

Analisis Peramalan Polutan PM_{2.5} Dan PM₁₀ Di DKI 4 Jakarta Dengan SARIMA

Haikal Fikri Rabani¹, Renzie Aditya Fattah Hudiantoro², Siti Nur Salamah³, Sarini Abdullah⁴

¹Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Indonesia, Depok, email: haikal.fikri@ui.ac.id

²Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Indonesia, Depok, email: renzie.aditya@ui.ac.id

³Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Indonesia, Depok, email: siti.nur26@ui.ac.id

⁴Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Indonesia, Depok, Kota, email: sarini@sci.ui.ac.id

Corresponding Author: Haikal Fikri Rabani

ABSTRAK — Polusi udara di DKI Jakarta, khususnya polutan PM_{2.5} dan PM₁₀ terus menjadi perhatian serius karena dampaknya yang signifikan terhadap kesehatan dan lingkungan. Wilayah DKI Jakarta Timur (DKI 4) telah teridentifikasi sebagai daerah dengan konsentrasi polutan PM_{2.5} dan PM₁₀ tertinggi dibandingkan wilayah DKI lainnya. Penelitian ini berfokus untuk memahami kondisi polusi di wilayah tersebut dengan menggunakan model yang dapat menjelaskan perkembangan polusi serta meramalkan tren ke depannya. Penelitian ini menggunakan data historis polusi di wilayah DKI Jakarta dari Januari 2020 hingga Agustus 2024 dari Satu Data Indonesia. *Preprocessing* data dilakukan dengan mengonversi data harian menjadi bulanan untuk mengidentifikasi pola musiman. Eksplorasi data menunjukkan perbedaan tren antara PM_{2.5} dan PM₁₀, di mana PM_{2.5} menunjukkan fluktuasi lebih besar dibandingkan PM₁₀ yang memiliki tren lebih stabil. Model SARIMA digunakan karena kemampuannya menangkap pola musiman dan tren jangka panjang dalam data *time series* sehingga memudahkan dalam memahami pola pertumbuhan kondisi polusi. Pemilihan model SARIMA dilakukan melalui serangkaian uji statistik dan evaluasi menggunakan metrik *error* seperti MAE, RMSE, dan MAPE, dengan tetap mempertimbangkan kemudahan interpretasi untuk mendapatkan *insight* yang bermakna. Selanjutnya, dilakukan *K-folds cross validation* untuk memastikan reliabilitas hasil peramalan. Model SARIMA (3,0,2)(1,0,0)[12] dipilih untuk meramalkan polutan PM₁₀ dengan akurasi 91,13%, sedangkan SARIMA (1,0,0)(1,1,0)[12] dipilih untuk meramalkan polutan PM_{2.5} dengan akurasi 91,15%. Kedua model ini menunjukkan hasil peramalan yang mendekati data aktual. Hasil peramalan pada wilayah Jakarta Timur (DKI 4) untuk periode September hingga November 2024 akan memiliki risiko polusi udara yang tinggi terutama pada polutan PM_{2.5} karena nilai konsentrasi PM_{2.5} berada di angka 86 $\mu\text{g}/\text{m}^3$ hingga 105 $\mu\text{g}/\text{m}^3$ dimana angka tersebut melebihi batas normal sebesar 17 sampai 21 kali lipat dari aturan yang ditetapkan oleh WHO sebesar 5 $\mu\text{g}/\text{m}^3$ dan PM₁₀ berada pada kisaran 54 $\mu\text{g}/\text{m}^3$ hingga 64 $\mu\text{g}/\text{m}^3$ yang berada dalam batas normal yang ditetapkan oleh WHO sebesar 50 $\mu\text{g}/\text{m}^3$. Hasil penelitian ini diharapkan dapat memberikan kontribusi penting dalam mendukung kebijakan mitigasi polusi udara dan membantu menyediakan sistem peringatan dini yang efektif untuk mengurangi dampak kesehatan jangka panjang akibat polusi udara di Jakarta Timur (DKI 4).

KATA KUNCI — Metrik *Error*, Peramalan Kualitas Udara, Polusi Udara, Tren Musiman.

I. PENDAHULUAN

Kesehatan merupakan salah satu kunci keberlangsungan hidup bagi setiap manusia. Setiap manusia membutuhkan lingkungan hidup yang sehat agar dapat mempertahankan kehidupannya, maka dari itu kualitas udara yang sehat menjadi salah satu komponen terpenting bagi manusia di sebuah lingkungan [1]. Kualitas udara merupakan salah satu faktor terpenting bagi kesehatan manusia karena udara menjadi salah satu elemen terpenting bagi manusia agar dapat bernafas [2]. Indeks Standar Pencemaran Udara (ISPU) merupakan sebuah parameter untuk mengukur kualitas udara. Pengukuran ISPU tercatat enam parameter yang digunakan untuk pengamatan pencemaran udara, yaitu Sulfur Dioksida (SO₂), Nitrogen Dioksida (NO₂), Ozon Permukaan (O₃), Karbon Monoksida (CO), Partikel Debu (PM₁₀), dan Partikel Debu (PM_{2.5}). Polutan udara terutama PM_{2.5} dan PM₁₀ menjadi perhatian utama dalam penilaian kualitas udara karena ukurannya yang sangat kecil, sehingga dapat terhirup tanpa disaring oleh tubuh dan berdampak negatif bagi kesehatan manusia, terutama sistem pernafasan. Kedua polutan tersebut sering disebut sebagai

polutan yang paling berbahaya, seperti yang telah diungkapkan dalam berbagai studi terdahulu [3], [4], [5], [6].

Pada tahun 2023, Indonesia menempati peringkat ke-14 dengan tingkat polusi udara tertinggi di dunia dengan nilai PM_{2.5} sebesar 37.1 $\mu\text{g}/\text{m}^3$ dimana angka tersebut telah melebihi tujuh kali dari batas normal yang ditetapkan oleh WHO berdasarkan *AirVisual* oleh AQI. Masalah polusi udara ini semakin serius di Indonesia, khususnya di kota-kota besar seperti Jakarta yang dapat membawa dampak negatif terhadap keberlangsungan hidup manusia dan lingkungan di sekitarnya [7]. Referensi [8] menunjukkan bahwa DKI Jakarta memiliki jumlah penduduk sebesar 10.672.100 dan jumlah kendaraannya mencapai 12.035.299 kendaraan. Hal ini menjadi perhatian karena jumlah kendaraan di DKI Jakarta lebih banyak sebesar 1.12% sehingga DKI Jakarta memiliki risiko yang cukup tinggi terkena polusi udara akibat kendaraan bermotor yang mendominasi kota ini.

Indonesia telah mengalami peningkatan polusi partikulat selama satu dekade terakhir dan sekitar 93% dari 262 juta penduduk di Indonesia tinggal di daerah dengan kadar PM_{2.5}

melebihi batas normal yang telah ditetapkan oleh WHO [9]. Oleh karena itu, penelitian ini bertujuan untuk menemukan model yang dapat memberikan wawasan mendalam mengenai konsentrasi polutan $PM_{2.5}$ dan PM_{10} di wilayah DKI Jakarta dengan konsentrasi polutan tertinggi. Penelitian ini tidak hanya berfokus pada akurasi peramalan, tetapi juga bagaimana model tersebut dapat memberikan informasi yang bermanfaat dan mudah diinterpretasikan untuk kebijakan mitigasi polusi udara.

II. KAJIAN TERKAIT

Penelitian mengenai prediksi polusi udara telah berkembang pesat dengan berbagai model prediktif yang diterapkan selama beberapa dekade terakhir. Salah satu pendekatan yang sering digunakan dalam analisis deret waktu adalah *Autoregressive Integrated Moving Average* (ARIMA). Referensi [10] menunjukkan bahwa model ARIMA memberikan hasil yang akurat dalam memprediksi kualitas udara di Jakarta Pusat dengan data $PM_{2.5}$ dan PM_{10} dari tahun 2015 hingga 2020. Hasil penelitian tersebut menunjukkan tingkat akurasi lebih dari 90%, menekankan keandalan model ARIMA dalam menangani data polusi udara dengan pola musiman yang jelas. Referensi [11] juga menunjukkan bahwa model ARIMA mampu memprediksi $PM_{2.5}$ di Semarang dengan tingkat kesalahan Mean Absolute Percentage Error (MAPE) sebesar 13.71% yang menunjukkan performa prediksi yang cukup baik.

