dan model terbaik untuk PM10 pada Tabel 2 adalah SARIMA (5,1,4)×(1,1,1)12.

4.3. Hasil Prediksi Polusi Udara

A. Prediksi Polusi Udara PM2.5 Juli-Desember 2020

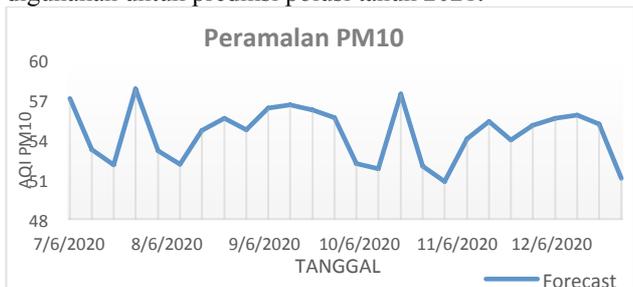
Model yang digunakan untuk prediksi Juli-Desember 2020 yaitu SARIMA(5,1,5)×(1,1,1)12 yang hasilnya di tampilkan dalam Gambar 10. Hasil prediksi PM2.5 akan digunakan untuk prediksi polusi tahun 2021.



Gambar 10. Grafik Prediksi PM2.5 Juli - Desember 2020

B. Prediksi Polusi Udara PM10 Juli-Desember 2020

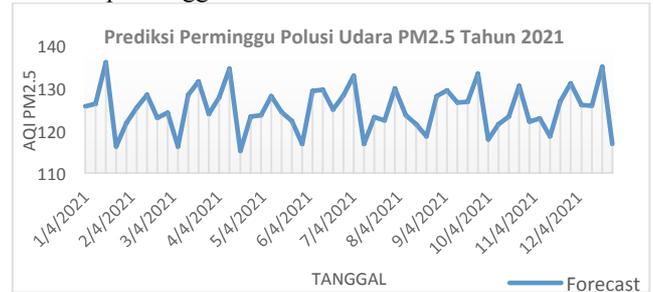
Model yang digunakan untuk prediksi Juli-Desember 2020 yaitu SARIMA(5,1,4)×(1,1,1)12 yang hasilnya di tampilkan dalam Gambar 11. Hasil prediksi PM10 akan digunakan untuk prediksi polusi tahun 2021.



Gambar 11. Grafik Prediksi PM10 Juli - Desember 2020

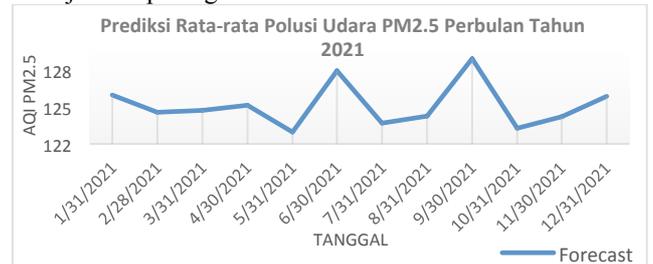
C. Prediksi Polusi Udara PM2.5 2021

Prediksi PM2.5 dilakukan dengan langkah-langkah yang sama saat prediksi polusi udara untuk paruh kedua tahun 2020. Seperti *preprocessing*, *resample*, *decompose*, melihat plot ACF dan PACF, kemudian memilih model terbaik yang digunakan untuk prediksi atau peramalan. Model yang digunakan untuk prediksi atau peramalan kedua ini sama dengan model yang digunakan pada prediksi polusi bulan Juli - Desember 2020 yaitu model SARIMA(5,1,5)×(1,1,1)12. Penggunaan model tersebut berdasarkan pola plot ACF dan PACF dan juga dibuktikan dengan menguji beberapa model. Berikut adalah hasil Prediksi perminggu.



Gambar 12. Prediksi Perminggu Polusi Udara PM2.5 2021

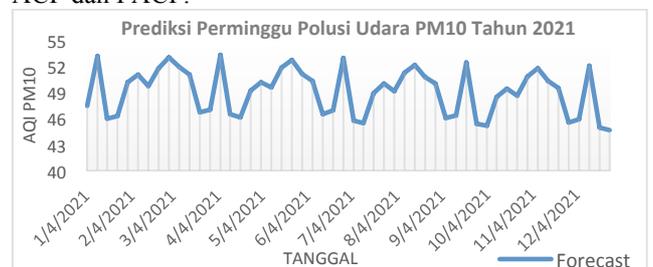
Akurasi prediksi yang dihasilkan adalah 94,67% dengan eror atau kesalahan 5,33%. Rata-rata rentang nilai polusi PM2.5 berada pada angka 115 untuk nilai terendah dan 135 untuk nilai yang tertinggi. Grafik Dalam Perbulan ditunjukkan pada gambar berikut.



Gambar 13. Prediksi Per bulan Polusi Udara PM10 2021

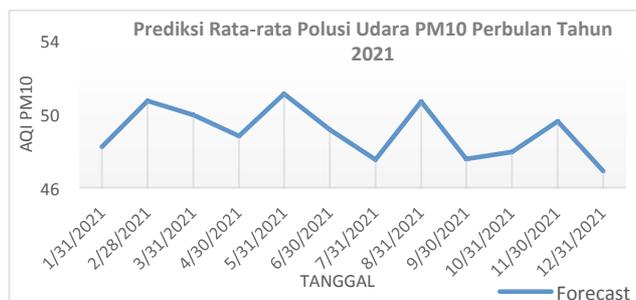
D. Prediksi Polusi Udara PM10 2021

Model yang digunakan untuk prediksi atau peramalan kedua ini berbeda dengan model yang digunakan pada prediksi polusi paruh kedua tahun 2020 yaitu SARIMA(4,1,4)×(1,1,1)12. Hal ini berdasarkan pola plot ACF dan PACF.