Akan tetapi, model ARIMA juga memiliki beberapa keterbatasan, terutama ketika diterapkan pada data yang memiliki sifat non-linear atau fluktuasi yang sangat dinamis. Referensi [12] di Hyderabad, India, menunjukkan bahwa ARIMA efektif dalam memprediksi NO_2 dengan proyeksi prediksi hingga 30 tahun ke depan. Meskipun akurat, model ini lebih cocok digunakan untuk data yang memiliki tren jangka panjang yang stabil dan kurang optimal ketika data bersifat lebih variatif.

Untuk mengatasi keterbatasan tersebut, beberapa penelitian telah menerapkan model lain yang lebih kompleks, seperti *Recurrent Neural Network-Long Short Term Memory* (RNN-LSTM). Referensi [13] dalam penelitiannya di Surabaya, menemukan bahwa LSTM lebih unggul dibandingkan ARIMA, terutama dalam memprediksi polutan seperti CO dan SO_2 dengan kesalahan prediksi yang lebih rendah. Hasil ini mengonfirmasi bahwa model *deep learning* lebih mampu menangani data dengan variabilitas tinggi dan pola yang lebih kompleks. Meskipun LSTM unggul dalam menangani data *time series* yang kompleks dan memberikan akurasi prediksi yang tinggi namun kelemahan utamanya adalah ketidakmampuannya untuk memberikan interpretasi yang mendalam terhadap data. Selain hanya menghasilkan prediksi tinggi dengan akurasi tinggi, model ini tidak dapat diinterpretasikan dengan baik yang menjadi masalah dalam konteks *data mining*, di mana pemahaman yang bermakna dari data sangat penting. Jika fokusnya hanya pada prediksi, LSTM adalah pilihan yang baik, tetapi untuk *data mining* yang bertujuan memahami dan mengeksplorasi pola data, kelemahan ini menjadi signifikan. Berbeda dengan SARIMA yang tidak hanya memberikan

prediksi akurat, tetapi juga memberikan wawasan mengenai seberapa pengaruh dari *series* (melalui nilai p , d , q), efek musiman, serta periode musiman dari data tersebut. Hal tersebut menjadi kekuatan SARIMA yang mampu menjelaskan dan menceritakan pola yang terjadi di data, serta memberikan *insight* lebih mendalam. Di sisi lain, LSTM membutuhkan sumber daya komputasi yang besar dan waktu pelatihan yang lebih lama, tetapi alasan ini lebih bersifat tambahan dibandingkan kelemahan utamanya, yakni non interpretabilitas. Kemudian, akurasi adalah fokus utama dalam *machine learning* termasuk *deep learning* seperti LSTM.

Selain model LSTM, metode seperti ARIMAX yang menggabungkan variabel eksogen seperti faktor meteorologi juga terbukti mampu meningkatkan akurasi prediksi. Referensi [14] menemukan bahwa ARIMAX menghasilkan prediksi yang lebih baik untuk Air Quality Index (AQI) di Jakarta dengan memperhitungkan variabel suhu udara dan kecepatan angin. Hal ini menunjukkan bahwa menambahkan faktor eksternal dapat membantu menangkap kompleksitas data yang tidak dapat ditangani oleh model ARIMA murni. Namun, model ini memiliki kekurangan, yaitu ketergantungan yang tinggi pada kualitas data eksternal yang digunakan serta peningkatan kompleksitas dalam pemodelan.

Pada penelitian ini tetap memilih model SARIMA karena kemampuannya dalam menangkap pola musiman, tren, dan perubahan jangka panjang yang sangat penting dalam peramalan konsentrasi polutan PM_{10} dan $PM_{2.5}$. Kontribusi utama dari penelitian ini adalah pemanfaatan data terkini yang mencakup periode lebih panjang, yaitu dari Januari 2020 hingga Agustus 2024, serta fokus mencari model SARIMA terbaik yang dapat menjelaskan perkembangan kondisi polusi tersebut hingga dapat diramalkan kondisi polusi ke depannya dengan perhatian khusus pada wilayah DKI Jakarta yang memiliki konsentrasi tertinggi dari kedua polutan tersebut. Penelitian ini juga berhipotesis bahwa konsentrasi PM_{10} dan $PM_{2.5}$ mungkin dipengaruhi oleh faktor musiman yang akan diuji melalui hasil prediksi dari model SARIMA. Penelitian ini juga melakukan evaluasi mendalam dengan membandingkan hasil peramalan dari model SARIMA dengan data aktual melalui metode *k-folds cross validation* untuk memastikan akurasi dan keandalan model yang jarang diterapkan dalam penelitian sebelumnya. Hasil penelitian ini diharapkan dapat menjadi dasar rekomendasi bagi pemerintah dalam sistem peringatan dini dan kebijakan mitigasi polusi udara yang lebih efektif.

III. SOLUSI DAN USULAN

Penelitian ini akan menggunakan metode analisis berbasis *time series* untuk melakukan peramalan konsentrasi polutan udara. Proses peramalan ini mengacu pada pendekatan *data mining* yang dirancang untuk mengidentifikasi pola dan tren dalam data historis untuk menghasilkan peramalan yang akurat mengenai konsentrasi PM_{10} dan $PM_{2.5}$.

A. Data Mining

Data Mining adalah proses kompleks yang bertujuan mengekstraksi informasi tersembunyi dari kumpulan data besar.

Teknik ini mengidentifikasi pola atau hubungan yang tidak mudah diperoleh melalui kueri biasa, serta digunakan untuk membantu organisasi dalam pengambilan keputusan atau memaksimalkan keuntungan. *Data Mining* pada dasarnya merupakan upaya menemukan pengetahuan baru dalam basis data [15].

B. Time Series Analysis

Analisis deret waktu atau *time series analysis* adalah metode penelitian yang mempelajari data berurutan dari waktu ke waktu untuk menggambarkan struktur data serta memprediksi nilai masa mendatang [16]. Dalam analisis *time series*, terdapat beberapa komponen utama, yaitu:

- Level (L)*: Rata-rata dalam *time series*
- Trend (T)*: Pola peningkatan atau penurunan dalam data
- Seasonality (S)*: Pola berulang dengan siklus jangka pendek
- Noise (N)*: Variasi acak dalam data

Keempat komponen ini mempengaruhi sifat stasioneritas data. Apabila *time series* mengandung *seasonality*, *trend*, dan *level* yang terus berubah, maka diperlukan proses *differencing* untuk mengatasi ketidakstasioneran data [17], [18].

C. ARMA (Autoregressive Moving Average)

Model *Autoregressive Moving Average* (ARMA) merupakan metode prediksi untuk data deret waktu (*univariate time series*) yang mengombinasikan pendekatan *Autoregressive* (AR) dan *Moving Average* (MA) [19]. Model ini memiliki dua parameter utama, yaitu p dan q sehingga diekspresikan sebagai ARMA (p, q) atau ARIMA ($p, 0, q$) dengan formula matematis sebagai berikut [20]:

$$Y_t = \mu + \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (1)$$

Dengan Y_t adalah nilai peramalan di waktu t , μ mewakili konstanta, $\phi_1, \phi_2, \dots, \phi_p$ adalah koefisien untuk komponen *autoregressive*, $\theta_1, \theta_2, \dots, \theta_q$ adalah parameter *moving average* sedangkan ε_{t-q} merupakan *error* atau sisa pada waktu $t - q$ dan $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ adalah nilai-nilai sebelumnya yang digunakan dalam proses peramalan.