Gambar 14. Prediksi Perminggu Polusi Udara PM2.5 2021

Rentang polusi PM10 berada pada angka 42 dan 54. Akurasi prediksi yang dihasilkan adalah 88,49% dengan tingkat eror 11,51%. Grafik dalam per bulan ditunjukkan dalam grafik berikut.



Gambar 15. Prediksi Perbulan Polusi Udara PM10 2021

V. KESIMPULAN

1. Prediksi dilakukan menggunakan data historis polusi udara yang telah melalui tahap *preprocessing* maupun *resample*, dan diproses dengan ARIMA salah satu model *time series analysis* yang dibangun menggunakan metode *Data Mining*. Data yang belum melewati tahap *preprocessing* tidak dapat digunakan karena masih terdapat data yang tidak lengkap atau bernilai *null* dan menguji kestasioneran *time series* sangat diperhatikan dalam model ARIMA ataupun SARIMA.
2. Keakuratan pada prediksi didapatkan dengan melakukan "multipel hipotesis" dengan menguji beberapa model yang memiliki skor AIC terkecil. Sehingga didapatkan akurasi tertinggi pada prediksi PM2.5 tahun 2020 dan 2021 dengan model SARIMA(5,1,5)(1,1,1)12 yaitu 90,05% dan 94,67%, sedangkan untuk PM10 prediksi pertama menggunakan model SARIMA(5,1,4)(1,1,1)12 dengan akurasi 85,36 % dan prediksi kedua dengan model SARIMA(4,1,4)(1,1,1)12 dengan akurasi 88,49%.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Allah SWT, kedua orang tua, saudara, dan terkhusus kepada kedua dosen pembimbing, serta seluruh dosen prodi Teknik Komputer dan Jaringan serta sahabat-sahabat penulis.

REFERENSI

- [1] M. A. Esfandani and H. Nematzadeh, "Predicting air pollution in Tehran: Genetic algorithm and back propagation neural network," vol. 4, no. 1, pp. 49–54, 2016.
- [2] K. G. Ri, R. Manimegalai, G. D. M. Si, R. Si, U. Ki, and R. B. Ni, "Air Pollution Analysis Using Enhanced K-Means Clustering Algorithm for Real Time Sensor Data," no. August 2006, pp. 1945–1949, 2016.
- [3] R. Raturi and J. R. Prasad, "Recognition of Future Air Quality Index Using Artificial Neural Network," pp. 3–6, 2018.
- [4] H. Widowati, "Berapa Jumlah Kendaraan di DKI Jakarta?," 2019. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2019/08/02/berapa-jumlah-kendaraan-di-dki-jakarta>. [Accessed: 19-Jul-2020].

- [5] S. G. A. J. H. Sulistyarto, "Strategi Pengembangan Kebijakan Penurunan Emisi Kendaraan di Kawasan Senayan, Jakarta," vol. 8, no. 2, 2020.
- [6] S. Taneja, N. Sharma, K. Oberoi, and Y. Navoria, "Predicting trends in air pollution in Delhi using data mining," *India Int. Conf. Inf. Process. IICIP 2016 - Proc.*, pp. 1–6, 2017.
- [7] S. Barak and S. S. Sadeh, "Forecasting energy consumption using ensemble ARIMA-ANFIS hybrid algorithm," *Int. J. Electr. Power Energy Syst.*, vol. 82, pp. 92–104, 2016.
- [8] S. Siami-Namini and A. S. Namin, "Forecasting Economics and Financial Time Series: ARIMA vs. LSTM," pp. 1–19, 2018.
- [9] C. M. Davis, J. M. Hardin, T. Bohannon, and J. Oglesby, "Data mining applications in higher education," *Data Min. Methods Appl.*, pp. 123–148, 2007.
- [10] S. S. N. A. S. Namin, "Forecasting Economic And Financial Time Series: ARIMA Vs. LSTM," pp. 1–19, 2018.
- [11] D. Hatidja, "Penerapan Model ARIMA untuk memprediksi Harga Saham PT. Telkom Tbk.," 2011.
- [12] J. Brownlee, "How to Decompose Time Series Data into Trend and Seasonality," 2017. [Online]. Available: <https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/>. [Accessed: 04-Jun-2020].
- [13] M. As'ad, S. S. Wibowo, and E. Sophia, "Peramalan Jumlah Mahasiswa Baru Dengan Model Autoregressive Integrated Moving Average (Arima)," *J I M P - J. Inform. Merdeka Pasuruan*, vol. 2, no. 3, pp. 20–33, 2017.
- [14] Aasim, S. N. Singh, and A. Mohapatra, "Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting," *Renew. Energy*, vol. 136, pp. 758–768, 2019.
- [15] A. I. M. Keith W. Hipel, Ed., "Chapter 12 seasonal autoregressive integrated moving average models," in *Developments in Water Science*, 1994, pp. 419–462.
- [16] P. M. Swamidass, Ed., "MAPE (mean absolute percentage error) Mean Absolute Percentage Error (MAPE)," in *Encyclopedia of Production and Manufacturing Management*, Boston, MA: Springer US, 2000, p. 462.
- [17] P. Ramos, N. Santos, and R. Rebelo, "Performance of state space and ARIMA models for consumer retail sales forecasting," *Robot. Comput. Integr. Manuf.*, pp. 1–13, 2015.