D. ARIMA (Autoregressive Integrated Moving Average)

Autoregressive Integrated Moving Average (ARIMA) adalah sebuah model peramalan untuk data *time series* yang bersifat univariat dimana model ini menggabungkan model *Autoregressive* dan *Moving Average* untuk data yang bersifat stasioner [21],[22]. Secara matematis, model ARIMA (p, d, q) dapat dinyatakan secara berikut

$$Y_t = \mu + \sum_{i=1}^p \phi_m Y_{t-1} + \sum_{j=1}^q \theta_n \varepsilon_{t-j} \quad (2)$$

Dengan μ adalah rata-rata dari *time series* stokastik, ϕ_m adalah parameter *autoregressive* dari AR, θ_n adalah koefisien dari parameter *moving average*, Y_t adalah waktu prediksi pada waktu t , y_{t-1} adalah nilai *time series* dari waktu $t - 1$, dan ε_{t-j} adalah error dalam nilai yang diramalkan.

E. SARIMA (Seasonal Autoregressive Integrated Moving Average)

Seasonal Autoregressive Integrated Moving Average (SARIMA) adalah sebuah model yang berguna untuk memodelkan deret waktu musiman di mana rata-rata dan

statistik lainnya dapat berubah-ubah sepanjang tahun, sehingga tidak stasioner pada setiap musim. SARIMA merupakan pengembangan langsung dari model ARIMA *non seasonal* dan ARIMA yang memperhitungkan komponen musiman dalam analisis deret waktu [23].

Bentuk umum model ARIMA *seasonal* dilambangkan sebagai ARIMA (p, d, q) \times (P, D, Q) s , di mana p mewakili urutan AR *non seasonal*, d adalah tingkat *differencing non-seasonal*, q adalah urutan MA musiman, dan s merupakan periode dari pola musiman yang berulang [24].

F. Metrik Evaluasi Kinerja Model

Evaluasi kinerja model dalam penelitian ini dilakukan menggunakan metrik *error* MAE, RMSE, dan MAPE. Metrik-metrik ini digunakan untuk mengukur akurasi peramalan model dengan membandingkan hasil peramalan dengan data aktual [25].

1) Mean Absolute Error (MAE)

MAE adalah metrik yang mengukur rata-rata selisih absolut antara nilai peramalan dan nilai aktual, dirumuskan dengan:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

dengan n adalah jumlah observasi, y_i adalah nilai aktual pada observasi ke- i , dan \hat{y}_i adalah nilai prediksi pada observasi ke- i . Semakin kecil nilai MAE, maka semakin baik kinerja model dalam meramalkan.

2) Root Mean Squared Error (RMSE)

RMSE mengukur akar kuadrat dari rata-rata selisih kuadrat antara nilai peramalan dan nilai aktual, dirumuskan dengan:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

dengan n adalah jumlah observasi, y_i adalah nilai aktual pada observasi ke- i , dan \hat{y}_i adalah nilai prediksi pada observasi ke- i .

3) Mean Absolute Percentage Error (MAPE)

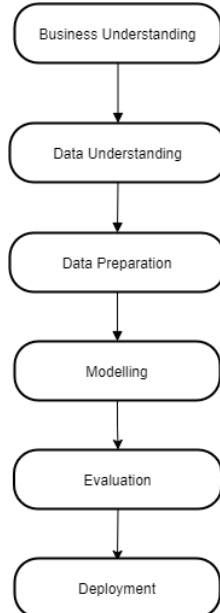
MAPE mengukur rata-rata persentase kesalahan peramalan dengan membandingkan selisih absolut antara nilai peramalan dan aktual dibagi dengan nilai aktual kemudian dikalikan 100 untuk mendapatkan persentase, dengan rumus:

$$MAPE = \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (5)$$

dengan n adalah jumlah observasi, y_i adalah nilai aktual pada observasi ke- i , dan \hat{y}_i adalah nilai prediksi pada observasi ke- i . MAPE sering digunakan untuk mengevaluasi seberapa baik model dalam meramalkan nilai dalam konteks persentase. Nilai MAPE yang lebih rendah menunjukkan model yang lebih akurat.

G. Metode Penelitian

Metode penelitian yang digunakan pada penelitian ini adalah CRISP-DM yang disajikan dalam bentuk diagram alur di bawah ini untuk memudahkan pemahaman serta memberikan gambaran yang jelas mengenai setiap tahapan yang dilakukan. Berikut adalah tahapan-tahapan penelitian yang dilaksanakan.



Gambar 1. Diagram alur tahapan penelitian

1) Business Understanding

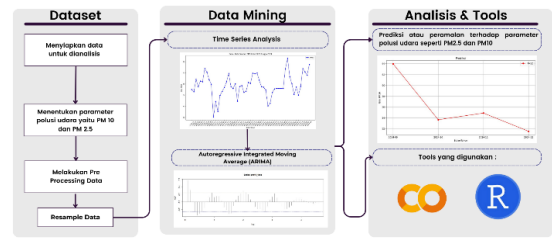
Business Understanding pada tahap ini dilakukan untuk memahami konsep *Data Mining*, yaitu memahami kondisi polusi udara di wilayah DKI Jakarta bagian timur pada tahun 2020 hingga Agustus 2024, dan menentukan model ARIMA terbaik yang dapat memberikan hasil peramalan untuk bulan September hingga November 2024 serta mengevaluasi kinerjanya dalam meramalkan konsentrasi polutan udara PM_{10} dan $PM_{2.5}$ yang telah diidentifikasi sebagai polutan udara paling berbahaya.

2) Data Understanding

Data yang digunakan dalam penelitian ini adalah data harian Indeks Standar Pencemaran Udara (ISPU) dari Januari 2020 hingga Agustus 2024 yang bersumber dari Satu Data Indonesia. Data diperoleh dari lima Stasiun Pemantau Kualitas Udara (SPKU) di DKI Jakarta. Data yang digunakan memiliki 6573 observasi untuk PM_{10} , 4898 observasi untuk $PM_{2.5}$, 9 variabel, dan tipe data yang bersifat numerik.

3) Data Preparation

Data Preparation pada penelitian ini bertujuan untuk memberikan pemahaman yang jelas mengenai tahapan yang dilakukan dalam menganalisis data polutan udara, seperti ditunjukkan pada Gambar 2 berikut:



Gambar 2. Diagram gambaran Data Preparation

Gambar 2 menunjukkan alur penelitian yang dimulai dari proses pengolahan *dataset*, yaitu mempersiapkan data polutan udara PM_{10} dan $PM_{2.5}$ untuk dianalisis. Data yang diperoleh akan melewati beberapa tahapan, termasuk *preprocessing* yang hanya berfokus pada variabel PM_{10} dan $PM_{2.5}$ mencakup penyatuan data (*data integration*) dan penanganan nilai yang hilang (*missing value*) [26], serta *resampling* untuk menyesuaikan data dengan kebutuhan analisis *time series*. Pada tahap *Data Mining* dilakukan analisis menggunakan model SARIMA untuk meramalkan polusi udara berdasarkan data historis, di mana model ini dipilih karena kemampuannya menangani data berurutan dengan pola berulang. Setelah peramalan dilakukan, hasilnya dianalisis dan dibandingkan dengan data aktual untuk mengukur akurasi model. Seluruh proses, mulai dari pengolahan data hingga penerapan model SARIMA dilakukan menggunakan *Google Colab (Python)* dan *R Studio*.

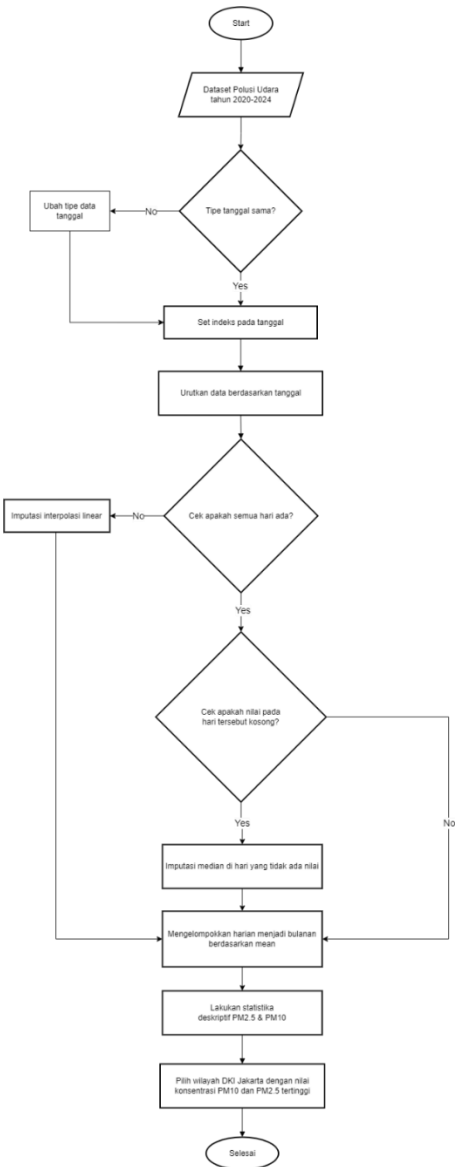
Preprocessing dalam penelitian ini merupakan tahap penting untuk memastikan data siap digunakan dalam analisis *time series*. Berikut adalah langkah-langkah yang dilakukan dalam melakukan *preprocessing* data polusi udara untuk *time series* berdasarkan Gambar 3.

Proses dimulai dengan pengumpulan *dataset* polusi udara yang mencakup data dari tahun 2020 hingga 2024. Langkah pertama adalah pengecekan tipe data untuk tanggal, memastikan bahwa tipe data tersebut konsisten. Jika ditemukan ketidaksesuaian, dilakukan perubahan tipe data agar sesuai. Setelah tipe data tanggal sudah sesuai, data diatur berdasarkan tanggal sebagai indeks dan diurutkan untuk mempermudah analisis.

Data yang sudah diatur berdasarkan tanggal sebagai indeks kemudian akan diperiksa untuk memastikan bahwa semua tanggal dalam *dataset* tercatat secara lengkap. Jika ada tanggal yang hilang, digunakan metode interpolasi linear untuk mengisi tanggal yang kosong tersebut. Selain itu, pengecekan juga dilakukan untuk memastikan apakah terdapat nilai yang hilang pada hari tertentu. Jika ditemukan nilai yang hilang, metode imputasi median digunakan untuk mengisi kekosongan nilai tersebut karena metode ini cocok untuk data yang tidak terdistribusi normal.

Data harian yang sudah diolah dikelompokkan menjadi data bulanan berdasarkan rata-rata nilai harian. Tahap berikutnya adalah melakukan analisis statistik deskriptif untuk melihat distribusi, tren, dan pola dari data PM_{10} dan $PM_{2.5}$. Terakhir, dipilih daerah di wilayah DKI Jakarta dengan konsentrasi

polutan PM_{10} dan $PM_{2.5}$ tertinggi yang akan dianalisis lebih lanjut.

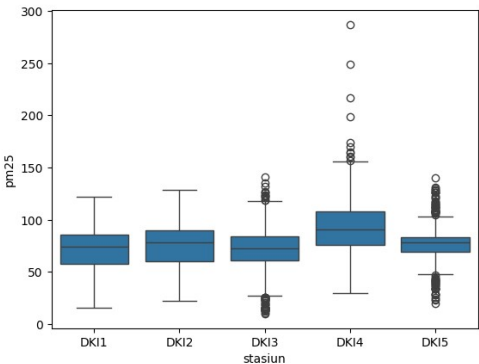


Gambar 3. Diagram *preprocessing* data time series

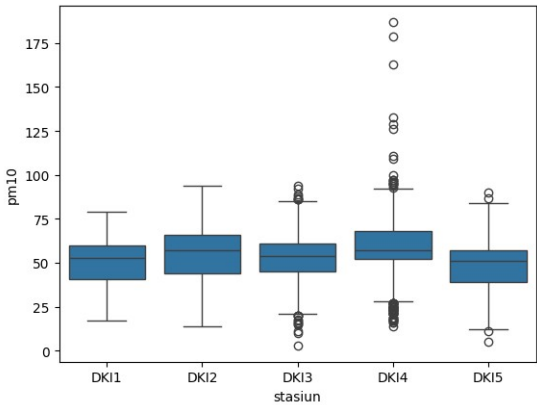
Data seluruh wilayah DKI Jakarta yang sudah melewati *preprocessing* akan dilakukan analisis statistik deskriptif untuk menentukan wilayah DKI Jakarta yang paling berpolusi pada polutan PM_{10} dan $PM_{2.5}$ seperti tersaji pada Tabel I dan Tabel II. Gambar 4 dan Gambar 5 menunjukkan hasil *box plot* untuk polutan PM_{10} dan $PM_{2.5}$ di wilayah DKI Jakarta.

TABEL I HASIL STATISTIKA DESKRIPTIF UNTUK $PM_{2.5}$					
	DKI 1 (Pusat)	DKI 2 (Utara)	DKI 3 (Selatan)	DKI 4 (Timur)	DKI 5 (Barat)
COUNT	978,0	1013,0	1007,0	1239,0	661,0
Mean	71,36	75,18	71,69	91,66	76,72
Std	19,24	20,02	20,84	25,86	17,60
Min	16,0	22,0	10,0	30,0	20,0
25%	58,0	60,0	61,0	75,5	69,0
50%	74,0	78,0	73,0	91,0	78,0
75%	86,0	90,0	84,0	108,0	83,0
Max	122,0	129,0	141,0	287,0	140,0

TABEL II HASIL STATISTIKA DESKRIPTIF UNTUK PM_{10}					
	DKI 1 (Pusat)	DKI 2 (Utara)	DKI 3 (Selatan)	DKI 4 (Timur)	DKI 5 (Barat)
COUNT	1313,0	1348,0	1342,0	1574,0	996,0
Mean	50,09	55,03	52,43	58,82	47,61
Std	12,93	15,27	12,90	15,96	13,62
Min	17,0	14,0	3,0	14,0	5,0
25%	41,0	44,0	45,0	52,0	39,0
50%	53,0	57,0	54,0	57,0	51,0
75%	60,0	66,0	61,0	68,0	57,0
Max	79,0	94,0	94,0	187,0	90,0

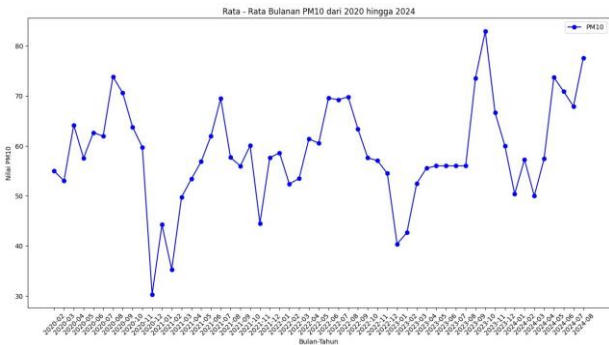


Gambar 4. Box plot polutan $PM_{2.5}$ di wilayah DKI Jakarta.

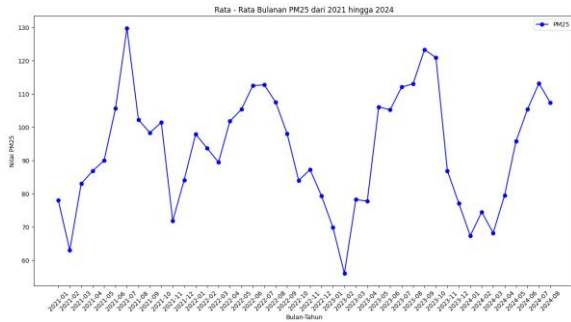


Gambar 5. Box plot polutan PM_{10} di wilayah DKI Jakarta.

Berdasarkan hasil pada Tabel I dan Tabel II, dapat dilihat bahwa DKI 4 memiliki nilai mean dan max tertinggi dibanding wilayah DKI lainnya. Gambar 4 dan Gambar 5 menunjukkan bahwa DKI 4 memiliki banyak *outlier*. Hal ini menunjukkan bahwa DKI 4 memiliki tingkat polusi yang ekstrem dibanding wilayah DKI lainnya.



Gambar 6. Tren PM_{10} dari tahun 2020 - 2024

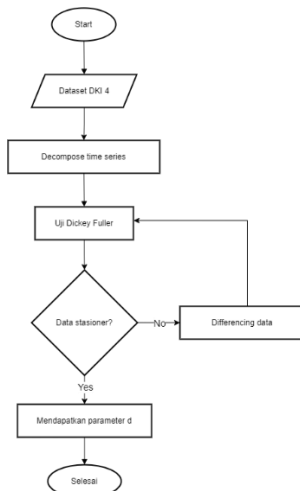


Gambar 7. Tren $PM_{2.5}$ dari tahun 2021 – 2024

Berdasarkan Gambar 6 dan Gambar 7, PM_{10} menunjukkan fluktuasi signifikan dari 2020 hingga 2024, serta beberapa penurunan tajam pada awal 2021. Sementara itu, $PM_{2.5}$ juga fluktuatif dengan puncak tertinggi pada pertengahan 2023 dan penurunan tajam pada akhir 2021. Hal tersebut mengindikasikan potensi adanya pengaruh eksternal yang lebih signifikan, seperti kebijakan pembatasan aktivitas selama pandemi.

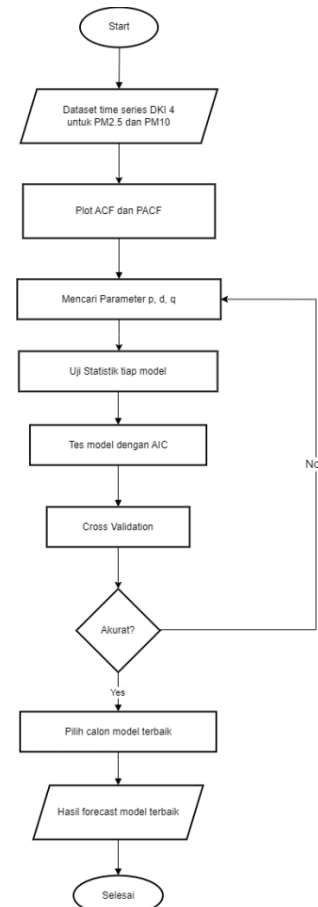
4) Modelling

Model yang digunakan pada penelitian ini adalah SARIMA. Data yang akan digunakan untuk analisis *time series* dengan menggunakan SARIMA perlu diperhatikan apakah datanya stasioner atau tidak. Berikut adalah diagram alur untuk memeriksa apakah datanya stasioner atau tidak. Berikut diagram alur uji data stasioner.



Gambar 8. Diagram alur uji data stasioner.

Penentuan parameter SARIMA (p, d, q) $(P, D, Q)[s]$ dalam penelitian ini mengikuti langkah-langkah seperti yang ditunjukkan pada Gambar 8. Dimulai dari analisis *dataset time series* untuk $PM_{2.5}$ dan PM_{10} di wilayah Jakarta Timur (DKI 4) dan dilanjutkan dengan plot ACF dan PACF untuk menentukan parameter p, d, q .



Gambar 9. Diagram memilih model SARIMA terbaik

5) Evaluation

Pada tahap ini, model akan melalui serangkaian uji statistik seperti Uji Ljung-Box dan Uji Saphiro Wilk. Uji Ljung-Box merupakan uji statistik yang digunakan untuk memeriksa apakah residualnya saling independen sedangkan Uji Saphiro-Wilk digunakan untuk memeriksa apakah residualnya berdistribusi normal. Setelah model memenuhi asumsi dari uji statistik seperti Uji Ljung-Box dan Uji Saphiro-Wilk, model tersebut akan dievaluasi berdasarkan nilai metrik *error* terkecil agar bisa mendapatkan model dengan nilai akurasi terbesar sehingga hasil peramalan akan menjadi lebih akurat.

6) Deployment

Pada tahap ini, model terbaik yang telah diperoleh untuk polutan $PM_{2.5}$ dan PM_{10} akan diterapkan untuk melakukan peramalan pada bulan September hingga November 2024. Hasil dari peramalan tersebut nantinya akan dianalisa informasi penting mengenai polusi udara di wilayah DKI Jakarta Timur (DKI 4) pada bulan tersebut.

IV. HASIL EKSPERIMEN DAN PENGUJIAN

Metode yang digunakan adalah metode SARIMA dengan melakukan beberapa tahapan seperti *differencing* hingga datanya menjadi stasioner dan dapat diimplementasikan di model SARIMA.

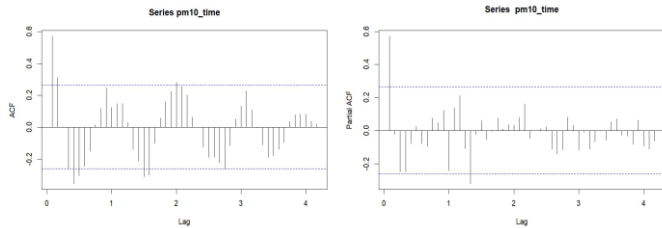
A. Differencing dan Uji Augmented Dickey-Fuller

Data yang telah melewati tahapan *preprocessing* selanjutnya akan diperiksa apakah data tersebut bersifat stasioner atau tidak. Dengan melakukan Uji Augmented Dickey-Fuller (ADF Test) dengan hipotesis sebagai berikut,

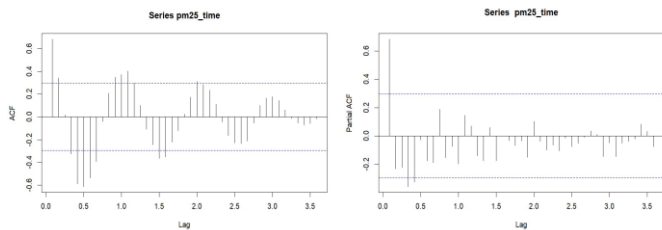
H_0 : Data tidak stasioner

H_1 : Data stasioner

Gambar 9 dan Gambar 10 menunjukkan *plot* ACF dan PACF untuk masing-masing polutan



Gambar 10. Diagram ACF (kiri) dan PACF (kanan) polutan PM₁₀



Gambar 11. Diagram ACF (kiri) dan PACF (kanan) polutan PM_{2.5}

Tabel III menunjukkan hasil *Augmented Dickey-Fuller* untuk masing-masing PM_{2.5} dan PM₁₀

TABEL III

HASIL STATISTIKA DESKRIPTIF UNTUK PM₁₀

Hasil Uji				
Polutan	Dickey-Fuller	Lag Order	p-value	Keputusan
PM _{2.5}	-4,4437	3	0.01	H_0 ditolak
PM ₁₀	-4,1204	3	0.0108	H_0 ditolak

Berdasarkan Gambar 10 dan Gambar 11, hasil ACF dan PACF untuk kedua polutan memiliki *lag* yang berpola, hal tersebut menunjukkan bahwa datanya sudah bersifat stasioner. Tabel III juga menghasilkan Uji Dickey-Fuller yang dilakukan terhadap data polutan PM_{2.5} dan PM₁₀, didapatkan keduanya memiliki nilai *p-value* kurang dari 0,05. Akibatnya, H_0 ditolak dan data tersebut tidak perlu dilakukan *differencing* karena sudah stasioner.

B. Implementasi Pada Model Terbaik

Parameter p, d, q pada SARIMA dapat ditentukan menggunakan *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF). Seluruh model terbaik yang dihasilkan oleh SARIMA akan dicari nilai *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE) untuk menentukan model terbaik dalam meramalkan PM_{2.5} dan PM₁₀.

Data yang telah dilakukan *preprocessing* akan dilakukan uji statistik dengan melihat nilai dari Uji Ljung-Box dan Uji Shapiro-Wilk. Tabel dibawah ini menunjukkan hasil *p-value* dari semua Uji Statistik yang digunakan dan nilai *p-value* agar H_0 diterima untuk kedua uji statistik tersebut harus lebih besar

dari 0,05. Berikut adalah hipotesis untuk uji-statistik yang digunakan :

- 1) Hipotesis Uji Shapiro-Wilk

H_0 : Residualnya berdistribusi normal

H_1 : Residualnya tidak berdistribusi normal

- 2) Hipotesis Uji Ljung-Box

H_0 : Antar residualnya saling independen

H_1 : Antar residualnya tidak saling independen

TABEL IV

HASIL Uji STATISTIK MODEL UNTUK PM₁₀

Hasil Uji Statistik				
Model SARIMA	L-jung Box	Keputusan	Shapiro -Wilk	Keputusan
(1,0,1)(1,0,0)[12]	0,0254	H_0 ditolak	0,0112	H_0 ditolak
(0,0,2)(0,0,1)[12]	0,0221	H_0 ditolak	0,0062	H_0 ditolak
(2,0,0)(0,0,1)[12]	0,0303	H_0 ditolak	0,0108	H_0 ditolak
(1,0,2)(1,0,1)[12]	0,0156	H_0 ditolak	0,0234	H_0 ditolak
(2,0,1)(1,0,1)[12]	0,1143	H_0 tidak ditolak	0,0211	H_0 ditolak
(1,0,2)(0,0,1)[12]	0,0411	H_0 ditolak	0,0386	H_0 ditolak
(3,0,2)(1,0,0)[12]	0,1973	H_0 tidak ditolak	0,3563	H_0 tidak ditolak
(2,0,3)(0,0,1)[12]	0,0201	H_0 ditolak	0,5872	H_0 tidak ditolak
(1,0,1)(1,0,1)[12]	0,0103	H_0 ditolak	0,0050	H_0 ditolak
(2,0,2)(1,0,0)[12]	0,2369	H_0 tidak ditolak	0,234	H_0 tidak ditolak

TABEL V

HASIL Uji STATISTIK MODEL UNTUK PM_{2.5}

Hasil Uji Statistik				
Model SARIMA	L-jung Box	Keputusan	Shapiro -Wilk	Keputusan
(2,0,0)(1,0,0)[12]	0,0185	H_0 ditolak	0,8517	H_0 tidak ditolak
(1,0,1)(1,0,0)[12]	0,0187	H_0 ditolak	0,8409	H_0 tidak ditolak
(0,0,2)(1,0,0)[12]	0,0072	H_0 ditolak	0,6869	H_0 tidak ditolak
(2,0,1)(1,0,0)[12]	0,0955	H_0 tidak ditolak	0,4997	H_0 ditolak
(1,0,2)(1,0,0)[12]	0,0090	H_0 ditolak	0,8492	H_0 tidak ditolak
(2,0,2)(1,0,0)[12]	0,0073	H_0 ditolak	0,7262	H_0 tidak ditolak
(1,0,1)(0,0,1)[12]	0,0149	H_0 ditolak	0,8884	H_0 tidak ditolak
(2,0,0)(0,0,1)[12]	0,0166	H_0 ditolak	0,8783	H_0 tidak ditolak
(2,0,1)(0,0,1)[12]	0,0894	H_0 tidak ditolak	0,4297	H_0 tidak ditolak
(1,0,0)(1,1,0)[12]	0,0787	H_0 tidak ditolak	0,0528	H_0 tidak ditolak

Model SARIMA pada Tabel IV menunjukkan ada dua model yang memenuhi *p-value* yaitu model SARIMA (3,0,2)(1,0,0)[12] dan SARIMA (2,0,2)(1,0,0)[12]. Sedangkan,

pada Tabel V, model SARIMA yang memenuhi p -value adalah $(2,0,1)(0,0,1)[12]$ dan $(1,0,0)(1,1,0)[12]$.

TABEL VI
HASIL ERROR MODEL UNTUK PM_{10}

Error				
Model	AIC	RMSE	MAE	MAPE
$(3,0,2)(1,0,0)[12]$	392,78	6,218	4,792	8,874%
$(2,0,2)(1,0,0)[12]$	395,82	6,904	5,234	9,637%

TABEL VII
HASIL ERROR MODEL UNTUK $PM_{2.5}$

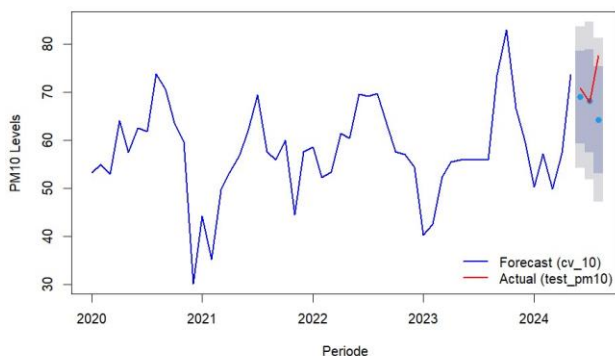
Error				
Model	AIC	RMSE	MAE	MAPE
$(1,0,0)(1,1,0)[12]$	263,77	10,32	7,87	8,85%
$(2,0,2)(1,0,0)[12]$	361,92	12,05	9,87	11,33%

Berdasarkan hasil *error* yang diperoleh pada Tabel VI, model SARIMA $(3,0,2)(1,0,0)[12]$ memiliki nilai *error* RMSE, MAE, dan MAPE, serta nilai AIC terkecil dibanding model SARIMA $(2,0,2)(1,0,0)[12]$ sehingga model $(3,0,2)(1,0,0)[12]$ merupakan model terbaik untuk PM_{10} . Sedangkan, untuk $PM_{2.5}$ yang dihasilkan pada Tabel VII, model SARIMA yang memiliki nilai *error* dan AIC terkecil adalah model SARIMA $(1,0,0)(1,1,0)[12]$.

C. Hasil K-Folds Cross Validation

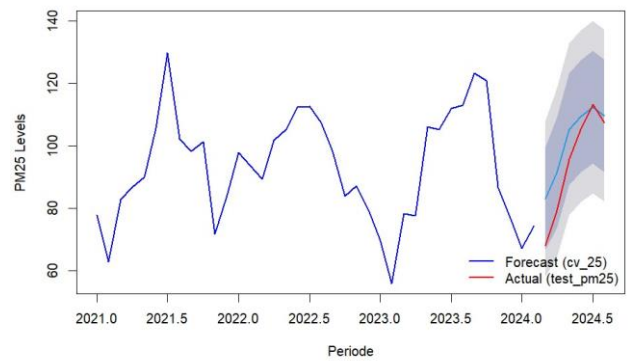
Model terbaik yang telah didapatkan akan dilalui proses *k-folds cross validation* untuk memverifikasi model dan melihat seberapa akurat model sebelum dilakukan *peramalan*. Gambar 12 menunjukkan hasil *3-folds cross validation* untuk model SARIMA $(3,0,2)(1,0,0)[12]$ serta Gambar 13 menunjukkan hasil *6-folds cross validation* untuk model SARIMA $(1,0,0)(1,1,0)[12]$. Warna abu-abu terang menunjukkan selang kepercayaan 95% sedangkan warna abu-abu gelap menunjukkan selang kepercayaan 80%.

3-Fold Cross Validation Forecast vs Actual



Gambar 12. Hasil 3-folds cross validation polutan PM_{10}

6-Fold Cross Validation Forecast vs Actual



Gambar 13. Hasil 6-folds cross validation polutan $PM_{2.5}$

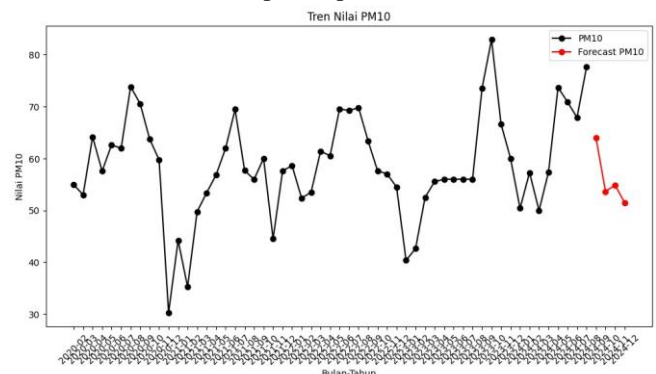
TABEL VIII
HASIL CROSS VALIDATION BULAN JUNI - AGUSTUS

		Aktual		Peramalan	
Daerah	Bulan	$PM_{2.5}$	PM_{10}	$PM_{2.5}$	PM_{10}
DKI 4	Juni	105,372	70,86	109,495	68,99
	Juli	113,218	67,90	112,426	68,26
	Agustus	107,352	77,61	109,653	64,25

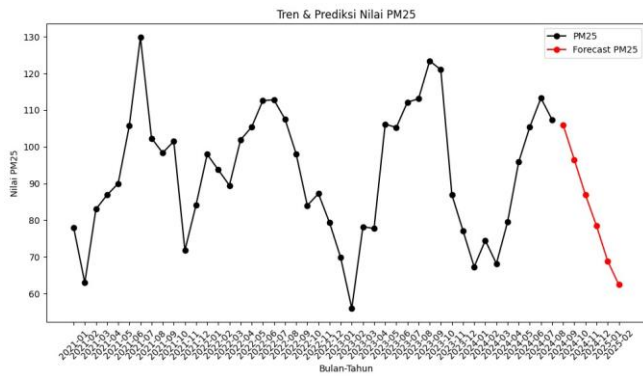
Berdasarkan hasil *cross validation* yang diperoleh pada Tabel VII, hasil peramalan yang dihasilkan oleh model tersebut mendekati dengan data aktual. Hasil peramalan tersebut juga dapat diverifikasi melalui diagram *cross validation* pada Gambar 12 dan Gambar 13 dimana data aktual tersebut berada di dalam selang kepercayaan peramalan yaitu 80% dan 95%.

D. Hasil Peramalan Polusi Udara di Wilayah DKI Jakarta Timur (DKI 4)

Hasil peramalan PM_{10} di wilayah DKI Jakarta Timur (DKI 4) dengan menggunakan model SARIMA $(3,0,2)(1,0,0)[12]$ yang ditampilkan pada Gambar 14 dan hasil peramalan $PM_{2.5}$ di wilayah DKI Jakarta Timur (DKI 4) dengan model SARIMA $(1,0,0)(1,1,0)[12]$ ditampilkan pada Gambar 15.



Gambar 14. Tren dan hasil peramalan polutan $PM_{2.5}$

Gambar 15. Tren dan hasil peramalan polutan PM₁₀

TABEL IX

HASIL PERMALAN BULAN SEPTEMBER - DESEMBER

Hasil Peramalan			
Daerah	Bulan	PM _{2.5}	PM ₁₀
DKI 4	September	105,925	63,95
	Oktober	96,481	53,64
	November	86,902	54,68

Berdasarkan hasil peralaman yang terlampir pada Tabel IX, polutan PM_{2.5} berada pada kisaran $86 \mu\text{g}/\text{m}^3$ hingga $105 \mu\text{g}/\text{m}^3$ dengan tingkat akurasi 91,13% dan polutan PM₁₀ berada pada kisaran $54 \mu\text{g}/\text{m}^3$ hingga $64 \mu\text{g}/\text{m}^3$ dengan tingkat akurasi 91,15%.

V. ANALISIS HASIL EKSPERIMEN DAN PENGUJIAN

Wilayah DKI Jakarta Timur (DKI 4) merupakan daerah yang paling berpolusi dibanding wilayah DKI lainnya berdasarkan analisa statistika deskriptif. Gambar 6 dan Gambar 7 menunjukkan tren polutan PM_{2.5} dan PM₁₀ dan dapat dilihat bahwa terdapat anomali pada tahun 2021 dimana grafik tersebut menurun. Hal ini dapat disebabkan karena adanya COVID-19 yang membuat aktivitas di luar sangat terbatas bagi masyarakat.

Perbedaan utama pada penelitian ini dibanding penelitian lainnya adalah tahap *preprocessing* dan *k-folds cross validation*. Salah satu tahap *preprocessing* yang berbeda dengan penelitian lainnya adalah data harian dikelompokkan menjadi bulanan berdasarkan rata-rata. Penelitian ini juga menggunakan metode *k-folds cross validation* untuk menguji keakuratan model yang telah dihasilkan sebelum dapat digunakan untuk melakukan peramalan pada bulan berikutnya. Data yang digunakan merupakan yang terbaru di tahun 2024 akan tetapi masih ada data yang hilang sehingga harus dilakukan metode imputasi. Penelitian ini tidak membandingkan model-model yang dihasilkan oleh berbagai teknik imputasi lainnya namun masih dapat memberikan hasil yang akurat seperti yang tercantum pada Tabel VIII.

Model terbaik yang telah diperoleh untuk meramalkan PM_{2.5} dan PM₁₀ di wilayah DKI Jakarta Timur (DKI 4) merupakan model SARIMA yang telah melewati *k-folds cross validation*. Model yang telah dibuat pada penelitian ini memiliki akurasi sebesar 91,13% untuk SARIMA (3,0,2)(1,0,0)[12] dan 91,15% untuk SARIMA (1,0,0)(1,1,0)[12]. Hasil *k-folds cross validation* untuk kedua model tersebut menunjukkan bahwa peramalan polutan mendekati dengan data aktual dan polutan

PM_{2.5} dan PM₁₀ dipengaruhi oleh data musiman. Hasil peramalan yang dihasilkan oleh Model SARIMA (3,0,2)(1,0,0)[12] untuk PM_{2.5} berada pada kisaran $86 \mu\text{g}/\text{m}^3$ hingga $105 \mu\text{g}/\text{m}^3$ dan SARIMA (1,0,0)(1,1,0)[12] untuk PM₁₀ berada pada kisaran $54 \mu\text{g}/\text{m}^3$ hingga $64 \mu\text{g}/\text{m}^3$ untuk bulan September hingga November. Berdasarkan hasil tersebut, DKI Jakarta Timur akan memiliki risiko polusi udara yang tinggi terutama polutan PM_{2.5} karena nilai konsentrasi PM_{2.5} berada di kisaran 17 sampai 21 kali lipat dari batas normal yang ditetapkan oleh WHO sebesar $5 \mu\text{g}/\text{m}^3$ dan untuk PM₁₀ masih berada di batas normal yang ditetapkan oleh WHO sebesar $50 \mu\text{g}/\text{m}^3$.

VI. KESIMPULAN

Penelitian ini berfokus pada menentukan model terbaik untuk meramalkan polutan PM_{2.5} dan PM₁₀ di wilayah DKI yang paling berpolusi. Penelitian ini telah mengidentifikasi wilayah DKI yang memiliki konsentrasi polutan PM_{2.5} dan PM₁₀ tertinggi yaitu DKI 4 berdasarkan analisa statistika deskriptif. Polutan PM_{2.5} dan PM₁₀ dipengaruhi oleh musiman berdasarkan diagram ACF dan PACF yang dimana diagram tersebut menunjukkan adanya *lag* berpola yang merupakan ciri khas dari data musiman. Penelitian ini juga menggunakan metode *cross validation* untuk menguji keakuratan model yang telah dibuat agar dapat memberikan hasil peramalan yang akurat. Hasil *cross validation* baik model SARIMA (3,0,2)(1,0,0)[12] maupun SARIMA (1,0,0)(1,1,0)[12] menunjukkan bahwa data aktual masih berada di selang kepercayaan 80% dan 95% dengan tingkat akurasi 91,13% dan 91,15%. Hasil peramalan menunjukkan bahwa PM_{2.5} berada pada kisaran $86 \mu\text{g}/\text{m}^3$ hingga $105 \mu\text{g}/\text{m}^3$ dimana angka tersebut melebihi batas normal sebesar 17 sampai 21 kali lipat dari aturan yang ditetapkan oleh WHO sebesar $5 \mu\text{g}/\text{m}^3$ dan PM₁₀ berada pada kisaran $54 \mu\text{g}/\text{m}^3$ hingga $64 \mu\text{g}/\text{m}^3$ untuk tiga bulan kedepan dengan PM₁₀ masih berada di batas normal yang ditetapkan oleh WHO sebesar $50 \mu\text{g}/\text{m}^3$. Hasil tersebut memberikan informasi bahwa wilayah DKI Jakarta Timur akan mengalami risiko polusi udara tinggi yang dapat berdampak terhadap kesehatan manusia apabila tidak ditangani dengan tepat. Risiko polusi udara tersebut dapat ditekankan dengan membuat aturan yang lebih ketat mengenai aktivitas yang berpotensi melibatkan polusi udara di wilayah DKI Jakarta Timur (DKI 4).

REFERENSI

- [1] A. Rifqi and R. T. Aldisa, "Penerapan Data Mining Untuk Clustering Kualitas Udara," *J. Sistem Komputer dan Informatika (JSON)*, vol. 5, no. 2, pp. 289-297, 2023.
- [2] K. Prabowo and B. Muslim, *Penyehat Udara*, 1st ed. Jakarta Selatan: Pusat Pendidikan Sumber Daya Manusia Kesehatan, 2018.
- [3] BMKG, "Informasi partikulat PM2.5," [Online]. Available: <https://www.bmkg.go.id/kualitas-udara/informasi-partikulat-pm25.bmkg>. [Accessed: 03-Oct-2024].
- [4] "Bahaya Polusi Udara bagi Kesehatan: Dampak, Penyebab dan Pencegahannya," *kemkes.go.id*.
- [5] I. Wellid, L. M. Simbolon, M. A. Falahuddin, N. Nurfitriani, K. Sumeru, M. F. bin Sukri, and N. Yuningsih, "Evaluasi Polusi Udara PM2.5 dan PM10 di

- Kota Bandung serta Kaitannya dengan Infeksi Saluran Pernafasan Akut," *J. Kesehatan Lingkungan Indonesia*, vol. 23, no. 2, pp. 129-137, 2024.
- [6] T. S. Maksun and S. F. N. Tarigan, "Analisis Risiko Kesehatan Akibat Paparan Partikel Debu (PM_{2.5}) dari Aktivitas Transportasi," *Jambura Health and Sport J.*, vol. 4, no. 1, pp. 19-28, 2022.
- [7] IQAir, "World's most polluted countries," [Online]. Available: <https://www.iqair.com/world-most-polluted-countries>. [Accessed: 03-Oct-2024].
- [8] Korlantas Polri, "Rekap laporan Polres Metro Jaya," [Online]. Available: http://rc.korlantas.polri.go.id:8900/eri2017/laprekappolres.php?kdpolda=6&pol_danya=METRO%20JAYA. [Accessed: 03-Oct-2024].
- [9] Air Quality Life Index (AQLI), "The 2021 Indonesia air quality life index report (version 9.7)," The Energy Policy Institute at the University of Chicago, 2021. [Online]. Available: https://aqli.epic.uchicago.edu/wp-content/uploads/2021/09/AQLI_IndonesiaReport-2021_IND-version9.7.pdf. [Accessed: 03-Oct-2024].
- [10] H. D. Nasir, D. Nur, and Z. Saharuna, "Prediksi Tingkat Polusi Udara Dengan Data Mining," in *Seminar Nasional Teknik Elektro dan Informatika (SNTEI)*, pp. 90-95, Oct. 2020.
- [11] N. F. Khusna, S. Aulia, S. Amaria, A. Rahmah, S. A. Sanmas, and F. Fauzi, "Peramalan Kualitas Udara di Semarang Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA)," in *Prosiding Seminar Nasional UNIMUS*, pp. 426-435, 2023.
- [12] P. Gopu, R. R. Panda, and N. K. Nagwani, "Time series analysis using ARIMA model for air pollution prediction in Hyderabad city of India," in *Soft Computing and Signal Processing: Proceedings of 3rd ICSCSP 2020*, vol. 1, Singapore: Springer, pp. 47-56, 2021.
- [13] M. A. Faishol, "Analisis Data Runtun Waktu Prediksi Polusi Udara di Kota Surabaya menggunakan Deep Learning RNN-LSTM," Doctoral dissertation, Institut Teknologi Sepuluh Nopember, 2020.
- [14] N. F. Muzakki, A. Z. Putri, S. Maruli, and F. Kartiasih, "Forecasting the Air Quality Index by Utilizing Several Meteorological Factors Using the ARIMAX Method (Case Study: Central Jakarta City)," *J. JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 8, no. 3, pp. 569-586, 2024.
- [15] S. Agarwal, "Data mining: Data mining concepts and techniques," in *2013 International Conference on Machine Intelligence and Research Advancement*, IEEE, pp. 203-207, Dec. 2013.
- [16] S. S. N. A. S. Namin, "Forecasting Economic and Financial Time Series: ARIMA Vs. LSTM," 2018, pp. 1-19.
- [17] D. Hatidja, "Penerapan Model ARIMA untuk Memprediksi Harga Saham PT. Telkom Tbk.," 2011.
- [18] 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: 03-Oct-2024].
- [19] V. Pawestri, A. Setiawan, and L. Linawati, "Pemodelan Data Penjualan Mobil Menggunakan Model Autoregressive Moving Average Berdasarkan Metode Bayesian," *J. Sains dan Edukasi Sains*, vol. 2, no. 1, pp. 26-35, 2019.
- [20] E. Catria, A. A. Putra, D. Permana, and D. Fitria, "Adding Exogenous Variable in Forming ARIMAX Model to Predict Export Load Goods in Tanjung Priok Port," *UNP J. Stat. Data Sci.*, vol. 1, no. 1, pp. 31-38, 2023.
- [21] S. S. Wulandari and S. Y. Sufri, "Penerapan Metode ARIMA dalam Memprediksi Fluktuasi Harga Saham PT Bank Central Asia Tbk.," *J. Ilmiah Matematika dan Pendidikan Matematika*, vol. 11, no. 1, 2021.
- [22] S. N. A. S. Namin, "Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting," *Renewable Energy*, vol. 136, pp. 758-768, 2019.
- [23] K. W. Hipel and A. I. McLeod, "Seasonal autoregressive integrated moving average models," in *Developments in Water Science*, vol. 45, pp. 419-462, 1994.
- [24] S. S. N. A. S. Namin, "Forecasting Economic and Financial Time Series: ARIMA Vs. LSTM," 2018, pp. 1-19.
- [25] I. Amansyah, J. Indra, E. Nurlaelasari, and A. R. Juwita, "Prediksi Penjualan Kendaraan Menggunakan Regresi Linear: Studi Kasus pada Industri Otomotif di Indonesia," *Innovative: J. Social Sci. Res.*, vol. 4, no. 4, pp. 1199-1216, 2024.
- [26] R. F. Putra, R. S. Y. Zebua, B. Budiman, P. W. Rahayu, M. T. A. Bangsa, M. Zulfadhilah, and A. Andiyan, *Data Mining: Algoritma dan Penerapannya*, PT. Sonpedia Publishing Indonesia, 2023.